THE READABILITY OF TEXTUAL MATERIALS IN BIOLOGY

BY: WILLIAM SAMUEL GRAHAM

FOR THE DEGREE OF MASTER OF PHILOSOPHY AT

THE UNIVERSITY OF ASTON IN BIRMINGHAM

DATE OF SUBMISSION: MAY 1981

SUMMARY

THE READABILITY OF TEXTUAL MATERIALS IN BIOLOGY

WILLIAM SAMUEL GRAHAM

MASTER OF PHILOSOPHY 1981

The study was primarily concerned with the readability of textual material used by fourth form, 'O' level ability pupils. The techniques used to measure readability can be divided into two:-

first, traditional measures (Fry's Graph, cloze procedure, subjective judgement) and second, methods of determining readability based on a psychological model of comprehension.

The traditional measures were applied to two texts from a textbook by D.G. Mackean and to two versions of the same texts which had been re-written so as to modify the language in the direction of increased readability. It was concluded that the two modified texts were easier to read with understanding. However, the measures did not agree as to the level of difficulty of each passage and reasons for this are discussed. The three traditional measures were critically evaluated and it was concluded that their reliability and even their validity is open to question; a major criticism being their limited consideration of psychological theory.

The goal of the final part of this study was to match a rank order of difficulty for the four texts based on evidence from the traditional measures, with rank orders of difficulty produced by measures based on a psychological model of comprehension. Features of the text (number of propositions per 100 words, number of arguments per proposition) did not predict the expected rank order and further research is required to identify other determinants of reading difficulty. However, an index of reading time and recall successfully predicted the expected rank order. It was concluded that in biology, where the language of a text is not the only important determinant of reading difficulty, this index should be adopted as a new measure of readability.

Key words: biology, comprehension, readability.

ACKNOWLEDGEMENTS

Many people helped me shape my ideas but I am indebted to the late Clarry Greig for his encouragement when I first showed an interest in the readability of biology textbooks. I also owe grateful thanks to Dr. Norman Graham, Department of Educational Enquiry at The University of Aston in Birmingham who was of particular assistance during this study.

Contents

Tables

Figures

Outline and content of the study

There is a growing interest in the relationship between language and learning in science education concerning generally the language used by teachers and in particular the demands made on pupils by textual material. The potential value of textual material depends on at least three groups of factors: the ability of the pupils to read and understand; the method of presenting the text within a lesson; and factors associated with the text itself. While the first two areas will be referred to in the present study, it is the third group of factors on which attention will be focussed. The factors associated with the text itself affect what is generally described as the 'readability' of a passage.

Readability was chosen as an area of study for three main reasons:

(1) Practical experience in the classroom has indicated that many pupils have difficulty understanding the textual material that is available to them.

(2) Despite the widespread use of conventional methods of measuring readability, the accuracy and even the validity of many of these techniques have been questioned; a major criticism being their limited consideration of psychological theory.

(3) An examination of the psychological research literature suggested a number of techniques, which in spite of claims of educational relevance, have not yet been fully assessed as measures of readability.

The study will report on four areas of research.

The first area of research was to predict the difficulty of a sample of texts taken from five typical biology textbooks using a readability formula. This established the extent to which pupils have difficulty with the language of biology textbooks. It also provided the basis for the selection of two typical passages taken

from a textbook by Mackean.

The second area of research used a technique known as cloze procedure to measure the interaction between the reader and text. The two passages from Mackean were compared with two passages that had been re-written in simpler language yet retained similar content.

The third area of research determined the readability of the same four passages by asking the pupils about linguistic complexity, understanding of the subject matter and interest. Pupils were also required to visualise the main ideas in one of the passages and represent them graphically as a model. A comparison of the models provided further information concerning the relative difficulty of the four passages.

The fourth area of research used measures of readability based on a psychological model of comprehension. These measures were applied to the four passages and the results were compared with the data collected from the other three areas of research. The comparison allowed the author to comment on whether these new measures are likely to improve the assessment of readability.

CHAPTER I

THE ASSESSMENT OF READABILITY

1.1 Issues in Educational Research

It is a paradox, that almost all the whole of what is taught in science courses from primary to tertiary level is research based, whereas few of the processes by which we teach science have an equivalent scholarly foundation (Whitfield, 1979). There are many reasons for this, notably the relative youth of science teaching compared to the activity of science itself and the relative delay in the emergence of disciplines, particularly psychology, upon which measurement of processes must depend.

Psychologists have often left themselves open to the criticism that the results of their laboratory experiments cannot be directly or easily applied to the classroom (e.g. Carroll, 1963; Kintsch, 1974). The advantages of control and manipulation have meant that experimental studies have often tended to largely ignore the complexities of what goes on in schools. It is a sad fact that "educational implications" often appear as a short section tacked onto the end of a paper.

If we take a look at the history of psychology and education, we find a very different picture to the one prevailing today. Nowhere is this clearer than in the work of psychologists such as Herbart (1893), Sully (1886), and Thorndike (1906). For Herbart, education took its aims from ethics; whereas psychology was there to show education the means and hindrances to these aims. Central to Herbart's psychology was what one could call a theory of organized thinking. Thinking implied organization and structure, discrimination between what is important and unimportant, a conscious deliberate association of what belongs together. The goals of instruction was thus not a mere accumulation of knowledge. Instead instruction was seen as the complement of experience - it arranges and connects the formless and dispersed fragments collected by experience. In order to achieve this Herbart said learning must always proceed from the known. That is,

children can interpret and respond to a new situation only if they have a stock of recalled, related ideas. His doctrine known as "apperception" had a strong impact upon education. It led to the idea of the curriculum: subject matter organized so that the pupil passes from familiar to closely related but unfamiliar content.

Today, these ideas sound both commonsense and very modern. Indeed one could characterize the efforts of some contemporary cognitive psychologists as an attempt to reformulate Herbart's theories and to re-establish them as guidelines for education. This does not mean that psychologists should, or could ever tell educators what to do. The psychologists' job is to state the laws that education must use to achieve its own ends. "Education", said Sully (1886), "needs continually to revert to psychology" (p.17).

Thorndike has probably had the greatest effect on psycho-educational thinking this century. Thorndike (1906) believed there were two main ways in which psychology could help education become more efficient. First, particular teaching methods could be rigorously evaluated by using the results of teaching (e.g. test scores, answers to comprehension questions). Second, teaching methods should be consistent with the child's nature and the laws of learning. Education and psychology are closely related because these laws of learning are supplied to education by psychology. According to Thorndike (1906): "To change what is into what ought to be, we need to know the laws by which the changes occur" (p.60).

From Herbart to Thorndike there has thus existed a well-articulated programme for the co-operation between psychology and education. Yet educational research, instead of being based on psychology, has become empirical, data rich and theory poor (Kintsch and Vipond, 1979). Teachers are at fault because they are often too "achievement orientated" (Williams, 1960). McDonald (1963) suggested that, while psychologists should try to relate their work to the

classroom, it was also up to teachers to become more like experimental psychologists themselves, carrying out small experiments rather than relying on commonsense. The value of the small-scale experiment is that there is no one "best method" of learning for all occasions and that the complexities of the classroom can best be examined by controlled experiments on specific situations, especially if the teacher takes into account actual classroom practice (Carroll, 1963).

Control is obviously vital in educational research; for example, experimenters who wish to emphasise factors such as sentence length or word concreteness (e.g. Stromnes and Nyman, 1974) construct very different passages to those who wish to avoid the effects of prior knowledge on comprehension (e.g. Rohwer and Harris, 1975). Faced with the problem of studying a single variable, experimenters can be in danger of creating "artificial" conditions which make any claims of educational implications less valid. Hence the importance of various factors needs to be assessed using "real" materials and their target audience as the subjects. This issue of educational relevance is one which affected each stage of the design of the present study.

Today there is a resurgence of interest on the part of both teachers and psychologists. An example of this trend can be seen in the increasing number of contributions concerning the "science of teaching science" in educational journals such as the "School Science Review".

The present study described a line of research potentially useful for solving a practical educational problem; that of readability. A model of text comprehension based on modern psychological theory might provide us with some interesting insights into this long-standing, important, and recalcitrant applied problem. The rest of this chapter describes the traditional approaches to readability and discusses the reasons why readability is of considerable practical significance to education.

1.2 The assessment of readability

a) Definitions of readability

Readability has been variously defined, but the two typical definitions given below indicate the essential attributes which are usually included.

> 'The quality of a written or printed communication that makes it easy for a given class of persons to understand its meaning or that induces them to continue reading.' (English and English, 1958).

> 'In the broadest sense, readability is the sum total (including interactions) of all those elements within a given piece of printed material that affects the success which a group of readers have with it. The success is the extent to which they understand it, read it at optimum speed, and find it interesting.' (Dale and Chall, 1948).

The most frequest and heavily emphasised element of readability has been comprehension which is of major relevance to education. Comprehension is very hard to define because it is not a label or description of anything that the reader actually does; it is a statement about how well the thing is done. Lunzer et al (1979) defined comprehension in the following way:

> 'To penetrate beyond verbal forms of text to the underlying ideas, to compare these with what one already knows and also with one another, to pick out what is essential and new, to revise one's previous conceptions.'

The textual material needs to be in a form from which we can comprehend the meaning of the words, sentences and passages. Only then can we appreciate the intention of the writer and succeed in relating his message to the larger context of our own system of knowledge.

The second attribute usually included in definitions of readability is optimum reading speed. This emphasises the perceptual aspects of reading such as the ease with which letters can be identified. It relates to primary reading skills, i.e. the skills needed in learning to read and their relationship with the syntactic complexity of the text.

Interest and motivation is the final element to be included in most definitions of readability. It is an aspect of readability which has come to be regarded more and more seriously by researchers. A child who is not a good reader can often surprise its teacher by reading and comprehending difficult texts. This occurs when the child is enthusiastic about a particular topic and motivation is high.

b) Factors influencing readability

There has been a long and sustained interest in the assessment of the effectiveness of the spoken word. Klare (1963) reports cases of 19th Century educationalists who related counts of vocabulary and familiar words to reading difficulty. Their investigations can be thought of as being the first objective attempts to measure the difficulty of texts.

The beginning of this century saw the development of readability formulae which incorporate textual factors that can be easily measured. Later developments of the formulae were due to the changing aims of educators and the demands of the people using the formulae (Klare, 1963). These later formulae were devised to deal with the characteristics of particular kinds of reading material such as the level of abstractness in a passage (Gillie, 1957).

More recently, methods such as charts and sentence completion procedures have stimulated new interest in the measurement of readability. The most recent influence on understanding reading comprehension comes from the academic disciplines of cognitive psychology and artificial intelligence (computer simulation of mental processes). In these areas (e.g. Anderson, 1977) reading comprehension is viewed as a process subject to the same constraints as the human memory and problem solving. In fact, even as early as 1908, Edmund Burke Huey believed that if we could understand reading we would understand the mysteries of the human mind. Current research will permit a more coherent and theoretically supported approach to the practical problems involved in measuring reading comprehension.

The factors which influence comprehension can be conveniently separated into two categories - factors associated with the reader and factors associated with the text. The factors associated with the reader refer to things such as linguistic competence, interest and motivation. The outside factors fall into two categories - the elements on the page and the qualities of the reading environment.

(i) Factors associated with the reader

Linguistic Competence

When a child learns language, it learns three systems: phonological, semantic and syntactic.

The phonological system includes knowledge of the different phonemes (individual sounds) in the language and knowledge of how they are blended together to create words. While phonological knowledge is a prerequisite to comprehension, once developed, it does not seem to play so major a role in comprehension as do the syntactic and semantic systems (Pearson and Johnson, 1978).

The syntactic system refers to the orderly arrangement of words in sentences. A child's knowledge of syntax is remarkably sophisticated by the time he or she enters school.

Semantic knowledge refers to knowledge of word meanings, that is, the concepts that underly the labels we call words. It also includes our knowledge of the relationships between words. For example, dogs are mammals; dogs have ears; dogs do bark; dogs and cats are both pets.

Of the three aspects, the semantic system is thought to be the most important. Later in the study there will be further discussion of how the semantic system works in comprehension. However, it is worth noting that the closer the match between the syntactic and semantic information in a text and in the reader's mind, the greater the likelihood that the text will be understood.

Interest and Motivation

A pupil is more likely to understand a passage better if he or she is suitably interested in the topic. Indeed, the research suggests that students comprehend better when they read passages that are of interest to them (Shayer, 1969; Estes and Vaughan, 1973). Unfortunately, there is little detailed or systematic research on the effect of interest on comprehension which has isolated the influence of prior experience. That is, whether students better comprehend high interest rated topics because of their motivational value or because the students know more about those topics. There have not been many attempts to define and quantify the degree to which a given text is compelling to a given reader. The simplest measure of interest is by subjective report. Asking the reader can be a helpful indicator of interest but its reliability is highly suspect (Carroll, 1972).

Motivation is in the reader, but external factors such as incentives and reinforcers will influence the attitude adopted towards a passage. As a somewhat negative example, fear and anxiety can improve comprehension scores: when a subject knows that he or she is to be tested, performance is sometimes beneficially altered (Thompson, 1976).

Interest and motivation are probably the most important aspects of readability, but they are also the least tractable. Harrison (1977) considers interest to be a particularly important factor to take into account when choosing textbooks for schools. However, the subjective nature of interest does not permit the teacher to make accurate predictions.

Reading Ability

It is almost tautological to say that the more reading ability one possesses, the better one will comprehend. The generalization can in part be explained by the notion of "automaticity". La Berge and Samuels (1973) use 'automacity' to describe the fact that some readers are able to achieve word identification automatically while others struggle with words as they read. Their contention is that

a novice or less-able reader will devote much of his or her attention to word identification having little attention for processing the meaning of the message. Hence comprehension is minimal.

Comprehension is easier if the words can be read accurately and automatically but reading the words is easier if the message can be understood. In short, there is also evidence that context can help to short circuit the amount of attention required to read the print (Pearson and Studt, 1975; Tulving and Gold, 1963).

(ii) Factors associated with the text

Legibility

In contrast to interest and motivation, research into legibility of print has been undertaken on a relatively large scale. Legibility can be defined as:

> 'The ease and accuracy with which meaningful written material is perceived and comprehended.'
> (O.U. P.E. 231).

A study of legibility would need to consider how factors such as the following, affect 'ease and accuracy'.

a) Type face choice.

b) Length of line, type size, spacing between lines (all interconnected).

c) Spacing between letters and between words.

d) The use of capital letters as opposed to lower case. The use of bold type, italic type or underlining for emphasis.

e) Margins - their width and position.

f) Colour of type and colour or tone of paper.

Concise reviews of how these factors affect legibility have been written by Gilliland (1972) and Watts and Nisbet (1974). Both reviews point out that much of the research has produced inconclusive results, due to poor control over the many variables being studied. The most

useful work has concentrated on a small number of variables.

Carrick (1978) lists information about the print used in a number of 'O' level biology textbooks. All of the books employed small print for labels and some other parts of the books, and this may pose problems for poorer readers in the class. Most used italics or bold print for emphasis of important ideas.

It is recognised that the legibility of a text is important but it was not included in the present study for the following reasons. Firstly, control of many variables is outside the sphere of influence of teachers. All the texts examined in the present study were already in use. Secondly, recent developments do not have the same application to readability measures as others mentioned in this review. Thirdly, the technical nature of the forms of measurement of such perceptual processes as visual acuity and saccadic movements involved in legibility means that they cannot be readily utilized. Fourthly, there is no ready-made procedure for assessing the general legibility of a text. Indeed there is little agreement about the importance of various factors which affect legibility. If an accurate assessment of legibility of children's texts is to be made, tests need to be carried out on the varieties of print which resembles those which they encounter. If a significant number of children were found to be less skilful in making certain kinds of discrimination, the problems could be identified and ways of overcoming them studied: but because of the absence of research in this area, it is not known whether there is a problem or not.

Language

It is perhaps inevitable that the study of language should provide so many variables. More than 150 factors have been examined and found to be related to reading difficulty (Gilliland, 1970). Luckily, most of the variables can be grouped into those acting at word, sentence and beyond the sentence levels.

WORDS. Attempts to characterise difficult versus easy words have concentrated on word frequency, word length and abstractness. Words differ in terms of how commonly they are used in the language. "The" is the most common word in the language; while "ubiquitous" is less common. There is good evidence that people have more difficulty understanding passages composed of infrequently used words. Frequency counts based on word lists such as those compiled by Thorndike and Dale give us a simple estimate of readability. However, Gilliland (1970) suggests that a frequency count does not reflect the word frequencies peculiar to an individual which leads to an inaccurate estimation of the level of comprehension.

The length of words has been incorporated in many measures of readability because of its ease of calculation (often as the number of syllables) and its acceptable level of reliability. McLaughlin (1969) uses a polysyllabic count as the sole variable for his formula but it has not been widely accepted.

Some words have concrete referents in the world: dog, cat, pig. Others have quite abstract referents: love, pity, shame. There is evidence that readers have more difficulty selecting the meanings of abstract words (Greiling, 1973) and that they have more difficulty understanding text that has many abstract words (Thorndyke, 1977). No completely acceptable method of measuring abstraction has been devised though there have been a number of attempts (e.g. Flesch, 1950; Gillie, 1957).

SENTENCES. There is an impressive body of evidence to suggest that longer sentences and more complex sentences tend to appear in passages that people have difficulty understanding. Conversely, shorter and simpler sentences tend to be associated with passages that are easy to understand (Bormuth, 1966). But the problem with short simple sentences is that they may not make the relationship between two ideas clear and explicit. For example, the sentence: "The peasants revolted because the king raised taxes" can be simplified to become "The king raised taxes. The peasants revolted." but the cause-effect relation between the ideas has become less explicit.

At the sentence and word levels of reading difficulty most measures have been expressed as formulae which predict a reading grade (e.g. Mugford, 1972; Flesch, 1948; and Gunning, 1952). They do not take into account the order of words in a sentence although some interesting work by Yngve (1960) provides a method. A detailed consideration of formulae can be found in Chapter 2.

BEYOND THE SENTENCE. Several theoreticians (e.g. Rumelhart, 1975) have developed story grammars that are analogous to sentence grammars. There appear to be some particular story structures that create more difficulty than others. For example, passages that proceed in a more-or-less cause-effect fashion appear to be more comprehensible than those which are characterized by detail without strong causal links between the ideas in the stories (Thorndyke, 1977).

Recent studies (Bransford and Johnson, 1973; Thorndyke, 1977) have verified what English teachers have been recommending to pupils for decades, that providing a statement of the theme at the outset of a passage increases comprehension and memory of the passage.

Such findings support the work of Ausubel (e.g. 1960) who claims that "advance organizers" facilitate comprehension. An advance organizer is a summary statement of the more general set of concepts

that will follow in the text. The rationale is that advance organizers help a reader to tap appropriate information in his or her head - information that will facilitate comprehension of the new data in the passage. Yet as is often the case in educational research, there is another body of counter evidence (e.g. Capponecchi, 1973; Feller, 1973) suggesting that they do not help at all.

Writers often use tables, graphs, diagrams and pictures based on the adage that a picture is worth a thousand words. Bransford and Johnson (1973) have provided a clever demonstration of how a picture can make an otherwise nonsensical passage completely reasonable and comprehensible.

Reading Environment

It is a depressing fact that the home environment can influence comprehension. The language and concept base that a pupil brings to the printed page are critical if one regards comprehension as a direct function of prior knowledge.

In the classroom, the physical and the emotional environment will influence the reading comprehension of pupils. The influence of teachers can be positive by e.g. promoting curiosity but can of course be negative by threatening punishment for failure.

The factors mentioned here have a general effect on learning and achievement and, hence, on reading comprehension, but to exclude them would have been inappropriate in a section on reading comprehension.

c) Conclusions

In sum, the study of readability involves the consideration of an alarming number of variables, ranging from a pupil's motivation to the number of syllables in a 100 word passage. Only some of these variables have been incorporated into conventional measures of readability, while the most frequently quoted and widely accepted methods are based upon an analysis of only a few easily identifiable aspects of the text.

The various measures even reflect in some way the aspects of readability which were considered earlier. Readability was seen to involve: (1) ease of reading; (2) interest and motivation; and (3) ease of understanding. To some extent these three aspects have been considered as alternatives and as a result each measure reflects only one of them.

When referring to ease of reading, readability has come to be measured by word recognition speed, error rates, number of eye fixations per second and other variables which relate to primary skills (Spencer, 1969; Tinker, 1965). Defined as interest or motivation, readability has been measured by subjective assessments, often based on questionnaires. Finally, when readability is referred to as ease of understanding, measures have usually incorporated characteristics of words and sentences in a formula.

None of the measures take into account all the ingredients that affect the readability of a passage. For example, the conceptual difficulty of a passage could greatly affect the pupil's comprehension. Moreover, the variables are almost certain to interact and only large-scale experiments such as those sponsored by the Schools Council can examine more than a few of them at one time. It is also regrettable that many of the criteria used in determining readability are not comparable since they reflect different reading skills (Gilliland, 1972).

Thus, while speed of word recognition is an element affecting ease of understanding it cannot be directly compared with a measure of comprehension based on question and answers. There is however, a technique described as cloze procedure (Taylor, 1953) which provides a global measure to which all readability variables contribute.

Cloze procedure involves the deletion of a number of words at fixed intervals. Subjects are then asked to complete the passages and the number of correct responses is scored. A high score can be regarded as indicating that a passage is more readable than a passage on which lower scores were obtained (Taylor, 1953).

Gilliland (1970) considers the advantage of cloze procedure to be that it uses the readers for whom readability is being measured. To date, it is the measure which most adequately reflects the elements of readability as defined by Dale and Chall (1948) but others have their reservations about this technique (e.g. Carroll, 1972; Carver, 1973; Neville and Pugh, 1974). Moreover, although the technique is quite widely used in experiments, it appears to be quite rare in schools, even though it has several practical advantages such as ease of preparation and simplicity in scoring, which are likely to make it of interest to teachers.

1.3 Readability as an educational problem

> "In the eyes of the vast majority of biologists the time factor was finally ousted from the arena of evolution by the concept of the generation, which, in addition to its original meaning of a step in the pedigree covered by one individual from its birth to the birth of its offspring, acquired a chronological significance."

There are few people who would not have some difficulty making sense of the opening quotation (Flood, 1958). This point could have been made quite adequately, but much more simply, by saying that when modern biologists discuss evolution they prefer to talk about "generations" rather than "years". Difficulties of this sort permeate textbooks turning science into a subject potentially too difficult for many children. If the materials used are within the grasp of the children, they will make sense out of what they read, will profit from reading, and will be interested in reading again. However, if the materials cannot be comprehended, children will become frustrated, turn away from reading, and concern themselves with more profitable enterprises. So teachers are continually faced with the problem of choosing material appropriate to the needs of a group of children.

There has been a growing interest in language in science education culminating in the report of the Bullock Committee (1975) and more recently the Schools Council Project, "The Effective Use of Reading" (Lunzer and Gardner, 1979). The Bullock Report was concerned with language across the curriculum and stated that:

> 'All subject teachers need to be much more aware of the linguistic demands their specialisations make upon pupils.' (para.20.12).

If the policy of language across the curriculum is to be developed an understanding of the potential and the problems arising from the reading demanded of the pupils is necessary. The contribution that readability could make in science has been pointed out by Gilbert (1973):

> 'By being able to effectively read the materials presented (by the text) students will be more able to associate the direct experience of experimentation with the vicarious experiences of reading, thereby developing strong meaningful relationships.'

On the whole, scientists have been reluctant to question the language of their subject which has evolved through centuries of experience to meet their special needs. Rosen (1967) suggests that two assumptions about this kind of language are prevalent. First, it is thought many pupils will never be able to acquire it so that no special steps need to be taken to help them to do so. They should be trained in the 'practical' uses of language. Secondly, when pupils are thought to be suitable subjects for advanced learning, it is assumed that such difficulties as they encounter will be of some pure intellectual kind and that language will look after itself. The problem does not exist (Savory, undated):

> 'Science textbooks are only exceptionally written in prose of the highest quality, and more often they are written in prose that may be described in contrary terms, yet the students who use these books do not in fact find them difficult to understand. They may be strange, because they deal with unfamiliar matters; they may be voluminous, making great demands on the memory; but they are not as obscure, even to the inexperienced reader, as the so-called average man is led to believe.'

Most work on the language of science has concentrated on the use of technical vocabulary. (e.g. Evans, 1976; Flood, 1958; Gardner, 1972; Rosen, 1972.) Knowing a subject's technical terms is a pre-requisite of understanding it, as hierarchial schemes of educational objectives demonstrate. Lewis (1965) in his review of literature on science objectives assigned "knowledge of terminology" to a position of primary importance. Although the development of intellectual ability and skills are generally considered to be more important than merely adding to knowledge, it has now become evident that there can be no such progress until at least some facts and terms have been mastered.

In a series of studies, Evans (1973, 1974a, 1974b, 1975, 1976) has examined the technical vocabulary of textbooks. There was great variation in the use of technical terminology, but in general the more concise the textbook, the more specialised words were employed. Evans advocated careful exploration of new terms since elementary textbooks may be used by pupils without assistance from a teacher. His articles do not assess the difficulty which technical vocabulary presents for pupils. In Gillard's study (Gillard, 1975) of textbooks, pupils, when asked to list unfamiliar words in typical passages of biology, chemistry and physics books, selected most words from chemistry and fewest from biology books.

Concern about technical vocabulary has led to the development of lists of 'core words' (e.g. Flood, 1957) but publishers seem wary of controlled vocabularies for 'O' level textbooks (Pinnock, 1977). Evans (1976) points out that synonyms (e.g. auricle/atrium; scapula/shoulder blade) serve no useful purpose and are the cause of real confusion. A code is required so that only one of the words is used and the others discarded.

Many teachers are aware that technical terms are an important part of language; but the work of Barnes et al (1969) demonstrates how they often fail to perceive that other aspects of language play a part in learning. Barnes and his co-workers cite examples from biology lessons recorded on tape where the teacher is so concerned to impart the technical vocabulary that he or she fails to see that the pupils cannot understand the subject matter. 'The biological terminology', they write, 'seems to take a value of its own'.

Does the language of science texts match the linguistic competence of the pupils? Barker (1977) suggests that the language in textbooks available in most schools may be more advanced than the majority of pupils can readily understand. Publishers sometimes try textbooks in

manuscript stage in a wide range of schools, using pupils of the appropriate age and ability. The feedback obtained from the schools gives a guide to the suitability of the book (Grieg, 1977). Yet matching a book to a child can hardly be reliable when based upon comments such as "for pupils aged twelve to sixteen years". Personal reading offers some basis for judgement, but the unreliability of subjective estimates in many spheres has been established (Chall, 1958; Klare, 1975; Harrison, 1979).

The use of combined judgements has also been tried in the assessment of readability. Moyle (1971) lists two studies involving the grading of books by committees of experienced teachers. The results of these studies show a pattern typical of assessment by panels - the grading done by committees is much more consistent than by individuals but this improvement in the consistency of judgements is only relative, since the studies reported above have demonstrated the unreliability and inconsistency of even expert judgement. One would not want to infer from this, however, that anyone is suggesting that teachers are wholly incapable of choosing suitable material for their students to read. It is rather that the unreliability of subjective assessments should lead to the utilisation of other objective methods employed for the conventional measurement of reading comprehension.

The small number of objective assessments of biology text that have been made, confirm the view that they are too difficult. (Carrick, 1978; Gillard, 1975; Gould, 1977; Graham, 1978.) Some of the studies used a traditional approach to readability assessment by using a formula (e.g. Carrick, 1978). It is arguable, however, whether this measure can be regarded as adequate in the light of more recent linguistic analyses (Gilliland, 1970). Other studies measuring the readability of biology texts have been based on cloze procedure (e.g. Gould, 1977; Graham, 1978) but their approaches to a solution of the readability problem were limited in scope. Moreover, as stated earlier, there

have been reservations about the assessment of texts by the cloze technique (e.g. Carroll, 1972; Carver, 1973; Neville and Pugh, 1974).

To sum up, it is clear that there is growing anxiety about the quality of textual material in biology and concern that the readability measures which have been used are not adequate. In a society becoming increasingly aware of the necessity for efficient and pleasurable communication through print, the control of difficulty of comprehension is critical, particularly in educational situations. The first step in the present study was to determine the extent to which biology texts could cause difficulty.

Summary

The first chapter began with a look at issues in educational research. It was argued that educational research should be based on sound psychological theory if it is to be educationally valid. Such research also has to take into account not only the control and manipulation of variables but also the reality of the classroom. It was proposed that a model of text comprehension based on modern psychological theory could have implications for the assessment of readability.

The concept of readability was defined and the factors which influence readability were reviewed. These factors were conveniently divided into: those inside the reader - linguistic competence, interest and motivation, reading ability; and those outside the reader - legibility, language, reading environment. The traditional approach to the assessment of readability has been through the production and use of formulae. More recently, alternatives such as cloze procedure have been used to improve the assessment of readability.

The remainder of the chapter reviewed the evidence that readability is an educational problem. More specifically it was claimed that biology texts are generally too advanced for the majority of pupils who use them. However, the extent of the problem has not been fully researched, while the procedures and even the validity of the readability measures that have been used are open to question.

CHAPTER II

A PRELIMINARY SURVEY OF BIOLOGY TEXTBOOKS

Introduction

In Chapter I, it was pointed out that many educationalists and teachers consider the language of many biology textbooks to be too difficult. This difficulty is providing a barrier to reading comprehension and thus learning (e.g. Gilbert, 1973; Mallinson, 1958; Rosen, 1967). Moreover, these reports support the author's classroom observations concerning the interaction between pupil and his or her textbook. However, it must be noted that subjective jusdgments can be unreliable (Harrison, 1979). The simplest alternative is to use a readability formula which provides an objective index of the prose difficulty in a textbook.

Objective surveys of biology textbooks using readability formulae are limited in number but do seem to confirm that the language presented can be too difficult (Carrick, 1978; Gould, 1977). There are three reasons why it was decided to apply a readability formula to some commonly used 'O' level and 'CSE' biology textbooks. First, there is doubt concerning the reliability of sampling procedures in previous studies. Secondly, there is no information concerning the variability of reading difficulty levels across pages within biology textbooks. Thirdly, the results would provide the basis for the selection of test passages to be used in later stages of this study.

The first part of this chapter is a general review of some readability formulae and then more specifically a method known as Fry's Readability Graph is described.

The second part of this chapter reports on the application of Fry's Readability Graph to passages taken from five commonly used 'O' level and CSE biology textbooks. An optimum sampling method is established and the implications of the results are discussed. The chapter finishes with criticisms of readability formulae and an alternative measure of readability is suggested.

2.1 Readability Formulae

Readability formulae are the most frequently used and most widely accepted measures of readability. They are concerned with a narrower concept of readability related to linguistic variables. Readability formulae give warning of potential difficulties by predicting a reader's reaction to the fairly broad areas of vocabulary and syntax complexity.

Many measures of this type have been put forward. Klare (1963), in a thorough survey of readability formulae, lists thirty-one established procedures, and admits that others exist which were not listed specifically. To this list must be added other formulae published since that date. A selection of typical and popular formulae can be found in Table 2.1.

Table 2.1 Some examples of readability formulae.

1. Dale and Chall (1948)

$$X_{C50} = 0.1579X_1 + 0.496\,X_2 + 3.6365$$

X_{C50} - reading grade score of pupil who can answer correctly one half the McCall Crabbs test questions on a passage.

X_1 - percentage of words outside the Dale list of the 3,000 words familiar to at least 80% of 9 year olds.

X_2 - average sentence length.

2. Farr - Jenkins - Patterson (1951)

New Reading Ease Index = 1.599 s - 1.015 sl - 31.517

s - number of one syllable words per 100 words

sl - average sentence length

3. Flesch (1948)

Reading Ease (RE) = 206.835 - 0.846 wl - 1.015 sl

wl - number of syllables per 100 words

sl - average number of words per sentence

4. Gunning's FOG formula (1952)

Reading Grade Level - 0.4 (sl + ps)

ps - percentage of words with three syllables or more

sl - average sentence length

5. McLaughlin's SMOG formula (1969)

Reading Grade = 3 + p

p - number which is the nearest perfect square to the number of three-or-more syllable words in thirty sentences

6. Powers, Sumner and Kearl (1958)

Reading Grade = 8.4335 + 0.0923 sl + 0.0648 s

s - percentages of monosyllables

sl - average sentence length

A number of workers such as Klare (1963), Stokes (1978) and Harrison (1979) have investigated the reliability of many of the readability formulae. Harrison found the Dale - Chall (1948) word list to be the most accurate in predicting difficulty but it is also the most cumbersome to use. Each word has to be compared against a list of 3,000 words compiled by Dale. In addition there are nearly thirty rules giving details of how suffixes, irregular verb forms, abbreviations and proper nouns are to be treated. In the context of this study it has two further disadvantages; namely that the words in the list are American rather than English and general rather than specialist. Brown (1965) hypothesizes that when applied to science books, Dale - Chall results would be too high. However, there have been attempts to revise this formula for science materials (e.g. Holonquist, 1968).

Harrison found the Flesch Index (1948) to be the second best correlate of difficulty. The index was produced after an extensive analysis of linguistic variables which are associated with comprehension. Flesch found that the average number of syllables per word and sentence

length seemed to correlate most highly with difficulty. The formulae involves a calculation (see Table 2.1) which results in a reading ease score from 0 to 100.

Most of the readability formulae were developed in America and produce a reading grade which is inappropriate for this country. However, a reading grade can be converted to a reading level by adding five. It should be noted that "reading ages" as normally measured by standardised tests have little meaning beyond a reading age of about twelve. At that point all the basic reading skills are developed but that does not, of course, mean that reading development is complete. Experience of a wider vocabulary and more complex textual material becomes of foremost importance and it is then normal to speak of a reading level. The interpretation of Dale - Chall's reading grade score and Flesch's reading ease score have to be transposed to reading levels using a set of tables (Table 2.2).

Table 2.2 Tables for transposing Dale - Chall and Flesch scores to reading levels.

Dale - Chall Formula Score	Reading Level (yrs)
4.9 and below	9 and below
5.0 to 5.9	10 to 11
6.0 to 6.9	12 to 13
7.0 to 7.9	14 to 15
8.0 to 8.9	16 to 17
9.0 to 9.9	18 to 20
10.0 and above	21 and above

Flesch Formula Score	Reading Level (yrs)
91 to 100	10
81 to 90	11
71 to 80	12
61 to 70	13 to 14
51 to 60	15 to 17
31 to 50	18 to 21
1 to 30	22 or above

Almost every formula proposed has a word variable and/or a sentence variable. This reflects the conclusion of researchers that reading difficulty is centred around factors at word and sentence levels. Gilliland (1972) suggests that more recent work indicates that this is an over-simplification. A further problem is that both word length and sentence length can be unreliable as indices of readability. For example, despite their length, words such as 'germination' and 'elephant' will be familiar and thus more readable than 'palp' or 'gene' for the average reader. Similarly, at the sentence level, a short sentence of unusual structure may be more difficult to read than a longer, more familiar sentence. Consider the two sentences quoted by Yngve (1960) in an article concerning the structure of language:

1. 'If what going to a clearly not very adequately staffed school really means is little appreciated, we should be very concerned.' (21 words)

2. 'We should be very concerned if there is little appreciation of what it really means to go to a school that clearly isn't very adequately staffed.' (26 words)

Using sentence length as a measure, one might be led to expect sentence one to be easier to read and understand than sentence two. These comments should not be interpreted as a complete rejection of sentence length as a relevant factor. As longer sentences do tend to be more complex than shorter sentences, sentence length measures

reflect the effect of memory span upon readability. To sum up, Klare (1974/1975) says the evidence is "now quite conclusive" that both word and sentence variables are satisfactory for predictive purposes.

Some formulae have been developed using measures of grammatical complexity such as clause structure, number of T-units per sentence, or percentage of propositional phrases. Other formulae attempt to measure factors such as abstraction in writing (e.g. Flesch, 1950; Gillie, 1957). However, partly because not all teachers are reliable in terms of their knowledge of English grammar, and partly because these formulae are complex, they have found less favour than the traditional formulae.

Some readability formulae have been incorporated into tables and graphs to facilitate the assessment of readability. The case for tables and graphs is given by Gilliland (1972) who considered that they have two advantages. First, they require little or not calculation since the results are related to a set of previously prepared tables. Secondly, they are an easier and more familiar technique for preparing and evaluating data.

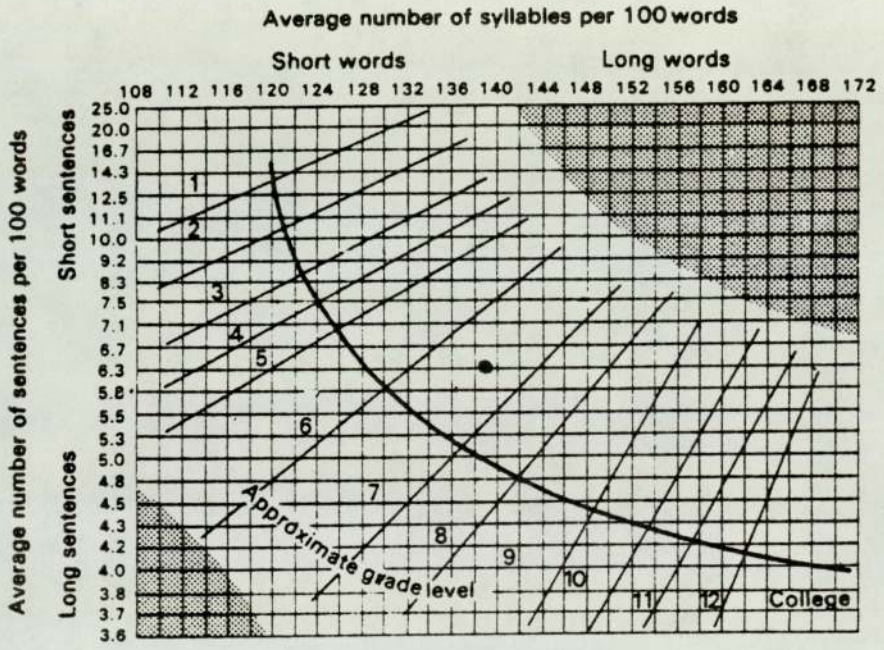
As yet, few tables and charts have been produced (e.g. Mugford, 1969; McLeod, 1962; McLaughlin, 1969) and there are few details of their validity and reliability which are necessary before they can be properly compared with other measures. An exception is the Fry Readability Graph (Fry, 1968; 1977) which has been found to be reliable yet is quick and simple to use. Fry's Readability Graph was chosen as a means of making the preliminary survey of biology textbooks so greater attention will be paid to this method in the following section.

2.2 The textbook survey

Fry's Readability Graph

The measure of readability used in the preliminary survey of biology textbooks was a graphical method devised by Fry (1968). The graph and expanded directions for its use can be found in Fig.2.1. Counts of the average number of sentences and average number of syllables per 100 words are cross-tabulated to place the text on the curve of grade levels.

Fig.2.1 Fry's Readability Graph



For example, a passage with 6.3 sentences per 100 words and 139 syllables per 100 words would be given a grade level of 7 (see dot plotted on graph).

Expanded Directions for Working Fry's Readability Graph.

1. Select sample passages and count out exactly 100 words each, starting with the beginning of a sentence. Do not count proper nouns, initializations and numerals.

2. Count the number of sentences in the hundred words, estimating length of the last sentence to one decimal point.
3. Count the total number of syllables in the hundred words, an easy way is to simply count every syllable over one in each word and add 100.
4. A word is defined as a group of symbols with a space on either side; thus Amoeba, DNA, 1979, and ∝ are each one word.
5. A syllable is defined as a phonetic syllable. Generally, there are as many syllables as vowel sounds. For example, Amoeba is three syllables. When counting syllables for numerals and initializations, count one syllable for each symbol. For example, 1979 is four syllables, DNA is three syllables, and ∝ is one syllable.
6. Plot on the graph sentence length and number of syllables; the area where the point is plotted gives the approximate grade level.
7. To convert a grade level to an approximate reading level, add 5.

Although Fry (1968) advised that an average value should be taken for only three passages from a book, it was considered desirable in the present survey to apply the method to a larger sample, since there could be considerable variation in the syllable and sentence counts. Carrick (1978) for example, applied Fry's Readability Graph to five percent of the pages in each biology textbook.

Comparisons with Dale - Chall and Flesch formulae produced high correlations (.94 and .96 respectively) and led Fry to conclude that the Readability Graph is equally reliable. Fry's method is based on the same criteria as the Flesch Reading Ease Formula and Harrison (1974) concluded that this non-arithmetic aid can be said to give scores which would be as valid as those obtained from the Flesch formula.

Methods and Results

In this study, five of the most popular "traditional" school biology textbooks for the 14 - 16 year old age group were examined. Table 2.3 names the books, while Table 2.4 gives details of the publisher's and/or the author's policy with regard to the language.

Table 2.3 Textbook details

Author/s	Title	Publisher	Date	No.of text pages	Level
Beckett, B.S.	Biology: a Modern Introduction	Oxford University Press	1976	294	O CSE
Brewer, C.V. and Burrow, C.D.	Life: Form and Function	Macmillan	1972	276	O
Brocklehurst, K.G. and Ward, H.	A New Biology	English Universities Press	1968	263	O CSE
Head, J.J.	Science through Biology	Edward Arnold	1976	293	O
Mackean, D.G.	Introduction to Biology	John Murray	(5th Ed) 1973	209	O CSE

O - Ordinary level of the General Certificate of Education

CSE - Certificate of Secondary Education

After Carrick (1978).

Table 2.4 Author's and Publisher's comments on language

Author/s	Comments
Beckett	Publisher: Accounts are "simply worded"
Brewer and Burrow	Authors: 'for reference...not intended to be read page by page...' 'minimum of text'. Publisher: Clear and concise text.
Brocklehurst and Ward	Authors: To be used as a 'textbook or for reference'.
Head	Publisher: '...highly readable and fascinating book which will capture the student's enthusiasm' '...clear explanations...' 'The author does not shrink from using correct terminology'.
Mackean	Author: '...meant as book suitable for the average 'O' level candidate'. Publisher: 'pupils reference book' 'brevity and directness primary objectives' 'clearly explained, fully illustrated information'.

After Carrick (1978)

In some cases the publisher or the author has made no clear expression of intention so they have been omitted from the list.

The survey only sampled passages taken from pages containing factual information. There are a number of reasons for omitting passages containing practical instructions or questions:

(i) Practical instructions and questions make up only a small proportion of the textual material. Smith (1974) has found that in 'Introduction to Biology' by D.G. Mackean, factual information takes up 45% of the space in the book while practical instructions and questions take up 3% and 2% respectively. The remaining 50% of the coverage is used for photographs and diagrams.

(ii) It is difficult to find 100 word long passages from sections on practical instructions and questions. Thus it is difficult to apply Fry's Readability Graph.

(iii) Experience has shown that the structure of practical instructions and questions, such as very short sentences, results in abnormally low

reading levels, when the Fry Readability Graph is used. (Graham, 1977). The problems associated with reading and understanding this material is a separate issue and new techniques and approaches will have to be developed to explore these content areas.

Before surveying all the textbooks it was necessary to establish the number of passages that were required in order to obtain an accurate estimate of a textbook's overall difficulty. Thus a pilot study was conducted using 'Introduction to Biology' by D.G. Mackean. Reading levels were determined by the application of Fry's Readability Graph to passages randomly taken from 5%, 10% and 20% of the pages.

Table 2.5 Details of reading levels obtained by the application of Fry's Readability Graph to 5%, 10% and 20% of pages taken from 'Introduction to Biology'.

% of pages sampled	number of passages analysed	Reading Level (yrs) Mean	Range	S.D.
5	11	16.2	13-18	1.81
10	21	15.9	13-18	1.79
20	41	15.85	12-18	1.89

The results in Table 2.5, clearly indicate that the mean reading level is similar for all of the sample sizes. If one takes into account that the accuracy of Fry's Readability Graph is approximately ± 1 reading level (Harrison, 1978) then it can be concluded that there is no real benefit in using a sample size greater than 5% of the pages. However, it must be pointed out that this conclusion may not be valid for other types of textbook which have a different style to that found in biology. Also there may be a minimum number of passages that have to be used to obtain an accurate estimate of the overall reading level. Sampling 5% of the pages in a small book of e.g. 60 pages would only give one 3 passages to analyse. Nor do the present results probably apply to other readability formulae. Burkhead and Ulferts have examined the effect of sampling frequency on the reliability of the Dale-Chall formulae.

Samples taken at 20, 30, 40 and 50 page intervals were found to be as reliable as samples taken at 10 page intervals. (Burkhead and Ulferts, 1977).

The five biology textbooks were surveyed by the application of Fry's Readability Graph to passages randomly taken from 5% of the pages.

Table 2.6 Details of reading levels obtained by the application of Fry's Readability Graph to 5% of pages taken from 5 school biology textbooks.

Author/s	Reading Levels (yrs) Mean	Range	S.D.
Beckett	15.6	12-18	2.0
Brewer and Burrow	16.8	12-18	2.0
Brocklehurst and Ward	15.6	12-18	2.1
Head	15.8	12-18	1.9
Mackean	16.2	13-18	1.8

It can be seen from the results in Table 2.6 that most of the biology textbooks have an overall reading level of 16 years; the exception was Brewer and Burrow which has a mean reading level of 17 years. An important finding was that the reading levels within each book were far from consistent, usually ranging from 12 to 18 years.

One final question arises from the work reported above; "Does the reading level of textual material within individual chapters remain fairly constant or vary?" In order to find out, three chapters were randomly chosen from 'Introduction to Biology'. Fry's Readability Graph was then applied to passages randomly chosen from every column of text to be found in each chapter.

Table 2.7 Reading levels obtained by the application of Fry's Readability Graph to 3 chapters taken from 'Introduction to Biology'.

Chapter	Number of passages	Reading Levels (yrs) Mean	Range	S.D.
5	7	14.1	12-16	1.65
12	9	15.8	14-17	1.20
20	5	16.8	15-18	1.18

The results in Table 2.7 establish that the textual material within a chapter also has a wide range of reading levels. It is obvious that Mackean has not aimed particular chapters at a particular age group.

Discussion

The measurement of texts by Fry's Readability Graph indicated that four of the five biology textbooks had mean reading levels of 16 years; that is, they are only suitable for use by some fifth form pupils. Teachers should be aware that these results do not indicate that the language of the textbooks is suitable for the less able 16 year old taking for example a C.S.E. exam - even though some of these books are described as being suitable for this exam by the publishers (see Table 2.4). The textbook by Brewer and Burrow had a mean reading level of 17 years making it unsuitable for virtually all 'O' level and CSE groups. An interesting point is that the Bullock Report (Bullock Committee, 1975, p.10). appears to suggest that a value of 6 should be added to American grade levels in order to obtain a reading level. If this report is correct, then all the mean values given should be raised by one year. All the textbooks used in the survey would then become unsuitable for the majority of fifth form pupils.

Harrison (1979) has looked at the reading levels of printed texts in use in a number of subject areas. Science and social science contained the most difficult texts, with first year science standing out as particularly difficult. It is slightly disturbing to note that seven of the nine samples of first-year science prose were from teacher produced worksheets. These findings are consistent with a number of studies such as the one by Johnson (1979) and those reported by Klare (1963) in which science texts were rated as being too difficult for the pupils who read them.

On the whole, the present survey obtained results which were similar to those reported by Carrick (1978) and Gould (1977). However, Gould (1977) reports that the textbook by Mackean has a mean reading level of 18 years which is two years greater than the mean value found by Carrick (1978) and in the present survey. Gould (1977) gives no indication of the number of samples he took from each book and may have

followed Fry's (1968) original recommendations to take only three passages from the textbook. This may explain his rather high estimate of reading level for the textbook by Mackean.

Perhaps the most important result was the range of reading levels obtained for passages within each textbook. It appeared to be quite random within a textbook whether a passage was suitable for 12 year old pupils or 18 year old pupils. There is an argument for writing chapters at different reading levels; for example, a chapter on classification aimed at young pupils could have a low reading level; while a chapter on genetics aimed at older pupils could have a higher reading level. Unfortunately, none of the authors of the textbooks used in this survey can claim to have used such an approach.

The situation may be worse still, because of the way in which texts are used in science. Harrison and Gardner (1977) distinguish between the supported and unsupported use of texts. In subject areas such as English, teachers tend to support the text by using it in the classroom, often reading it aloud and using it as the basis for discussion. The teacher is present and actively explaining the text, or is at least immediately available to answer any questions. Biology texts, by comparison, are used for reference purposes or for note making set as homework (Gould, 1977). This unsupported use of texts is related to the need of biology teachers to use class time for practical work and discussions. American research work suggests that the reading levels of science texts ought to be about two years less than those of a supported text. On this basis none of the textbooks analysed are suitable for use in an unsupported context with 'O' level or 'CSE' pupils.

It is suggested that the results of the present survey, taken together with findings from similar studies, give firm evidence that many pupils, perhaps even most pupils, are experiencing difficulties with their textbooks because of the way in which they are written. The implications are therefore:

1. The mean reading level of textbooks in biology is too high, making them hard to read.
2. Special attention should be paid to the distribution of results within a textbook. The potentially wide differences between passages in the same chapter of a textbook makes the task of selecting material at a suitable reading level difficult.
3. Children should not be expected to use these textbooks to work effectively on their own, e.g. at home.

Having criticised the present generation of biology textbooks it is useful to consider the practicality of applying a readability formula during textbook authorship. Readability research has consistently pointed out that there is no causal relationship between difficulty level as predicted by a formula and the actual difficulty a reader is likely to encounter. Formulae simply measure certain correlates of text difficulty and it is invalid to assume that short sentences, for example, will guarantee that a passage is easier to read (Klare, 1963).

The danger of writing to a formula has been experienced by the Resources for Learning Development Unit based in Avon, which has used readability formula scores to assist them in producing materials aimed at different reading levels for mixed-ability groups. Harrison (1979) reports that initially the formula results were a valuable guide, but gradually the writing team became aware of what factors they could alter in order to affect the measured readability score. Gradually, the team realized that they were tending to write stilted and incoherent prose. In a survey of publishers' attitudes towards readability, Greig (1977) found that in general, publishers seem wary of applying formulae since these may encourage clumsy writing.

So far, the readability of texts has been discussed without considering the limitations of readability formulae. Formulae serve a purpose but they leave a lot unexplained. For example, it is said

that an abstract, complex discussion is accompanied by many conjunctions (Kintsch and Vipond, 1979). Clearly the conjunctions do not cause the difficulty - they are merely an index of it. A formula that includes conjunctions as one of its factors is at the "symptom level", that is, it fails to deal with the deeper levels of text comprehension. There are further limitations of readability formulae, which will be outlined below:

1. Formulae sometimes give perplexing results. Lockman (1956) computed the Flesch Reading Ease scores for nine sets of instructions for psychological tests. He then had 171 naval cadets rate the instructions on "understandability". However, the rank-order correlation between the two sets of measurements was -0.65 ($p < 0.05$). With reserve, Lockman concluded: "It would appear that reading ease scores and understandability ratings were not measuring the same thing".
2. Klare (1963) notes that formulae measure only style, not content. They measure only one aspect of style - difficulty - ignoring dramatic effectiveness, appropriateness and so on. Nor are they perfect indicators of style difficulty and formulae are certainly not measures of good style.
3. Higher-level features of text organization are not included in readability formulae. These features are the "tips" and "cues" which allow the reader to organize and integrate sequential material. Readability formulae cannot discriminate between scrambled and well-ordered words; much less can they detect scrambled sentences or paragraphs. A standard set of comprehension passages from a large-scale test, the McCall - Crabbs Standard Test Lessons in Reading have been used in a large number of American validation studies (McCall-Crabbs, 1925, 1950, 1961). Chall (1958) points out that these tests are too short to reveal the higher-level features of a text. In this sense there was from the beginning a bias against including these features in readability formulae.

4. Taylor (1953) maintains that formulae are particularly insensitive to the effects of textual factors upon specific individuals or small groups. He also argues that formulae are insensitive firstly, to the effect of previous knowledge of the book; secondly, the measurement of the readability of passages involving the non-idiomatic use of language; and thirdly, collections of words which have little meaning for an individual. These criticisms reflect the point that readability formulae only measure some of the factors associated with the text. The researcher also requires measures which reflect the other factors that affect readability.

5. Stokes (1978) critically reviews the reliability and validity of readability formulae, and considers it unsatisfactory to validate formulae by means of correlation with other formulae. As Manzo (1970) aptly puts it:

> "readability can only rank-order materials, that is, compare them on the same linguistic variables. If there is any validity to this process it is only to the extent to which there is agreement with existing standards (as when early formulae compared their rankings with those derived from basal readers as a means of designating appropriateness for given grade levels). This is incestuous.....(and) makes readability research a construct without a point of reference."

6. Most formulae involves the use of a word variable and a sentence variable, reflecting the conclusions of researchers that they correlate highly with reading difficulty (e.g. Klare, 1974/1975). The evidence for a relationship between word or sentence variables and psychological theory is scanty and mostly indirect (Harrison, 1979). Clearly, the existing work on readability formulae is close to useless from a psychological perspective.

Although readability formulae are by no means useless for practical purposes they are far from adequate. Attempts to overcome some of the criticisms outlined above have led to the development of alternative measures. Of these new measures of readability a technique known as cloze procedure has proved most interesting to researchers.

Summary

A preliminary survey of the reading difficulty of 5 school biology textbooks was made by applying Fry's Readability Graph to a 5% sample of pages in each book. The mean, range and standard deviation of reading levels for these pages was calculated for each book. Results indicate that the mean reading difficulty of all the texts was at least 16 years, that is they are only suitable for use by fifth form pupils. The variability across the sampled pages indicated by the wide range and distribution of reading levels within any one book suggest that whole sections of text are unsuitable for even the 16 year old pupil.

The use of readability formulae can be criticised on the grounds that they are not based on psychological theory and only measure a limited number of the factors which influence readability. Existing words on readability propose an alternative measure known as cloze procedure which is said to overcome the criticisms outlined above.

CHAPTER III

COMPARING THE READABILITY OF BIOLOGY TEXTS BY CLOZE PROCEDURE

Introduction

The measurement of texts by readability formulae indicated that biology textbooks have reading levels which are higher than the age of the pupils that read them, but formulae only measure a limited number of factors that affect the readability of a text. A technique was required which would go beyond the measurement of language and include other elements of readability. Taylor (1957) suggests that cloze procedure reflects the sum total of all influences which interact to affect readability. When the cloze test is applied, both reader and text are assessed simultaneously by use of the one measure which undoubtedly gives this procedure a greater face validity than other procedures referred to.

The value of the cloze procedure as a measure of readability has been stressed by writers such as Bormuth, (1963), Rankin, (1959), and Klare, (1966), though there is some disagreement on the possible extent of its valid use (e.g. Carroll, 1972, Schlesinger, 1968). At the present time there is no clearly established procedure for the use of the cloze technique. An evaluation of the research literature was thus required to produce a method which may be used to optimise the effectiveness of cloze procedure.

Two texts were chosen from the textbook by Mackean to represent material which pupils have typically to cope with in an 'O' level biology course. In order to make a comparison, two texts on the same subjects with similar content were written in a simpler manner.

The first section of this chapter is a review of the research literature in order to establish a reliable method of approach. The second section begins with a description of the original and modified test

passages which were presented as cloze tests to fourth form 'O' level ability pupils. The results of these tests are analysed and their implications discussed.

3.1 Cloze Procedure

The procedure is to delete certain words either randomly or at fixed intervals from a passage. The mutilated passage can then be presented to a subject who has to supply the appropriate words for the gaps. In order to supply the word "likes" for example in "The cow _____ to eat grass", the subject has to:

(i) select a word that fits in grammatically

(ii) select a word with the correct meaning

(iii) follow the language pattern and vocabulary employed by the author.

If the sentence above had been preceded by "Last week" the word "likes" would no longer be acceptable for the omission must be the past tense "liked". The synonym "enjoys" could have been employed but it is perhaps less likely to occur in the given text.

Cloze procedure not only demands fluency and a knowledge of grammatical structure but also requires the pupil to understand the text. As such, it is claimed that cloze procedure measures total readability much more nearly than for example, readability formulae.

The term "cloze" is from the Gestalt concept of closure, which in crude terms is a tendency on the part of an organism perceptually to complete stimulus presentations which are physically incomplete. This statement does not mean that the missing parts of, say, a geometrical shape, are actually seen but merely that the parts are organized as belonging to a complete figure. Cloze procedure can be thought of as the predictability of words in a passage; the higher the score of correct responses the more readable a passage should be.

There have been a number of alternative approaches to the use of cloze procedure:

1. Specific lexical deletions can be used to diagnose or develop a particular reading skill. In lexical deletions, certain parts of speech such as nouns or verbs are omitted.

2. By deleting content words, a passage can be used to measure information gain at the end of a topic.

3. Rankin (1959) found that cloze scores correlated highly with reading comprehension scores. Since the early sixties cloze procedure has become a common correlate of the reader's comprehension replacing the use of, say, multiple-choice comprehension tests. Bormuth (1969) for example, has used the cloze comprehension test as criterion to develop 24 formulae with up to 20 different variables in each.

4. In a cloze test the reader's task is to work out from the surrounding context which word has been omitted and to insert the one which seems most appropriate in each blank space. The proportion of correctly-guessed words gives an indication of the extent to which the reader has understood the passage. The score a reader obtains on a number of passages will vary according to how difficult each one is to comprehend, and they enable the researcher to rank the passages in order of difficulty. The rationale behind this method of testing readability is discussed in papers by Bormuth (1967, 1968), and is widely accepted. Nevertheless, as will be emphasized later in this chapter, while one might endorse a wider use of the technique, caution is needed when cloze procedure is used as a test of readability.

There has been some disagreement concerning the preparation of cloze passages which are to be used as a test of readability. The first issue is to choose a suitable deletion rate. The work of MacGintie (1961) and Fillenbaum et al (1963) have independently come to the conclusion that the deletion of every fifth word in a passage is most satisfactory. However, McNinch et al (undated) and Klare et al (1972) both suggest that the deletion pattern should be varied depending upon the nature of the material to be tested. Harrison (1979) felt that although fifth-word deletions may increase the reliability of test scores, the effect on children would be to dampen their motivation; especially if the test

passage had a very high difficulty level. Smith and Dechant (1961) suggest that young children cannot understand a test passage if it has a deletion pattern greater than one in ten.

The second issue concerning the preparation of cloze passages is the choice of versions. If the deletion rate was, say fifth-word; then we can produce five different versions of a passage. For example:

Version 1. _____ second issue concerning the _____ of cloze passages is _____ choice of versions.

Version 2. The _____ issue concerning the preparation _____ cloze passages is the _____ of versions.

Version 3. The second _____ concerning the preparation of _____ passages is the choice _____ versions.

Version 4. The second issue _____ the preparation of cloze _____ is the choice of _____.

Version 5. The second issue concerning _____ preparation of cloze passages _____ the choice of versions.

Choosing one version of a test passage involves hazards. In the example above, it is clearly easier to select the correct words in version 1 than in version 5. Obviously a word such as "the" will be supplied without much difficulty, whereas the word "versions" makes greater demands on the subject's reading strategies. Thus the use of only one version may produce results which are not representative of the true difficulty of a test passage. Davies (1976) for example, has reported inter-version variations which were statistically significant. It is a pity that the study of British biology 'O' level textbooks using cloze procedure by Gould (1977) was based on results for only some versions of each test passage. His conclusions must therefore be treated with caution.

Choosing one "fair" version of a cloze passage would involve some form of content analysis. Bormuth (1968) and Davies (1976) have both identified a number of criteria which could be used to group cloze

items in terms of difficulty; the choice of criteria being relevant to the material being tested for readability. For example, words in a passage may be classified into structural versus content words or monosyllables versus polysyllables. To sum up, the burden of preparing and administering all versions of a cloze passage are considerably greater than for one version. However, there is a possibility that the chosen single version would differ significantly from another. It is the opinion of most workers in the field of readability (e.g. Gilliland, 1977; Harrison, 1978) that all versions should be used thus ensuring that responses to all deletions can be analysed.

The next problem is to set about producing an adequate sample of the materials to be used in the cloze test. At present, opinions differ as to what is "adequate". Harrison (1978) finds one passage of reasonable length (i.e. 200 words) from a text book to be acceptable. In view of the differences in subject matter in text books and their effects upon style of writing, one would wish to be cautious about generalising from it. Rankin and Culhane (1969) in their account of how cloze techniques might best be used, recommend the same approach as for readability formulae; that is, a number of samples of text are analysed and the results averaged. Taylor (1956) who was first to point out the value of cloze procedure as a measure of readability suggests a sample size of 20% of the pages in a book. However, Greig (1976) points out that test passages should be determined by considerations other than strictly number of page choice. He advises that the choice of text passage should be based on criteria such as the content or reading level of the textual material.

It would be useful to state that a cloze score of "x" indicated that a text was suitable for children of a particular age. Many workers, such as Rankin and Culhane (1969), Rankin (undated) and Bormuth (1969) have studied this problem. Bormuth (1967, 1968a) has reported that a 38% correct verbatim response correlates with a subject's ability to gain information effectively from a passage; in the latter of the two

studies referred to, he suggests an alternative criterion of 44%. Based on his findings, some educationalists have classified cloze scores in the following way (e.g. Thelen, 1974).

Score %	
0 - 39	Frustration level - passage is too difficult
40 - 59	Instructional level - assistance is required
60 - 100	Independent level - no assistance is required

There appears to be little justification for the use of Bormuth's findings as a criterion for passage selection. His results were based on large-scale research studies each involving a hundred subjects, and the figures of 38% and 44% were the average scores on a cloze test which were associated with a certain degree of success on a multiple-choice comprehension test. The problem is whether the cloze test scores of a class group would be a reliable indication of the difficulty level of a passage (Harrison, 1979). The strength of cloze procedure is to provide data which places texts in a rank order of difficulty.

3.2 The cloze tests

Method

Subjects

The requirement was for a sufficient number of fourth form pupils taking an 'O' level biology course, in order to allow for a statistical treatment of results. Six male grammar schools in the West Midlands were invited to take part in this experiment; four of the schools accepted. The author teaches at a grammar school which has links with the schools involved. The choice of these schools thus simplified the process of gaining permission from the headmasters. It is considered that this choice of one 'type' of pupil should not mean that the overall conclusions drawn from the experiment don't apply to children of either sex in other types of school.

In all, 140 subjects with a mean age of 14.6 years took part in the experiment. Three of the schools each provided one class group while the fourth school provided two class groups.

Materials

To familiarise the subjects with cloze procedure, an interesting pre-test story was written on sharks which had a reading level of 11 yrs. as determined by Fry's Readability Graph (Table 3.1). A single version of the pre-test was prepared using a fifth-word deletion pattern beginning with the fifth word.

Table 3.1 Pre-test story on sharks.

AN ENCOUNTER

I had never realised that the Mediterranean Sea could hold such terrors. I had been underwater for about half an hour when I decided to move back up to the surface.

As I slowly moved up, I saw a large grey shape in the distance. I was not sure what it was. So I stayed still as the grey shape swam towards me. With a sinking feeling in my stomach I realised that I was just about to meet my first shark.

Watching me with cold, hard eyes, he slowly swam around me. Each time he circled me the shark moved in a little closer. I could sense that he was waiting for a chance to attack.

Desperately I tried to think of ways to frighten off sharks. One of my friends had told me to shout at the top of my voice. So I shouted until I was hoarse. The grey shark seemed to be deaf.

As my grey friend began to glide towards me, I knew that my life was at stake. Those few minutes seemed to last a limetime, when suddenly I woke up!

Two test passages were selected from the textbook by Mackean to represent typical textual material. Two passages were chosen because of the dangers of generalizing from one. More test passages would have been desirable but this would have considerably increased the burden of preparation, administration and analysis.

Selection of the passages was determined by the content and reading level of the textual material. Both test passages were taken from topics that the subjects had already covered in their classwork. Thus the content and technical terms used in the passages should have been familiar to the pupils. Reading levels were used so as to select one difficult passage and one simpler passage. The difficult passage was from the topic 'exchanging gases' and had a reading level of 18 years. The simpler passage was from the topic 'lymphatic system' and had a reading level of 15 years which was similar to the age of the subjects used in this experiment. (Table 3.2).

Table 3.2 Test passages taken from "Introduction to Biology" by Mackean.

Lymphatic System

The capillaries are not the only route by which the tissue fluid returns to the circulation. Some of it returns via the lymphatic system. The proteins in the tissue fluid are unable to re-enter the capillaries but can drain into blindly-ending, thin-walled vessels which are found between the cells. These lymphatics join up to form larger vessels which eventually unite into two main ducts and empty out their contents into the large veins entering the right atrium.

The fluid in the lymphatic vessels is called lymph. Its composition is similar to plasma but it contains less proteins. It also contains a certain type of white cell, lymphocyte, which is made in the lymph glands.

Gaseous Exchange

The lining of the alveoli is covered with a film of moisture. The oxygen concentration in the blood is lower than in the alveolus, hence oxygen in the air space dissolves in the film of moisture and diffuses through the epithelium, the capillary wall, the plasma and into a red cell, where it combines with the haemoglobin. The capillaries reunite and eventually form the pulmonary veins which return the oxygenated blood to the left atrium of the heart. The low concentration of carbon dioxide in the alveoli stimulates the enzyme, carbonic anhydrate, in the blood to break down the bicarbonate salts and liberate carbon dioxide. This gas diffuses into the alveoli and is eventually expelled.

In addition, two test passages were written by modifying the language of the passages from Mackean. These modifications allowed for the comparison of texts with similar content yet different reading levels. In a previous study (Graham, 1977) it was found that the reading level of a text did not always predict its cloze score and it was suggested that features of the text other than language were important predictors

of cloze score. Similar discrepancies have been reported by a number of workers (e.g. Bormuth, 1966; Harrison, 1979; Miller, 1974). The modified passages were both written to give a reading level of 11 yrs as predicted by Fry's Readability Graph. Thus the author was able to compare cloze scores obtained for two passages with the same reading level.

The re-writing of the original passages from Mackean was necessarily a subjective process. If one is examining a single factor and its effect on readability, then a single set of rules can be objectively applied to the passages. However, the resulting passages are often less 'attractive' to read. In this study the objective was to communicate complex matters in relatively simple language and it was difficult to see how a systematic approach could help. Even so, studies by Gibson and Levin (1975), Grey and Leary (1935) and Peterson (1954), helped the author to formulate guidelines as to what constituted 'simple language'. The modified passages are presented in Table 3.3. and the guidelines can be found in Appendix 1.

Table 3.3 Modified test passages.

The Lymphatic System

Around the cells of our body is a liquid called tissue fluid. It keeps our cells damp and forms a link with the blood vessels.

The substances that our cells need pass out of the blood vessels and into the tissue fluid. These substances then pass into our cells. Anything made by our cells passes into the blood vessels in the same way.

Not all the fluid or substances made by our cells return to the blood vessels. Some of it drains into ducts between the cells. These ducts are part of the lymphatic system. The fluid inside the ducts is called lymph. Lymph is very similar to blood but contains less protein.

Table 3.3 contd.

Exchanging Gases

Inside your lungs there are small air sacs called alveoli. These sacs have walls which are damp. The gases dissolve in this damp layer.

Around the air sacs are many blood vessels. These blood vessels contain less oxygen than in the air sacs. So the oxygen moves from the air sacs into the blood.

The blood vessels around the air sacs also contain a lot of bicarbonate salts. An enzyme breaks these bicarbonate salts down to give carbon dioxide. There is less carbon dioxide in the air sacs than in the blood vessels. So the carbon dioxide passes out of the blood and into the air sacs.

The original and modified test passages were prepared using a fifth-word deletion pattern. All five versions of each test passage were used (see p.44 for details).

Procedure

The cloze tests were administered to five class groups of subjects during June 1979. Each school allocated a double period of classtime per group of subjects. The purpose of the experiment was explained by the author and it was pointed out that the experiment tested the textual material and not the pupils. The pre-test passage on sharks was handed out, verbal instructions on how to complete a cloze test were given and pupils' questions were answered. Having completed the pre-test, the pupils marked their own test sheets which generated further discussion about cloze procedure. Now familiar with the technique, each pupil in the class group was allocated at random one version of one of the four test passages. The cloze tests were completed in relaxed conditions, however provision was made to prevent cheating. Adequate time was given for all pupils to complete their cloze test passage. Finally, the test

sheets were collected in to be marked by the author.

Scoring

Scoring was carried out on a verbatim basis; that is a word was only scored as correct if it was the exact one which had appeared in the original passage. Minor mis-spellings were disregarded. A number of independent studies have demonstrated that the enormously lengthy process of using a scoring system that accepts synonyms as correct produces minimal increases in the correlations with passage difficulty. (Bormuth, 1968b).

Results

The results of this investigation are shown in Table 3.4 where mean cloze scores and standard deviations for each version of a passage are set out together with the overall means.

Table 3.4 Cloze scores and standard deviations for each version of a passage.

Test Passages		Versions 1	2	3	4	5	Mean Values	
AM	$\bar{x}$	44.8	52.2	53.7	44.8	33.7	45.9	
	SD	6.1	6.9	16.3	15.4	8.1	12.7	N = 33
	N	7	7	7	5	7		
BM	$\bar{x}$	35.5	41.2	34.0	50.9	54.6	43.2	
	SD	10.7	15.6	10.9	8.3	9.9	13.5	
	N	7	7	7	7	7		N = 35
AG	$\bar{x}$	77.3	46.1	57.2	67.5	76.2	64.8	
	SD	12.8	18.1	23.5	11.8	9.5	19.4	
	N	7	7	7	6	7		N = 34
BG	$\bar{x}$	54.6	41.3	56.1	33.3	56.1	48.3	
	SD	12.1	19.4	12.2	15.3	11.4	16.5	
	N	7	7	7	7	7		N = 35
Mean Values	SD	53.0	45.2	50.2	49.1	63.6		
		10.4	15.0	15.7	12.7	9.7		
		N = 28	N = 28	N = 28	N = 28	N = 28		

Author M - Original passage (Mackean) G - Modified passage (Graham)

Topic A - 'Exchanging gases' B - 'lymphatic system)

The results were subjected to a three-way analysis of variance for independent measures in order to test for differences between the versions, topics and authors. A summary of this analysis of variance is given in Table 3.5.

Table 3.5 Analysis of variance: Versions by Topics by Authors

Source of Variation	SS	DF	MS	F	Sig of F.
Main Effects					
Versions	1631.1	4	407.8	2.21	0.066
Topic	3122.9	1	3122.9	17.4	0.001
Author	4783.3	1	4783.3	26.6	0.001
2-way Interactions					
Versions - Topic	1168.2	4	292.1	1.6	0.172
Versions - Author	4377.0	4	1094.2	6.1	0.001
Topic - Author	1503.0	1	1503.0	8.4	0.005
3-way Interactions					
Versions - Topic - Author	4895.1	4	1223.8	6.8	0.001
Explained	21679.8	19	1141.0	6.3	0.001
Residual	21007.2	117	179.5		
Total	42687.0	136	313.9		

The presence of a significant three-way interaction makes the interpretation of this table difficult. Since the differences between versions were not of any practical importance to the main hypothesis it was decided to use combined versions for each passage. However, the difference between versions is a subsidiary result of considerable importance and will be dealt with later on in this section (see page 58). The mean values of the cloze scores are presented in Table 3.6. The results were subjected to a two-way analysis of variance for independent measures in order to test for differences between the topics and authors. A summary of this analysis of variance is given in Table 3.7.

Table 3.6 Percentage cloze scores for four passages.

		TOPIC		
		Exchanging gases	Lymphatic system	Total
AUTHOR	Original passages (Mackean)	45.9	43.2	89.1
	Modified passages (Graham)	64.8	48.3	113.1
	Total	110.7	91.5	

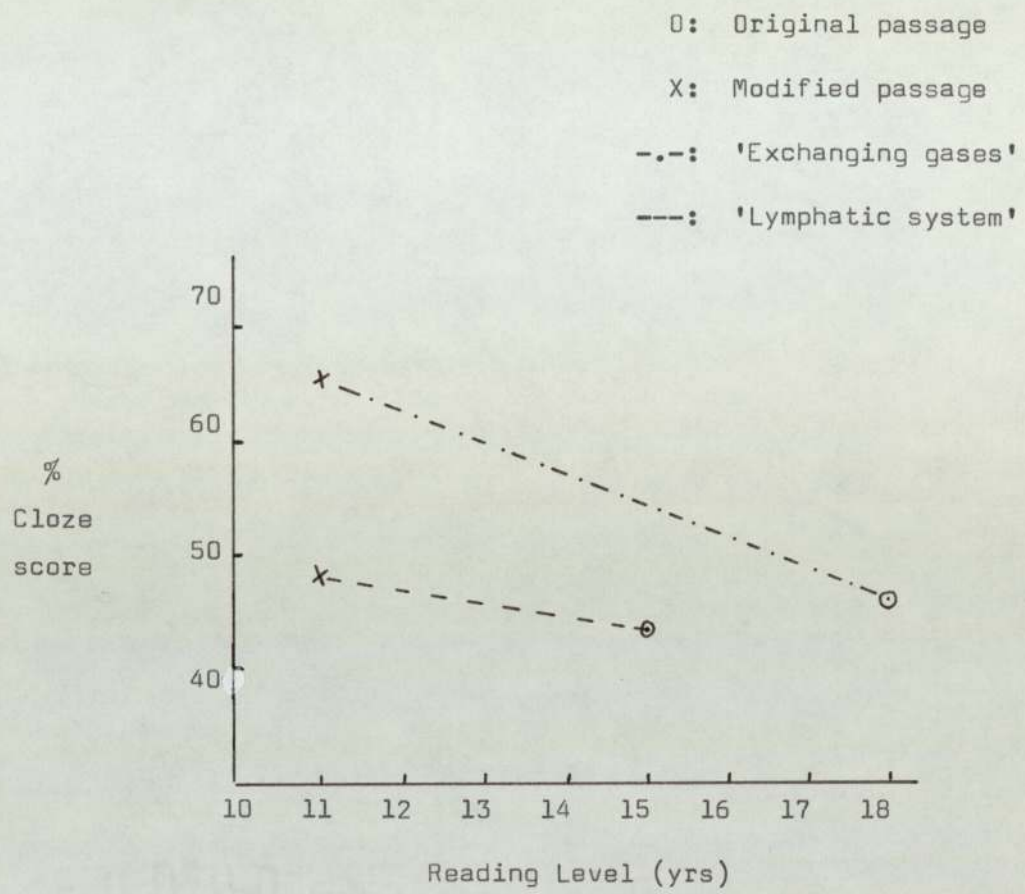
Table 3.7 Analysis of variance: Topics by Authors.

Source of variation	SS	DF	MS	F	Sig.of F.
Topic	3173.3	1	3173.3	12.8	0.001
Author	4766.2	1	4766.2	19.2	0.001
2-way interaction	1629.9	1	1629.9	6.6	0.012
Explained	9627.9	3	3209.3	12.9	0.001
Residual	33059.1	133	248.6		
Total	42687.0	136	313.9		

The two-way interaction between the topics and authors was statistically significant at the 0.05 confidence limits. However, an examination of Table 3.6 indicates that the differences between the authors is in the same direction and similarly, the differences between the topics is in the same direction. Hence there is an ordinal interaction and it is reasonable to interpret the results as indicating that the modified passages obtained higher cloze scores than the original passages. Table 3.6 shows that the modifications made to the original passage on the topic 'lymphatic system' only raised the cloze score by 5%. In contrast, the modifications made to the original passage on 'exchanging gases' resulted in an improved cloze score of approximately 20%. It can be concluded that there must be factors other than language which are important determinants of the readability of the topic 'lymphatic system'.

The relationship between cloze scores and reading level as predicted by Fry's Graph is illustrated in Fig.3.1. Note that there are differences between the results obtained using these measures. For example, passages BG, BM and AM all have similar cloze scores yet they were predicted by Fry's Graph to have reading levels of 11, 15 and 18 years respectively.

Fig. 3.1 Combined cloze figures and readability levels (Fry Graph) for original and modified passages on two biology topics.



The second major conclusion of this study is that Fry's Graph and cloze procedure have given widely different estimates of the readability of each passage.

There are a number of subsidiary results that are not concerned with the main experimental hypothesis, but they are nevertheless of considerable importance.

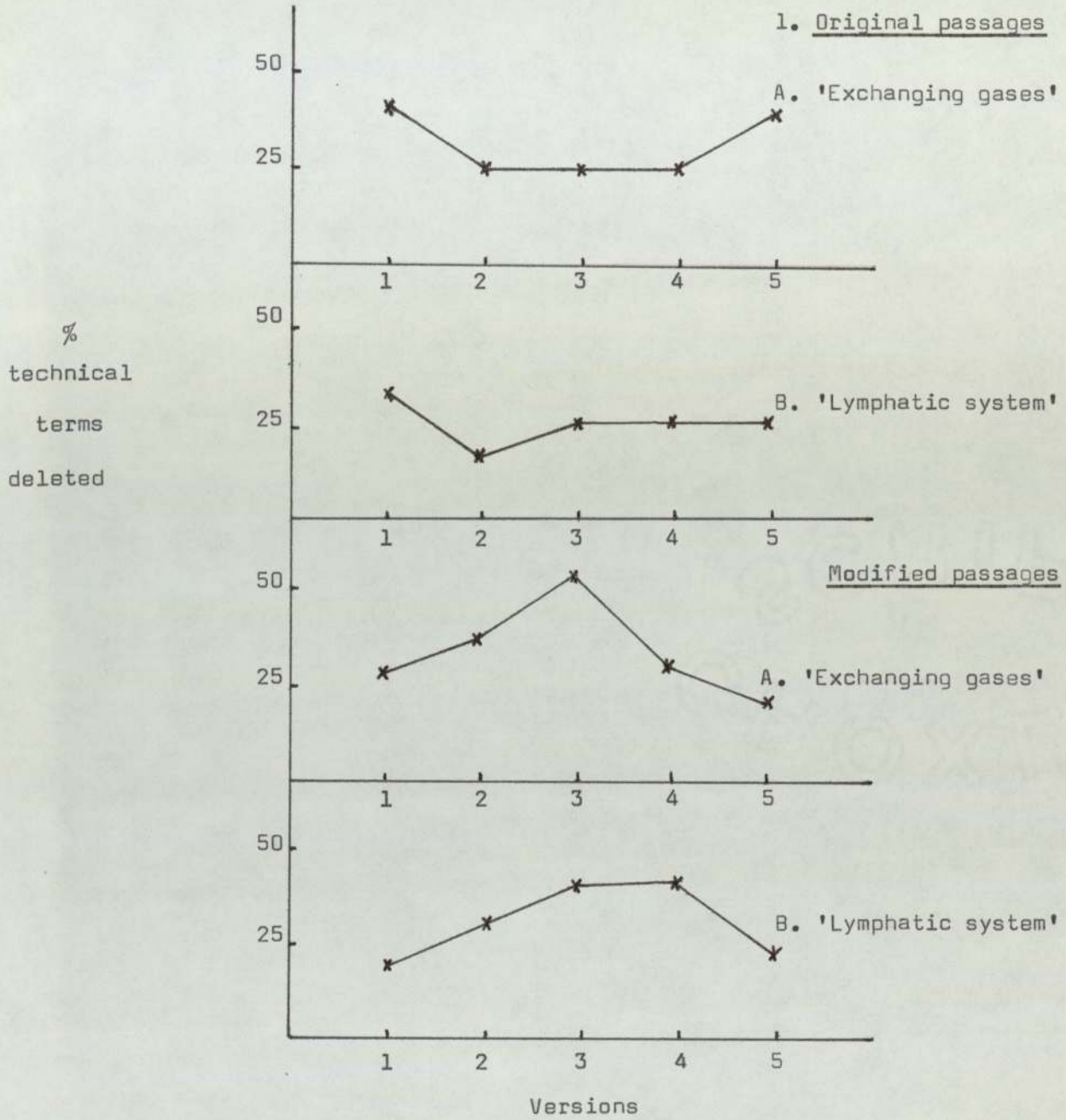
1. The burden of preparing and administering all versions of a cloze passage are considerably greater than for one version. Thus the author was prompted to ask the question: "Would the results have been the same if only one version of each passage had been used? Table 3.8 is a summary of analyses of variance, topics by authors, using values from individual versions. It can be clearly seen from Table 3.8 that the pattern of relationships differs from one version to another. In each case, the version has affected the relative influence of the topic and author on the cloze scores.

Version	Source of Variation	SS	DF	MS	F	Sig of F.
1	Topic	1784.0	1	1784.0	15.5	0.001
	Author	4667.2	1	4667.2	40.5	0.001
	Interaction	314.9	1	314.9	2.7	0.111
2	Topic	62.4	1	62.4	0.25	0.621
	Author	438.4	1	438.4	1.8	0.197
	Interaction	67.9	1	67.9	0.3	0.606
3	Topic	751.9	1	751.9	2.8	0.109
	Author	1140.5	1	1140.5	4.2	0.051
	Interaction	606.4	1	606.4	2.2	0.148
4	Topic	1398.9	1	1398.9	8.5	0.008
	Author	0.01	1	0.01	0.0	0.994
	Interaction	2471.3	1	2471.3	15.0	0.001
5	Topic	1.04	1	1.04	0.01	0.918
	Author	3374.8	1	3374.8	35.1	0.001
	Interaction	2937.6	1	2937.6	30.6	0.001

The implication of these results is that it is essential to use each possible version of a passage in a cloze test. A simple content analysis of the deleted words provides further insight into the reasons for the differences between versions and Figure 3.2 shows the percentage

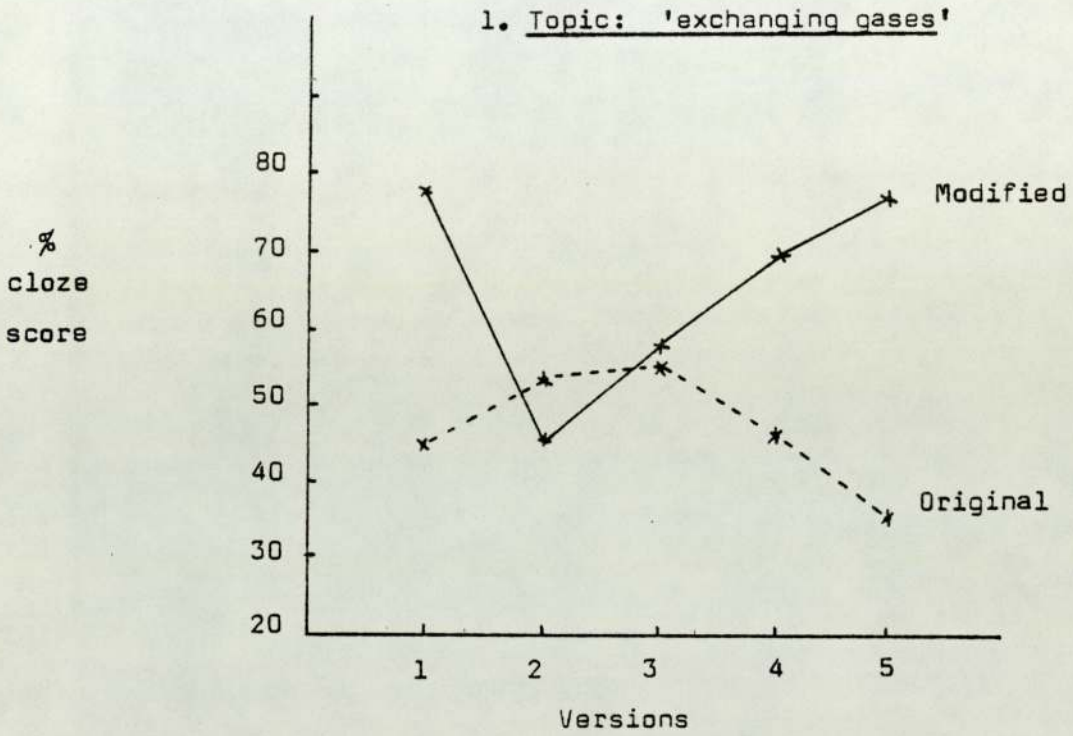
of technical terms that were deleted in each version of a passage. In order to make comparisons, the cloze scores for each version of the four passages are further illustrated in Figure 3.3.

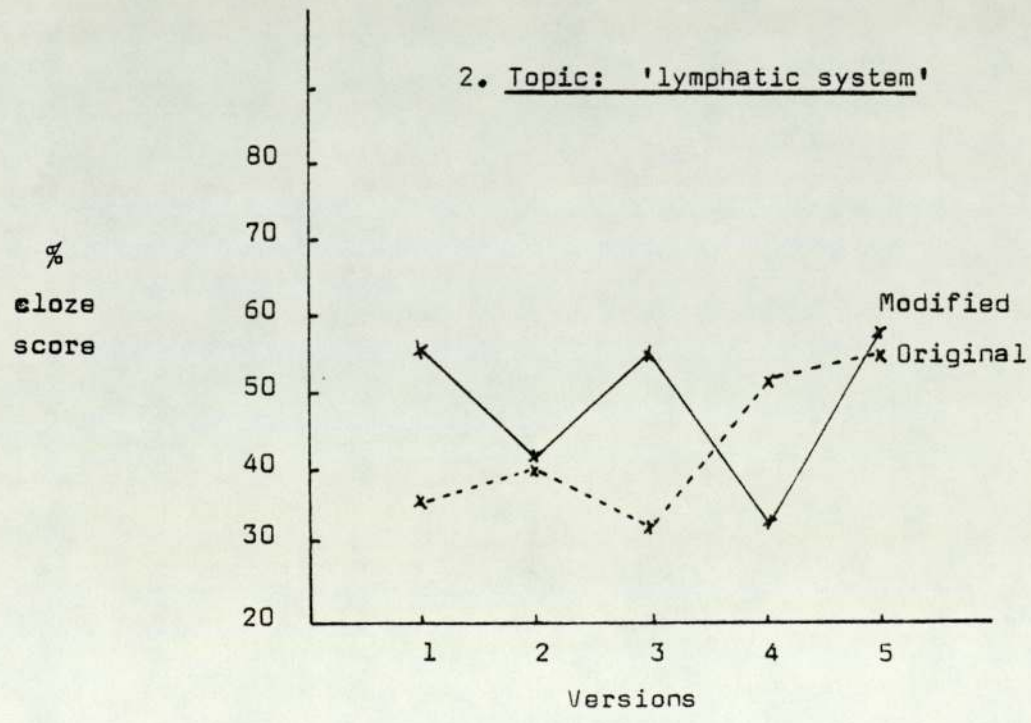
Fig. 3.2 Percentages of technical terms deleted from each version of the four passages.



For example, in Version 3 of the modified passage on exchanging gases, 53% of the deleted words were technical, while in Version 5 the figure was only 20%. Turning to Fig.3.3 it can clearly be seen that Version 3 was found to be more difficult than Version 5 with cloze scores of 57.2% and 76.2% respectively. Obviously, a technical word such as 'bicarbonate' will be harder to supply for a deletion than a non-technical word such as 'the'.

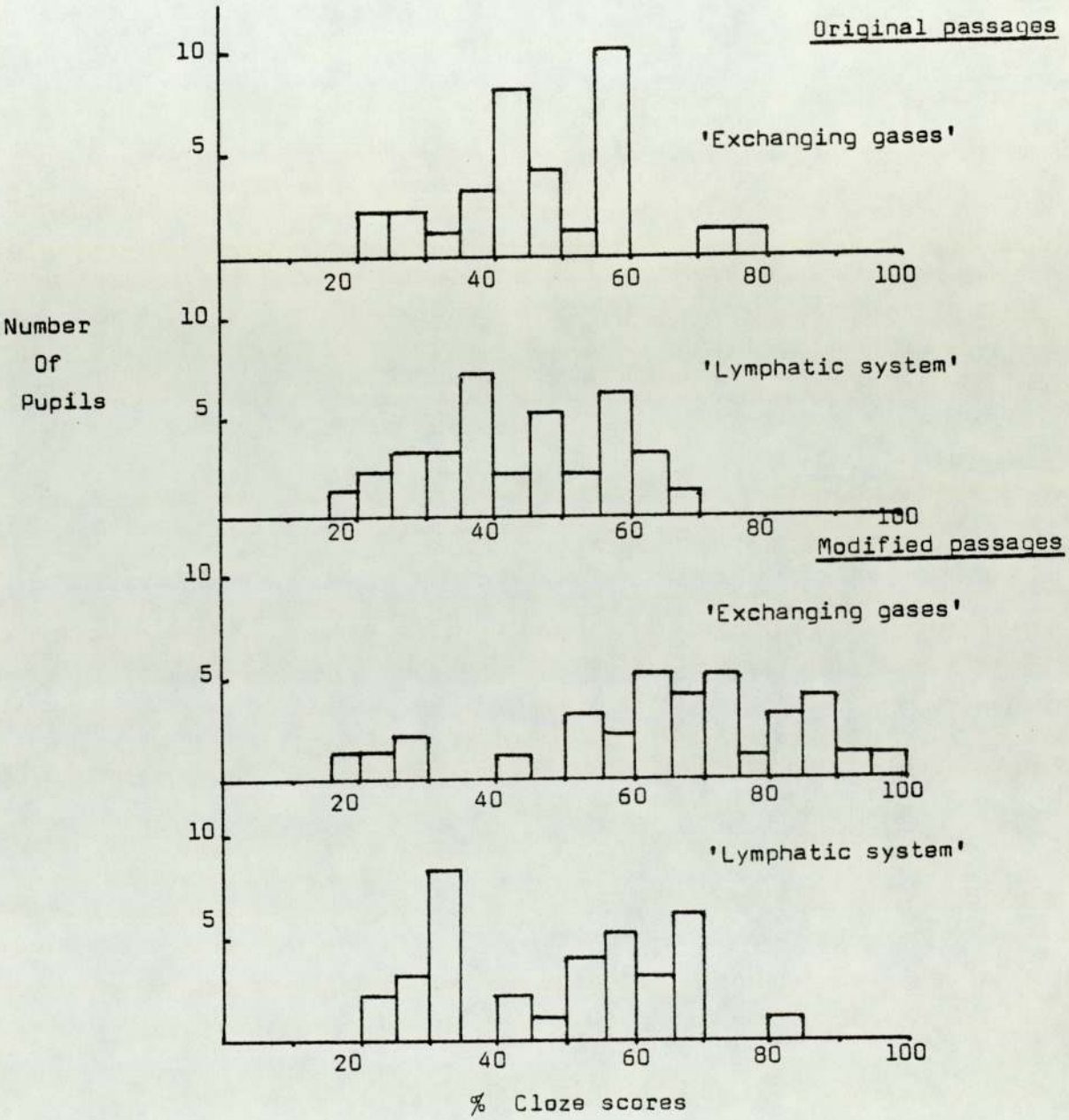
Fig. 3.3 Mean cloze scores for each version of the four passages.





2. On the crucial issue of reliability, the standard deviations for each passage were large in relation to the mean score (Table 3.4). In other words, individual results showed a good deal of variation. (Fig. 3.4).

Fig. 3.4 Distribution of cloze scores.



As can be seen from the graphs there was a wide range of performance. In all cases there was a group of pupils that obtained very low scores which is most marked in the modified passages where the distribution of cloze scores is almost bimodal. These low scores may have been obtained by pupils with very poor comprehension skills but there is an

alternative explanation. During the cloze tests it was observed that some of the subjects showed little enthusiasm towards the task and may have not tried very hard to complete the deletions. Clearly, if only a small group of subjects perform the cloze test it is unlikely that the result will be a reliable one.

Discussion

The cloze tests produced two results. Firstly, the textual modifications made to the two original passages were not equally successful. The modifications made to the passage on 'exchanging gases' significantly increased its cloze score which is a confirmation of the American work by Williams (1968) who analysed science material and rewrote it in a way that increased its readability. However, modifications made to the passage on the 'lymphatic system' did not significantly raise its cloze score. A similar result was found by Greene and Szabo (1979) who used 9 criteria to reduce reading levels in instructional material but found no significant improvement in achievement. Obviously there are factors other than linguistic which have affected the subject's ability to read the 'lymphatic system' with understanding.

The second perplexing result was that there were differences between the cloze scores and reading levels for each passage. Most studies have not found such large differences between the two measures of readability, for example Bormuth (1966) and Miller (1974) found correlations between formula and cloze scores of .8 and .9 respectively. However, there are some studies where the two measures have not been closely correlated (e.g. Graham, 1977; Kintsch and Vipond, 1979; Schief and Wood, 1974). Some of the reasons for these results are as follows:

1. Harrison (1979) suggests that low correlations between cloze and formula scores were the result of an interaction between the interest of a passage and its cloze score. This line of argument is supported by Schief and Wood (1974) who found poor to negative correlations between measures. They used newspaper articles which were found to be more interesting than passages from reading tests or information books. In this study it may be that the topic 'lymphatic system' was found to be less interesting than 'exchanging gases' which resulted in lower cloze scores.

Mobley (1981) in an investigation of school textbooks questions the interpretation of data from cloze procedure and strongly argues the value/subjective assessment. Despite the limitations of asking pupils about the test passages, a questionnaire would help to determine the role of interest in the readability of the test passages.

2. A second factor that may have affected the cloze scores is the level of abstraction for each passage. Abstract writing is rather easily defined as that which contains a relatively high proportion of generalities and concepts far removed from basic life experiences (Gillie, 1957). Gillie's simplified formula for measuring abstraction in writing is derived from the Flesch abstraction formula (1950). It involves a count of 'definite words', that is, finite verbs and definite articles and a count of 'abstract nouns' such as those with the endings -ness, -ment and -ion. Though atheoretical, the results of applying this formula appear to be meaningful if reported in terms of a verbal description, rather than by a numerical score. As is the case with any of the readability formulas, numerical scores often imply an unwarranted degree of precision.

All the test passages were found to qualify as 'fairly concrete'. It appears that the differences between the test passages are not due to surface features of text but must be psychological.

Earlier in the study it was stated that some workers in the field of readability have reservations about cloze procedure (e.g. Carroll, 1972; Carver, 1973; Harrison, 1979; Neville and Pugh, 1974). Criticisms about this technique can be divided into those concerning its reliability and its validity.

1. The reliability of cloze procedure is open to question because of the size of the standard deviations in relation to the mean score. For example, in studies by Bormuth (1967, 1968a) the overall standard deviation was 17% of the mean score. In this experiment, the standard

deviations were about 15% of the mean cloze score. In other words, results on a passage have a good deal of variation and approach both ends of the difficulty range. In the present experiment, individuals were found scoring from 15% up to 95% on the same passage.

2. The validity of cloze procedure as a measure of readability has been questioned by a number of authorities. Weaver and Kingston (1963) performed a factor analytic study which suggests that cloze scores are affected by a special aptitude or ability to utilize redundancy in a passage and supply the missing elements, independent of verbal ability. Coleman and Miller (1968) used the technique to compare subjects with prior knowledge of a passage with those having none. They found that cloze scores were hardly higher than when there was no prior inspection. Carroll (1972) states that it is even possible to secure cloze scores for meaningless material, so long as grammatical cues are present. He concluded that the cloze technique is too crude to permit measuring the degree to which the individual comprehends particular lexical or grammatical cues or possesses a knowledge of specified linguistic rules. Carroll assumes that the ability to supply the correct word for a specific deletion depends instead on inferential processes. Moreover, cloze procedure is totally severed from cognitive psychology concerned with the processes of reading comprehension.

To sum up, if one assumes that cloze procedure is a reliable and valid measure of readability it can be concluded that the test passages present reading difficulties which were not detected by Fry's Graphical Technique. Readability formulae have their value, however, variables such as sentence length and word length are not the only reasons for reading difficulty.

There appear to be two lines of study which may help to identify the variables which have affected the readability of the test passages. First, a questionnaire should help to provide information concerning

the attitude of pupils towards the test passages. Second, a technique based on sound theory should give us some psychological reasons why the passages are hard or easy to read with understanding.

Summary

Two passages were chosen from a textbook by Mackean to represent typical textual material that fourth form 'O' level ability pupils have to read. Two additional passages were produced by modifying the originals in the direction of greater readability. The modifications were the result of applying a set of simple, subjective rules. Five versions of each test passage were prepared using a fifth-word deletion pattern and administered to groups of fourth year male pupils in four secondary grammar schools.

Significant differences between the deletion versions of all four test passages emphasize the methodological problems of relying on single versions of a text when measuring readability. As was predicted, significantly higher mean scores were found on the two modified passages than on the originals. However, it was also found that there were discrepancies between the cloze scores and Fry Graph results. For example, the modified passage on the topic of 'lymphatic system' had lower scores than that on the topic of 'exchanging gases' though these passages were of the same reading level according to Fry graph results.

A number of reasons were put forward to explain the differences between the cloze scores and Fry graph results; for example, the reliability and validity of cloze procedure as a measure of readability is questioned. It is concluded that the levels of interest for each test passage should be determined by a questionnaire and that further aspects of the readability of the test passages should be investigated within the framework of a psychological model of text comprehension.

CHAPTER IV

COMPARING THE READABILITY OF BIOLOGY TEXTS BY SUBJECTIVE REPORT

Introduction

In the preceding chapters, two traditional measures were used to describe the readability of four test passages. However, Fry's graphical method and cloze procedure produced contradictory results and a number of reasons for these discrepancies have been put forward. There is one further approach, that of subjective report, which has traditionally been used to assess the readability of textual material. In spite of the limitations of subjective reports, asking pupils about the test passages provides results which may be tentatively used as a measure of readability. The present study used a questionnaire which included questions on a number of aspects of readability such as linguistic complexity, understanding of the subject matter and interest. Pupils were also required to visualise the main ideas in one of the test passages and represent it graphically as a model.

The chapter begins with a short review of work on subjective report and modelling main ideas as measures of readability. In the second part of this chapter the experiment is described and the implications of the results are discussed.

4.1 Subjective report and modelling main ideas as measures of readability

In the absence of convenient quantitative methods, most teachers are usually obliged to choose books on behalf of others by making subjective estimates. Knowing about their pupils' interests and capabilities the teacher may be able to make accurate and appropriate judgements. However, studies of teacher judgements have shown that individual estimates are often inadequate and can be improved upon by using group estimates (e.g. Harrison, 1979). Moyle (1971) lists two studies involving the grading of books by committees of experienced teachers. The grades given by these panels were much more consistent than by individuals. Yet it must be remembered that these judgements are still based on subjective assumptions and values.

Instead of the teacher making an estimate, a second approach is to ask the pupils about the readability of a text. Classified by Carroll (1972) as a subjective report he considers its reliability to be highly fallible. However, if the honesty and attention of the pupils can be assured, then the subjective report becomes a valid and useful measure of comprehension. The potential of the subjective report is that particular elements which cause difficulty in comprehension can be explored. For example, by varying the nature of a passage it is possible to relate subjective ratings to characteristics such as grammatical complexity and vocabulary (e.g. Peterson, 1954).

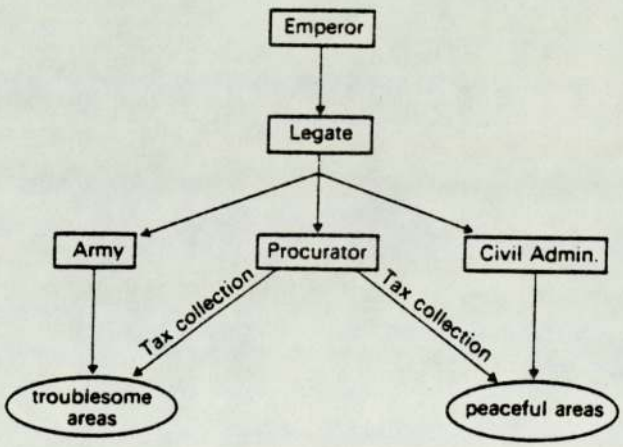
A simple exercise was included in the questionnaire in which pupils were encouraged to try to visualise the main ideas in a passage and represent them as a model. In intensive reading, the ideas which are generated can often be seen, in the mind's eye, as an image. The ideas are translated and organized into a variety of forms such as a picture, chart or flow diagram to suit the interests and purposes of the reader. A model of the ideas in a text is often superior to a prose summary because it provides a more direct representation of the concepts to be comprehended. A more detailed account of the practical aspects

of modelling has been provided by the Open University (1977), Reading Development Course. The course booklet included a particularly useful series of examples of comprehension outcomes which pupils might achieve on texts of various kinds. (Fig. 4.1).

Fig. 4.1 A range of comprehension outcomes.

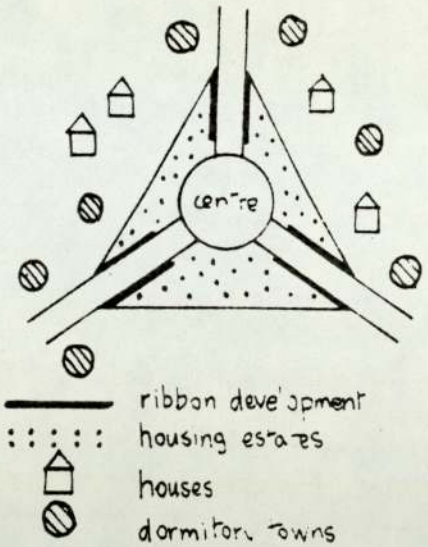
The following illustrations are the kind of models that might be produced for each passage, and are intended to provide you with a basis for comparison. They also contain some errors, which you can look for.

Roman rule



As a province of the Roman Empire, Britain was under the direct rule of the Emperor who appointed a governor or Legate. This Legate was in charge of the military forces and civil administration. Taxes were collected under the authority of procurators who were the Emperor's stewards. The most troublesome part of Britain remained under the control of the army. Other parts were ruled as the Emperor's estates but most of Britain was peaceful and was governed by the local people.
(*In History – Government*, David Smith and Derek Newton, Schofield and Sims Ltd. 1972.)

The changing pattern of settlement after 1900



Since 1900 the tendency to move outwards to the suburbs has led to ribbon development along the main roads and to the growth of housing estates in the wedges formed by the roads. The relatively wealthy have continued to move outwards to houses built in the country or surrounding villages which, through their expansion, have acquired the characteristic appearance of 'dormitory towns'. In this way thousands commute daily into the metropolis of New York from settlements in the Eastern part of Long Island just as thousands travel daily into London from the outskirts of Hertfordshire and Kent (*Introduction to Human Geography*, D. C. Money, University Tutorial Press, 1957)

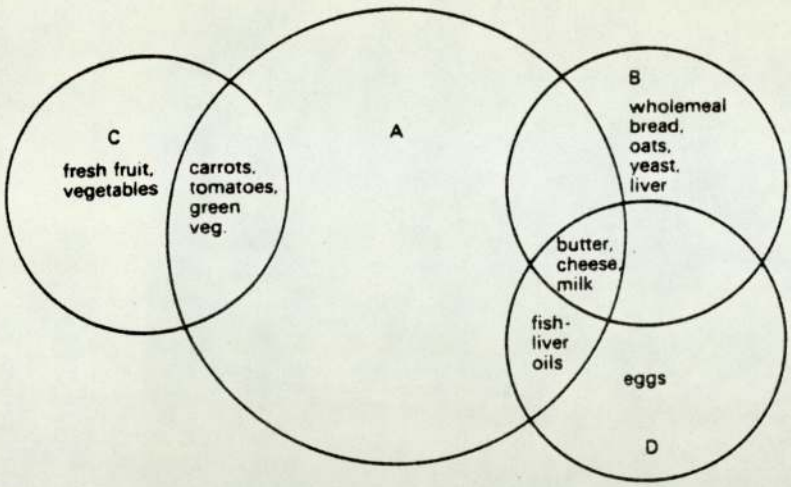
The dairy farmer's year

Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Feed animals indoors			plough	plant	make hay hoe				harvest root crops		put animals indoors

Throughout the day, Dad's many jobs vary somewhat according to the season of the year. The area in which we live is very wet so farmers try to keep their stock off the land from December until March and have to feed them indoors during these months. Ploughing is begun in April. In May kale, cabbage, swedes, turnips, mangolds and grass seed are put in the ploughed fields. During June and July we are busy hay-making, and hoeing the root crops. August and September are fairly quiet months on a dairy farm, but during October and November the root crops are harvested.

(*Living Geography*, Book 1, Mackintosh and Thomson, Holmes McDougall, Edinburgh, 1967)

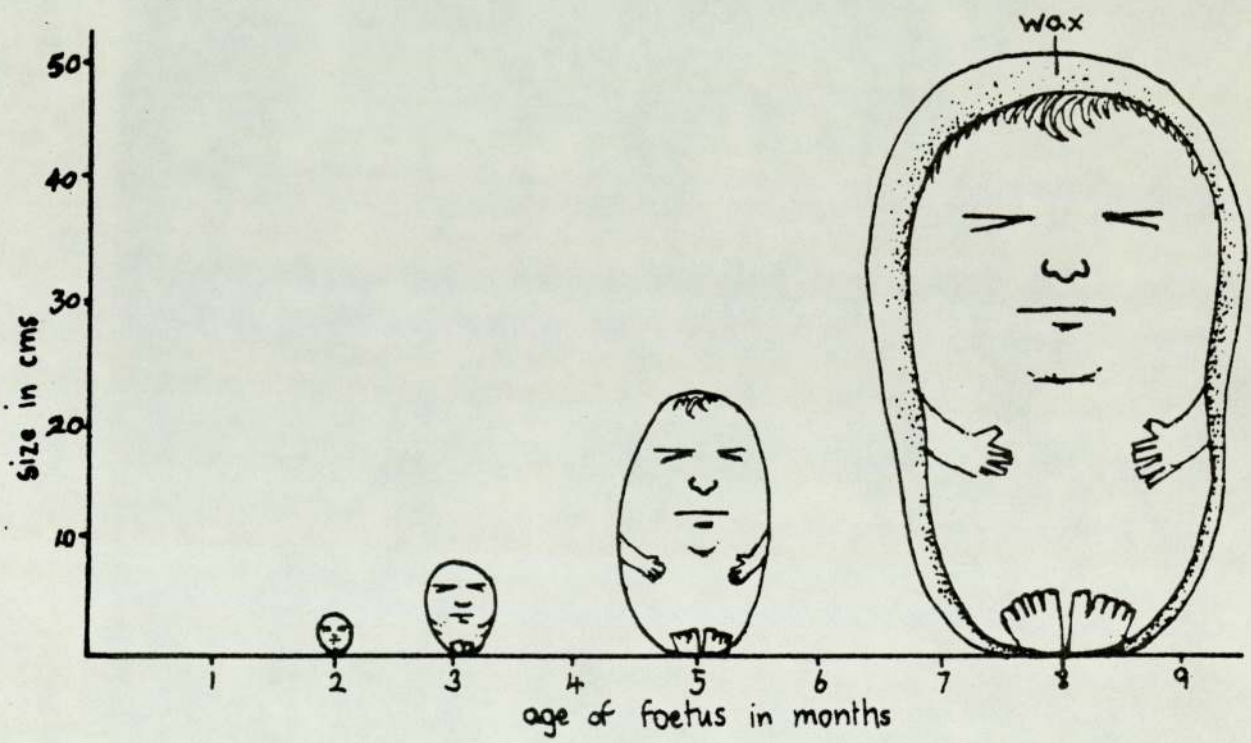
Vitamin sources



Vitamin A is found mainly in fats and the fatty parts of some foods, so plenty of milk and butter will help to provide it. Other valuable sources are fish-liver oils and certain vegetables, especially carrots, tomatoes and dark green leafy vegetables. Vitamin D is also found in butter, cheese, milk and eggs but the richest source is fish-liver oils. Sunlight acting on the skin produces vitamin D in the tissue. Vitamin C is found in fresh fruit and vegetables so plenty of these should be served. In addition orange juice should be given every day. The B vitamins are found in wholemeal bread, oats, yeast, liver and dairy foods. One pint of milk a day will supply all the riboflavin (B_2) which a child under five needs.

(Adapted from *Feeding the under Fives,* issued by the National Dairy Council)

The development of the embryo



Two months old The embryo at two months old is only 2.5 cms. long, has rather a large head – already has tiny ears, fingers, toes and eyelids,

Three months old The foetus is at three months, 9 cms. long and weighs 14 gms. Nails have grown on the fingers and toes, the foetus can move its arms and legs.

Five months old The foetus at five months old is about 25 cms. long and weighs about 300 gms. Some hair has grown on the head, and eyebrows have developed. The baby is ready to be born at nine months. It is about 50 cms. long and weighs about 330 gms. Its head is about one quarter of the body length and has a good coating of hair. The eyes can open, the skin is red and wrinkly and covered with a protective waxy substance.

(Source unknown.)

The result of modelling main ideas is a measure of comprehension and an analysis of the models provides information about the specific difficulties a particular text presents to an individual. Apparent difficulty in modelling the ideas in a passage may indicate that it does not contain all the necessary elements. Clearly, if a pupil cannot represent the ideas in a text as an accurate model, it is likely to be difficult for that pupil to produce a clear comprehension outcome in his or her own mind. A comparison of models for different passages allows one to comment on their relative difficulty.

4.2 The Use of the Questionnaire

Method

Subjects

The requirement was for fourth form pupils taking an 'O' level biology course whose honesty and attention could be on the whole assured. It was thus decided that the most suitable subjects would be pupils at the author's own grammar school. All the pupils knew the author and had been taught by him during the year of this experiment or during the previous year. In all, three classes of male fourth form pupils with a mean age of 14yrs 6 months completed the questionnaire giving a sample size of 78.

Materials

The questionnaire asked the pupils' opinions about three aspects of readability - linguistic complexity; understanding of the subject matter; and interest. They were also required to model the main ideas from passages. The whole questionnaire as presented to the pupils appears in Fig.4.2.

The pupils were assigned one of the four test passages described and used in Chapter III (see Tables 3.2 and 3.3).

Fig. 4.2 The Questionnaire

NAME FORM PASSAGE

Your opinion is required on the passage you have just read. Place a mark on the line that most accurately expresses your feeling.

1. How difficult was the vocabulary of this passage?

many hard words	several hard words	most words clear	very easy words	childish

2. How difficult was the sentence structure?

long hard sentences	some difficult sentences	not aware of hard or easy sentences	some sentences too simple	childish sentences

3. How meaningful were the details?

many confusing details	some details not clear	enough details to make it clear	too few details to explain ideas	no helpful details

4. How well did you understand the main ideas?

main ideas very clear	most ideas clear	could follow the main ideas	had trouble getting the main ideas	confusing not understood

Explain why the main ideas were easy or hard to understand.

5. How interesting was the passage as compared with other biology passages?

good reading	fairly interesting	neither enjoyed or disliked it	dull reading	disliked reading it

Explain why you found the passage interesting or dull.

6. When you read the passage, could you visualise its meaning as a model (diagram)?

a clear impression of a model	fairly clear view	could model the main ideas	had trouble visualising a model	couldn't make a model

On the reverse side of this sheet attempt to draw a model of the main ideas in your passage. You may also include any further comments you have about the passage.

Procedure

The questionnaire was administered to 3 class groups during the last term of the academic year 1979-1980. A short informal introduction was given by the author in which it was explained that he was conducting a survey and that the questions were in no way a test. Instructions were given on how to construct models from the main ideas in a passage using the examples of comprehension outcomes described in Fig.4.1. Any questions arising from the instructions were answered.

Each pupil was allocated at random, one of the four test passages and a copy of the questionnaire. The questionnaires were completed in relaxed conditions, however provision was made to prevent the children looking at their neighbour's answers. Adequate time was given for all pupils to complete their questionnaire. Finally, the printed sheets were collected in to be examined by the author.

Results

The responses of the subjects to the questionnaire can be divided quite conveniently into four sections - linguistic complexity, understanding of the subject matter, interest and the ability to model ideas. Each section will consist of a table summarising the responses to the various questions which are then discussed.

a) Linguistic complexity

Table 4.1 Responses to questions on linguistic complexity.

Question	Passage	% subjects				
		many hard words	several hard words	most words clear	very easy words	childish
1. How difficult was the vocabulary?	AM	-	64	27	9	-
	BM	5	-	63	32	-
	AG	-	-	44	56	-
	BG	-	-	33	67	-
		long hard sentences	some difficult sentences	not aware of hard or easy sentences	some sentences too simple	childish sentences
2. How difficult was the sentence structure?	AM	9	55	36	-	-
	BM	10	50	30	10	-
	AG	-	-	100	-	-
	BG	-	22	33	33	12

A - 'exchanging gases' B - 'lymphatic system'

M - original passage (Mackean) G - modified passage (Graham)

Table 4.1 indicates that the pupils considered the vocabulary of the original texts to be harder than the modified texts and that sentence structure of the original texts was considered to be more difficult that the modified texts. Despite both modified passages having the same reading level, the topic 'lymphatic system' was thought to be more difficult than the topic 'exchanging gases'. It is interesting to note that some of the pupils commented that the short sentences used in the modified passages were too simple and childish. A surprising result was that whilst 5% of

the subjects claimed there were 'many hard words' in the original passage on gaseous exchange (BM), no-one considered there to be 'several hard words'. This result may have been due to confusion over the definitions of 'many' and 'several'.

To sum up, the pupils consider a text to be more readable if:

1) simple vocabulary is used

2) the length of sentences is reduced but not to make the text childish.

b) Understanding of the subject matter

Table 4.2 Response to questions on understanding of the subject matter.

Question	Passage	% subjects				
		many confusing details	some details not clear	enough details to make it clear	too few details	no helpful details
3. How meaningful were the details?	AM	-	71	24	5	-
	BM	-	22	45	33	-
	AG	-	11	89	-	-
	BG	10	40	50	-	-
		main ideas very clear	most ideas clear	could follow main ideas	had trouble getting main ideas	confusing not understood
4. How well did you understand the main ideas?	AM	10	27	27	27	9
	BM	20	20	30	20	10
	AG	50	39	11	-	-
	BG	50	17	22	11	-

A - 'exchanging gases' B - 'lymphatic system'

M - original passage (Mackean) G - modified passage (Graham)

Many more pupils appear to have found the details less meaningful in the original texts in comparison with the modified passages. However, Q.3 was poorly constructed and many subjects indicated that they were not too sure how to respond. Q.4 was more successful and responses indicate that the main ideas were better understood in the modified passages. Again, the topic 'lymphatic system' appears to be slightly more difficult than 'exchanging gases'.

A number of features were identified by the pupils which they considered to help or hinder their understanding of the main ideas. Typical comments were:

1. Original passages.

'The rather long sentences seemed to stretch your mind and memory.'

'I could not understand all the different words.'

'.....there was too much information in the sentences.....'

2. Modified passages.

'.....the sentences were not too long.'

'Easy to understand because it explained them in easy stages and familiar words.'

The most helpful subjects were those with an average reading ability because the superior readers identified few difficulties. For example, one able pupil made the following comment about the original passage on the 'lymphatic system' (BM):

'Most sentences were clear and not too long.'

A final interesting point is that the subjects considered the textual material to be at fault rather than their own poor reading skills.

To sum up, the pupils considered the subject matter to be easier to understand if:

1) terms are explained
2) the sentences are short and do not contain too much information
3) the main ideas are logically presented.

c) Interest

Table 4.3 Responses to the questions on interest.

Question	Passage	% subjects				
		good reading	fairly interesting	neither enjoyed or disliked it	dull reading	disliked reading it
5. How interesting was the passage?	AM	-	45	15	15	25
	BM	-	-	60	10	30
	AG	11	61	17	11	-
	BG	-	-	72	28	-

A - 'exchanging gases' B - 'lymphatic system'

M - original passage (Mackean) G - modified passage (Graham)

There was a marked difference in interest between the two topics. The topic 'lymphatic system' was considered neutral or dull evoking the typical response:

'It was rather dull as it was useless information to know.'

The topic 'gaseous exchange' generated more interest, probably because it was considered to be relevant to themselves:

'.....it is what is happening in our body.'

There were also differences in interest between the original and modified passages. For most pupils the modified passages were interesting because:

'.....it was well explained in easy terms and therefore held interest.'

However, the most able pupils again found the modified passages to be:

'.....childish and thus less interesting.'

To sum up, most pupils considered a passage to be interesting if:

1) the topic is relevant to themselves

2) the language is not too complex

3) style and presentation is not formal.

d) Ability to model ideas

Table 4.4 Responses to question on modelling.

Question	Passage	% subjects				
		clear impression	fairly clear	model main ideas	trouble visualising	couldn't make a model
6. Could you visualise its meaning	AM	-	-	41	32	27
	BM	-	33	11	45	11
	AG	22	44	12	22	-
	BG	11	45	33	11	-

A - 'exchanging gases' B - 'lymphatic system'

M - original passage (Mackean) G - modified passage (Graham)

Nearly all the pupils had difficulty visualising the main ideas in the original passage on 'gaseous exchange' and slightly less trouble with the topic 'lymphatic system' (Table 4.5). In comparison, the pupils considered the ideas presented in the modified passages to be easier to visualise.

Table 4.4 gives the percentages of pupils that constructed models which successfully represented the main ideas in a passage, while examples of these models can be seen in Fig.4.2 a,b,c and d.

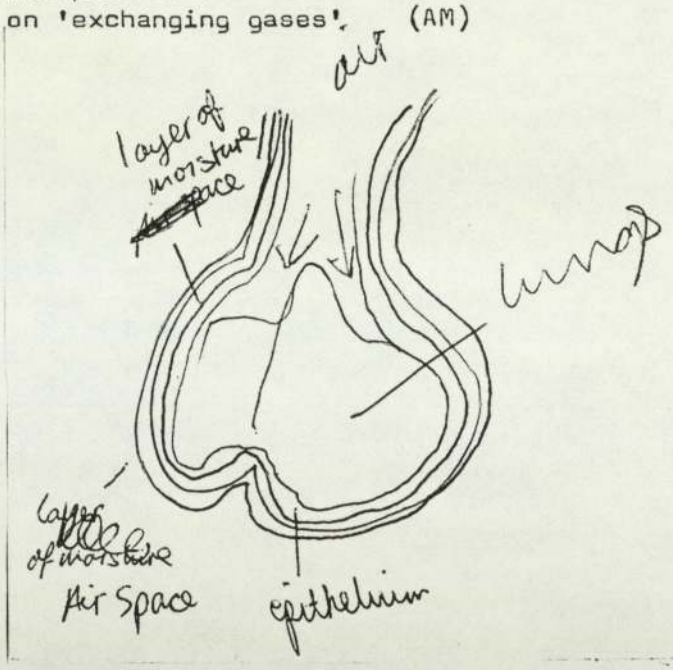
Table 4.5 Percentage of pupils successfully constructing models.

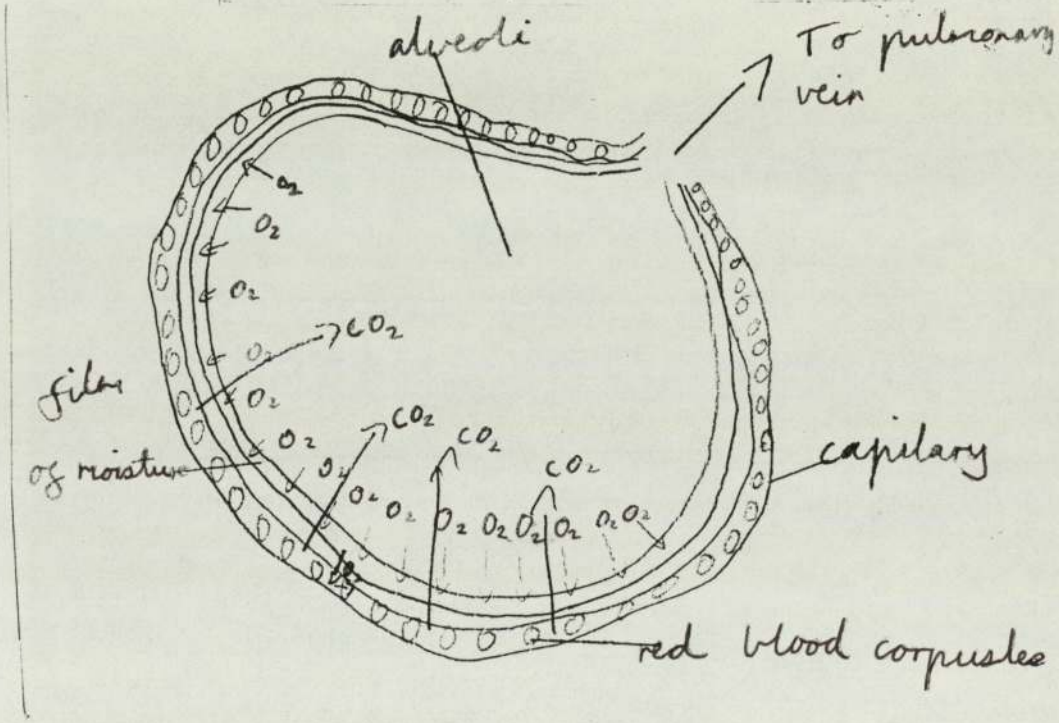
Passage	% subjects
AM	63
BM	70
AG	89
BG	78

A - 'exchanging gases' B - 'lymphatic system'

M - original passage (Mackean) G - modified passage (Graham)

Fig.4.3a Examples of the models constructed for the original passage on 'exchanging gases'. (AM)





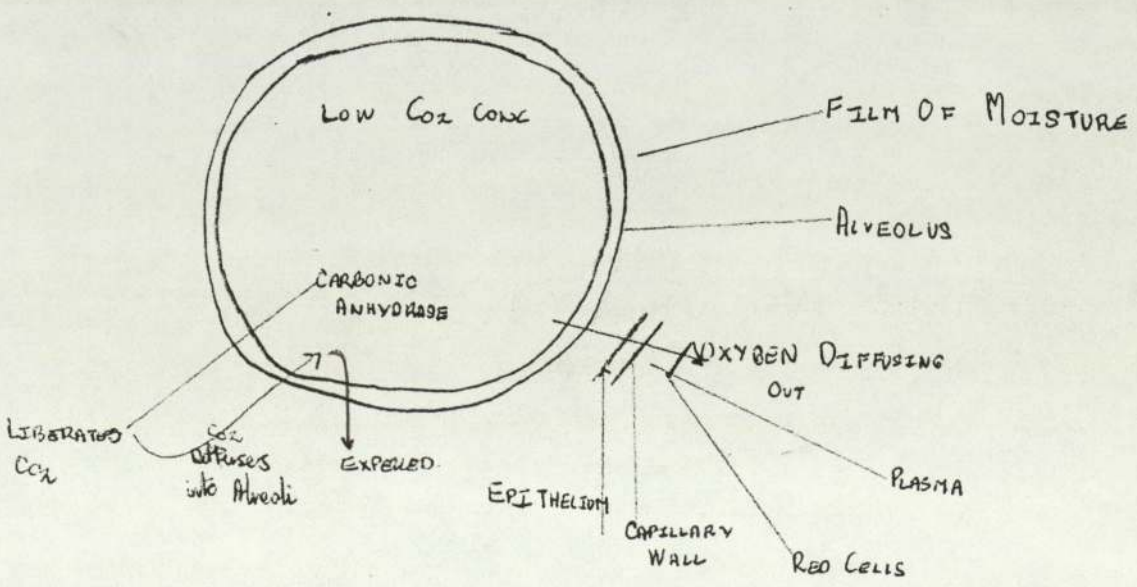
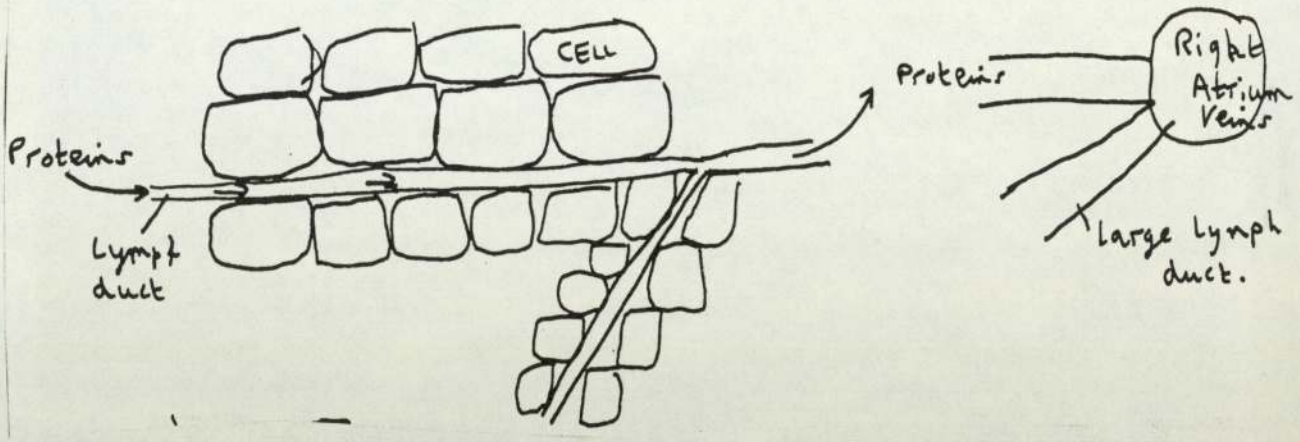
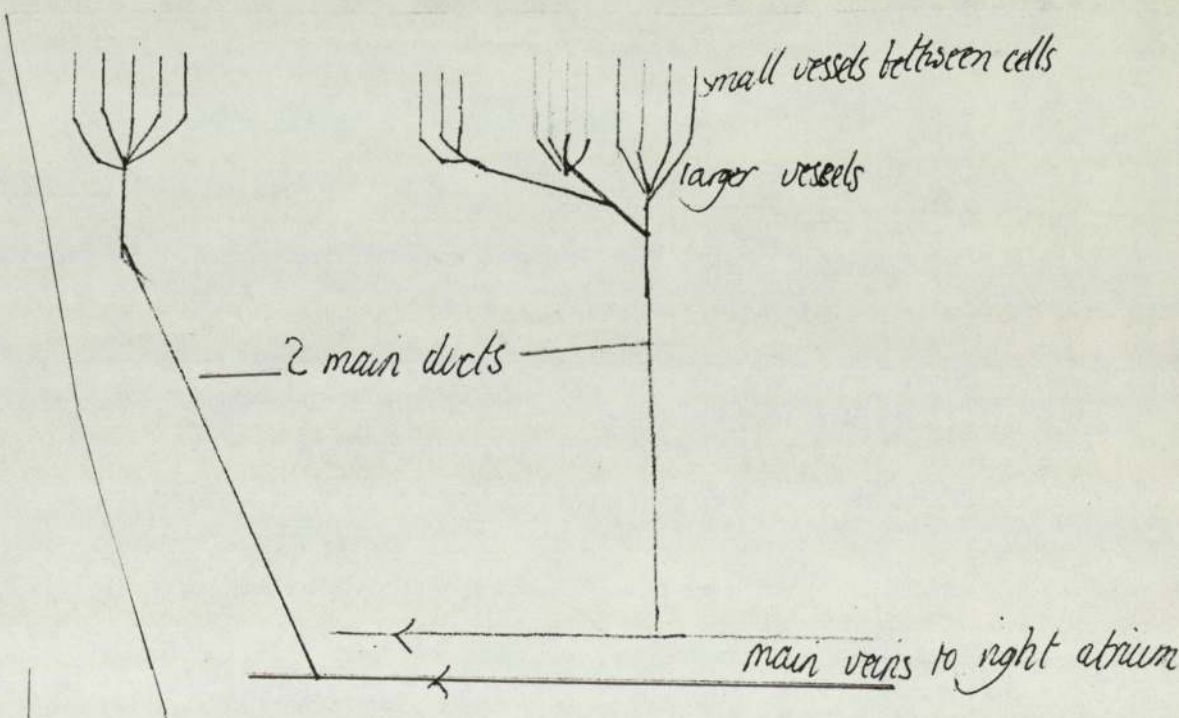


Fig.4.3b. Examples of the models constructed for the original passage on the 'lymphatic system' (BM)





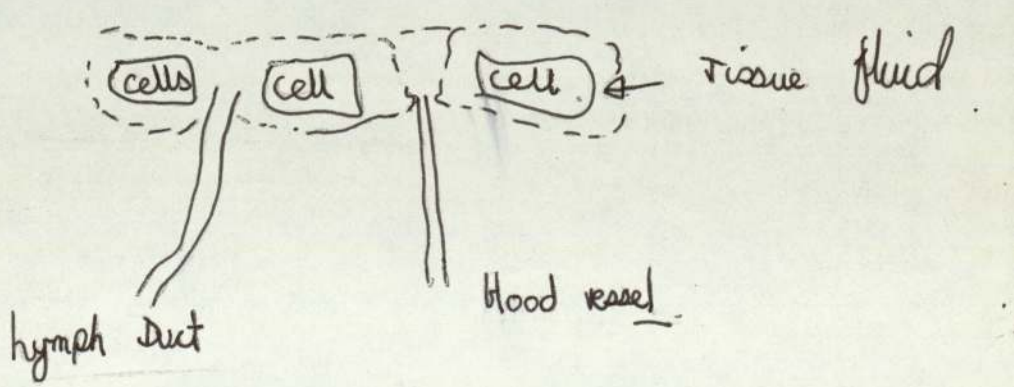
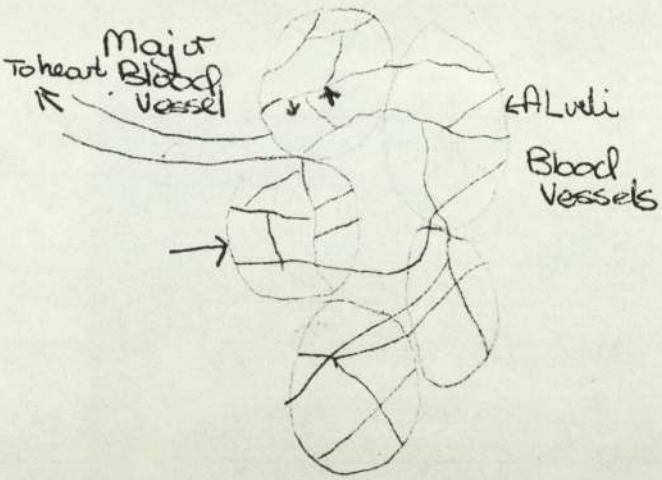


Fig.4.3c. Examples of the models constructed for the modified passage on 'exchanging gases' (AG)



↑ Direction of Air movement from sacs to blood vessel.

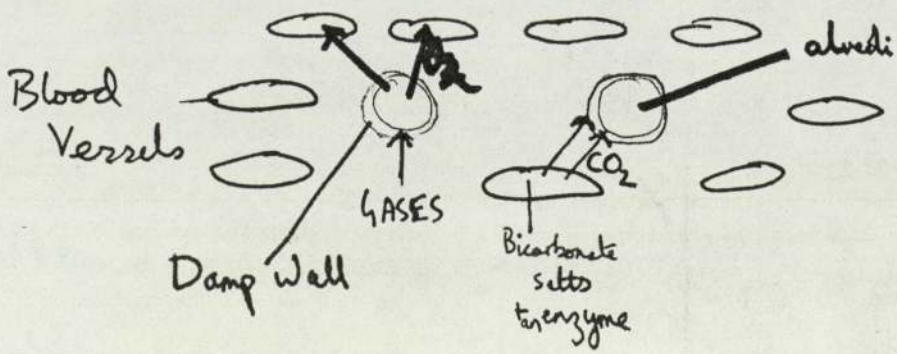
↑ main movement of oxygen into alveli

↑ movement of Carbon Dioxide into sacs from blood.

lungs → inside alveoli → gas exchange

↓

oxygen moves into blood vessels



bicarbonate salts around air sacs

↓

enzymes break down salts.

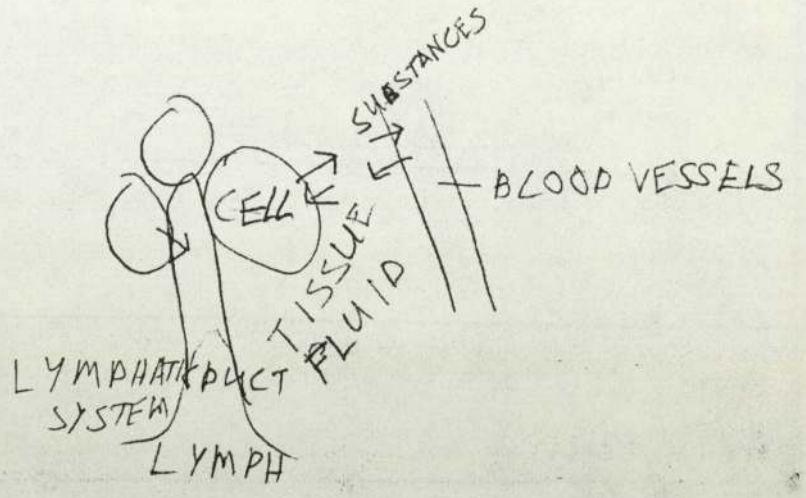
↓

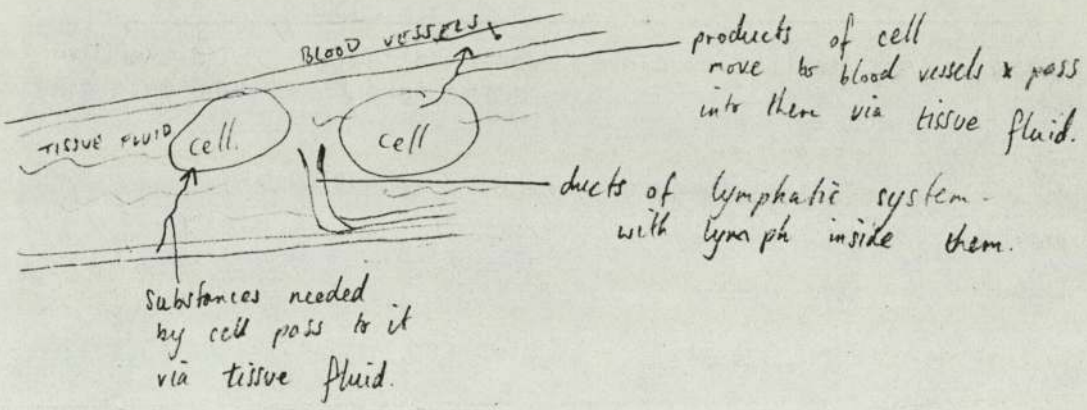
CO_2

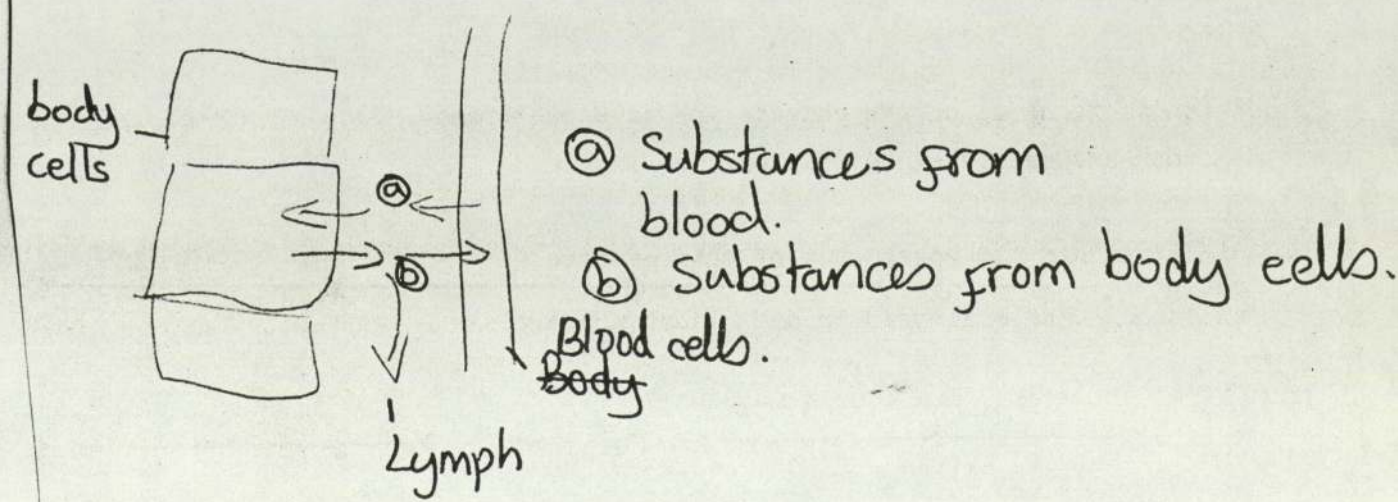
↓

moves into air sacs

Fig.4.3d. Examples of the models constructed for the modified passage on the 'lymphatic system' (BG)







Overall, the quality of the models was poor and the best representations were constructed by the more able pupils. If we first consider the differences in success for the original and modified passages it is clear that textual modifications helped the pupils to construct a model. If we secondly consider the differences in the models constructed for the two topics, it can be seen that 'exchanging gases' was easier to represent than the 'lymphatic system'. 'Exchanging gases' was represented basically as a two-way movement of oxygen and carbon dioxide across the walls of an air-sac (alveolus). (see Fig. 4.2 a and c). In contrast, the 'lymphatic system' is a cycle of events, so most pupils only modelled one part of the system. The diagrams also tended to lack detail and contain many errors (see Fig. 4.2 b and d).

Discussion

The questionnaire provided valuable information about the attitudes of pupils towards the test passages. It was clear that the pupils on the whole favoured the simpler modified passages. Klare (1963) found that preferences for reading material were governed by the simplicity of the text. He also reported that the students generally read more of simpler texts than those which are harder.

The answers to the questions on interest indicated that there were differences in attitude towards the test passages. Most pupils found the original passages difficult and thus lost interest in them but there were some pupils who considered the modified passages too childish and also lost interest. These results have implications for the classroom because the teacher has to take into account not only the text but also the particular individual.

Differences in interest were also found for the two topics. The 'lymphatic system' was not considered to be directly concerned with themselves and thus uninteresting. In comparison, 'exchanging gases' was thought to have interest value because they could relate to the topic. It is generally acknowledged that pupils favour the human element of biology and are more willing to participate in classwork.

These findings may help to explain the differences between reading levels and cloze scores for the test passages (see Chapter 3). The Fry Reading Graph uses a limited number of variables to determine a reading level while cloze procedure reflects the sum total of all influences which interact to affect readability. Unfortunately cloze procedure does not tell us which factors have had a major influence on the subject. By asking the subjects about the test passages it was found that interest was an important factor. It is suggested that the less interesting topic 'lymphatic system' reduced the pupils' performance in the cloze tests while the high interest topic 'exchanging gases' promoted student comprehension which raised their cloze scores. Research on

interest suggests that students comprehend better when they read passages that discuss topics of high interest to them (Estes and Vaughn, 1973). However, like most previous studies on interest there is the problem of prior experience. It may be that the pupils knew more about 'exchanging gases' before the present research and thus indicated greater interest in it.

Interest has played a part in the comprehension of the test passages and this has implications for the classroom. If the teacher's goal is to get students to understand seemingly uninteresting material, he or she can try to generate interest through a variety of means - such as movies or field trips related to the topic. However, it must be pointed out that while such activities promote enjoyment, it does not necessarily help students to understand a topic for which they do not possess intrinsic interest. Another approach is to present material with a level of difficulty which minimises the amount of effort necessary to obtain a certain goal.

Chall (1958) has collected evidence from various sources and reached some helpful, though possibly anticipated, conclusions. She quotes a survey by Gray and Leary (1935) in which the content of the book was found to be the most important factor, with style of writing, format and organization also affecting interest in decreasing order of importance. Chall reports a study by Strang in which high school and college students were asked to list factors which they thought made books readable. Stylistic factors, such as 'plain everyday English' or 'easy simple vocabulary'and 'short paragraphs and sentences' were found to be most common. Content and format were ranked second and third most important respectively. Chall also reports a study by Engelmans in which children were found to prefer a conversational style of writing to a narrative expository style. The format was found to be read more quickly when reading speed tests were given.

The findings of such studies are useful in providing general bases

on which to judge the readability of books. However, Gilliland (1972) states that the subjectivity of interest and the lack of direct control over it reduces the usefulness of the above conclusions and it is to more precisely measurable aspects of books that attention must be directed.

The overall ability of the pupils to graphically model comprehension outcomes from the test passages was poor. This was surprising because biology uses many forms of illustration to represent subject matter. Modern maths provides the basis for the development of greater proficiency in the use of logic which should help pupils to convert the arguments in a text into a form that makes the relationships clear. It is suggested that comprehension outcomes in the form of graphic models could enrich the learning experiences of a child. What is required is realistic and appropriate training whcih can be incorporated into the curriculum.

The failure of the children to effectively model the main ideas in the test passages may have been due to:

(1) inexperience in constructing models

(2) poor draftsmanship

(3) low motivation towards the task

(4) lack of comprehension because of the nature of the text.

Keeping in mind the first three factors, the final factor provides us with a crude measure of text comprehension. A comparison of the quality of the pupil's outcomes for each of the test passages offers interesting possibilities.

A comparison of the models constructed for the original and modified passages indicates that the latter helped pupils to reach a satisfactory comprehension outcome. An assessment of the models constructed for the two topics suggests that pupils found it harder to graphically represent the 'lymphatic system'. Possibly this topic does not lend itself to modelling because it is hard to illustrate a complex and dynamic set of ideas. If the 'lymphatic system' is hard to model it is likely that a reader will find difficulty representing the ideas

as a comprehension outcome in his or her mind.

To sum up, the questionnaire was of an exploratory nature only. It was not intended for use with a large stratified sample, examining the responses to different questions in detail and applying statistical analyses. However, in spite of its limitations, the questionnaire did at least help to answer some of the questions raised earlier in the study. Modelling the main ideas graphically is a little explored area that seems to offer interesting possibilities and it is suggested that more research time could be profitably spent on this work.

The questionnaire and modelling exercise have helped in making a comprehensive assessment of the readability of four test passages. It is important here to emphasize the value of this more general and less formal method when used in conjunction with other measures. However, all these traditional techniques have been shown to have faults. To effectively evaluate the readability of a text new methods are required which have been designed within the context of an explicit model of the processes involved in text comprehension.

Summary

A questionnaire and a modelling exercise were designed to complement the readability results obtained by wholly objective techniques. The questionnaire asked the pupils about linguistic features of the text, interest and the ability to graphically model a comprehension outcome.

Results confirm that features of the text such as sentence length make a passage less readable. Interest was considered to be greater for the topic 'gaseous exchange' than for the 'lymphatic system'. Reasons for this difference in interest were put forward and it was suggested that this factor may have affected scores obtained for the test passages when cloze procedure was used. The ability to model a comprehension outcome was poor due to inexperience with the nature of the task. A comparison of models constructed for the test passages indicated that there was greater failure with the original passages and with the topic on the 'lymphatic system'. It was suggested that the 'lymphatic system' was difficult to represent as a model because it is a complex and dynamic set of ideas.

CHAPTER V

AN APPROACH TO READABILITY BASED ON A PSYCHOLOGICAL MODEL OF TEXT COMPREHENSION

Introduction

Existing research has produced techniques for measuring readability but there is little understanding of what has been measured. The traditional approach is an atheoretical one and relies on correlating plausible text properties with plausible criteria of readability. The trouble is that typically, neither set of variables is adequate.

Chapter two discussed the problems associated with readability formulas which are concerned with word and sentence properties at a superficial level, and so do not directly reflect either the content or the organization of a text. Reading comprehension is ultimately a process of acquiring information and the nature and structure of that information - that is, the characteristics of the meaning of a text - are the real determinants of readability. It is not doubted that long sentences, unfamiliar words and abstractness produce reading difficulties; but a deeper understanding of reading processes should identify other important variables. A more complex approach to text structure and text processing requires a model of comprehension.

Cloze procedure was used in Chapt. three which produced results that did not correlate with results obtained by a readability formula. Although cloze procedure offers an attractive 'window' into the response of a reader towards a text, the use of cloze tests as a measure of readability still requires a fair amount of fundamental research. J.B. Carroll (1972) in a seminal survey emphasised that there remains a good deal of doubt about what precisely cloze procedure is testing.

Chapter four described the use of a questionnaire technique to assess the readability of the test passages. The results of this survey helped to indicate which factors the pupils judged to be important determinants

of textual difficulty. The section on graphically representing comprehension outcomes provided valuable information and appears to be a relatively unexplored avenue of research. However, on the crucial issue of the reliability of subjective judgements, too much reliance on people's intuitions is risky and more direct procedures should be preferred.

This chapter begins with a look at some of the psychological models of comprehension that are now available, and then more specifically a model devised by Kintsch (1974) is described. The chapter continues with an attempt to apply measures of readability based on aspects of the model by Kintsch to the same four test passages used in previous chapters. Results are presented, analysed and their implications discussed.

5.1 The Nature of Comprehension

Psychologists, primarily in an attempt to explain human memory, have contributed greatly to our understanding of the comprehension process. Reading research appears to have shifted away from an emphasis on decoding and methods of teaching reading towards an emphasis on understanding how readers comprehend and how to help students comprehend better (National Institute of Education, 1976). The major goal of this section is to illustrate the ways in which the structure of a text can be represented as an outline model of comprehension.

The plan for this section consists of two parts. First, the way in which people organize their minds to relate old and new concepts (simple ideas) will be discussed. For the sake of convenience, it seemed appropriate to use the words "concept" and "word" interchangeably, recognizing of course, that concepts are meanings associated with the surface form of the word as it arises in print or speech. Secondly, there will be a jump from understanding concepts to understanding longer units of discourse such as propositions, sentences and stories.

Understanding concepts

Given the word "dog", the following words may have come to mind: cat, bark, pet, Rover (or any dog's name), alsation (or any breed's name), bite, animal.

Each of these word associations represents a predictable type of relation to the stimulus word "dog". For example:

1. The word "animal" is the class of things to which dogs belong. Class relations imply that the stimulus concept belongs to the class of things denoted by the associative response.
2. "Rover" is a particular individual while "alsation" is a type of dog. Both represent examples of dogs and are thus designated example relations.

3. "Bark" and "bite" are two attributes of dogs and are thus called property relations.

Relationships such as these have been graphically portrayed in what is referred to as a semantic network by Collins and Quillian (1969) and Lindsay and Norman (1972) amongst others. Semantic maps capture the character of intuitive notions about the similarities and differences among concepts.

The construction of a semantic map can be approached by giving students a word association test in which they are asked to write down the words brought to mind by each of the concepts being studied. The list of words produced as responses to a given concept constitute the associative meaning of the stimulus word for that individual. The closeness in meaning of two stimulus words can then be measured by the two response lists. In this way it is possible to discover relationships among concepts. Physics has been a popular area of science to explore, using concepts from mechanics (e.g. force, energy, time) and electromagnetism (e.g. current, resistance). (E.g. Preece, 1976; Shavelson, 1974).

Using word association tasks can also give a quick diagnostic picture of what pupils already know and need to know about a set of concepts. The unfamiliar concepts can then be approached through the known related concepts which the pupils do possess.

Semantic maps have proved to be a useful construct for illustrating certain basic relations among concepts. However, class, example and attribute relations cannot be used to express all relations among concepts; at present, the relations among events or episodes that occur cannot be represented. Since people understand and remember events or episodes, semantic maps will be incomplete until a way is found to represent them. Fortunately, several researchers have developed models which attempt to take into account the relationships between events. (E.g. Fillmore, 1968; Frederiksen, 1975; Lindsay and Norman, 1972).

Understanding longer discourse

Meyer and McConkie (1976) approached the problem by dividing passages into "idea units", that is, they represented concepts as broadly defined units; each "idea unit" was felt to be a meaningful piece of information. Their organization and utilization of passages though reliable, was intuitive. In fact, the authors later found that their organization agreed well with a scheme of discourse analysis being developed by the linguist Grimes. The semantic grammar produced by Grimes has a tree structure whose modes contain content words from the prose passage. Lines among the nodes show how the content is organized and labels in the tree structure classify the relationships among modes.

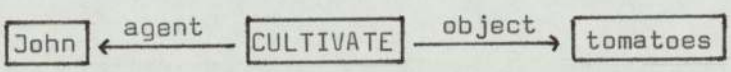
Lindsay and Norman (1972) have utilized Fillmore's (1968) case grammar to expand the number and kind of relations that can exist in a semantic network. In a case grammar analysis of an event, step one is to identify the basic action; step two is to identify the agent - who caused the action to occur; and step three is to identify the object - what was affected? An example may help to clarify these relations.

John is cultivating tomatoes.

Action: cultivating

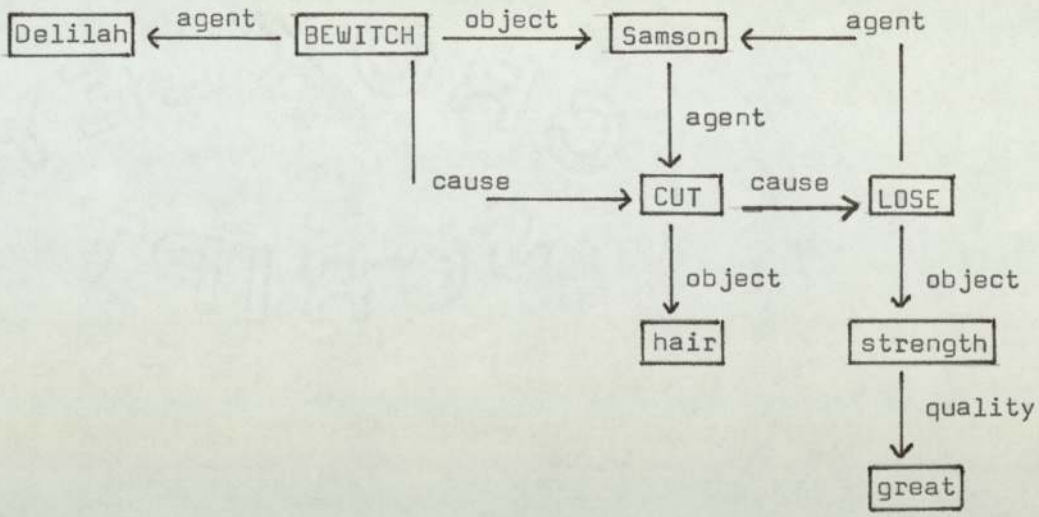
Agent: John

Object: tomatoes



The action has become the focal point around which all other concepts in the event revolve. Fillmore (1968) and Lindsay and Norman (1972) have added several other possible relations among parts of an event in order to account for normal discourse. For example, consider the relations which are components of the following sentence (after Pearson and Johnson, 1978).

Because Samson was bewitched by Delilah, he cut his hair and lost his great strength.



Notice that there are three basic actions - "BEWITCH", "CUT" and "LOSE" which are linked by causal relations. Notice also that Samson is the object of one of the actions but the agent of the other two.

Each of the parts of the sentence - "BEWITCH", "CUT" and "LOSE" - plus the associated agents and objects, and so on, is a proposition. A proposition can be thought of as a unitary statement about the world. In simpler terms, a proposition can be thought of as being clauses (dependent or independent).

Understanding the relations among propositions appears to be critical to good comprehension. In empirical studies, Kintsch (1974) and Meyer (1977) have provided evidence to support the conclusion that propositions are basic units of thought while the relations among propositions such as cause, purpose, condition and time, carry the thread of a story or passage in a text.

To review briefly: networks of modes (representing concepts) and links (representing relations) can be used to illustrate the relations among concepts in a proposition and the relations between propositions. To quote Lindsay and Norman, (1972, p.400-401):

> "We now have the basic design for the data base underlying human memory. The memory system is an organized collection of pathways that specify possible routes through the data base. Retrieving information from such a memory is going to be like running a maze. Starting off at a given node, there are many possible options available about the possible pathways to follow. Taking one of these paths leads to a series of crossroads, each going off to a different concept. Each new crossroad is like a brand new maze, with a new set of choice points and a new set of pathways to follow. In principle, it is possible to start at any point in the data base and, by taking the right sequence of turns through successive mazes, end up at any other point. Thus, in the memory system all information is interconnected."

A number of models are now available which attempt to represent text structure. However, a model first proposed by Kintsch (1974) was chosen because it suggests a number of text and reader characteristics which may help to determine readability. The theory described in the next section will help one to understand the implications of his model for the concept of readability.

5.2 The Kintsch model for the representation of meaning

The model of text comprehension first proposed by Kintsch (1974) assumes that the basic units of meaning are propositions. Propositions consist of a number of word concepts, one of which serves as a predicate while the remaining ones act as arguments. The predicate specifies a relationship among the arguments in a proposition. Kintsch describes the ordered list of propositions as being the text base which represents the underlying meaning of the passage. A full description and justification of the theoretical system is given by Kintsch and van Dijk (1978).

It is not a very elegant procedure and it certainly would not be suitable if one was concerned with the finer points of logic or semantic analysis. However, it is simple and relatively robust and in the present study the procedure was simply accepted as it is. What is gained by analyzing paragraphs in this way? The reason is that two texts of equal length are not always equally easy or hard to read even when characteristics of the words and word sequences are controlled (e.g. same reading level). It appears that there are also properties of the text base which have to be taken into account.

First, texts may be comparable in word length, but they may differ in the number of propositions expressed by these words. Kintsch and Keenan (1973) have systematically varied the number of propositions in a text base while keeping constant the number of words in the text. They observed that reading time increased regularly as a function of the number of propositions in the text base. Reading time depends not so much on the number of propositions that are expressed in a text but rather on the number of propositions that a reader actually derives from the text. Indeed, for the range explored by Kintsch and Keenan, each proposition processed sufficiently so that it could be produced on an immediate recall test added about 1.5 seconds to the reading time. The longer the texts were, the more time was required for the processing of each additional

proposition. Furthermore, scientific prose required more processing time per proposition than simple narratives. Obviously, the number of propositions in a text base is not the only determinant of processing difficulty, but it is an important one.

A second property of the text base which affects readability was identified by Kintsch et al (1975). It is the number of different arguments that are used in a text base. Texts were constructed which had equal numbers of words and propositions; but in some passages the same concepts were repeated whereas in others new concepts were continuously introduced. Thus, the first kind of text had few arguments, whereas the latter had many arguments. Kintsch et al found the more different arguments were used in a text base the longer it took subjects to read the passage for each proposition recalled. Thus, a basic difficulty encountered in comprehension appears to be establishing new concepts, whereas repeating an established concept appears to involve simpler operations.

So far the text base has been described as a list of unrelated propositions. Clearly, this is not the case. In terms of the processing model proposed by Kintsch (1974), the text base provides a framework for the construction of a graph which describes the interactions within the text. Kintsch and Vipond (1979) tentatively suggest that properties of this graph structure may also predict reading difficulty. However, these properties have not yet been proven to affect readability and it is suggested that they may be a promising area for future research.

Thus, two text base statistics have been shown to affect readability: the propositional density of a text and the number of new concepts per proposition. If comprehension really involves the construction of a text base as proposed in this processing model, these two properties should be able to predict the level of difficulty for a text. However, readability is not just an inherent property of texts but is the result of the interaction between a particular text and particular readers.

Kintsch and Vipond (1979) suggest that normally, comprehension makes low demands on a person's resources. Sometimes, however, the normal processes of comprehension are blocked and the reader has to carry out resource-consuming operations before understanding the information. Kintsch and Vipond assume that each one of these extra operations disrupts the automatic comprehension process and adds to the difficulty of reading. Texts requiring many operations that make high demands on resources, should yield either increased reading times or low scores on comprehension tests. Comprehension of a text could therefore be evaluated in a way that considers comprehension time and test performance because of the trade-off between the two. For example, Jackson and McCelland (1975) used reading speed adjusted for comprehension as a measure of comprehension difficulty. Thus an index which combines reading time and amount of recall should provide a satisfactory index of processing difficulty.

To sum up, the processing model proposed by Kintsch (1974) offers us at least three ways of measuring the difficulty of a text. The first two ways - propositional density and number of new concepts per proposition - should be able to predict readability based upon considerations about text structure. The third way - an index of reading time and recall - should give a measure of reading difficulty, based upon considerations about comprehension processes.

Fry's Graph, cloze procedure and subjective judgements have been used in the previous chapters to describe the readability of four test passages. The main goal of this chapter was to compare the results obtained using measures based on Kintsch's model applied to these four test passages with the description of their readability based on traditional techniques.

5.3 Analysis of Test Passages

Method

Subjects

The requirements for this experiment was subjects who were similar to those described in earlier chapters; that is, pupils who were male and in the fourth form of a grammar school. It was decided to choose pupils that the author taught himself thus ensuring 'feedback' concerning the tasks to be set. Two classes of pupils completed the tasks giving a sample size of 56 with a mean age of 14yrs 5 months.

Materials

The original and modified test passages described earlier in this study were used for the present experiment. Table 5.1 summarises the data which has been collected for these passages, while Fig. 5.1 further illustrates the differences.

Table 5.1 Test statistics for the four test passages.

	AM	BM	AG	BG
No. Words	115	113	107	113
Reading level (yrs)	18	15	11	11
Cloze score (%)	45.9	43.2	64.8	48.3
Rank order based on subjective judgement (4 most difficult)	4	3	1	2
Ability to model ideas from passage (%)	63	70	89	78

A - 'exchanging gases' B - 'lymphatic system'

M - original passage (Mackean) G - modified passage (Graham)

Fig. 5.1 The rank order of readability for four passages.

	Rank order			
	Difficult			Easy
Reading level	AM	BM	(AG/BG)	
Cloze procedure	BM	AM	BG	AG
Subjective judgement	AM	BM	BG	AG
Ability to model main ideas	AM	BM	BG	AG

A - 'exchanging gases' B - 'lymphatic system'

M - original passage (Mackean) G - modified passage (Graham)

First of all, note that cloze scores do not correspond well with the other data and passages BM, AM and BG are not clearly distinguished. Note also that the Fry Graph does not predict differences in reading level between the modified passages AG and BG. The measures however, can be used to suggest an "average" rank order of the four passages. Beginning with the most difficult the order is AM > BM > BG > AG.

It is further suggested that the differences between the passages are approximately the same. The author proposes to regard this rank order, based upon four rather different readability measures, as an index of the readability of the four passages. The aim of the present experiment is to match this rank order, based upon considerations about text structure and comprehension processes that are described in the next part of this chapter.

Procedure

Section 5.2 of this chapter suggests two main ways of measuring the difficulty of a text. The first way - propositional density and number of new concepts per proposition - are attempts to predict readability based on an analysis of the text. The second way - an index of reading time and recall - measures the interaction between subject and text, that is, comprehensibility.

1. Text analysis

The approach is based upon the model for the representation of meaning in memory developed by Kintsch (1974). The meaning of a text is represented by its text base, which is a structured list of propositions. A proposition consists of a predicate with one or more argument; arguments are word concepts or propositions themselves. For convenience, the concept is denoted by its corresponding word; however, it is important to remember that the analysis if not dealing with words at all, but with abstract concepts. The convention of writing the predicate of a proposition first has been adopted. Furthermore, the propositions are numbered so that if one proposition serves as the argument of another, it can be simply referred to by number and need not be written out again.

Table 5.2 shows the text base for a short passage. The 9 propositions are ordered in the same way as they are expressed in the text (i.e. a proposition "occurs" when its predicate or one of its arguments first occurs). The number of word concepts in the passage is 5.

Table 5.2 The text base for a passage on 'tissue-fluid'.

Around the cells of our body is a liquid called tissue-fluid.
It keeps our cells damp and forms a link with the blood vessels.

1. AROUND, (3), (4)
2. OUR, body
3. OF (location), cells, (2)
4. CALLED, liquid, tissue-fluid

---

5. OUR, cells
6. DAMP, cells
7. KEEPS, (4), (5), (6)
8. FORMS, tissue-fluid, (9)
9. LINK, cells, blood-vessels.

No. propositions - 9

No. word concepts - 5

This procedure is explained in more detail by Turner and Greene (1977), but one modification to this procedure was to consider double nouns such as tissue fluid as one word concept. Kintsch and Vipond (1979) suggest that modifications and improvements will not destroy the system; that is, the same pattern of results should be obtained with a somewhat modified system, though numerical values would of course be changed.

The author found this procedure initially difficult to apply and many subjective judgements had to be made concerning the interpretation of the rules. To overcome the latter problem, a second teacher was asked to analyse independently the four passages and arrived at substantially the same results. The text bases presented in this study were the result of discussing the independently produced versions and agreement on an optimum version. It is considered that a consistent approach towards the analysis of the four passages at least ensured that the results were comparable.

2. The interaction between reader and text

The interaction between the subject and test passage was measured in two ways - time taken to read the passage and ability to recall the same passage. The tests were administered on separate occasions to 2 class groups during June 1980.

A short, informal introduction was given by the author in which it was explained that he was carrying out an experiment that tested the passages and not the pupils. Each pupil was told that at the start of the test he was to start the stopwatch that was provided, read the passage until he fully understood it and then stop the stopwatch. The test passage was to be immediately turned over and on the reverse side they were to write all that they could recall of the passage. The pupils were familiarised with the procedure by a trial run with a practise passage.

Each pupil was allocated at random, one of the four test passages. The experiment was conducted in relaxed conditions, however provision was made to prevent the pupils from cheating. Adequate time was given for all the pupils to finish the task. Finally, the author recorded the reading time for each pupil and collected in the test papers for analysis.

Recall of complete propositions was scored with no distinction being made between verbatim or paraphrase recall. Partial recall, inferences and errors were neglected. When a subject made an error in one proposition which affected a related proposition, the related proposition was scored as correct. This prevented the counting of errors more than once. For instance, a pupil may have written:

'Around the cells of our body is a liquid called serum.
It keeps our cells damp and forms a link with the blood vessels.'

The first proposition would have been scored as incorrect because of the substitution of "serum" for "tissue fluid". However, the next proposition would be correct because whatever the liquid is called, it keeps our cells damp.

Results

1. Text analysis

The analysis translated the four passages into text bases which are presented in Appendix 3. Table 5.3 gives details of the number of propositions which occured in each text base. Because the texts differ in length, values per 100 words are reported rather than actual numbers to facilitate comparison. A second text base statistic that is said to affect readability is presented as the number of concepts per proposition. The predicted rank orders of difficulty based on this data can be found in Fig.5.3.

Table 5.3 The number of propositions and word concepts in the text bases of the four passages.

	Test Passage			
	AM	BM	AG	BG
Number of propositions per 100 words	21.7	28.3	22.4	23.9
Number of arguments per proposition	0.76	0.56	0.625	0.48

A - 'Gaseous exchange' B - 'Lymphatic system'

M - Original passage (Mackean) G - Modified passage (Graham)

Fig. 5.2 Rank order of readability as predicted by characteristics of the text base.

	Difficult			Easy
No. of propositions per 100 words	BM	BG	AG	AM
No. of arguments per proposition	AM	AG	BM	BG

A - 'Gaseous exchange' B - 'Lymphatic system'

M - Original passage (Mackean) G - Modified passage (Graham)

The results suggest that neither the number of propositions per 100 words nor the number of word concepts per proposition predict the same rank order of reading difficulty as suggested by the readability data presented in Table 5.1.

The two text base characteristics described above are both said to play an important part in determining readability (Kintsch and Vipond, 1979). However, it must be appreciated that these two characteristics probably interact with each other to affect readability.

A number of attempts were made to express the relationship between the number of propositions per 100 words and the number of word concepts per proposition in the form of an index. Unfortunately none of these attempts produced an index which could successfully predict the expected rank order of the four test passages.

2. The interaction between reader and text

Table 5.4 presents the details of the mean times taken for the subjects to read the test passages with understanding and their standard deviations. Similar statistics are presented for the percentage recall of propositions.

Table 5.4 Mean reading times and percentage recall for the four passages.

		Test Passages			
		AM	BM	AG	BG
Time per 100 words (seconds)	$\bar{x}$	221.0	180.5	146.0	169.0
	SD	83.8	80.1	91.3	58.4
	N	14	14	14	14
Percentage recall of propositions	$\bar{x}$	31.4	31.2	44.95	37.0
	SD	15.7	14.9	14.6	14.9
	N	14	14	14	14

A - 'Exchanging gases' B - 'Lymphatic system'

M - Original passages (Mackean) G - Modified passages (Graham)

A one way analysis of variance, treating the test passages as random variables was carried out for both reading time and percentage recall. (Table 5.5).

Table 5.5 One way analysis of variance for reading time and percentage recall of propositions (full details can be found in Appendix 4).

Analysis	Source of Variation	SS	DF	MS	F	Sig of F
Reading time	texts	41449.6	3	13816.5	2.0	0.124
Recall	texts	1743.65	3	581.2	2.6	0.063

Table 5.5 indicates that there were no significant differences between the test passages for either measure. The results were confounded by the large amount of variation to be found within the groups of scores for each passage. For example, a multiple R squared analysis for time, indicated that only 10.4% of the variation was between the test passages while the remaining 89.6% described the variation within the passages.

Kintsch et al (1975) suggest that there is a trade-off relationship between reading times and the recall scores, making both measures difficult to interpret. A more suitable index of how difficult a particular text is, would be a measure that combines both reading time and recall. Table 5.6 presents the number of seconds of reading time per proposition recalled (Tcall), while Table 5.7 presents details of a one way analysis of variance treating the test passages as independent variables.

Table 5.6 Average reading time in seconds per proposition recalled (Tcall).

	Test Passage			
	AM	BM	AG	BG
Tcall	7.0	5.7	3.25	4.6

A - 'Exchanging gases' B - 'Lymphatic system'

M - Original passage (Mackean) G - Modified passage (Graham)

Table 5.7 One way analysis of variance for Tcall. (full details can be found in Appendix 4).

Analysis	Source of Variation	SS	DF	MS	F	Sig of F
Tcall	Texts	223.9	3	74.6	2.2	0.099

Unfortunately much of the variation was within the test passages and the differences between the mean Tcall scores were not significant. However, the data obtained by the measure Tcall places the test passages in the following rank order of difficulty:

$$AM > BM > BG > AG$$

which is the same rank order as predicted on the basis of evidence collected earlier on in this study (Fry's Graph, cloze procedure, subjective report).

Discussion

The processing model described by Kintsch allowed the author to ask a number of interesting questions. However, the purpose of this experiment was not to justify the theoretical system but to:

1) attempt to predict textual difficulty by the number of propositions per 100 words and number of word concepts per proposition in the text base of each passage;

2) determine the comprehension of a passage by an index which incorporated reading time and recall of propositions.

1. Text analysis

A number of studies (e.g. Kintsch and Keenan, 1973; Kintsch et al, 1975) have identified two text base statistics that appear to affect readability: the propositional density of a text and the number of concepts per proposition. If comprehension does involve the construction of a text base, it is easy to see why the number of propositions should play a role. Their second finding is not as obvious, but it has some far-reaching implications. Introducing a new concept is harder than repeating an old one which suggests that the repetition of concepts might serve an important function in the formation of a text base.

In this study, the number of word concepts per proposition and the number of propositions per 100 words did not predict the expected rank order of difficulty. However, there are a number of reasons for this result. The work of Kintsch and his associates used passages with large, clearly defined differences between the numbers of propositions and concepts; whereas the differences between the test passages used in this study were much smaller and less definite. For example, all the test passages had a ratio of concepts to propositions which would have placed them in a group of paragraphs classified by Kintsch et al (1975) as having many different concepts.

Another difference between the passages used in this study and those used in previous studies is that the former are "real" while

the latter were constructed so as to control all the features of the text other than the one to be tested. Even so, not all the differences reported by Kintsch et al have been statistically significant and they point out that:

> '.....there are numerous other, mostly unexplored factors that affect reading difficulty.'

Thus in order to obtain empirical evidence concerning the readability of "real" texts, there are a large number of subsidiary factors that have to be taken into account.

The analysis of a text can be taken a step further, because in terms of a model of comprehension, a text base must have some kind of property that distinguishes it from an arbitrary list of propositions. The problem of text cohesion is an old one to which a good deal of effort has been devoted by linguists and logicians. Clark (1976) talks about the importance of "bridging" in text comprehension while Haviland and Clark (1974) have shown that subjects can comprehend a sentence more rapidly if "connected" to a preceding sentence.

The model by Kintsch solves the problem of coherence by connecting propositions if they share a common argument. For example, (P,A,B) is coherent with (R,B,C) because the two propositions share the word concept B; (P,A,B) is also coherent with (Q,D, (P,A,B))because one proposition is embedded here as an argument into another. The sharing of a common argument is not a sufficient criterion linguistically; however, the fact that in many texts other factors tend to be correlated with it, makes it a useful indicator of coherence that can be checked easily, quickly and reliably. Thus the formation of a text base is formally equivalent to the construction of a graph in which propositions that share an argument are connected to the proposition that first introduced that argument. A full explanation and verification of the procedure, which also has to take into account inferences and memory storage, is given by Kintsch and van Dijk (1978).

The implications of the analysis outlined above for readability have been given by Kintsch and Vipond (1979), who propose that a number of variables might be important. For example, the number of separate graphs for each text that result when a coherent network is constructed, could influence readability. A text base that breaks down into several unconnected parts should be harder to recall than an otherwise equivalent but coherent one; the lack of coherence should produce retrieval difficulties.

In the context of this study it is suggested that the topic 'lymphatic system' was harder than expected because the reader had to carry out a greater number of resource consuming operations to determine which propositions to retain from each series of events in short term memory so as to make connections with the next set of input propositions. A reader with a poor selection strategy or a short memory retention when reading about a complex series of events might have had all kinds of problems.

It must be stressed that at present, it is not known which factors of the text base other than the density of propositions and the number of concepts per proposition are important. However, the author considers that other variables are worth investigating in a formal way and that some of the variance not accounted for might have its source in just these factors.

2. Measuring the interaction between reader and text

The goal of this part of the experiment was to match the rank order of readability based on data from measures described in previous chapters, with a rank order based on a measure of reading time and recall (Tcall). Tcall discriminated between the four passages and their rank orders were successfully matched. Thus it can be concluded that the goal has been achieved and that Tcall may now turn out to be a new powerful measure of readability. However, there are three criticisms of the present study which are outlined below:

1. The reliability of Tcall as a measure is open to question because of the variability of results obtained for each passage. In this experiment the differences in Tcall between the four passages were not statistically significant. Some of the variability of results may have been due to the procedure which was crude in comparison to the work of Kintsch and his associates. It is also suggested that future studies use a much larger sample size even though this would considerably increase the burden of administration and scoring.

2. In order to reduce the problem of analysing prose recall to manageable proportions, severe limitations were placed upon the approach to scoring. Recall was only looked at in terms of the reproduction of propositions actually stated in the text. Correct inferences, partial recall and the reconstructive aspects of recall were ignored.

A theory based approach that is more concerned with the reconstructive aspects of recall is that of Frederiksen (1972, 1975). He showed that when subjects were given a number of trials at reading and recalling the same passage, the number of overgeneralised and inferred concepts and relations increased. Such an increase in responses which do not reproduce text content, reflect processes fundamental to the acquisition process and not just attempts during recall to fill in gaps in a subject's memory structure.

One interesting area for further research would be to score recall using, for example, Frederiksen's method, which could then be incorporated into the measure Tcall. Research of this kind has numerous possibilities and could provide some interesting results.

3. Not everyone agrees that models of comprehension will help to improve educational technology. Marcel (reported in Underwood and Holt, 1979) at a symposium on 'Reading and Lexical Access' presented a paper which questioned the use of this type of model. He looked critically at models of comprehension and doubted their general usefulness and validity. This controversial view may be correct, but whether we can come to an understanding of what is involved in reading without a model remains to be seen.

In sum, this chapter has shown that recent research on text comprehension has many implications for the concept of readability. It is realized that by analyzing a few examples, nothing has been proven. However, these analyses suggest many areas which require research through a series of converging experiments within the framework of the model.

A text analysis of the passages indicated that the number of propositions in 100 words and number of concepts per proposition were not sufficient to predict difficulty. It is quite probable that further research will identify other factors that make a text easy or hard to read. However, readability is not just an inherent property of a text but is the result of the interaction between a particular text (with its text characteristics) and particular readers (with their information-processing characteristics). Thus, for example, the effect of motivation towards the topic 'lymphatic system' has to be taken into account. Readability must be defined for specific texts and specific readers.

The measurement of reading time and recall (Tcall) for the four passages provided encouraging results. Further research is clearly necessary to clarify the procedure so that the measure is a reliable one. However, the evidence is that Tcall can be considered a useful measure of the readability of a text. Its usefulness is that like cloze procedure it measures the interaction between reader and text; and unlike cloze procedure, Tcall is based on sound psychological theory. The measure was also able to distinguish between the passages in a way that required a number of traditional measures of readability. In subjects such as biology where much of the difficulty is not just due to the language, this tool might prove of invaluable assistance to the research worker, author and teacher.

Summary

The chapter began with the arguments for measures of readability based on a psychological model of comprehension processes. The representation of the structure of a text as an outline model of comprehension was discussed. More specifically, a model of comprehension devised by Kintsch was described and its implications for the assessment of readability were pointed out.

The results of readability measures described and used in previous chapters provided a rank order of difficulty for the four test passages. The aim of this chapter was to match this rank order by measuring textual difficulty using two measures based on Kintsch's model. The first measure predicted readability through the analysis of the passages for the number of propositions and the number of new concepts. The second measure was an index of reading time and recall (Tcall) which takes into account the interaction between reader and text.

The result of the textual analysis did not successfully predict the rank order of the passages. Further research is required to determine which other factors of the text base are also important determinants of reading difficulty. The measure Tcall successfully placed the passages in the same rank order; however, the differences between the passages were not statistically significant. It was concluded that Tcall should be adopted as an alternative approach to the measurement of readability.

CHAPTER VI

CONCLUSIONS

Introduction

> 'Miss Marple gave her attention first to the main news on the front page.....; from that she passed to the Court Circular, on which page today's news from the sale room could also be found. A short article on science was often placed there, but she did not propose to read that. It seldom made sense for her.'
>
> Agatha Christie (Nemesis).

Even though new educational procedures and materials are being developed, the textbook remains the basic tool of the biology teacher. Gould (1977) reports that textbooks are used in approximately 60% of biology lessons, which confirms the view that teachers still regard it as an important teaching aid. This widespread belief in the value of the textbook prompted the author to ask three questions:

1. How readable is textual material in biology and can the factors that contribute to reading difficulty be identified?
2. What are the optimum conditions required to apply traditional measures of readability to biology texts and how useful are the results that are obtained?
3. Can new measures of readability be developed which are based on a psychological model of comprehension?

In this chapter, the author will attempt to answer in turn, the three questions outlined above.

6.1 The readability of biology texts

The most important implication of the findings in this study for biology teachers and textbook authors is that the demands made on children by the texts currently in use are excessive. The results suggest that teachers should be aware of the possible reading difficulties imposed by the textbooks they use even with 'O' level ability pupils and should make allowances in the ways they employ them in lessons. Certainly, the teacher who knows of these problems can guard against such undesirable practice as uncomprehending, wholesale copying of sections of the book instead of note taking. Particular attention should be paid to the difficulties for average and below average ability candidates for the 'CSE' exam and for all pupils working at home without aid. There is a danger that with a wider spectrum of ability now common in many classes and less money to spend on textbooks, teachers will purchase class sets of books that only cater for the more able pupils. The long term goal must be for authors and publishers to provide the schools with textbooks whose level of difficulty is more carefully controlled. Indeed, Gould (1977) suggests that publishers are beginning to recognize the need for this with some of their most recent books. In the short term, the present shortage of textbooks in many of the schools may not, in fact, be such a serious problem - without them the pupils may be making more sense of their biology classes.

The second implication of this investigation is that readability is not somehow an inherent property of texts but is the result of the interaction between a particular text (with its text characteristics) and particular readers (with their individual characteristics). It is clear that language alone does not always correctly predict the readability of a biology text. In this study the topic 'lymphatic system' was found to be more difficult than predicted by Fry's Graph. The topic consisted of a set of complex and dynamic ideas which caused the pupil

to experience conceptual difficulty. Pupils showed little interest in the topic, possibly due to frustration, or because they did not consider the 'lymphatic system' to be relevant to themselves. It is suggested that topics such as the lymphatic system should be taught in the final year of the exam course. This means that the pupils, being older and well motivated, are more likely to cope with difficulties.

Gould (1977) found, in his survey of the use of books during lessons, that 36.5% of the use of textbooks was for the examination of illustrations. Diagrams are an important part of any biology textbook and children appreciate them because they aid comprehension. One pupil, when asked about the readability of the topic 'lymphatic system' made the following comment:

> "You can't understand that without pictures. They would make the reading more clear."

This study did not consider the contribution of illustrations even though it is an important issue. It would be interesting to present to parallel groups of pupils, matched for reading ability, the same passage without illustrations and with various kinds of illustration in order to assess the contribution that the illustrations make to understanding of written material. There is some evidence that there is greater learning from a line diagram. (Dwyer, 1967).

Is readability important? Quoting from Chall (1958):

> "Readability is now the concern of all those who depend upon communication through the printed word."

Klare (1975) has conducted a number of carefully controlled studies in which he has shown that if you take two groups of readers who are equal in reading ability and give them exactly the same comprehension test, those given a more readable version of the test passage will learn and understand more.

Thus pupils will read with greater understanding if the readability of textual material is improved. However, it must be pointed out that there will also be greater understanding if the pupils are encouraged to

improve their reading ability. Gillard (1975) found that reading skills in science did not improve as skill in reading for English lessons improved, and he urged that these special science skills should be taught. Various recommendations by, for example, Fay (1965), Niles (1969) and Bullock (1975) include:

(a) learning to make a preliminary survey of the reading matter

(b) knowing how to establish a purpose for the reading to be done, by asking appropriate questions

(c) learning to interpret the graphical and illustrative content

(d) developing the ability to summarize

(e) learning to scan and skim, so that the most important material receives most attention

(f) acquiring the ability to find, understand, and combine information from several sources.

(After Carrick, 1978).

It is hoped that there will be increasing awareness by teachers of the difficulties arising from reading, and that this will lead them to help to make biology more accessable to their pupils. Teachers can help by improving the readability of the textual material presented and by developing pupils' reading skills. In this way children may spend less time reading like Coleridge's sponges, sand glasses or strain bags, and more time reading like his mogul diamonds.

6.2 An Evaluation of Traditional Measures of Readability

Three traditional measures of assessing readability have been used in this study - Fry's Graph, cloze procedure, subjective judgement. The purpose of this section is to comment on the optimum conditions required to apply these measures and critically evaluate the usefulness of the results that are obtained.

Fry's Graph

Fry's Graphical technique was found to be speedy and easy to use, while the work of Harrison (1979) suggests that it is reliable. However, there is the danger that the technique could provide an inaccurate estimate of readability if the following points are not kept in mind. First, results from this study suggest that 10% of a textbook's pages should be sampled. Second, an average reading level is probably meaningless because the present generation of biology textbooks have passages which range from 11 to 18 years. Third, it is essential to remeber that the formulae take account of the number of syllables and the sentence only. These are known to be closely related to overall complexity of language structure but the contribution of other factors are being ignored.

Despite the limitations of Fry's Graph, the author considers it to have an enormous value as an objective measure of the prose difficulty of a passage and thus assisting the individual teacher in the important task of choosing suitable material.

Cloze Procedure

When cloze procedure is applied, both reader and text are assessed simultaneously by the use of one measure. This undoubtedly gives the procedure a greater face validity than other traditional measures of readability. There are, however, a number of important reasons why the author has misgivings about this technique. First, it is clear from the results of this study that it is essential to use each possible version of

a passage. Only in this way can one be certain that every word has been taken account of, and that an unreliable result will not be obtained by a single version which happens to delete a particularly difficult (or easy) set of words. Second, it is unlikely that a cloze test would produce a reliable result if only used with a small group of children. In this study, individual results on any one of the test passages showed a good deal of variation. Third, it is hard to draw many conclusions about the readability of a textbook when only a small sample of texts have been subjected to cloze procedure. The problem is that it is difficult to have a large sample size because of the burden of preparing and administering each cloze test.

To sum up, these considerations seem to suggest that unless cloze procedure is used with great care and in a systematic way, as an indicator of readability, cloze scores will be unreliable and may be misleading. A fair amount of fundamental research will be required before cloze procedure can be considered a reliable and valid measure of readability.

Subjective judgement

Despite the many disadvantages of the questionnaire technique, the author found that the results made an important contribution towards the understanding of the readability of the test passages. By classifying the pupils' responses to the questions, the author was able to guard against sweeping and ambiguous statements. However, the questionnaire was of an exploratory nature only and the small sample size would have made a statistical analysis of the responses impractical. In fact, it was found that the "open ended" questions provided the most useful information because the pupils were able to give their own opinions about any aspects of the text that they found difficult. In sum, even though the type of questionnaire described in this study may be considered to be an un-refined technique by many researchers, it is of value to the teacher. Armed with the opinions of the pupils, the teacher can be consequently more accurate with his or her choice and presentation of suitable material.

The questionnaire also included an exercise which tested the pupils' ability to translate a set of ideas from one of the passages into a graphic model. The results provided the author with comprehension outcomes that were easily analysed. At present, modelling comprehension outcomes is a little explored area but it offers interesting possibilities for the future. First, with no convenient tools for measuring conceptual difficulty, the construction of models could go a long way towards helping teachers made an assessment of a text. Second, the training needed to construct models seems likely to be of value in improving pupils' comprehension of biology texts.

It is not doubted that the traditional measures can be of practical assistance to the teacher but there is little understanding of what has been measured. Throughout this study the reliability and even the validity of these traditional measures has been questioned. New methods are required which have been designed within the context of an explicit model of the processes involved in text comprehension.

6.3 Measures of readability based on a psychological model of comprehension

For nearly ten years, psychologists showed little interest in readability. The number of references per year made to this subject in the Psychological Abstracts from 1965 to 1975 are close to zero. In contrast, readability has been of considerable practical significance to educationslists. During the same ten years, the number of references per year made to readability in the Education Index was high - unfortunately, there was little theoretical progress. However, recent advances have resulted in a number of models of reading comprehension and there is renewed interest amongst psychologists because these models appear to have some powerful implications for the measurement of readability.

This study chose a model of comprehension devised by Kintsch (1974) because it offers two new ways of measuring readability:- first, by analysing the text and second, by measuring the interaction between reader and text.

Text analysis

The model by Kintsch suggests that there are a number of characteristics of the text which determine readability. Of these, the number of word concepts and propositionswere chosen to measure the readability of the four test passages. However, neither measure, nor an index of the two could correctly place the passages in the correct rank order of difficulty. There are two main reasons for this result.

First, the two characteristics of the text base are probably not the only important determinants of reading difficulty, and further research is required to identify these content variables. Second, Kintsch and his associates wished to emphasise the characteristics of the text that they were studying and thus created 'artificial passages'. This study used 'real' passages that is, there was no control of the many textual factors that determine readability.

To sum up, text analysis is not yet a reliable measure of readability because not enough is known about the content variables that

significantly affect comprehension. However, when more is known about these content variables the approach outlined above could have powerful implications for a redefinition of the problem of readability.

The interaction between reader and text

Unlike most measures of readability, the use of reading time and recall is theory dependent. Recall of a text (or summarizing, if the text is long) has not been used as much as it should be as a criterion variable in readability studies. The reason lies in the impossibility of scoring recall protocols without some theoretical foundation. Scoring for meaning (if it is not to be arbitrary) presupposes a model for the representation of meaning. A number of such models are now available and a wider use of the recall technique is expected. However, if one is interested in comprehensibility, recall scores can never be considered in isolation, because they depend on reading times (Kintsch et al, 1975). Thus, neither reading time nor amount of recall alone provides a satisfactory index of processing difficulty, but only their combination.

When the index Tcall (reading time per proposition recalled) was determined for the four test passages, the results placed the passages in their correct rank order of difficulty. In fact, the scores also correctly indicated how difficult each passage was in comparison to the others. The author realizes that by analyzing a few examples nothing has yet been proven. But a research programme to investigate readability by the measure Tcall would be promising. If the measure is found to be reliable and is subsequently adopted by educationalists, then this study will have made a small contribution towards improving the disturbed relationship between education and psychology.

Appendix 1

Subjective guidelines used to modify texts taken from a textbook by D.G. Mackean in the direction of increased readability.

A. Organization

1. A logical sequence of events
2. Reduction in sentence length
3. Utilization of sentences and paragraphs to express concepts
4. Minimisation of detail contributing little to the main ideas

B. Vocabulary

1. Replacement of difficult terms
2. Cueing of significant and difficult terms
3. Defining difficult terms
4. Repetition of new terms once introduced

C. Interest

1. A natural manner of presentation

Appendix 2

Further details of readability statistics reported in Chapter 3.

Table 3.6 Analysis of variance, topics by authors using values from individual versions.

Version 1

Source of Variation	SS	DF	MS	F	Sig of F.
Topic	1784.0	1	1784.0	15.5	0.001
Author	4667.2	1	4667.2	40.5	0.001
2-way interaction	314.9	1	3.4.9	2.7	0.111
Explained	6766.1	3	2255.4	19.6	0.001
Residual	2766.6	24	115.3		
Total	9532.7	27	353.1		

Version 2

Source of Variation	SS	DF	MS	F	Sig of F.
Topic	62.4	1	62.4	0.25	0.621
Author	438.4	1	438.4	1.8	0.197
2-way interaction	67.9	1	67.9	0.3	0.606
Explained	568.7	3	189.6	0.8	0.526
Residual	5964.6	24	248.5		
Total	6533.4	27	242.0		

Version 3

Source of Variation	SS	DF	MS	F	Sig of F.
Topic	751.9	1	751.9	2.8	0.109
Author	1140.5	1	1140.5	4.2	0.051
2-way interaction	606.4	1	606.4	2.2	0.148
Explained	2498.8	3	832.9	3.1	0.47
Residual	6505.3	24	271.1		
Total	9004.1	27	333.5		

Version 4

Source of Variation	SS	DF	MS	F	Sig of F.
Topic	1398.9	1	1398.9	8.5	0.008
Author	0.01	1	0.01	0.0	0.994
2-way interaction	2471.3	1	2471.3	15.0	0.001
Explained	3873.5	3	1291.2	7.8	0.001
Residual	3465.2	21	165.0		
Total	7338.7	24	305.8		

Version 5

Source of Variation	SS	DF	MS	F	Sig of F.
Topic	1.04	1	1.04	0.01	0.918
Author	3374.8	1	3374.8	35.1	0.001
2-way interaction	2937.6	1	2937.6	30.6	0.001
Explained	6313.5	3	2104.5	21.9	0.001
Residual	2305.4	24	96.1		
Total	8618.9	27	319.2		

Appendix 3 Proposition lists for the four test passages.

1. ORIGINAL PASSAGES

A. 'EXCHANGING GASES' (AM)

1. LINING, alveoli
2. COVERED (with), (1), (3)
3. FILM, moisture

4. CONCENTRATION, oxygen
5. IS LOWER (than), (4), alveolus, blood
6. HENCE, (5)
7. IN (location), oxygen, air-space
8. DISSOLVES (in), (7), (3)
9. DIFFUSES (through), oxygen, epithelium, capillary-wall, plasma

10. DIFFUSES (into), oxygen, red-cell
11. COMBINES, oxygen, haemoglobin

12. KEVNITE, capillaries
13. FORM, (12), pulmonary-veins
14. RETURN, pulmonary-vein, (15), (16)
15. OXYGENATED, blood
16. LEFT, atrium, heart

17. LOW CONCENTRATION, carbon-dioxide, alveoli
18. STIMULATES, (17), (19)
19. CALLED - enzyme, carbonic-anhydrase
20. IN (location), blood
21. BREAKDOWN, (19), bicarbonate-salts
22. LIBERATE, bicarbonate-salts, carbon-dioxide

23. DIFFUSES (into), gas, alveoli
24. IS EXPELLED, gas
25. EVENTUALLY (time), (24)

B. 'LYMPHATIC SYSTEM' (BM)

1. ONLY, route
2. ARE NOT, capillaries, (1), (3)
3. RETURN, tissue-fluid, circulation

4. SOME, tissue-fluid
5. RETURNS, (4), lymphatic-system

6. IN, proteins, tissue-fluid
7. UNABLE (to), (6) (8)
8. RE-ENTER, capillaries
9. SO DRAIN (6), (10), (11)
10. BLINDLY-ENDING, vessels
11. THIN-WALLED, vessels
12. FOUND, (10), (11), (13)
13. BETWEEN, cell

14. JOIN UP, lymphatics
15. FORM, (14), (16)
16. LARGER, vessels
17. UNITE, (16) (19)
18. MAIN, ducts
19. TWO, (18)
20. EVENTUALLY, (17)
21. LARGE, veins
22. ENTERING, (21), right-atrium

23. IN, fluid, lymphatic-vessels
24. IS CALLED, (23), lymph

25. COMPOSITION, lymph
26. SIMILAR TO, plasma (25) (27)
27. CONTAINS LESS, proteins

28. CONTAINS, lymph, (29)
29. CERTAIN, (30)
30. TYPE, white-cell
31. NAME, (29), lymphocyte
32. MADE IN, lymphocyte, lymph-glands

2. THE MODIFIED PASSAGES

A. 'EXCHANGING GASES' (AG)

1. INSIDE, (2), (3), (4)
2. YOUR, lungs
3. SMALL, air-sacs
4. CALLED, air-sacs, alveoli

5. HAVE, sacs, walls
6. DAMP, walls
7. DISSOLVE, gases, (8)
8. DAMP, layer

9. AROUND, air-sacs (10)
10. MANY, blood-vessels

11. CONTAIN (less), blood-vessels, oxygen
12. CONTAIN (more), air-sacs, oxygen

13. SO, (11), (12)
14. MOVES, oxygen
15. INTO (location), air-sacs, blood

16. AROUND, blood-vessels, air-sacs
17. CONTAIN, blood-vessels, (18)
18. LOT (amount), bicarbonate-salts

19. BREAKS DOWN, enzyme, bicarbonate-salts
20. GIVE (result), bicarbonate-salts, carbon-dioxide
21. LESS (amount), carbon-dioxide, air-sacs
22. LESS (than), blood-vessels, carbon-dioxide

23. SO, (21), (22)
24. PASSES, carbon-dioxide, blood, air-sacs

B. 'LYMPHATIC SYSTEM' (BG)

1. AROUND, (3) (4)
2. OUR, body
3. OF (location), cells, (2)
4. CALLED, liquid, tissue-fluid

5. OUR, cells
6. DAMP, cells
7. KEEPS, (4), (5) (6)
8. FORMS, tissue-fluid (9)
9. LINK, cells, blood-vessels

10. NEED, (5), substances
11. PASS (out), (10), blood-vessels
12. PASS (into), (10), tissue-fluid

13. PASS (into) substances, cells

14. MADE, (5), anything
15. PASS (into), (14), blood-vessels
16. SAME WAY, (11), (12), (13)

17. MADE, (5), substances
18. NOT RETURN, blood-vessels, fluid

19. DRAINS (INTO), (20), ducts
20. SOME, fluid, (17)
21. BETWEEN, ducts, cells

22. PART OF, ducts, lymphatic-system

23. INSIDE, fluid, ducts
24. CALLED (23), lymph

25. SIMILAR TO, lymph, blood
26. VERY, (25)
27. CONTAINS (less), lymph, protein

Appendix 4

Further details of readability statistics reported in Chapter 5.

Table 5.5

(a) One way analysis of variance for reading time.

Source of variation	SS	DF	MS	F	Sig of F.
texts	41449.6	3	13816.5	2.0	0.124
residual	356951.8	52	6864.5		
total	398401.4	55	7243.7		

(b) One way analysis of variance for recall.

Source of variation	SS	DF	MS	F	Sig of F.
texts	1743.6	3	581.2	2.587	0.063
residual	11682.5	52	224.7		
total	13426.1	55	244.1		

Table 5.7

One way analysis of variance for Tcall.

Source of variation	SS	DF	MS	F	Sig of F.
texts	223.9	3	74.6	2.2	0.099
residual	1761.0	52	33.9		
total	1984.9	55	36.1		

References

Anderson, R. (1977) The Notion of Schemata and the Educational Enterprise. In Schooling and the Acquisition of Knowledge, eds. Anderson, R., Spiro, J. and Montague, W.E., Hillsdale, N.J.: Lawrence Erlbaum Associates.

Ausubel, D.P. (1960) The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 51, 267-272.

Barker, J.A. (1977) Comment: Biology books in Britain. Journal of Biological Education, 11, 161.

Barnes, D., Britton, J. and Rosen, H. (1969) Language, the learner and the school. Harmondsworth: Penguin.

Bormuth, J.R. (1963) Cloze as a measure of readability. International Reading Association Conference Proceedings, 8, 131-134.

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

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

Bormuth, J.R. (1968) Cloze test readability: criterion reference scores. Journal of Educational Measurement, 5, 189-196.

Bormuth, J.R. (1968b) The cloze readability procedure. Elementary English, 45, 429-436.

Bormuth, J.R. (1969) Empirical determination of the instructional reading level. In Reading and Realism: Proceedings of the International Reading Association, ed. Figurel, J.A., International Reading Association.

Bransford, J. and Johnson, M.K. (1973) Considerations of some Problems of Comprehension. In Visual Information Processing, ed. Chase, W.G., New York: Academic Press.

Brown, W.R. (1965) Science Textbook Selection and the Dale-Chall Formula. School Science and Mathematics, 65, 164-167.

Bullock, A. (1975) A Language for Life. London: H.M.S.O.

Burkhead, M. and Ulferts, G. (1977) Sample frequency in application of Dale-Chall Readability Formula. Journal of Reading Behaviour, 9, (3), 287-290.

Capponecchi, W.P.A. (1973) A Comparative Study of an Advance Organizer in Mathematics to Determine its Effectiveness on Knowledge Acquisition and Retention. Unpublished doctoral dissertation reported in Teaching Reading Comprehesnion, by Pearson and Johnson. Holt, Rinehart and Winston (1978).

Carrick, T. (1978) Readability of biology textbooks. Journal of Biological Education, 12, (2), 113-122.

Carroll, J.B. (1963) — Research on Teaching Foreign Languages. In Handbook of Research on Teaching, ed. Gage, N.L. Rand McNally.

Carroll, J.B. (1972) — Defining Language Comprehension. In Language Comprehension and the Acquisition of Knowledge, ed. Carroll, J.B. and Freedle, R.O., Winston and Wiley.

Carver, R.P. (1973) — A critical review of mathemogenic behaviours and the effect of questions upon the retention of prose materials. Journal of Reading Behaviour, 4, 93-119.

Chall, J.S. (1958) — Readability: An appraisal of research and application. Columbus: Ohio State University.

Clark, H.H. (1976) — Inferences in comprehension. In D. La Berge and S.J. Samuels (Eds), Basic processes in reading: Perception and Comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Coleman, E.B. and Miller, G.R. (1968) — A measure of information gathered during prose learning. Reading Research Quarterly, 3, 369-386.

Collins, A. and Quillian, R. (1969) — Retrieval Time From Semantic Memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Dale, E. and Chall, J.S. (1948) — A formula for predicting readability: instructions Educational Research Bulletin, 27, 11-29.

Davies, P. (1976) — The development of cloze tests. UKRA 13th Annual Conference.

Dwyer, F.M. (1967) — Adapting visual illustrations for effective learning. Harvard Educational Review, 37, 250-263.

English, H.B. and English, A.C. (1958) — A Comprehensive Dictionary of Psychological and Psychoanalytical Terms. London: Longmans.

Estes, T. and Vaughan, J. (1973) — Reading Interests and Comprehension: Implications. The Reading Teacher, 27, 149-153.

Evans, J.D. (1973) — Towards a Theory of Technical Communication. School Science Review, 55, (191), 233-241.

Evans, J.D. (1974a) — Vocabulary problems in teaching science. School Science Review, 55, (192), 585-595.

Evans, J.D. (1974b) — The teacher and his text: problems for research, School Science Review, 55, (1973), 807-811.

Evans, J.D. (1975) — Technical terms used in school textbooks of human biology. Journal of Biological Education, 9, 118-122.

Evans, J.D. (1976) — The treatment of technical vocabulary in textbooks of biology. Journal of Biology, 10, 19-30.

Farr, J.N., Jenkins, J.J. and Patterson, D.G. (1951) — Simplification of the Flesch Reading Ease Formula. Journal of Applied Psychology, 35, 333-337.

Fay, L. (1965) Reading Study Skills: math and science. In Reading and Inquiry, ed. Figurel, J.A., 92-94. Newark, Deleware: International Reading Associatio

Feller, W.A. (1973) The Effects of Two Types of Advance Organizers and Two Types of Spaced Questions on the ability of a Selected Group of Tenth Grade Biology Students to Recall, Comprehend and Apply Facts from Written Science Material. Unpublished doctoral dissertation reported in Teaching Reading Comprehension by Pearson, P.D. and Johnson, D.D., Holt, Rinehart and Winston (1978).

Fillenbaum, S., Jones, L. V. and Rapaport, A. (1963) The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. Journal of Verbal Learning and Verbal Behaviour, 2, 186-194.

Fillmore, C.J. (1968) The Case for Case. In Universals in Linguistic Theory. eds. Back, E. and Harms, R.G., New York: Holt, Rinehart and Winston.

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

Flesch, R. (1950) Measuring the level of abstraction. Journal of Applied Psychology, 34, 384-390.

Flood, W.E. (1958) The Problem of Vocabulary in the Popularization of Science. Faculty of Education Publication, University of Birmingham.

Frederiksen, C.H. (1972) Effects of task-induced cognitive operations on comprehension and memory processes. In Language Comprehension and the Acquisition of Knowledge, ed. Carroll, J.B. and Freedle, R.O., Washington, D.C Winston.

Frederiksen, C.H. (1975) Effects of Context-Induced Processing Operations on Semantic Information Acquired from Discourse. Cognitive Psychology, 7, 139-166.

Fry, E.A. (1968) A readability formula that saves time. Journal of Reading, 11, 513-516.

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

Gardner, P.L. (1972) Words in Science. A.S.E.P., Melbourne.

Gibson, E.J. and Levin, H. (1975) The Psychology of Reading. MIT Press.

Gilbert, C.D. (1973) An examination of readability levels for selected basic science texts. School Science and Mathematics, 73, 747-758.

Gillard, H.C. (1975) Factors affecting the efficient reading of science textbooks: a pilot study. In The Road to Effective Reading, ed. Latham, W., 130-158, London: Ward Lock.

Gillie, P.J. (1957) — A simplified formula for measuring abstraction in writing. Journal of Applied Psychology, 41 214-217.

Gilliland, J. (1970) — The assessment of readability - an overview. In Reading and the Curriculum, ed. Merritt, J., London: Ward Lock Educational, 144-158.

Gilliland, J. (1977) — Personal communication.

Gilliland, J. (1972) — Readability. London: Hodder and Stoughton.

Gould, C.D. (1977) — The readability of school biology textbooks. Journal of Biological Education, 11, 248-252.

Graham, W.S. (1977) — Comparing the readability of two biology textbooks. Unpublished P.G.C.E. study, Faculty of Education, University of Birmingham.

Graham, W.S. (1978) — Readability and science textbooks. School Science Review, 59, (208), 545-550.

Gray, W.S. and Leary, B.E. (1935) — What makes a book readable: An initial study. University of Chicago Press.

Greene, C.E. and Szabo, M. (1979) — Effects of Reduced Reading Level on Achievement in ISCS. Science Education, 63, (1), 37-44.

Greiling, M.G. (1973) — Recognition and Comprehension of Lexicol Words Used Alone or in Context as a Function of Spelling Pattern Predictability, Word Frequency and Word Abstractions. Unpublished M.A. paper reported in Teaching Reading Comprehension by Pearson, D.D. and Johnson, D.D., Holt, Rinehart and Winston (1978).

Greig, C. (1976) — A Practical Introduction to Cloze Procedure. Faculty of Education, University of Birmingham.

Greig, C. (1977) — Educational Publishers, School-books and Authors. Faculty of Education, University of Birmingham.

Gunning, R. (1952) — The technique of clear writing. New York: McGraw-Hill.

Harrison, C. (1974) — Readability and Schools. Schools Council Project discussion document, University of Nottingham School of Education.

Harrison, C. (1977) — Assessing the readability of school texts. In Reading: Improving Classroom Practice and Research, ed. Gilliland, J., London: Ward Lock Educational.

Harrison, C. (1978) — Personal Communication.

Harrison, C. (1979) — Assessing the readability of school texts. In The Effective Use of Reading, ed. Lunzer, E. and Gardner, K., Heinemann.

Harrison, C. and Gardner, K. (1977) The place of reading. In Language Across the Curriculum, ed. Marland, M., Heinemann.

Haviland, S.E. and Clark, H.H. (1974) What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 13, 515-521.

Herbart, J.F. (1893) The science of education (H.M. Felkin and E. Felkin, trans.) Boston: Heath.

Holonquist, J.B. (1968) A Determination of Whether the Dale-Chall Readability Formula May Be Revised to Evaluate More Validly the Readability of High School Science Materials. Doctoral dissertation, Colorada State University.

Jackson, M.D. and McCelland, J.L. (1975) Sensory and cognitive determinants of reading speed. Journal of Verbal Learning and Verbal Behavior, 14, 565-574.

Johnson, R.K. (1979) Readability. School Science Review, 60, (212), 562-568.

Kintsch, W. (1974) The Representation of Meaning in Memory. Hillsdale, N.J.: Erlbaum.

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

Kintsch, W., Kozminsky, E., Streby, W.J., McKoon,G., and Keenan, J.M. (1975) Comprehension and recall of text as a function of content variables. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 196-214.

Kintsch, W. and van Dijk, T.A. (1978) Towards a Model of Text Comprehension and Production. Psychological Review, 85, (5), 363-394.

Kintsch, W. and Vipond, D. (1979) Reading comprehension and readability. In Perspectives on Memory Research, ed. Nilsson, L.G., 329-365. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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

Klare, G.R. (1966) Comments on Bormuth's readability: a new approach. Reading Research Quarterly, 1, (4), 119-25.

Klare, G.R. (1974/1975) Assessing readability. Reading Research Quarterly, 10, 62-102.

Klare, G.R. (1975) Judging readability. Instructional Science, 6, 55-61.

Klare, G.R., Simarko, H.W. and Stolurow, L.M. (1972) The cloze procedure: a convenient readability test for training materials and translations. International Review for Applied Psychology, 21, (2), 77-106.

La Berge, D. and Samuels, S.J. (1973) Towards a Theory of Automatic Information Processing in Reading. In Theoretical Models and Processes of Reading, eds. Singer, H. and Russell,R Neward: International Reading Association (1976).

Lewis, D.G. (1965) Objectives in the Teaching of Science. Educational Research, 7, 186-199.

Lindsay, P. and Norman, D. (1972) Human Information Processing. New York: Academic Press.

Lockman, R.F. (1956) A note on measuring "understandability". Journal of Applied Psychology, 40, 195-196.

Lunzer, E. and Gardner, K. (1979) The Effective Use of Reading. Heinemann Educational Books.

MacGintie, W.H. (1961) Contextual constraint in English prose paragraphs. Journal of Psychology, 51, 121-130.

Mallinson, G.G. (1958) Textbook and reading difficulty in science teaching. The Science Teacher, 25, 474-475.

Manzo, A. (1970) Readability: a postscript. Elementary English, 47, (7), 962.

McCall, W.A. and Crabbs, L.M. (1925), (1950), (1961) Standard Lessons in Reading. New York: Teachers College Press.

McDonald, F.J. (1963) The Teacher and the Improvement of Educational Practice. In Human Learning in the School, ed. De. Cecco, J.P., Holt, Rinehart and Winston.

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

McLeod, J. (1962) The estimation of readability of books of low difficulty. British Journal of Psychology, 32, 112-118.

McNinch, G., Kaselskis, R. and Cox, J.A. (undated) Appropriate cloze deletion schemes for determining suitability of college textbooks. In Interaction: Research and Practice for College - Adult, ed. Nacke, P.L., Twenty-third yearbook of National Reading Conference, 249-253.

Meyer, B.J.F. and McConkie, G.W. (1976) What is recalled after hearing a passage. Journal of Educational Psychology.

Meyer, B.J. (1977) Organization in Prose and Memory: Research with Application to Reading Comprehension. In Reading: Research, Theory, and Practice, the Twenty-Sixth Yearbook of the National Reading Conference. ed. Pearson, P.D., Clemson, S.C.: National Reading Conference, 214-220.

Miller, L.R. (1974) Predictive powers of the Dale-Chall and Bormuth readability formulas. Journal of Business Communication, Winter, 11, 21-30.

Mobley, M. (1981) The Readability of School Text Books. Language for Learning, 2, (1), 11-19.

Moyle, D. (1971) Readability: the use of the cloze procedure. In The Reading Curriculum, ed. Melnik, A. and Merritt, J., University of London Press.

Mugford, L. (1972) A new way of predicting readability, Reading, 4/2, 31-35.

National Institute of Education (1976) Request for Proposal for a National Center for the Study of Reading. Washington, D.C.: Dept. of Health, Education and Welfare.

Neville, M.H. and Pugh, A.K. (1974) Context in Reading and Listening: a Comparison of Children's Errors in Cloze Tests. British Journal of Educational Psychology, 44, 224-232.

Niles, O.S. (1969) Reading skills common to the content areas. In Fusing Reading Skills and Content, ed. Robinson, H.A. and Thomas, E.L., 1-16. Newark, Delaware: International Reading Association.

Open University (1977) Developing Fluent Reading: Course PE231, Block 1, Unit 4. The Open University Press.

Pearson, P.D. and Johnson, D.D. (1978) Teaching Reading Comprehension. Holt, Rinehart and Winston.

Pearson, P.D. and Studt, A. (1975) Effects of Word Frequency and Contextual Richness on Children's Word Indentification Abilities. Journal of Educational Psychology, 67, 89-95.

Peterson, E.M. (1954) Aspects of readability in the social studies. Bureau of Publications, Columbia University, N.Y.

Pinnock, K. (1977) Personal communication.

Powers, R.D., Sumner, W.A. and Kearl, B.E. (1958) A recalculation of four readability formulas. Journal of Educational Psychology, 49, 99-105.

Preece, P.F.W. (1976) Associative structures of science concepts. British Journal of Educational Psychology, 46, 174-183.

Rankin, E.F. (1959) The Cloze Procedure - its validity and utility. Eighth Yearbook of the National Reading Conference, 131-144.

Rankin, E.F. (undated) Grade Level Interpretation of Cloze Readability Scores. Source unknown.

Rankin, E.F. and Culhane, J.W. (1969) Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 13, 193-198.

Rohwer, W.D. and Harris, W.J. (1975) Media Effects of Prose Learning in Two Populations of Children. Journal of Educational Psychology, 67, 651-657.

Rosen, H. (1967) The Language of textbooks. In Talking and Writing. ed., Britton, J. Methuen.

Rosen, H. (1972) The language of textbooks. In Language in Education - a Source Book, ed. Cashdan, A. and Grugeon, E., 119-125. London: Routledge and Kegan Paul and Open University.

Rumelhart, D.E. (1975) Notes on Schema for Stories. In Representation and Unstanding: Studies in Cognitive Science. eds. Bobrow, D. and Collins, A., New York: Academic Press.

Savory, T.H. (undated) The Language of Science. Reported in The Language of Textbooks by Rosen, H. (1972)

Schief, M. and Wood, R.W. (1974) A comparison of procedures to determine readability level of non-text materials. Reading Improvement, 11, 57-64.

Schlesinger, I.M. (1968) Sentence Structure and the Reading Process. The Hague: Morton.

Shavelson, R.J. (1974) Methods for examining representations of a subject-matter structure in a student's memory. Journal of Research in Science Teaching, 11, (3), 231-249.

Shayer, S.W. (1969) Relationships between reading interest and reading comprehension. In Reading and Realism 1968 proceedings, Vol.13, Part 1, ed. Figurel, J.A. Newark, Delaware: International Reading Association 698-702.

Smith, G. (1974) Analysis of "Introduction to Biology" (D.G. Mackean) University of Sussex - Centre for Educational Technology - Volkswagen Curriculum Analysis Project.

Smith, H.P. and Dechant, E.V. (1961) Psychology in Teaching Reading. New York: Prentice-Hall International Inc.

Spencer, H. (1969) The Visible Word: problems of legibility. London: Lund Humphries.

Stokes, A. (1978) The reliability of readability formulae. Journal of Research in Reading, 1, 21-34.

Stromnes, F.J. and Nyman, J. (1974) Immediate and Long-Term Retention of Connected Concrete Discourse as a Function of Mnemonic Picture-Type Sequence and Context. Scandanavian Journal of Psychology, 15, 197-202.

Sully, J. (1886) Outlines of psychology. New York: Appleton.

Taylor, W.L. (1953) Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 30, 415-433.

Taylor, W.L. (1956) Recent developments in the use of cloze procedure. Journalism Quarterly, 333, 42-48 and 99.

Taylor, W.L. (1957) Cloze readability scores as indice of individual differences in comprehension and aptitude. Journal of Applied Psychology, 41, (1)

Thelen, J.N. (1974) Using the cloze test with science textbooks. Science and Children, 12, (3), 26-27.

Thompson, M. (1976) Trait, State and Academic Test Anxiety: Their Relationship to Reading Performance. Unpublished Ph.D. dissertation reported in Teaching Reading Comprehension by Pearson, P.D. and Johnson, D.D., Holt, Rinehart and Winston (1978).

Thorndike, E.L. (1906) The principles of teaching. In Psychology and the science of education: Selected writings of Edward L. Thorndike, ed. Joncich, G.M., New York: Teachers College, Columbia University.

Thorndyke, P. (1977) Cognitive Structures in Comprehension and Memory of Narrative Discourse. Cognitive Psychology, 9, 77-110.

Tinker, M.A. (1965) Bases for Effective Reading. University of Minnesota Press.

Tulving, E. and Gold, C. (1963) Stimulus Information and Contextual Information as Determinants of Tachistoscopic Recognition of Words. Journal of Experimental Psychology, 66, 319-327.

Turner, A. and Greene, E. (1977) The construction of a propositional text base. Technical report, University of Colardo.

Underwood, G. and Holt, P. O'B. (1979) Cognitive skills in the reading process: a review. Journal of Research in Reading, 2, (2), 82-94.

Watts, L. and Nisbet, J. (1974) Legibility in Children's Books. Windsor: National Foundation for Educational Research.

Weaver, W.W. and Kingston, A.J.A. (1963) A factor analysis of the cloze procedure and other measures of reading and language ability. Journal of Communication, 13, 252-261.

Whitfield, R.C. (1979) Educational Research and Science Teaching. School Science Review, 60, (212), 411-430.
Williams, J.D. (1960) Teaching Problem-Solving. Educational Research, 3,

Williams, D.L. (1968) Rewritten Science Materials and Reading Comprehension. Journal of Educational Research, 61, 204-206.

Yngve, H.V. (1960) A Model and an Hypothesis for Language Structure. Proceedings of the American Philosophical Society, 104, (5), 444-466.