

**READING AS A PROBLEM
SOLVING ACTIVITY**

Trevor Sim

University of London, Institute of Education

**Thesis submitted in part fulfilment of the requirements for the degree
of Doctor of Philosophy, Spring 1994.**



ACKNOWLEDGEMENTS

I would like to thank the pupils and staff of the schools visited for the purpose of this study. I am indebted to Dr Richard Cowan for his help and encouragement, to Eileen Arnup, Margaret Reeves and Heather Sim who patiently typed successive drafts and also to those friends and colleagues who have by word or deed helped complete this work.

ABSTRACT

Reading comprehension in pupils in Years Four and Five was investigated in a series of experiments predominantly using error detection in short, narrative passages. The claim that readers of this age may exclusively accept decodability as sufficient to judge text as non-problematic was investigated. Little evidence was found in support of this claim. However evidence was obtained that consonant strings were more often detected than nonsense words and more nonsense words were detected than real words of inappropriate meaning when they were substituted into the same passages at the same points. Fewer real words of inappropriate meaning were detected when they were the same part of speech as the word they replaced than when they were a different part of speech. No evidence was obtained that significant numbers of children exclusively detected only nonsense words. However decreasing the readability of the passages significantly reduced the detection rate for real words of inappropriate meaning while the detection rate for consonant strings and nonsense words remained both higher and more stable. It was suggested that children who are asked to read passages of too low a readability for them may be more likely to exclusively employ a lexical standard of comprehension.

No evidence was obtained that asking children to read or listen to a passage a second time before completing an error detection task improved their performance. Moreover no difference in semantic comprehension monitoring was found to be dependent on whether the material was presented orally or visually. Better comprehenders were better than less good comprehenders on both error detection and cloze tasks. However there was no difference in the relationship between performance on prompted e.g. cloze, as compared to unprompted e.g. error detection,

comprehension tasks between better and less good comprehenders. Both groups performed better on the prompted comprehension tasks. Better performance was maintained on cloze tasks even when the subjects were not only alerted to having to read the passage for meaning but knew they were to be asked questions on it. The extensive use of unprompted comprehension tasks with feedback was proposed as a method of closing the gap between students' performance on unprompted as contrasted with prompted measures of comprehension.

Better comprehenders were better at sequencing sentences to make a story but did not perform better than less good comprehenders at recognising sentences from stories they had just read. Both better and less good comprehenders were less good at rejecting as having just been read sentences semantically congruent with the stories as contrasted with sentences semantically incongruent with the stories. This was consistent with most readers engaging in constructive processing of short stories.

The results of this series of experiments were compatible with and discussed in terms of comprehension involving the construction of a mental model of what is heard or read while listening to or reading short stories. Suggestions for further experiments were made.

CONTENTS

	<u>Page No.</u>
ACKNOWLEDGEMENTS	2
ABSTRACT	3
CHAPTER 1 Factors Influencing the Comprehension of Text	7
CHAPTER 2 Comprehension of Text at the Single Word Level	49
CHAPTER 3 Reading as a Piecemeal Process	64
CHAPTER 4 Standards for Evaluating Comprehension of Text	77
CHAPTER 5 The Influence of Reading Level on Error Detection	91
CHAPTER 6 Reading and Re-reading	105
CHAPTER 7 Do Readers Know When They Don't Understand?	117
CHAPTER 8 The Word Omission Task and the Syntactic Standard of Comprehension	133
CHAPTER 9 The Influence of Density of Substitutions on their Detection	142

CHAPTER 10	
The Relationship of Comprehension Level to Performance on Measures of Prompted and Unprompted Comprehension	151
CHAPTER 11	
Reading Comprehension and Comprehension of What is Read to One	160
CHAPTER 12	
Reading Comprehension and Memory	173
CHAPTER 13	
Summary, Conclusions, Implications for Teaching Comprehension of Text and Suggestions for Further Research	198
APPENDICES	
A Readability	214
B Materials Used	217
C A Review of Empirical Studies of Training to Enhance Comprehension of Text	239
D Relationship Between Error Detection and Performance on a Standardised Comprehension Test	283
E Controlling Individual Differences in Psychological Experiments	284
REFERENCES	287

CHAPTER ONE

FACTORS INFLUENCING THE COMPREHENSION OF TEXT

Reading is a form of problem solving; the reader has to solve the problem of what successively encountered words, phrases and sentences in a written text mean. (Daneman 1987)

Introduction

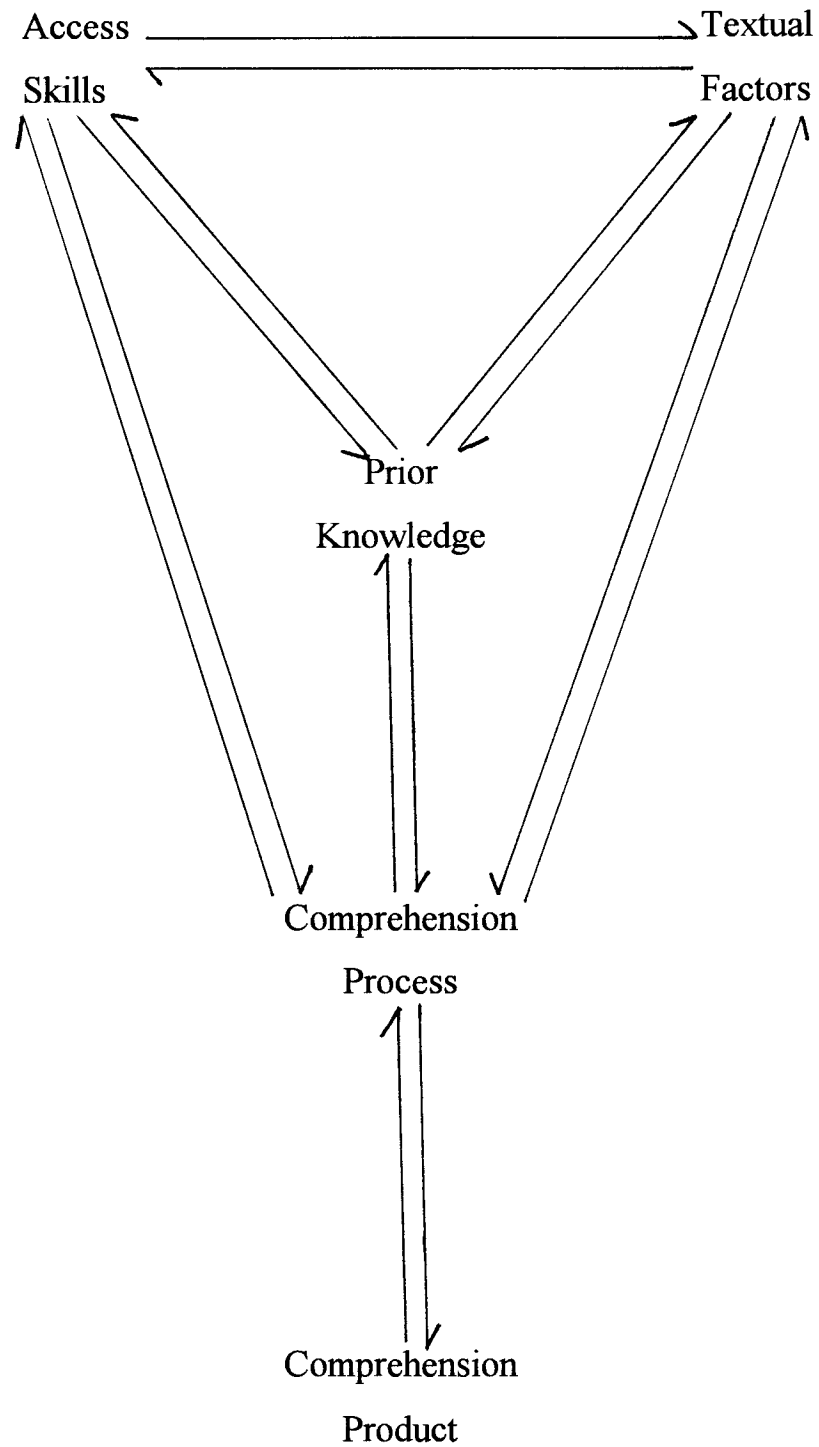
The standard of reading in British schools is a matter of considerable importance to the public. When in 1990 data was released by an educational psychologist (Times Educational Supplement 29 June 1990) purporting to show a decline in reading standards it received national radio, T.V. and press coverage. Responding to the considerable public concern, which these claims aroused, the then Secretary of State for Education commissioned a survey on reading standards. This and other opinion was then considered by the Education, Science and Arts Committee of the House of Commons who published their own report (Standards of Reading in Primary Schools 1991), which concluded that the claim that reading standards had fallen in recent years had not been proved.

Underlying this concern was the belief that reading is still an extremely important skill, possession of which gives independent access to a wide range of information which would not otherwise be independently accessible by the individual. Reading in this sense refers to more than the recognition of individual words in context e.g. the EXIT sign above a door. The reading which is the focus of this thesis is that of understanding the meaning of written words arranged in at least a sentence. That this is regarded as important even at an early stage of reading is reflected by the inclusion of a

comprehension test for children who reached Level Two in Reading in the National Curriculum Standard Assessment Tasks (SATS) carried out for the first reported assessment at the end of Key Stage One in 1992. Though reading comprehension can be assessed as early as age seven a significant minority comprising 28 per cent of pupils who took the 1991 SATS at the end of Key Stage One did not achieve the standard of reading accuracy required to 'read aloud a simple unseen piece of text with reasonable accuracy' (D.E.S. 1991). There is a body of British reading comprehension research (Yuill and Oakhill 1991) based on children of average reading age 8 years 1 month on the Neale Analysis of Reading Ability. However given Oakhill's (1993b) caution that 'instruction in skills such as comprehension monitoring may need to be restricted to children who have learned the "basic skills" of reading' the studies reported in this thesis will concentrate on children aged between eight and ten. Moreover it is within this age range that the majority of children will be on or working towards Level 3 of the Reading Attainment Targets in the English National Curriculum which emphasise making inferences and deductions from stories and using written material as a source of evidence.

Though the emphasis of this study is on comprehension of what is read rather than on the reading of the words at some level the child's access to the text and his knowledge of the subject matter about which he is reading will influence the comprehension process. These influences may interact in complex ways as summarised in Figure 1.1 and they will be discussed in the sections that follow.

Figure 1.1 A diagrammatic representation of the major factors influencing comprehension of text



Access to the text

i) Access skills

The way meaning is represented on a surface has varied in form across time and between cultures. This study is not concerned with written languages where each word is represented by its own distinctive picture or logogram, but with one particular written language English in which sounds are represented by letters of an alphabet. The sounds represented by combinations of letters in English are neither consistent nor unique. For example, the **f** sound in fish is represented by **f**, but in enough by **gh**, while the **ough** letter group in bough makes the same sound as the **ow** group in bow, but not when you shoot an arrow from it. English may be like this because unlike some other European languages, it has evolved over many hundreds of years and avoided the systematic overhaul of its spelling system to which for example, French in 1736 and 1990, German in 1901 and Spanish in 1959 have been subjected. This may have contributed to the irony identified by Upward (1992) that English university students studying German made fewer spelling errors in German than English in their exams.

A word in English therefore, has characteristics which can be described in terms of the alphabet, letter sound relationships, the overall shape of the word, its frequency of occurrence in English, the syntactic constraints operating in the text of which it is part and the meaning it contributes to the text.

Approaches to teaching reading have often emphasised the reader focusing on one particular characteristic when reading. Approaches to the teaching of reading include the alphabetic, the phonic, the whole word, the sentence and the 'real books' approach.

The alphabetic method assumed familiarity with the form and names of letters would help children to recognise and pronounce words. It was felt constant repetition of letter names in spelling out words would not only enable the learner to become familiar with the letter names but also to recognise letter strings. The main emphasis was on the recognition of new words rather than understanding their meaning (Goodacre 1972). It was still common in the United States in the first half of the nineteenth century for children to have to achieve a certain standard of spelling before they were taught to read (Fitzgerald 1990). This practice continued as an alternative mode of teaching in some American primers as late as 1893. Even by the 1950's, there were still at least 4 schools in Kent using this method (Diack 1965).

As Diack (1965) has pointed out, it is not very far from teaching the names of the letters to teaching their sounds and the phonic approach to teaching reading relies on teaching the correspondence between letters and letter combinations and sounds. One problem with this method is a tendency to add 'uh' so that when the letters are blended together, distortions are introduced so 'man' sounds more like 'manner'. Certainly, too rigid an adherence to a strategy of single letter sound correspondence can lead to phenomenon such as children reading 't' 'h' 'e' 'the'.

Limitations of the phonic approach such as those outlined above caused some theorists such as Schonell (1961) to argue in favour of using introductory reading books with a strictly controlled vocabulary for beginning readers. By constant exposure to the same words, it was theorised that word shapes would be recognised. While this method has virtues within a reading scheme with a strictly controlled vocabulary or in the reading of common words, it offers little help in deciphering words

unfamiliar to the child. Though Carroll, Davies and Richman (1971) in a study of American school books from Grade 3 to Grade 9 found 50 per cent of their sample of over 5,000,000 words consisted of just 109 words nevertheless, the total number of different words used was over 86,000 of which virtually all must have been expected to be understood by the books' readers. It would be very laborious, even if possible, to teach the recognition of all of those words by sight.

An extension to the whole-word approach is the sentence method. This approach emphasises more the word's meaning within a sentence, and tends to make use from the beginning of reading material made up of groups of words which make sense. It was hoped that the use of continuous prose led to children reading more fluently and rapidly. Repetitions, stumbles and errors when reading out loud were seen as providing information as to the reader's level of skill.

The real books and language experience approaches make even greater use of continuous prose with a particular emphasis on material which is relevant or topical for the reader. It frequently involves the reader in the production of his or her own reading material, for example, by the child dictating his 'news' to the teacher who writes it out for the child to read and/or copy and/or illustrate. The real books approach is often associated with the work of Goodman (1982) and Smith (1973). They argued that reading was a psycholinguistic guessing game involving an interaction between thought and language. Smith (1973) stated 'that sound, if it is produced at all comes only after the comprehension of meaning in reading.' and that 'decoding skills are used only to a very limited extent, and then primarily because a good deal of instructional effort is expended on impressing such methods upon children.' Amongst evidence he referred to was that Smith (1969)

found under conditions of reduced lighting adults more quickly identified the first letter in a highly predictable sequence than in a less predictable sequence. He argued that 'if the identifiability of a letter varies with the identifiability of the word in which it occurs, then word identification cannot depend on preliminary letter identification.' (Smith and Holmes 1971). Instead he claimed readers 'mediate the identification of individual letters by discriminating properties of the sequence as a whole.' (Smith and Holmes 1971).

There is however considerable controversy over the use of the real books approach. In the debate over reading standards referred to earlier, the educational psychologist who provided the press with the test results purporting to show a decline in standards, also attributed blame for this to teachers using the real books approach (Turner 1990, 1991). The Education, Science and Arts Committee of the House of Commons found no evidence to attribute any overall decline in reading standards to the real books movement. They noted HMI had concluded children taught wholly or mainly by the real books approach were likely to make poor progress, but that this approach was only used by around 5 per cent of teachers.

Oakhill and Garnham (1988) also questioned the theoretical basis for the 'real books' approach. They noted that Smith's (1973) argument that as adults can access meaning directly, so can children, was flawed as the research he quoted is based on experiments with text made physically difficult to read. Under these unusual circumstances, adults do make informed guesses, but Oakhill and Garnham concluded that there was no evidence that they do so to help them read clearly printed words during everyday reading.

Oakhill and Garnham were not arguing that prediction from context was unimportant in reading, but they observed that there must be some decoding of the printed text so that guesses could be confirmed or unconfirmed. Adams (1990), following an extensive review of methods of teaching initial literacy arrived at a similar conclusion but with a different emphasis. She concluded that skilled readers visually process virtually every letter but that they are not perceived independently of each other. Instead, she believed their efficient perception depended on word knowledge and context. That in general, word processing skills are important in reading is supported by the finding of Evans and Carr (1985), that of children in their sample starting school with greater than average language skill, those whose reading curriculum largely avoided analytic word processing skills ended the year with poorer reading comprehension i.e., negative correlations were obtained between four out of six language measures and reading performance. Evans and Carr concluded that 'Development of efficient print-specific skills endows a beginning reader with the resources to use knowledge and inference flexibly rather than slavishly.....' (Evans and Carr 1985 P.346).

Though these authors arrived at similar conclusions, it would be wrong to imply there was a consensus. For example in Adams' own book, there is a note of dissent from two of those empanelled to advise Adams by the Centre for the Study of Reading who commissioned the book. In particular, Professors Cullinan and Strickland were concerned to emphasise the developmental nature of learning to read and that phonics was best learned in the context of reading and writing and in particular, each day reading good stories that engaged the reader. In essence, their concerns related to 'not if, but how and when spelling sound information is made available to learners' (p.433).

It would also be wrong to suggest there was a unified view of what constitutes phonological skills. Traditional phonics is concerned with grapheme/phoneme relationships. However, Bryant and his co-workers (Bryant and Bradley 1985, Goswami and Bryant 1990) have distinguished what they claim to be the separate skills of segmenting syllables into initial sound or onset and the remainder or rime. They argue that knowledge of these onsets and rimes can be generalised through analogical reasoning to enable new words to be read or written.

The view advanced here of reading comprehension proceeding through access to the text is consistent with the outcomes of component skills analyses of reading. These identify word recognition and language comprehension as the major components of reading comprehension (Levy and Carr 1990). There is some evidence that decoding is the more important for younger children (Forrest-Pressley and Waller 1984, Hoover and Tunmer 1993). A number of authors have emphasised the importance of rapid, fluent or automatic decoding for comprehension (LaBerge and Samuels 1974, Perfetti 1985). These theories assume that decoding and comprehension are competing for a fixed amount of processing capacity and efficient decoding would leave more capacity for comprehension. There are three reasons why differences in decoding skill are unlikely to provide a complete explanation of children's comprehension of text. Attempts to enhance comprehension through training in rapid decoding have not proved successful (Fleisher, Jenkins and Pany 1979, Yuill and Oakhill 1991). Decoding is itself influenced by context as children read the same words in stories more accurately than from lists (Goodman 1965, Nicholson 1993). Finally Cromer (1970) and Yuill and Oakhill (1991) have identified groups of readers matched on reading accuracy who nevertheless differed in comprehension while Snowling and Frith (1986) have identified 'hyperlexic'

children with advanced decoding skills but poor comprehension. The existence of this latter group of children indicates that single word reading skills can develop to some extent separately from comprehension. Though some decoding is necessary for comprehension and proficient decoding is likely to facilitate comprehension this does not preclude how (or how well) text is written from influencing comprehension.

ii) Textual Factors

In addition to the individual's reading accuracy the readability of the text in terms of the words used (Marks, Doctorow and Wittrock 1974), the grammatical features of words within clauses and sentences (Coleman 1968), sentence length (Harrison 1977) and the structure of the text (Thorndyke 1977) all affect access to and so ultimately comprehension of the information in the text.

The relationship between textual factors and readability is not simple. For example Beck, McKeown, Omanson and Pople (1984) found increases in recall and marginally significant increases in comprehension of two stories they had modified to improve coherence. The modification they made increased the total number of words, average sentence length and number of polysyllabic words. The relationship may also not be one way in that Hare, Rabinowitz and Schieble (1989) have suggested that students taught only using basal readers will have difficulty transferring skills such as main idea comprehension to naturally occurring texts. Nevertheless the significant point is that modifications to the text of a story can influence how well it is understood.

The Comprehension Process

i) Introduction

In a situation where access to the text is not problematic, the question of what is unique if anything about reading comprehension as distinct from comprehension of what is read to someone should be addressed. Sticht and James (1984) concluded following a review of studies of listening and reading that listening and reading build upon a common language and knowledge base and that a person's listening comprehension skills establish their reading potential. There are however differences between reading and listening which may account for the generally superior performance of younger children on listening comprehension as compared to reading comprehension tasks (Sticht, Beck, Hauke, Kleiman and James 1974). Listeners and speakers often share a common temporal or spatial context so that nods, looks and pointing together with facial expression and context dependent references such as now, over here, this big can convey significant information (Rubin 1980). Moreover prosodic cues can emphasise salient aspects of discourse and facilitate understanding. In contrast readers may have the opportunity to reread or pace their reading of text. If access to text is not problematic and prosodic and deictic factors are controlled one would predict no difference in outcome dependent on whether the material was read or heard and this prediction will be empirically evaluated. However it would be important to ensure that the readers and listeners knew the purpose of them reading or listening was to make sense of what they read or heard.

ii) The purpose of reading

Forest-Pressley and Waller (1984) provided some evidence that how children read is related to their perception of the purpose for which they are reading. Forrest-Pressley and Waller investigated whether children's ability

to answer 14 comprehension questions on hard or easy 500 word texts differed depending on whether they were told they were reading: (a) for fun; (b) to decide on the title for the passage; (c) to skim the passage to find the answer to a special question or (d) to study the passage. The subjects were 105 third grade (8 year olds) and 122 sixth grade (11 year old) children attending one of three suburban schools in Waterloo, Ontario. From this pool, 72 third graders and 72 sixth graders were selected. Twenty four in each age band were good readers, 24 average readers and 24 poor readers. When reading easy passages, 8 year old readers retained more information in the study condition than in the skim condition, and more information in the fun than in the skim condition. Eleven year old poor readers, performed better in the study than in the skim condition when reading hard passages, while eleven year old good readers performed better in the study condition than in the skim condition when reading easy passages. Generally, the children did best in the study condition and least well in the skim condition; the results of the fun and title condition being hard to interpret unambiguously. It would seem likely from this result that children can use different strategies when reading.

Kieras (1981) found variation in reading times for text when students were asked to either just read, read to devise a topic title or read to later recall the text. The students took longer when reading to recall than when reading to devise a topic title for which students in turn took longer than when just reading. This suggests that different reading processes were in operation. Masson (1982) also provided evidence that students encouraged to read at a normal rate (about 225 words per minute), a skimming rate of 375 words per minute or a fast skimming or scanning rate of 600 words per minute recalled or recognised less as reading rate increased.

Carver (1990a) distinguished five basic reading processes which are tabulated below. This study is however principally concerned with normal everyday reading for meaning; which corresponds to the reading process Carver calls rauding.

Table 1.1 Carvers' Five Basic Reading Processes

Process	Goal	Objective Consequences	Rate for College Students
Scanning	Find target word	Correctly identifying target word	600 wpm
Skimming	Find anomalous word	Correctly identifying anomalous words in passage	450 wpm
Rauding	Understanding the complete thoughts the writer intended to communicate	Correctly identifying incomplete thoughts or anomalous sentences	300 wpm
Learning	Know the information	Answer multiple choice questions on the passage	200 wpm
Memorising	Recall the facts	Write down exact words from the passage	138 wpm

Wpm = standard words, ie., six character spaces per minute.

If children do have a variety of reading processes or strategies open to them, then it becomes important that they are made aware of the purpose for which they are reading. Failure to do this may well prejudice their performance. For example, Paris and Myers (1981) asked two groups, one of 16 good readers, and one of 16 poor readers, all fourth graders, to underline words or phrases they did not understand in a story appropriate for their reading level. Each story was modified by replacing two nouns with phonologically

acceptable nonsense words and by rearranging the words within two clauses to produce non-meaningful phrases. The children were told underlining might help answer the questions about each story, ie., the underlining was possibly seen as a means to an end, rather than an end in itself. Moreover, the children were not told to expect nonsense words. Only a quarter of the nonsense words were underlined by good fourth grade readers and even fewer by the poor fourth grade readers. In contrast, Baker (1984b) as one part of an experiment asked 54 fourth graders to underline anything they thought was a problem in the passages they were reading. Half of the children were told the exact nature of the problems and given two examples of each type. The better fourth grade readers, underlined 71% of nonsense words if given specific instruction, and only 33% if not, while poorer fourth grade readers underlined 38% of nonsense words if given specific instruction and 42% if not. Whilst generally in Baker's study, more nonsense words were underlined than in Paris and Myer's study, nearly three times more were underlined by the good readers given specific instruction to look out for them in Baker's study than by good readers of similar age in Paris and Myer's study. The important point is that knowing clearly what was expected of them may well have enhanced the children's performance.

Markman and Gorin (1981) examined the ability of 7 year old and 9 year old pupils, to detect falsehoods or inconsistencies while listening to brief passages. In each age group, a third of the children were only given brief instructions, a third were given brief instructions and examples of falsehoods and a third were given brief instructions and examples of inconsistencies. Those children given specific instruction to find falsehoods found significantly more than those children set to find only non-specific problems. Those given specific instructions to find inconsistencies found significantly more than those only given non-specific instructions. Children set to find

inconsistencies did so significantly more often than those set to find falsehoods and vice versa. It appears therefore, that children can adjust their standard of evaluation of text read to them. It is also clear, given the low rate of detection of comprehension failure of under 50% in this study as well as in the earlier studies, that even given specific instruction and practice in the detection of one error type, children of primary age clearly find detecting errors very difficult.

As students' performance can be altered by altering their knowledge of the purpose of the reading they were being asked to carry out, then closely allied to this is the readers interest or motivation to perform optimally. Lunzer and Gardner (1979) report a study by Shnayer (1969). He took a group of about 500 11 to 13 year olds and on the basis of their performance on a test of reading comprehension, split them into seven ability groups. He then gave each subject a number of comprehension tests on passages which had been assessed as being two grade-levels, above the child's own reading age. The children were also asked to rate the interest level of the passages. One important finding of this study was that there was (with the exception of the lowest ability group), no significant differences between the groups in comprehension of stories rated as high interest. Unfortunately, unambiguous interpretation of this result is not possible as children may prefer a passage they have understood well to one they understood less well.

Asher (1980) summarised a series of experiments he and his colleagues carried out to investigate the relationship between interest and reading comprehension. These involved a procedure whereby children's interests were assessed by having them rate each of 25 colour photographs representing a wide array of topics on a seven point interest scale. No mention of reading was made during the interest assessment phase. A week

later, the child received from a second experimenter, a set of six passages: three corresponding to the child's three highest rated topics and three corresponding to the child's lowest rated topics. All of the passages came from the Britannica Junior Encyclopaedia and were in cloze format with every fifth word missing. The child's task was to read the passage and supply as many of the missing words as possible. In the three studies of fifth and sixth grade children, there was a greater desire to read passages on highly rated topics. The children's comprehension was superior on high-interest passages. In two of the studies, boys' performance was facilitated more than girls' performance by high-interest material. This result would suggest there is an interaction between a reader's interest and their comprehension of the text.

iii) Standards of comprehension

Comprehension of text may not be a unitary construct. Collins and Smith (1982) identified four different types of failure to comprehend a text.

1. Failure to understand a word either because it is new or doesn't make sense in the context.
2. Failure to understand a sentence, either because one can find no interpretation, or can only find a vague interpretation or there are several possible interpretations.
3. Failure to understand how one sentence relates to another either because the interpretation of one sentence conflicts with another, or one can find no connection between sentences or that there are several possible connections between the sentences.
4. Failure to understand how the whole text fits together, either because one can find no point to all or part of the text, or one cannot understand why

certain episodes occurred, or one cannot understand the motivations of certain characters.

Baker (1985a) argued that there are three standards that readers use to evaluate their understanding of the text: lexical, syntactic and semantic. The lexical standard operates at the level of an individual word where context can be ignored, for example, realising the meaning of a word is unknown to the reader. The syntactic standard judges conformity to the grammatical constraints of a language, for example, realising the order of words in a sentence is jumbled. The semantic standard requires that the passage is continuous and is unambiguous and complete.

Barrett (quoted in Clymer 1972) devised a five category taxonomy of reading comprehension:

- a) literal comprehension
- b) reorganisation
- c) inferential comprehension
- d) evaluation
- e) appreciation

later reduced to 4 categories by the elimination of b) reorganisation (Smith and Barrett 1974).

His literal category would involve at the simplest level, locating a name of a character whilst his appreciative category would involve sensitivity to the emotional, aesthetic, psychological and artistic elements of a work. In other words, the spectrum of comprehension categories covered in Barrett's taxonomy ranges from identifying single words to literary criticism.

These are only three of numerous classification systems. In that they all segment reading comprehension into categories they have implications for the measurement of comprehension. Many tests of comprehension do not make any distinction between categories of comprehension. For example, a frequently used individual measure of reading comprehension is the Neale Analysis of Reading Ability. Both in its second edition (Neale 1966) and new revised edition (Neale 1989) it does not differentiate between categories of comprehension. Though used by Yuill and Oakhill (1991) as a measure of inferential comprehension, it also contains questions which deal only with surface features of the text. For example, in one story about a surprise parcel which arrived one Saturday, one of eight questions on that story asks on what day the parcel was delivered. As no day other than Saturday is mentioned, purely lexical monitoring would have enabled this question to be correctly answered. In contrast, another question asked how the reader knew the recipients were not expecting the parcel. This requires them to infer that as it was a surprise, they had no prior knowledge of its arrival. Most measures of comprehension used with primary and young secondary pupils are of this mixed variety giving a global measure of the reader's lexical, syntactic and semantic understanding of the prose. The evaluative and appreciative elements whilst by no means being absent from the primary curriculum, are notable by their absence from normative measures of reading comprehension for this age range.

The standard of comprehension of particular interest to this study is the semantic standard. This involves integrating the meaning of the sentences which comprise the text. The inappropriate application of a lexical or syntactic standard of comprehension is likely to lead to inadequate semantic comprehension. In particular the typical classroom practice of monitoring primary aged children's reading by listening to them read may

overemphasise word pronunciation at the expense of passage comprehension (Frederiksen 1979, Goodman and Goodman 1979, HMI 1989).

This idea of reading each word with little or no reference to what has gone before, or comes after, as a reason for poor comprehension is not new. In 1908, Huey in his *Psychology and Pedagogy of Reading* wrote 'until the insidious thought of reading as word-pronouncing is well worked out of our heads, it is well to place the emphasis strongly where it really belongs, on reading as thought-getting independently of expression.' (page 350). Thorndike (1917) also noted 'The vice of the poor reader is to say the words to himself without actively making judgements concerning what they reveal.' Thorndike regarded reading as 'a very elaborate procedure, involving a weighing of each of many elements in a sentence, their organisation in the proper relations one to another, the selection of certain of their connotations and the rejection of others, and the co-operation of many forces to determine final response. In fact, we shall find that the act of answering simple questions about a simple paragraph includes all the features characteristic of typical reasonings.' (Thorndike 1917 p323)

The possibility that some children experience poor comprehension because they read each word with little or no reference to the rest of the text continues to be advanced (Baker 1985a, Cromer 1970, Garner 1981, Isakson and Miller 1976, Paris and Myers 1981). The most direct evidence of children reading word by word is provided by Baker (1984b).

Baker (1984b) prepared 15 passages so that 12 of them contained a single word which was either a nonsense word, violated common world knowledge or created an internal inconsistency. Baker individually asked 54 nine year olds and 54 eleven year olds to silently read the passages and underline any

problematic sections as they went along. The subjects had previously been selected as better or less good readers on the basis of their reading scores on the California Achievement Test. Half the children were only given general instructions while half were given specific instructions and two examples of each type of problem.

Better readers identified more problems than less good readers and eleven year olds also identified more problems than nine year olds. Children receiving specific instructions about the problems identified more than those children only given general instructions. More nonsense words than either prior knowledge violations or internal consistencies were identified. Thirty five percent of nine year olds and seventeen percent of eleven year olds identified only nonsense words. This latter result suggested that one in three nine year olds in this study were only applying a lexical standard of comprehension. However the younger children missed about 50 per cent of the nonsense words and in her commentary on this experiment Baker (1985a) speculated that these children may have been applying a decodability standard i.e. accepting any word they could sound out as being non-problematic.

The first experiment in this study will investigate whether nine year old readers apply a standard of decodability and whether a significant proportion employ this standard exclusively. The following two experiments will investigate children's application of lexical, syntactic and semantic standards of comprehension. Baker (1984b) found nonsense words which can be identified by the application of a purely lexical standard were more frequently identified than prior knowledge violations or internal inconsistencies which for their detection would require the application of a semantic standard. However in this experiment Baker's students identified

relatively few nonsense words. For example 11 year olds in her 1984a study identified 94 percent of nonsense words as compared to 60 percent in her 1984b study. It may be the passages used by Baker (1984b) were too hard for many of her subjects. However the detection of nonsense words which can be achieved through the application of a lexical standard is likely to be less strongly influenced by giving students too hard passages than the detection of real word substitutions which would involve the use of the surrounding context. If fewer real word substitutions whose identification requires the application of a semantic standard are detected than nonsense words and giving children harder versions of passages decreases their detection rate for real word errors more than nonsense words then a possible alternative explanation for the large number of students whom Baker (1984b) found to be only employing a lexical standard of comprehension would be that the passages were too hard. The fourth set of experiments in this study will investigate the effect of changes in passage readability on the detection of different classes of substitution.

iv) Prior knowledge

One factor in the absence of the appreciative category from normative measures of reading comprehension for primary age pupils may be that while most categorisations of reading comprehension acknowledge the role of prior knowledge, this is much more evidently a factor when one is contrasting two stories or putting them within a social or literary context.

The following excerpt from Schank and Abelson (1977), illustrates the role of knowledge in interpreting a passage. 'While giving his order to the waiter, at Mama Leone's one evening, Spillane was approached by the owner, a notorious Mafia figure'. Our knowledge of both life (general knowledge) and of the Spillane genre (specific knowledge), enable us to

flesh out this scene to convey much more than the few lines actually say. We can guess that it is an Italian restaurant at which he is known, that he's probably sitting at the table and ordering pasta and something of significance is about to pass between Spillane and the restaurant's owner.

Owings, Petersen, Bransford, Morris and Stein (1980) investigated recall and found information consistent or incongruent with past experience is harder to recall for both better and poorer fifth-grade students. It is notable that despite this, only the better students identified less congruent passages as difficult to learn and adjusted their study times accordingly. But if they were specifically asked, the poorer students could identify whether the passages made sense e.g., they could recognise it is more sensible for a hungry boy to eat than to go to bed. Ceci, Caves and Howe (1981) investigated seven and ten year olds' memory for information inconsistent with their knowledge. For example, "The Six Million Dollar Man" was said to be too weak to carry a can of paint. A story containing inconsistent information on characters familiar to children was read to 14 seven-year-olds and 15 ten-year-olds. Some of these children were asked to rate the characters immediately after hearing the story and some were asked to rate them three weeks later. Those asked for immediate ratings showed no systematic distortion towards prior knowledge. However, those asked to rate the characters three weeks later, displayed considerable shifts in memory in the direction of their pre-experimental knowledge of the characters' attributes. This suggests prior knowledge influences long-term memory. Recht and Leslie (1988) investigated how prior knowledge influenced the ability of seventh and eighth grade students of both high and low reading ability to recall or summarise a passage on baseball knowledge. Students with greater knowledge of baseball recalled more than students with lesser knowledge. Moreover, students with high reading ability and

high knowledge did not recall more, or summarise better than students with low reading ability and high knowledge.

Theories of Comprehension

i) Introduction

A number of factors have been identified as influencing comprehension. A problem is that these factors may interact in a complex way. For example, on encountering what Armbruster (1984) refers to as inconsiderate text a reader may lose interest and fail to persevere or alternatively may shift down to a slower reading rate involving rereading and so preserve comprehension. This latter strategy is likely to be more successful in the presence of prior knowledge of the subject matter of the text. This does not imply a strength in one area will necessarily by itself enhance comprehension, nor that a weakness in one aspect of reading will necessarily result in a weakness in comprehension. For example Yuill and Oakhill (1991) gave case studies of children who read accurately but who were able to answer fewer questions correctly than other children who read the same passage haltingly and made more mistakes.

No unified theory of reading comprehension which takes into account all the factors identified above has been developed. However one approach to reading that emphasising metacognition (e.g. Garner 1987) has features which may make its use productive.

Metacognitive approaches to reading arose from work on metamemory pioneered by Flavell (Flavell, Friedrichs and Hoyt 1970, Flavell 1971). Flavell (1985) defined metamemory as 'an individual's knowledge or

cognition about anything pertaining to memory (e.g. that certain kinds of information are harder to learn and remember than others)'.

Research by Flavell, Friedrichs and Hoyt (1970) on developmental changes in the memorisation process indicated that only older children of 7 and 9 years of age used a specific strategy over time. Older children were also far better than younger children of 4 and 6 at estimating their own memory span in advance and also at assessing their recall readiness after attempting memorisation.

Flavell and Wellman (1977) constructed a model of what a person might come to know, or know how to find out concerning memory in the course of cognitive growth and learning experience. They identified two categories of metamemory:

1. Sensitivity to the need for effort to prepare for future retrieval.
2. Knowledge of what variables interact to affect how well an individual will perform on a particular retrieval problem.

Their model also proposed three classes of variable within the latter category of metamemory: person, task and strategy:

- i) Person - what the person is like as a rememberer.
- ii) Task - what the memory problem is like.
- iii) Strategy - what behaviour is engaged in to remember.

Research into these areas has informed research in other domains. For example, the three classes of metamemory variables proposed by Flavell and

Wellman (1977) were used by Myers and Paris (1978) to investigate children's metacognitive knowledge about reading.

Myers and Paris interviewed children in the second and sixth grade on their metacognitive knowledge about reading. Person, task and strategy knowledge was assessed. Compared to the sixth grade readers, the younger readers did not know:

- a) readers have special skills
- b) motivation is linked to reading performance
- c) reading silently is faster than reading aloud
- d) the first and last sentences of a paragraph are particularly important
- e) retelling a story is more efficiently done by conveying the gist than by verbatim recall
- f) skimming is reading the words that yield most information
- g) rereading is an important strategy for resolving comprehension failure.

The younger children also focused on decoding rather than comprehension. Whilst there may be some debate as to whether items a) to g) do represent veridical knowledge about reading, this study nevertheless suggests that older children's metacognitive knowledge about reading differs from that of younger children. In particular, Myers and Paris found that 39 per cent of second grade students could not report any strategy to employ when a sentence was not understood. In a later study of good and poor fourth grade readers, Paris and Myers (1981) found that poor readers engaged in significantly less monitoring behaviour and were often unaware of the negative influences of some reading strategies. Comprehension monitoring and active control of the reading process are key issues within the metacognitive approach to reading.

The skills of selecting and monitoring the reading process are a subset of those skills identified by Brown (1978) as skills used in controlling and coordinating deliberate attempts to solve problems. These include:

1. predicting the system's capacity limitations
2. being aware of available strategies and their usefulness
3. defining the problem
4. planning appropriate problem solving strategies
5. monitoring the effectiveness of those strategies
6. evaluating operations in the face of success or failure.

Brown suggests these skills are the basic characteristics of thinking efficiently in a wide range of learning situations. A metacognitive approach to reading would regard it as an active thinking or problem solving process.

In addition to providing a framework for examining reading as a problem solving activity, the metacognitive approach to reading has utilised as its dominant research technique error detection. Error detection differs from other measures of comprehension such as cloze, sentence completion, short questions or multiple choice in which the element where meaning is lost or needs to be constructed is identified for the reader. Error detection, which requires the reader to identify loss of comprehension may offer a more satisfactory measure of the reader's habitual standard of comprehension when reading material such as newspapers, magazines or stories. That is error detection may offer a method of assessing comprehension of reading material as it is used outside of academic situations, e.g., in everyday life.

Winograd and Johnson (1982) believed the error detection approach 'is at its best when it is used to assess a reader's ability to overtly report the effects of

embedded errors'. They did however, also identify 12 possible alternative explanations other than that the reader's comprehension monitoring abilities are poor. These were:

1. The reader's lack of relevant background knowledge may cause him to overlook the error.
2. Readers may suspend disbelief because they have read so much that is unbelievable.
3. Older readers may have an overriding faith that speakers and writers intend their messages to be truthful, relevant and unambiguous.
4. The readers may not believe that texts can, and do contain errors.
5. Readers may make inferences that seem to resolve the errors and discrepancies.
6. Young children may not recall inconsistent information presented earlier.
7. Subjects may lack the logical capacity to make the necessary inferences.
8. Subjects may be hesitant to criticise the experimenter in a test situation.
9. Subjects may draw upon prior knowledge to supplement explicitly presented information.
10. Subjects may assign alternative meanings to the text.
11. Subjects may assume the writer has made a mistake and ignore it.
12. Subjects may notice the error but assume that subsequent information will resolve the problem.

Except for the logical capacity explanation the other alternatives are amenable to experimental control. For example, the first series of experiments reported in this study will use the error detection task to investigate the application of different standards of comprehension. As potential alternative hypotheses can be made to bear equally on the detection rates for different classes of errors the differences that emerge should relate

to the class of errors. The influence of readability on different standards of comprehension will also be investigated.

The particular standard of comprehension of interest to this study is the semantic standard, ie., judgements as to the internal and external consistency of the text. This can only be judged in relation to the text as a whole, ie., it involves intersentential judgements. It is difficult to conceive of an efficient system for making these which does not rely on the construction or use of an overarching framework. For example, a sequence of separate judgements only as to the consistency of adjacent sentences would fail to detect the inconsistency in the short story:

Jack was all alone on the hill. He had gone up the hill to fetch a pail of water. Some water slopped out of the pail. Jill slipped on the water and fell against Jack.

A number of different frameworks have been described in the literature. One involves the use of schema and another the construction of mental models. Schema theorists have proposed that comprehension involves the instantiation of a pre-existing schema using information from the text. The instantiation of an existing global schema cannot however explain how readers understand texts about unfamiliar objects and events (Brewer 1987, Johnson-Laird 1983). The second more general approach which can accommodate novel situations described in text is to build a mental model of a situation from information in the text (McNamara, Miller and Bransford 1991). Garnham (1987) has demonstrated that readers can choose to emphasise schematic or mental model processing. When subjects in Garnham's experiment were not warned they would have a memory test, they seemed to use mental model processing as they were unable to

distinguish whether they had heard a sentence in the story or a semantically identical one. Subjects warned about having a memory test had no trouble with making these distinctions which suggests that their text representations were similar to the text structure.

ii) Yuill and Oakhill (1991)

A recent mental model theory of reading comprehension based on experimental work carried out in schools in England is that of Yuill and Oakhill (1991). They advanced a model of reading comprehension which involved the reader building 'a mental model of the situation described in a text' (p13) by recognising individual words, understanding the grammatical and semantic relations between the words, integrating the ideas in the text and making inferences to aid this integration as they go along.

Yuill and Oakhill (1991) were particularly interested in children who were fluent readers but showed little evidence of understanding what they read.

The methodology they employed in their supportive studies usually involved comparing good comprehenders with poor comprehenders of seven to eight years of age with whom they were matched for decoding and vocabulary skills. The test they used for measuring reading accuracy and comprehension was the Neale Analysis of Reading Ability Second Edition (Neale 1966). The comparison groups were matched on vocabulary, accuracy of reading and age whilst differing on comprehension score. It is therefore not unexpected that they found no difference between the two groups in speed and accuracy of decoding or vocabulary. Nor is it surprising they found little benefit from training in rapid decoding. Indeed, as they themselves acknowledge, it would have been surprising if they had 'considering the manner in which the groups were selected' (p52).

The Neale Analysis also differs from a similar test the New MacMillan Reading Analysis (Vincent and de la Mare 1985) in that after reading the passage, it is unavailable for inspection when questions are asked on the passage. The reader is alerted to the need to remember the story at the start of the test. It would follow that one would expect the good comprehenders on this test to do well on tasks involving working memory. This was found in an experiment in which children had to recall the final digits of each set of three numbers making up lists of variable length with each set of three numbers being presented one at a time. Less predictable was the finding of no difference between the more and less skilled comprehenders on measures of short term memory.

Yuill and Oakhill (1991) carried out a replication of Isakson and Miller's (1976) experiment in which children read sentences which were either:

- i) meaningful, e.g., The old farmer planted the bean seeds in the rich, brown soil.
- ii) semantically anomalous, e.g., The old farmer paid the bean seeds in the rich, brown soil.
- iii) syntactically and semantically anomalous, e.g., The old farmer went the bean seeds in the rich, brown soil.

Unlike Isakson and Miller, Yuill and Oakhill matched their poor comprehenders and good comprehenders on decoding skill as described previously. Unlike Isakson and Miller, they found no difference between less skilled and more skilled comprehenders on oral reading errors on the different types of sentences. Their results provided no evidence that the skilled and less skilled comprehenders are differentially sensitive to semantic and syntactic features of text.

Yuill and Oakhill however, did find that whilst skilled comprehenders did not have better verbatim recall than less skilled comprehenders, they were able to recall the gist of sentences better. In general, they also found little difference between more skilled and less skilled comprehenders in their overall comprehension of sentences of varying grammatical complexity. The exception to this was that more skilled comprehenders appeared to better understand sentences containing pronouns.

Yuill and Oakhill found more skilled comprehenders were better at resolving to whom a pronoun referred in a sentence than less skilled comprehenders. Skilled comprehenders were also better at filling in a missing pronoun in a sentence as well as judging whether a statement based on the story which required the addition of the appropriate pronoun was true or not. Pronouns are only one type of cohesive device used in text. Yuill and Oakhill examined readers' comprehension of other anaphoric expressions in establishing text cohesion using a 700 word story read individually to 16 children in each reading comprehension group. The skilled comprehenders were better at identifying the meanings of the anaphors.

A key premise of Yuill and Oakhill's is that 'In order to understand stories adequately, readers must be able to draw appropriate inferences spontaneously when they hear or read a story, so that an integrated representation or mental model of the text as a whole can be built' (p65). They found that skilled comprehenders were more likely to accept that they had read a sentence which they hadn't but which could be validly inferred from the text they had read than less skilled comprehenders, ie., the more skilled comprehenders may have inferred more than the surface information in the text. When skilled and less skilled comprehenders were asked either literal or inferential questions on passages they had just read and could or

literal or inferential questions on passages they had just read and could or could not refer to, it was only on the inferential questions that the skilled comprehenders outperformed the less skilled comprehenders when the passage was present. Yuill and Oakill considered this was consistent with skilled comprehenders being 'more likely to use relevant general knowledge to make sense of information that is only implicit in a text' (p73).

A further example of this is that Yuill and Oakhill found skilled comprehenders were more likely to recall the gist of a sentence that had been read to them, e.g. The fish attacked the swimmer, when given as a clue a specific instance of this e.g., shark, than less skilled comprehenders. The argument advanced was that good comprehenders go beyond the immediately available information. Yuill and Oakhill also noted there was no difference between the different standards of comprehender when they were asked to give specific meanings to nouns in sentences they had heard, e.g., 'Do you think the fruit was an orange or a banana?' for the sentence, 'The fruit was full of juice.' suggesting it was not lack of knowledge that prevented the less skilled comprehenders making appropriate instantiations. Moreover as this occurred on material read to the children the less skilled comprehenders' difficulties were not solely related to reading but may be an aspect of a more general language problem.

Yuill and Oakhill also investigated children's ability to read stories of the type:

Paul has fifty pence to spend at the fair.

- 1) If Paul didn't buy a hot dog, he could ride on the big wheel.
- 2) Paul didn't buy a hot dog.
- 3) Later, Paul was very happy.
- 4) Why was Paul very happy?

They found that poor comprehenders gave fewer correct answers than good comprehenders. This as it stands gives little additional information as to the features distinguishing comprehenders selected as better on the basis of correctly answering more questions on passages they read than less good comprehenders. Were the experimental passages to have been presented orally one could then have drawn some inferences as to the generality of the comprehension performance tapped by the Neale Analysis.

Yuill and Oakhill also looked at metacognitive aspects of reading. In common with other studies (Myers and Paris 1978, Paris and Myers 1981) they found poor comprehenders viewed reading primarily in terms of decoding. However, while they found a difference between better and less good comprehenders on their performance on multiple choice questions on a passage they had read to them, there was no difference between the groups on the questions on which both groups did better. That is both groups answered correctly more questions on important parts of the story and fewer questions on less important parts. More skilled comprehenders were however better able to identify the main part in the story sequences shown

on cards making up the Picture Arrangement Test of the Wechsler Intelligence Scale for Children - Revised (Wechsler 1974) than less skilled comprehenders.

They also found more skilled comprehenders were better at identifying inconsistencies in stories than less skilled comprehenders and that their ability to resolve ambiguities in stories was greater than less skilled comprehenders when the anomaly and its resolution were not adjacent.

In an exploratory study of riddles, Yuill and Oakhill found better comprehenders were better at retelling and explaining riddles than less good comprehenders. However, this finding has to be assessed in the light of the selection of good comprehenders being based on their ability to answer questions on a passage they had read but could not refer to. Moreover, explaining riddles may relate not just to the listener's comprehension of it but their ability to articulate this. Again, the good comprehenders had shown themselves capable of articulating more acceptable responses to the questions posed them from the Neale Analysis. Though in a follow up experiment they found a significant correlation between recall of riddles and comprehension on the Neale this does not negate the plausible alternative hypothesis that the correlation relates to a commonality between the retelling task and the measure of reading comprehension used.

Yuill and Oakhill failed to find consistent differences between more and less skilled comprehenders in the way in which they told stories when presented with pictures though there was a tendency for more skilled comprehenders to make more use of the past tense. Better comprehenders were better able to correctly answer questions of the type 'Why did Mary spill the milk?' on being shown pictures of simple causal sequences, e.g., John bumping into

Mary who is carrying a milk jug and a puddle of milk on the floor. Better comprehenders were able to more accurately recall stories in their original format, or with order disrupted whilst retaining referential continuity but there was no difference between the skill groups in recall of fully randomised stories. This result could be used to support the assertion that better comprehenders can use a mental model to assist in recall of text. Alternatively, it may reflect that the good comprehenders were selected on their ability to respond to questions on a short story just read but unavailable for reference.

Overall, however, the good comprehenders produced a more coherent story on recall than the less skilled comprehenders.

In summary, Yuill and Oakhill found that good comprehenders were better at recalling the gist of text, resolving pronominal reference and understanding cohesive devices in text. They were better able to draw inferences from text and identify the main point of a story. They were also better able to identify inconsistencies in stories and resolve ambiguities.

Yuill and Oakhill also explored ways of enhancing readers' comprehension of text. In one study, they trained subjects in both of the comprehension skill groups to draw inferences from key words in stories, which gave a clue to the meaning of the story. This training significantly improved poor comprehenders' ability to answer comprehension questions on the passage. In a follow-up experiment with extended training over 7 sessions comparisons between the comprehension groups and treatments tended to be statistically non-significant making interpretation difficult. A similar pattern of non-significant findings arose from a further follow-up experiment. Confidence in their analysis of this last experiment is reduced by the failure

of the figures given in the summary table to tally. It is also of note that no significant difference was found between scores on a comprehension test administered 2 weeks after training and 6 months after training suggesting there was no carry over effect of training in this experiment.

Yuill and Oakhill found story titles describing the main consequences of a story prompted correct answering of comprehension questions more than titles listing the characters in the story. Single illustrations also assisted less skilled comprehenders in answering comprehension questions more than three separate pictures interspersed through the text, but had no significant effect for more skilled comprehenders. They also found training 9 and 10 year olds to use imagery when reading a story improved their performance on comprehension questions on the story. Poor comprehenders benefited more from the training than under control conditions. (See Appendix C for a review of other empirically evaluated interventions designed to improve reading comprehension.)

In conclusion, Yuill and Oakhill claimed their work demonstrated 'that comprehension problems are not necessarily due to difficulties in decoding words or accessing their meanings.' (p216). They claim that poor comprehenders who fall into this group have three main areas of weakness:

1. They are less likely than skilled comprehenders to integrate information from different parts of a text and make relevant inferences to help them integrate the text.
2. They have poor working memories.
3. They more often fail to notice inconsistencies in text and realise that they had failed to understand it.

They noted that their poor comprehenders appeared deficient in listening comprehension as well and that this may enable problems with comprehension of text to be identified earlier. They also suggest that training techniques could be adapted for listening comprehension tasks presumably with a view to improving the listener's ability to construct mental models; a skill which could then transfer to reading comprehension. In this, they follow Curtis (1980) who argued that 'Practice in comprehension independent of the demands of decoding has many practical as well as theoretical advantages. It would not only allow the child whose reading comprehension suffers because of inefficient word processing the opportunity to practise these skills at a level seldom reached during reading but it would also place the child in a situation where the probability of success at comprehending is much higher than in reading.' (p668).

iii) Carver's Rauding Theory

Yuill and Oakhill's (1991) model of reading comprehension has both contrasts and similarities with the principles put forward by Carver (1987) to explain performance differences on reading comprehension tasks. However, while Yuill and Oakhill make explicit their interest in semantic comprehension, Carver (1987) does not differentiate between different categories of comprehension. In general though, he proposed that one can increase the degree to which students will comprehend passages by using passages easier than those at their frustration level. This he called the Easiness Principle. While Yuill and Oakhill point out this could not explain their results as the children stop at their level of competence on the Neale, it would follow from Yuill and Oakhill's work that where ideas in a story are presented more simply, it would be easier to construct a mental model. This would enable inconsistencies in the text to be more easily detected so enhancing semantic comprehension monitoring. This would contrast with

lexical comprehension monitoring which as it would not require the construction of a mental model should remain relatively unaffected by changes to a story's readability.

Carver's second explanatory principle was the Reading Time Principle. He argued that students could improve the degree to which they comprehend a passage by spending more time reading and studying the passage. Yuill and Oakhill analysed the reading times of subjects in the second of their inference training experiments and found no difference in reading times between comprehension skill groups or treatment conditions. They therefore felt there was no evidence for there to be increased study time resulting from their intervention in this experiment. However, their model of reading comprehension as mental model building from the text by active monitoring of comprehension and drawing inferences does not necessarily have implications for increased study time enhancing comprehension. Indeed, they argue that 'the problem seems to lie not in the study time available but in how to use that time.' (p107) and that 'poor comprehenders do not seem to benefit from the opportunity of studying a text.' (p107). In this, they seem to be adopting a stronger position than Oakhill and Garnham (1985) who identified two different loci in the process of comprehension. They noted that sometimes the structure of a description made it difficult for the information to be integrated as it was being heard, or read and this could be compensated for by spending more time reading the description. For those descriptions difficult to process because they are hard to represent or hold in memory, they believed increased study time did not help. However, the continuity in the position of Oakhill and Garnham (1985) and Yuill and Oakhill (1991) is that they propose that encouraging children to spend more time reading or rereading a passage than they would normally choose to do would have no implications for their comprehension of the passage.

Carver (1977, 1987, 1990 a, b) in contrast quite explicitly predicts that spending more time reading and rereading a text will increase comprehension of that text. This position is based on Carver's (1990a) distinction between five different reading processes of scanning, skimming, rauding, learning and memorising.

The most powerful is the memorising process which is also the slowest as it involves rehearsal and repeat reading. The least powerful is the scanning process which is also the fastest and only involves the reader in looking for a target word. In moving up or down through these processes, one spends more or less time reading. Carver believes the rauding rate is the normal or typical reading rate for material at the reader's instructional reading level. However, when someone is given a passage which is hard for them to understand, they may shift down to a slower reading rate, consequently spending longer on the passage.

Carver (1990b) has defined three individual difference factors associated with the rauding process - rauding accuracy level (AL), rauding rate level (RL) and rauding efficiency level (EL). AL is similar to reading level and is the most difficult level of material individuals can accurately read when the material is read once at their rauding rate. The rauding rate (Rr) is the fastest rate that an individual can accurately read relatively easy material measured in sentences read per minute and the rauding rate expressed in grade equivalent, units is symbolised as RL. EL is a construct similar to general reading ability reflecting individual differences in accuracy and rate of comprehension. In the prediction equations which Carver has devised A, R and E without any qualifying subscript stand for accuracy of comprehension, rate of comprehension and efficiency of comprehension

respectively. When individuals read a text of difficulty DL, then Carver (1990b) has empirically derived a formula for predicting comprehension accuracy at the reading rate for that passage (Ar).

$$Ar = 0.03951 (AL - DL) + 0.6446 \text{ (Equation 1.1)}$$

Carver (1990a) gives an example where a new 3,000 sentence textbook with a DL of 5.8 grade equivalent units is being considered for a class of average AL = 7.6, Rr = 11.1 and RL = 7.1. Substituting into Equation 1.1 it can be estimated that typical students would comprehend 72% of it. However, if the group read the book twice, ie., spent twice as long on it then Carver (1990a) would predict from that reading accuracy level that the typical student would now comprehend 84% of the book.

While Carver's Reading Theory is precise at the level of predicting group comprehension product it is silent on the mechanisms by which this is achieved except in as much as he is clear that spending time reading and rereading a text will increase comprehension of that text.

Carver's (1987) third principle is the Practice Principle, which is that students 'ordinarily improve on any reading-related task simply by practising on that task' (p116). This would have to be taken into account in the interpretation of any experiment.

Carver's final principle is the Prior Knowledge Principle which is that students 'will usually comprehend more of a passage if they have, or are given, more prior knowledge' (Carver 1987 p122). This is a point of agreement between Carver and Yuill and Oakhill. Carver has more recently refined his position (Carver 1992). He argued that prior knowledge specific

to a passage was unlikely to substantially facilitate comprehension of that passage during typical or normal reading, but that general prior knowledge as measured by reading and answering questions on many short passages does influence how much of a longer passage will be comprehended.

A further commonality is that both quite explicitly acknowledge the importance of decoding skills for reading. However, they differ in the prescriptions they offer to improve students' comprehension. In contrast to the interventions described above as implemented by Yuill and Oakhill, Carver (1987) simply advocates students spend 'a great deal of time reading materials they can comprehend well.' (p125). Carver is not here referring to extra study time on one particular piece of text but to what Fitzgerald (1990) called sustained silent reading. Carver (1987) believed sustained silent reading practice would increase the students' vocabulary, their general knowledge and their decoding efficiency so improving their general reading ability. In the review of reading comprehension instruction edited by Fitzgerald (1990) sustained silent reading of a variety of texts over at least half a term was often successful in enhancing comprehension. Further, a recent study of 9½ and 10½ year olds by Cipelewski and Stanovich (1992), also found that the extent to which individuals engaged in reading as measured by their recognition of children's authors' names and the titles of children's books was a significant contributor to developing reading ability.

The Aims of this Thesis

This study will investigate comprehension of narrative prose in children aged between eight and ten using predominantly error detection tasks. Explanations of differences between children's reading comprehension in terms of the application of different standards of comprehension using Baker's (1985a) distinction between lexical, syntactic and semantic standards of comprehension will be investigated with particular reference to whether significant numbers of children exclusively use only lexical standards of comprehension. The influence of textual factors on the detection of different classes of errors will also be examined.

Yuill and Oakhill's model and Carver's theory differ in their predictions on the effects of spending more time studying a piece of text. This will be investigated experimentally for both reading and listening comprehension tasks. While Carver's theory is silent on the processes involved in comprehension, Yuill and Oakhill's model enables predictions to be made on the role of comprehension monitoring and working memory and experimental investigation of how comprehension monitoring and working memory differ in good and less good comprehenders will also be carried out.

CHAPTER TWO

COMPREHENSION OF TEXT AT THE SINGLE WORD LEVEL

If one can assume a child reader is motivated to perform optimally, has sufficient prior knowledge of the subject matter in the text and is aware of the specific purpose of the reading task, then two factors that seem likely to influence the outcome are the standards against which he evaluates his comprehension of the text and the strategies he adopts to regulate his comprehension.

The simplest level at which comprehension can break down is at the single word level. The word might not be able to be decoded, might not be previously known or not make sense in the context in which it is found. On coming across such a word, the reader could:

1. Ignore it and read on
2. Suspend judgement and hope further information will be provided
3. Form a tentative hypothesis
4. Reread the sentence, looking for a revised interpretation
5. Reread the previous context
6. Go to an expert source, e.g., a dictionary

But before any of these strategies can be applied, the child has to notice there is a problem.

It certainly seems to be the case that young children are surprisingly insensitive to the presence of nonsense words in prose. Miller and Isakson (1978) asked 108 children (36 each from grades one, two and three), to read

a story out loud. Each grade level had a different story, all of which were eight sentences in length. In each version of the story, two sentences had modified nouns, two sentences had modified verbs and four sentences were unmodified. The order of these modifications was rotated so the result was each story appeared in eight versions. The first grade children showed no significant disruption that was due to the appearance of a pseudoword in a sentence the children were reading. By grades two and three when a pseudoword appeared in the sentence, reading was disturbed and showed a higher proportion of errors on words surrounding the pseudowords. The pseudowords were decodeable.

Paris and Myers (1981) using the same paradigm as Miller and Isakson (1978), found that both good and poor fourth grade readers showed little awareness when reading aloud that phonologically acceptable nonsense words had been substituted for nouns in the passages they were reading. Paris and Myers (1981) were however, concerned that many checking behaviours may be subtle or covert and that hesitations and repetitions do not always reflect deliberate monitoring and correction. They also pointed out that spontaneous self-corrections could be contaminated by decoding and pronunciation errors. For these reasons they also asked the students to underline any of the words or sentences they did not understand in two stories. Good readers only underlined 25% of the nonsense words, while poorer readers only underlined 19% of the nonsense words, and this was true both for the easier (third grade level) and harder (fifth grade level) passages they read. In a replication, Baker (1984b) replaced words in a passage with two-syllable nonsense words which followed standard rules of English orthography. Half the fourth grade children were told specifically that nonsense words would occur in the passage and half were simply told to find problems. Those children receiving specific instructions about

nonsense words identified 58% of them, whilst those receiving general instructions identified 44%. Poorer readers from the fourth grade only underlined 38% of nonsense words after specific instruction, whilst the good readers given specific instruction managed to underline 71% of the nonsense words.

The fourth grade readers in the Paris and Myers (1981) and Baker (1984b) studies were failing to underline at least one in four nonsense words, even if they were good readers and were given specific instructions for what they were looking. One possibility is that rather than evaluating word understanding, the poorer readers in particular were evaluating decodability. Paris and Myers (1981) in a second experiment obtained some evidence to support this hypothesis. In this experiment, two groups, one of 14 good and one of 14 poor fourth grade readers were asked to read and study a story. They had the use of a dictionary and were able to ask questions. The passage they were asked to read contained several difficult vocabulary words. Six of the good readers looked up words in the dictionary and none of the poor readers did. The poor readers appeared more concerned with pronunciation than meaning and the four poor readers who asked about unknown words only wanted help in pronouncing them.

Further evidence that for some younger readers the goal of the reading may be about decoding words was provided by Myers and Paris (1978) who asked 20 second grade children and 20 sixth grade children, what they would do if they didn't understand a word they were reading. Only 19% of the sixth grade children said they would sound out the word, but 40% of the second grade children said they would sound it out. By contrast, 70% of sixth grade children (mean age 11 years 9 months) said they would seek help from an outside source; a dictionary or ask someone while only 40% of

second grade children (mean age 7 years 9 months) would seek assistance from an external agency. The trend for second graders to respond 'sound out' and sixth graders to respond 'dictionary' was statistically significant. This is not unambiguous evidence of younger children perceiving the goal of reading as decoding. For example, younger children may well know more words by sound than by sight, so sounding out may be a route to deriving meaning. Nevertheless, Myers and Paris (1978) investigation of children's metacognitive knowledge about reading led them to conclude that 'In general, second graders focused on decoding goals rather than semantically related goals for reading ...'

A distinction should however be drawn between perceiving the goal of reading as being decoding and being sensitive to the rules of phonics. Adults are unlikely to perceive reading as only sounding out but they appear to be very sensitive to phonological legality. This was demonstrated in a study of proof-reading amongst Finnish undergraduates by Niemi and Virjamo (1986) in which the frequency and length of altered words were controlled. They found misprints were hardest to detect in words which retained the original visual shape, but that nearly all the words containing phonologically illegal letter strings were detected. Visual deformation contributed only marginally to the detection of misprints in phonologically illegal strings, though the hit rate was so high for these, there was little room for improvement.

The studies reviewed have shown children to be quite insensitive to the presence of decodable nonsense words in passages. The low frequency of detection of these could be explained by the children using decodability as a standard by which they could monitor their comprehension. This would lead to the prediction that children would underline more undecodable nonsense

words, e.g., consonant strings than decodable nonsense words. However, unknown words are only of significance when they contribute to the meaning of the passage of which they are part. This raises the question as to whether some presumably good readers read in such a way as to automatically infer the meanings of unknown words given significant clues in the text and ignore unknown unimportant words and any investigation of error detection would have to take this possibility into account.

The question this study addresses is whether younger readers perceive the goal of reading as being sounding out rather than understanding words. Were this the case it would follow from this that junior aged children should identify more consonant strings than decodable nonsense words. Moreover, there should be no difference in the frequency with which important decodable nonsense words are identified as compared to decodable nonsense words which are not so important to understanding the passage in which they are embedded.

EXPERIMENT ONE

Subjects

The subjects were all selected from Year Five of an urban junior school catering for predominantly working class families. There were 59 possible subjects in this year in three classes. The pupils' teachers were asked to rate them on a scale of one (poor) to ten (excellent) for reading ability, following the suggestion of Francis (1992) that teachers can identify from their own monitoring children's reading progress.

Using each class's own mean rated reading ability score, they were divided into boys in the upper and lower half and girls in the upper and lower half. A sample consisting of 24 pupils from the lower half of their class and 24 pupils from the upper half of their class in each case consisting of 12 boys and 12 girls made up the total sample of 48 pupils. The final sample ranged in age from 9 years 1 month, to 10 years 1 month, with a mean age of 9 years 7 months and a standard deviation of three months.

Materials

Two passages each of 100 words were written. They are reproduced in Appendix B. Both experimental passages were of similar difficulty, and were capable of being read by an average reader of between 7½ and 8½ years of age. (For details of the calculation of the passages' readability please see Appendix A.)

Each passage was modified so that the same six words were changed in one version to decodable nonsense words of four letters in length and in the

other version, to consonant strings of four letters in length. Three of the words changed were important nouns or verbs, and three of the words changed were unimportant adjectives or adverbs. The important words were selected by a group of 3 educational psychologists from those nouns and verbs which occurred only once in the story and were central to the meaning of the story. The unimportant words were selected by the same group.

Procedure

Each subject was seen individually. The task was introduced by the experimenter saying, 'I would like your help. I have got two stories here which I would like to put in a book. I want to make sure the children reading them can understand them. Would you please read this story to yourself and underline any word you don't understand.' Each subject silently read both passages either one passage with decodable nonsense words as substitutions followed by the other passage with consonant strings as substitutions or vice versa. The order of presentation of the four possible sequences was randomised within each group of subjects. When the children finished reading the decodable versions of the passages, they were asked to read from flashcards all the decodable words they had not underlined and they were also asked what these words meant.

Results

A preliminary breakdown of the substitutions underlined by part of speech and type of substitution is given in Table 2.1.

Table 2.1 Substitutions underlined by part of speech and type of substitutions

Part of speech substituted	Type of substitution	
	Decodable Nonsense Word	Consonant String
Noun	44a	56
Verb	45	59
Adjective	45	60
Adverb	50	60
Total:	184	235

a out of a possible 72

The frequency of underlining of either important (noun or verb) or unimportant (adjective or adverb) substitutions was analysed for each passage by obtaining a score out of a possible three for each type of substitution for each subject and comparing the relative size of the scores using a Wilcoxon Matched-Pairs Signed-Ranks Test. In none of the passages in either decodable or consonant string form did the distribution of scores reach significance at the 0.05 level. Indeed, only in one passage was there a greater number of different scores between the two categories of substitutions than tied scores. Overall the consistency with which both important and less important substitutions within a given passage were underlined is compatible with the hypothesis that the children were reading for decodability rather than meaning. The relative importance of the substituted words will not be included in further analysis.

A Type V analysis of variance (Lindquist 1956) was conducted with three within subject factors: type of substitution, passage and session (See Appendix E). The results are summarised in Table 2.2.

Table 2.2 Summary of Analysis of Variance for Experiment One

Source	df	SS	MS	F	P
Between Subjects	47	268.74			
AB	1	1.26	1.26	0.22	
AC	1	11.334	11.34	1.99	
BC	1	5.51	5.51	0.97	
error	44	250.63	5.70		
Within subjects	48	85.5			
A	1	27.09	27.09	22.58	<0.01
B	1	0.84	0.84	0.7	
C	1	0.84	0.84	0.7	
ABC	1	3.77	3.77	3.14	
error	44	52.96	1.20		

A = substitution

B = passage

C = session

From the table it can be seen there is a significant main effect for type of substitution, which examination of Table 2.3 indicates is in favour of more consonant strings being underlined.

Table 2.3 Mean scores for each subject group

Group	Average scores out of six for decodable nonsense words (A ₁)	Average score out of six for consonant strings (A ₂)
Group One (A ₁ B ₁) then A ₂ B ₂)	4.25	4.92
Group Two (A ₂ B ₂ then A ₁ B ₁)	3.83	4.92
Group Three (A ₂ B ₁ then A ₁ B ₂)	3.33	4
Group Four (A ₁ B ₂ then A ₂ B ₁)	3.92	5.75
Combined mean	3.83	4.89

B₁ = Passage One

B₂ = Passage Two

While it can be concluded that the children underlined significantly more consonant strings than decodable nonsense words, this was only one of the ways in which the principal hypothesis that the children were reading for decodability rather than meaning was tested. The second approach was by asking the children about the decodable nonsense words they did not underline. The essential point of enquiring about these was to establish that the reader had not invested them with some meaning that made sense in the context of the story. If they had, this would have rendered invalid the conclusion that they were equating decoding the word with understanding it. From this point of view, the key statistic was the 49 out of 104 missed decodable words for which readers claimed to know the meaning. Out of the 49, only four when substituted in the passage made sense. Three of

these were reading **fest as fast** so the passage would read, '**Ben, talked fast about this morning on the way home.**' The other was believing **sant** was the name of somewhere, so the passage would read, '**he would go to sant.**' The rest of the meanings given made no sense in the context of the passage. Clearly then, it is not because decodable words were given a plausible meaning within the context of the passage that fewer of them were underlined.

Table 2.4 Crosstabulation of number of consonant strings and nonsense words detected

		Consonant Strings		
		Number of readers identifying within the range		
Decodable nonsense words		0	1-5	6
Number of readers identifying within the range	0	1	1	2
	1-5	0	16	14
	6	0	1	13

If the results are examined at the level of individual readers, most readers detected more consonant strings than decodable nonsense words. However, only two children identified all the consonant strings and none of the decodable words. So only two children are likely to see the goal of reading as just being able to sound out words.

It may be that decodability was necessary but not sufficient for a child to judge that he understood a word. It may also be that there is an interaction with the child's reading ability. If the overall detection of substitutions correlates with the teachers' ratings of the children's reading ability, then this might provide some support for this.

This was investigated by comparing within each class the rating given each child for their reading ability to their overall detection of substitutions. Because the numbers in some classes were small, because the rating was subjective and because correlations are only as good as the underlying data and are sensitive to discrepancies a categoric comparison was adopted. The results are presented in Table 2.5. Five children were omitted as they lay exactly on a median point.

**Table 2.5 Crosstabulation of teacher rating with overall score
summed across the classes**

		Substitutions Detected	
		Below Median	Above Median
Rating	Below Median	14	7
	Above Median	7	15

Two thirds of the children lie in the quadrant one would predict from the teacher rating. Those children rated better readers by teachers are more likely to do better on the error detection task used in this experiment ($\chi^2 = 3.92$, $df = 1$, $p < 0.05$).

Discussion

Taken overall, the results of this experiment were consistent with young readers sometimes accepting decodability as a sufficient criterion for judging they understood a word rather than as a necessary, but not sufficient criterion. The prediction that followed from this that children would underline more undecodable consonant strings than decodable nonsense words when asked to underline words they didn't understand, was supported by the results. Additionally, there was no evidence that the children were failing to underline the decodable nonsense words because they were investing them with plausible meanings. Moreover, the frequency of detection of substitutions within any passage did not seem to be dependent on the importance they had for understanding the meaning of the passage. This combined with the low rate of detecting consonant strings with roughly one in five going undetected, does not instil confidence that overall, the readers were good at monitoring their comprehension of the passage at the single word level.

There was at least one potential source of error in this experiment for which it is difficult to formulate a solution. This was the confounding of relative importance with part of speech: the important words for which substitutions were made were verbs and nouns while the unimportant substitutions were adjectives and adverbs. It may well be this is intrinsic to language, except where the adjectives or adverbs differentiate between nouns or verbs.

There are refinements that could be made to this experiment. The passages were a little too easy for the readers: 13 of them identified all the substitutions while only one failed to identify any. The passages may therefore be more suitable for Year Four children. Additionally, the control

over the visual similarity of the decodable nonsense words and the consonant strings could have been much improved. In this experiment, it was controlled by having all substitutions four letters in length, e.g., **colk** and **rmng**. However, if it is indeed just being decodable that is important, then three of the letters can be held constant and a vowel substituted for the fourth, e.g., **colk** and **crlk** so maintaining a much greater visual similarity between them.

It was argued that by asking children about a single word out of context, one might judge whether they knew what it meant or not. One might question whether it is a fair test to ask the child the word's meaning out of context. This could be overcome by asking the child what the word meant when presented in a sentence or in the passage. However, the child may know a word's meaning but not be able to express it. This constitutes a particular issue since if one accepts that the children assigned meaning to 49 of the missed decodable words and so didn't underline them, then these 49 plus the 184 underlined would make 233 decodable words in total that were identified. This is, virtually identical to the 235 consonant strings underlined. One way of examining this would be to also ask the pupils about the consonant strings they failed to underline. If they claimed not to have done so for some as they assigned them meaning at a rate equivalent to that of the decodable nonsense words, then it could be argued that the difference in rate of detection of decodable nonsense words compared to consonant strings still existed despite some uncertainty about any substitutions not underlined as the reader assigned them a meaning appropriate in the context of the passage.

These weaknesses in this experiment and ways in which it can be extended to throw more light on the reading process will be developed in Chapter Three.

CHAPTER THREE

READING AS A PIECEMEAL PROCESS

In Chapter Two, the idea that some children used decodability as the criterion against which they measured their comprehension of text was explored. This however, may just be one aspect of a way of reading which involves serial processing of text with each word being read in turn on a word-by-word basis, but not congealing to form the total meaning the writer of the text intended to convey.

Cromer (1970) obtained data from a study of 64 male students at college in the United States which provided support for the idea of some poor comprehenders reading in a word-by-word fashion. Cromer (1970) split his readers into good and poor comprehenders on the basis of their score on the Education Testing Service (1960) Cooperative English Test of Reading Comprehension. Both the poor comprehenders and their matched controls were split into deficit readers who obtained a low vocabulary score and difference readers who obtained an average vocabulary score. Each student was asked to read out loud stories arranged in one of four different ways: in sentences, as single words presented successively on a roll of paper, as meaningful phrases and in fragmented groupings of words which were not meaningful as units. A large number of findings were obtained. Amongst these, it was noted that all four groups of subjects made fewer errors in reading words when they were presented singly. However, there were differences between the different groups in their ability to respond correctly to comprehension questions on the texts when they were presented in different ways. In particular, the poor comprehenders with good vocabulary answered significantly more questions correctly when the text was presented

to them as meaningful phrases. Cromer interpreted this result combined with this group's comprehension not being disrupted by the single word or fragmented mode of presentation compared with sentences as demonstrating that these poor comprehenders typically read word-by-word.

Garner (1981) referred to this word-by-word reading as being 'piecemeal processing'. She investigated this idea using 20 students of mean age 10.8 years who had scored at least one year below grade level on the Maryland screening comprehension instrument. The students had all scored at least averagely for their age on a word recognition instrument, indicating decoding proficiency but as a check, students who failed to read one of the experimental passages without making three or fewer miscues were excluded, and four subjects were excluded through this latter process. The students were asked to silently read three passages. The passages were either as standard, with a last sentence one word alteration inconsistent with the preceding text or with two polysyllabic modifying words unfamiliar to the readers inserted, though these words added little to the sense of the passage and could be deleted without significantly altering its meaning. The subjects did not significantly differentiate between the standard and informationally inconsistent passages in terms of how sensible they found them. However, both of these types of passages were rated significantly more comprehensible than those containing long, unfamiliar but irrelevant words. All the 16 subjects indicated the long unfamiliar words were the source of concern. Garner felt this demonstrated that poor comprehenders who had no problems in just reading words are more concerned with long words within sentences than inconsistent information across sentences, that is poor comprehenders manage written language as bits and pieces, not as textual wholes.

Superficially, there appears to be a discrepancy between the findings of Garner (1981) of fifth and sixth grade students readily identifying difficult vocabulary in passages and the study by Paris and Myers (1981) discussed in Chapter Two in which fourth graders found considerable difficulty in identifying decodable nonsense words in passages. A possible explanation is that the nonsense words were short, e.g., kales, while the difficult words were long e.g., multifarious. Baker (1985b) as reported in Garner (1987) investigated this by asking third and fifth grade students to read a series of short paragraphs. Each child read eight paragraphs with one-syllable nonsense words inserted and eight paragraphs with three-syllable nonsense words inserted. The detection rate for one-syllable nonsense words was 38% while for three-syllable nonsense words it was 58%. The results suggest that longer words are more readily identified and if comparisons are to be made between rates of identification of different types of nonsense words, then their length should be held constant.

At the beginning of this Chapter, it was argued that using decodability as the standard against which to monitor comprehension was only one aspect of piecemeal processing of text. Garner (1980) investigated comprehension of text amongst 30 students from the seventh and eighth grades who were all judged to be good decoders, though we are not told by what criterion this was assessed. These students should therefore have had little difficulty in sounding out age appropriate text, however, half the students were said to experience some difficulty understanding what they read and half were said to experience little difficulty understanding what they read. There appears to have been no external standard for this discrimination. Parallel forms of two passages were prepared in one of each of which six critical words were changed to make key bits of information inconsistent, e.g., 'numbers' was altered to 'letters' in a passage about Thomas Jefferson's work to systematise

coinage. From the results, it seems clear that the good comprehenders rated the altered passages harder to understand, while poor comprehenders, though generally rating the passages harder to understand than the good comprehenders, made little distinction between altered and unaltered passages.

This design of Garner's may provide a more sensitive method than that used in Experiment One of investigating the role of decodability as a standard against which comprehension may be judged. If a person is reading on a word-by-word basis, then the process may still involve a number of stages of which the first may be whether the word can be decoded? Thereafter, the reader may check the sound of the word against those of words already known to him rejecting it as not comprehended if it is not already known. If this is the case, then one would expect a different detection rate for substitutions into passages which were non-decodable consonant strings or decodable nonsense words or common real but contextually inappropriate words, with progressively fewer of the substitutions being detected as one moved from consonant strings to real but contextually inappropriate words. This hypothesis will be investigated in the following experiment.

EXPERIMENT TWO

Subjects

The subjects were randomly selected from the three Year Four classes of an urban junior school so that each class contributed 9 boys and 9 girls. The sample of 27 boys and 27 girls ranged in age from 8 years 10 months to 9 years 10 months, with a mean age of 9 years 4 months and a standard deviation of 4 months.

Materials

Three passages each of 100 words, were written. The experimental passages were of similar difficulty and were capable of being read by an average reader of between 7½ and 8½ years of age. (For details of the calculation of the passages readability please see Appendix A.)

Each passage was modified so that the same six words were changed in one version to decodable nonsense words of four letters in length, in another version, to consonant strings of four letters in length and in a third version, to real but contextually inappropriate four letter words. In each version, three of the letters were held constant and only one was changed, e.g., **dert dert dart**.

In each passage, three of the words changed were important nouns or verbs, and three of the words changed were unimportant adjectives or adverbs. The important words were selected by a group of three educational psychologists from those nouns and verbs which occurred only once in the story and were central to the meaning of the story.

Procedure

Each subject was seen individually. The task was introduced by the experimenter saying, 'I would like your help. I have got some stories here which I would like to put in a book. I want to make sure the children reading them can understand them. Would you please read this story to yourself and underline any word that doesn't make sense in the story'.

Each subject was then presented in succession with all three passages so that each of the subjects was exposed to all three different types of substitution. The order of presentation and type of substitution was grouped into nine different sequences:

	First	Second	Third	Mean Age (Sd)
Group 1	A1 B1	A2 B2	A3 B3	9 yrs 2 mths (4 mths)
Group 2	A3 B2	A1 B3	A2 B1	9 yrs 5 mths (4 mths)
Group 3	A2 B3	A3 B1	A1 B2	9 yrs 3 mths (3 mths)
Group 4	A1 B2	A2 B3	A3 B1	9 yrs 6 mths (5 mths)
Group 5	A3 B3	A1 B1	A2 B2	9 yrs 6 mths (4 mths)
Group 6	A2 B1	A3 B2	A1 B3	9 yrs 4 mths (4 mths)
Group 7	A1 B3	A2 B1	A3 B2	9 yrs 6 mths (2 mths)
Group 8	A3 B1	A1 B2	A2 B3	9 yrs 4 mths (4 mths)
Group 9	A2 B2	A3 B3	A1 B1	9 yrs 2 mths (3 mths)

- A: Type of substitution
A1 = nonsense words
A2 = consonant strings
A3 = real, but contextually inappropriate words
- B: Passage

After reading each passage, each subject was asked to read from the passage any substitution they had not underlined and they were also asked what the word meant in the story they had just read.

Results

A preliminary breakdown of the substitutions underlined by passage and type of substitution is given in Table 3.1.

Table 3.1 Mean scores (and standard deviations) by passage and type of substitution

Passage	Type of substitution		
	Consonant String	Decodable	Real Word
One	5.78 (0.43)	3.78 (2.10)	2.33 (1.37)
Two	5.06 (1.73)	4.72 (1.67)	4.00 (2.14)
Three	5.83 (0.71)	5.39 (1.09)	2.72 (2.14)
All 3 passages	5.56 (1.14)	4.63 (1.77)	3.02 (2.01)

The frequency of underlining either important or unimportant substitutions was analysed by obtaining a score out of nine for each type of substitution for each subject and comparing the relative size of all the scores using a Wilcoxon Matched-Pairs Signed-Ranks Test. The distribution of scores did not reach significance at the 0.05 level ($z = -0.6185$, $n = 37$, n.s.). The consistency with which substitutions for important and unimportant words were underlined indicated that the relative importance of the substituted words need not be included in further analyses.

Further analysis using a Type V analysis of variance (Lindquist 1956) was conducted with three within subject factors: type of substitution, passage and session part. The results are summarised in Table 3.2.

It is clear from the means and standard deviations shown in Table 3.1 that the assumption of homogeneity of variance made in analysis of variance has been violated. However, Glass, Peckham and Sanders (1972) concluded that non-homogeneity of variance is likely to only have a very slight effect on the probability of committing a Type One error when the sample sizes are equal as in this case. Nevertheless, as a precaution, significance should be interpreted conservatively in the analysis that follows:

Table 3.2 Summary of analysis of variance for Experiment Two

Source	df	SS	MS	F	P
Between Subjects	53	259.59			
AB	2	40.01	20	4.88	0.05
AC	2	13.20	6.6	1.61	
BC	2	2.09	1.05	0.26	
ABC	2	19.79	9.89	2.41	
error	45	184.50	4.10		
Within subjects	108	369.33			
A	2	178.01	89	50.59	0.01
B	2	15.64	7.82	4.45	0.05
C	2	1.94	0.97	0.55	
AB	2	2.09	1.05	0.60	
AC	2	3.71	1.86	1.06	
BC	2	4.08	2.04	1.16	
ABC	6	5.53	0.92	0.52	
error	90	158.33	1.76		

A = type of substitution B = passage C = session part

From the table, it can be seen there is a significant main effect for type of substitution and for passage. However, there is also a significant interaction effect between passage and type of substitution. Tests of simple main effects (Winer 1962) were carried out. These indicated that significant differences for type of substitution were obtained for Passage One ($F = 21.21$, $df = 2,51$ $p < 0.01$) and Passage Three ($F = 20.07$, $df = 2,51$ $p < 0.01$) but not for Passage Two ($F = 2.06$, $df = 2,51$). Significant differences were also obtained between passages for decodable nonsense words ($F = 5.78$, $df = 2,102$ $p < 0.01$) and real word substitutions ($F = 6.71$, $df = 2,102$ $p < 0.01$) but not for consonant strings ($F = 1.67$, $df = 2,102$). While it cannot be unambiguously concluded that the children found it significantly easier to underline more consonant strings than decodable nonsense words or underlined more decodable nonsense words than real but contextually inappropriate words this was only one way of investigating whether the children were reading each word separately rather than reading for overall meaning. The children were also asked to give the meanings of any substitution they failed to underline. This was to establish whether the reader had invested them with some meaning that made sense in context. The results are given in Table 3.3.

Table 3.3 Substitutions not underlined

Type of substitution	Number not underlined	Number for which child proposed a meaning	Number for which child proposed a meaning which made sense in the sentence in which the substitution was made	Percentage of proposed meanings which made sense in the sentence in which the substitution was made
Consonant string	24	13	3	12.50%
Decodable nonsense word	74	39	11	14.86%
Real words	161	87	12	7.45%

The percentage of unidentified substitutions for which the child proposed a meaning was consistent across substitution type, and lay in the range from 52.7% to 54.04%. This is in line with the rate of 47% for which meanings were proposed for decodable nonsense words in Experiment One. Despite having the text available to them only between 7.45% to 14.86% of the proposed meanings made sense in the sentence for which they were proposed. In Experiment One in which the readers were asked to give meanings for nonsense words, they failed to underline when shown the word on a flash card 3.84% of the proposed meanings made sense in context. This does not suggest the readers were investing any type of substitution with plausible meaning in any significant way.

Examining the results on an individual rather than group basis indicates that seven people identified all 18 substitutions in the passages they read and none failed to identify any. Only four people identified more nonsense words than consonant strings and 26 people identified more consonant

strings than nonsense words. A Sign Test indicated this was highly significant ($Z = 3.83$, $p < 0.01$). Thirty five people underlined more nonsense words than real word substitutions, whilst only six people identified more real words than nonsense words ($Z = 4.69$, $p < 0.01$). The strongest comparison was between the 43 people who identified more consonant strings than real words and the single person who identified more real words than consonant strings ($Z = 6.18$, $p < 0.01$). However, on only one occasion did a child identify only all six consonant strings and no other substitutions and on three occasions were all six consonant strings and all six nonsense words underlined and none of the real word substitutions.

Discussion

The results of this experiment supported the hypothesis that progressively more consonant strings than decodable nonsense words than real but contextually inappropriate words when substituted into a passage would be underlined. However, the size of the effect seemed to be dependent on the passage into which the substitutions were made. Nevertheless, all differences were in the expected direction. Additionally, it was clear even quite young children are highly sensitive to phonological irregularities. However, they are not so sensitive at detecting when a word's meaning does not fit in with the overall meaning of the passage. There was also no evidence of a practice effect so they did not do better on the third passage than on the first passage once type of substitution had been taken into account. There was also no evidence the children were failing to underline a significant number of substitutions because they had assigned them a meaning; nor did they detect more substitutions made for words important to understanding the meaning of the passage than substitutions for unimportant words. Overall, the results are consistent with the children sometimes reading in a word by word fashion with violations of rules relating to individual words being readily detected but violations of overall passage meaning by individual words being less readily detected.

Though far fewer real word substitutions were detected, nevertheless, 46 of the 54 children identified at least one real word substitution. However, there was no control over the part of speech of the substitution. The real words may therefore have been detected not because they did not fit in with the passage semantically, but because they did not fit in with the passage syntactically. This possibility will be examined in Chapter Four.

CHAPTER FOUR

STANDARDS FOR EVALUATING COMPREHENSION OF TEXT

Baker (1985a) distinguished three standards that readers use to evaluate their understanding of text: lexical, syntactic and semantic. The lexical standard operates at the level of individual words and the surrounding context can be completely ignored. An example of a lexical standard would be judging that a string of letters was not a real word or that the meaning of a particular word was not known. Children might have applied a lexical standard in Experiments One and Two to identify consonant strings and decodable nonsense words. The syntactic standard operates at the level of grammatical rules. An example might be judging that the sequencing of subject, verb and object is scrambled. The semantic standard requires consideration of the meaning of the text. Baker (1985a) classified this into five categories:

1. Propositional cohesiveness, checking that the ideas expressed in adjacent propositions can be successfully integrated.
2. Structural cohesiveness, checking that the ideas expressed throughout the text are thematically compatible.
3. External consistency, checking that the ideas in the text are consistent with what one already knows.
4. Internal consistency, checking that the ideas expressed in the text are consistent with one another.
5. Informational clarity and completeness, checking that the text clearly states all of the information necessary to achieve a specific goal.

In experiments employing error detection, one could only obtain unambiguous information on children's use of semantic standards of comprehension if the substitution made in the text did not violate lexical or syntactic constraints. In Experiment Two, some of the real word substitutions were a different part of speech from the words for which they were substituted. It is possible therefore, that the use of syntactic standards of comprehension may have resulted in the detection of some real word substitutions.

Evidence for the use of a syntactic standard has been derived from studies of errors made in oral reading. A number of studies of errors made by readers indicate they are sensitive to part of speech (Clay 1968, Kolers 1970, Weber 1970). Most of the errors in these studies were substitutions which ranged from 73 to 82 percent of the errors made, though Beebe (1980) found only 40% of errors were substitutions. Kolers (1970) reported on an analysis he carried out on the errors college students made when reading geometrically transformed passages. About three-quarters of the errors made in reading nouns, verbs and prepositions were substitutions of other nouns, verbs or prepositions. For other parts of speech, this occurred for about half the substitutions. Weber (1970) reported on errors made by 6 year olds. Of these, just over 90% were grammatically appropriate to the preceding context and just over 60% were acceptable to both preceding and following grammatical context. Just over 60% of the substitutions made were the same part of speech as the word they replaced. The results of both Kolers and Weber are in line with those obtained by Clay (1968) who found just over 70% of substitutions were in the same morpheme class (nearly identical to part of speech) when she analysed reading errors made by 5 to 6 year old children. Carnine, Carnine and Gersten (1984) also found 79% of real word substitutions made by third grade students were grammatically correct

though they only found 20% were grammatically correct amongst first grade children. Overall, these results indicate that both young beginning readers and older competent readers are sensitive to grammatical category.

However, when it comes to manipulating text to obtain information on the relative contribution of syntactic and semantic components, Baker (1985a) pointed out a problem. 'By definition, syntactically anomalous text is also semantically anomalous, so it may be difficult to tell whether syntactic or semantic evaluation leads to detection of the anomaly', (Baker 1985a p163). This however, only applies to major syntactical aberrations. For example, the sentence 'Emerson once say that every man is as lazy as he dares' violates syntax but preserves meaning (Kolers 1970).

However, some information on the relative contribution of syntax and semantics can be gained using a design devised by Guthrie (1973). Guthrie offered 7 to 10 year old readers alternative words in sentences they were asked to read. The alternatives included one correct, one that was the same part of speech and one that was in a different form class to the correct word but had links with the content of the passage, e.g.

horses
Both flowers lifted their ears
talk

The child was asked to read the passage silently and circle the correct alternative. The words for which alternatives were offered fell into four categories: nouns, verbs, modifiers (adjectives and adverbs) and functions (prepositions, articles, conjunctions). Guthrie found the readers, when they made mistakes, were more likely to select an alternative that was the same part of speech as the correct word when it was a verb or function word and

less likely to select the alternative that was the same part of speech for nouns and modifiers.

The results of Guthrie's experiment may indicate differential sensitivity to part of speech. This could be further investigated by comparing alternatives that were semantically equivalent in that neither made sense in the passage in which they were substituted. If one was the same part of speech as the word it replaced, and one was a different part of speech, then if the child was operating a purely semantic standard of comprehension one would expect no difference in the rate of detection of the substitutions. However, if a mixed or purely syntactic standard of comprehension was being used, then one would expect significantly greater numbers of substitutions which were not the same part of speech to be detected.

One problem is that classification by part of speech can be ambiguous. For example, of the alternatives offered in the sentence taken from Guthrie's (1973) work both flowers and talk could be either a verb or a noun. Nevertheless, the robustness of the findings of Clay, Kolars and Weber regarding the consistency with which substitutions made by readers are the same part of speech as the word they replace suggests part of speech is a category which has empirical value. Given that some of the real word substitutions made in Experiment Two were not the same part of a speech as the word they replaced, the use of syntactic rather than semantic standards may have been sufficient to detect them. For this reason, it is proposed to compare children's ability to detect real word substitutions all of which are meaningless in the passage, but only some of which are the same part of speech as the words they replaced. If children are employing a syntactic standard of comprehension, one would anticipate substitutions which were a

different part of speech from the words they replaced would be detected more often.

EXPERIMENT THREE

Subjects

The subjects were randomly selected from the four Year Four classes of an urban junior school so that each class contributed six boys and six girls. The sample of 24 boys and 24 girls ranged in age from 8 years 6 months to 9 years 5 months with a mean age of 8 years 11 months and a standard deviation of 3 months.

Materials

Two passages of 100 words in length and capable of being read by an average reader of between 7½ and 8½ years of age were selected from those prepared for Experiments One and Two. Both these passages were modified so that the same six words were changed in one of two ways. In both forms, the substitutions were real words that made no sense in the context of the passage. However, in one form (same) the substitution was the same part of speech as the original word and in the other form (different), the substitution was a different part of speech from the original word. Alternative substitutions were equal in length and were either four or six letters in length. The words substituted were the original six previously modified in Passages Two and Three of Experiment Two of which three were important with regard to the passages' meaning and three were unimportant.

Procedure

Each subject was seen individually. The task was introduced by the experimenter saying, 'I would like your help. I have got some stories here

which I would like to put in a book. I want to make sure the children reading them can understand them. Would you please read this story to yourself and underline any word that doesn't make sense in the story.'

Each subject was then presented in succession with both passages so that each subject read one with substitutions which were the same part of speech and one with substitutions which were a different part of speech, from the words they replaced.

	First	Second	Mean Age (Sd)
Group One	B2 S	B3 D	9 yrs 2 mths (2 months)
Group Two	B2 D	B3 S	9 yrs (3 months)
Group Three	B3 S	B2 D	8 yrs 10 mths (3 months)
Group Four	B3 D	B2 S	8 yrs 11 mths (3 months)

B2 = Passage Two

S = Same part of speech

B3 = Passage Three

D = Different part of speech

After reading each passage, each subject was asked to read from the passage any substitution they had not underlined and they were also asked what the word meant in the story they had just read.

Results

A preliminary breakdown of the substitutions underlined by passage and type of substitution is given in Table 4.1.

Table 4.1 Mean scores (and standard deviations) by passage and type of substitution

Passage	Same	Different
Two	3.88 (1.54)	4.21 (1.56)
Three	3.08 (1.53)	4.13 (1.68)
Both Passages	3.48 (1.57)	4.17 (1.60)

Detections of substitutions for important and unimportant words were compared using a Wilcoxon Matched-Pairs Signed Ranks Test. There was no difference ($Z = -0.29$, $n = 32$, ns). This parallels the results obtained in Experiments One and Two and indicates that the relative importance of the substituted words need not be included in further analyses.

Further analysis using a Type V analysis of variance (Lindquist 1956) was conducted with three within subject factors: part of speech of substitution (same or different), passage and session. The results are summarised in Table 4.2

Table 4.2 Summary of analysis of variance for Experiment Three

Source	df	SS	MS	F	P
Between subjects	47	143.49			
AB	1	3.01	3.01	0.99	
AC	1	1.76	1.76	0.57	
BC	1	1.76	1.76	0.57	
Error	44	136.96	3.11		
Within subjects	48	104.5			
A	1	11.34	11.34	6.05	0.05
B	1	4.59	4.59	2.45	
C	1	3.76	3.76	2.00	
ABC	1	2.35	2.35	1.25	
Error	44	82.46	1.87		

A = same/different

B = passage

C = session

From Table 4.2 it can be seen there is a significant main effect for type of substitution. Examination of Table 4.1 indicates this is in favour of more substitutions being detected when they are a different part of speech from the word for which they are substituted. This suggests some contextually inappropriate real word substitutions are detected through the application of a syntactic standard of comprehension. However only two readers identified only substitutions of a different part of speech and no substitutions of the same part of speech.

As in Experiment Two, the readers were asked to give the meanings of words they did not underline. Of the 209 substitutions they did not

underline, the children proposed meanings for 127 or just over 60% which is in line with the 54% of real word substitutions for which meanings were proposed in Experiment Two. Of these only 11 or about 5% made sense when substituted into the passage. Again, this is of the same order as the 7% real word substitutions for which proposed meanings made sense in Experiment Two. In line with both previous experiments, this does not suggest that the readers are investing significant numbers of the substitutions with appropriate meaning. This combined with readers on average failing to detect over one in three real word substitutions suggests that they are failing to effectively employ a semantic standard of comprehension.

Discussion

The results of Experiment Three extend the results of Experiments One and Two. For convenience, the results of the three experiments are recorded in Table 4.3.

Table 4.3 Comparison of results of experiments on single word error detection in junior aged children

	Percentage of errors detected				
	Consonant strings	Pseudo-words	Real Words (Total)	Real Words (Different POS)	Real Words (Same POS)
Expt One	82	64			
Expt Two	93	77	50		
Expt Three			64	69	58
Paris and Myers (1981)		22			
Baker (1984a Expt 2)		94			
Baker (1984b)		53			38
Baker (1985b)		48			

POS = part of speech

It is clear from the results of Experiment One, Two and Three that eight, nine and ten year old children are very good at detecting consonant strings which cannot be decoded. They are however, much less good at detecting non-words which can be sounded out and less good still at detecting real words which do not make sense in the passage, particularly when they are the same part of speech as the word for which they were substituted. If one of the simplest standards of lexical comprehension monitoring is whether a word can be sounded out or not, one would expect the rate of detection of consonant strings to remain high however hard the passage. For real words, whose detection relies on the use of syntactic or semantic standards, one would expect the rate of detection to decrease for passages which were hard to understand.

It has been assumed in the discussion so far that the children were using decodability as the standard against which consonant strings were identified but it is also possible that some children may have used knowledge of orthographic structure to identify them as not being words. This interpretation assumes that over time readers can 'derive information about legal occurrences of letters or letter sequences from repeated exposures to printed words and that they employ this information to facilitate word recognition in normal reading tasks' (Massaro, Taylor, Venezky, Jastrzembski and Lucas 1980).

Information on orthographic structure falls into two categories statistical and rule governed. Statistical information is derived from the frequency with which letters, letter sequences and words occur. Rules of orthographic usage arise from phonological constraints or scribal conventions. An example of a phonological constraint is that *dl* never occurs as an initial consonant cluster in English. Scribal conventions include prohibitions on doubling initial consonants and separation of spellings of homophones e.g. *to*, *two*. That junior school children can write the correct form of homophones (Doctor and Coltheart 1980) indicates that some visual information must be available. However spellings of common words such as *cold* as *cld* (written by a 9 year old on the 17 March 1994), *stamp* as *stp* (Morris 1983) or *car* as *cr* (Goswami and Bryant 1990) indicate that awareness that consonant strings are not words is not universal. Even within consonant strings decisions by 6 to 10 year olds as to whether or not they are words are influenced by the likelihood of individual consonant pairs in the string. Moreover the presence of a vowel will slow the rejection by younger children of otherwise unlikely letter strings so that *k c i b f m* is rejected more slowly than *f d j c b l* (Henderson and Chard 1980). With respect to concluding which standard was used to identify the consonant

strings in Experiments One and Two there is insufficient evidence. Asking children why they underlined a consonant string to determine whether it was because it was recognised as not being a word or because it could not be sounded out would have provided information on which to make a judgement. Incorporating this control in future experimental work is identified as a refinement to the experiments reported on here in the concluding section of the last chapter.

When the results of Experiments One, Two and Three are placed in the context of previous experimental findings on the detection of single word errors in continuous prose, there is an overlap in the detection rates for real words and decodable non-words and also a wide range of detection rates. This suggests that factors unique to a particular experiment, e.g., age of reader, instructions given, reading ability, difficulty of passage and length of substitution may bring about substantial changes in the detection rate. This would suggest caution should be applied in generalising from the immediate concerns of the experiment in question to wider issues of rate of detection of errors in reading generally. But it does seem reasonable to conclude that in general, readers of junior age are not very good at detecting decodable errors in continuous prose.

Even though there is evidence that young readers are not good at detecting decodable substitutions in passages it is an open question as to what information the detection of errors gives about the reader's comprehension of the passage. This is less of an issue in the three experiments carried out in this study as the major comparison was between the frequency of detection of variations in substitutions for the same word. However, underlying this was the presumption that this gave some indication of the reader's comprehension of the passage. Winograd and Johnson (1982) list 12

reasons why one cannot say for certain that the subjects' comprehension monitoring abilities have failed because they do not indicate an error. One of these was that the subjects may assign alternative meanings to the text. However, in all three of these experiments, when the children were asked about the substitutions they did not identify, they were typically only able to give reasonable alternative meanings to those anticipated by the experimenter on between 4 and 15 percent of occasions. Moreover, the detection rate for substitutions appeared to be independent of the importance of the word substituted to the meaning of the passage. This does not lead one to feel confident that eight, nine and ten year old children accurately monitor their semantic comprehension of passages they are reading.

One factor that may influence their level of comprehension is the readability of the passage using sentence length and number of polysyllabic words as a measure of readability. However this may have a differential effect on the rates of detection of different types of substitution. This will be investigated in Chapter Five.

CHAPTER FIVE

THE INFLUENCE OF READING LEVEL ON ERROR DETECTION

In Chapter Four it was suggested that the reading level of a passage may have a differential effect on the detection of different types of errors in the passage. The level of difficulty or readability of a passage can be judged in a number of ways. Some methods measure the passage's readability by directly measuring a reader's performance while reading whilst other methods use indirect measures of difficulty, e.g. sentence length or number of polysyllabic words.

A commonly used set of categories for classifying reading material relative to a reader's performance was devised by Betts (1946). He divided reading material into three categories: Basal (better known as Independent), Instructional and Frustration. Amongst other factors, he identified comprehension criteria associated with these three levels. These are:

- Independent level:** A minimum comprehension score of at least ninety percent, based on both factual and inferential type questions.
- Instructional level:** A minimum comprehension score of at least seventy-five percent, based on both factual and inferential questions.
- Frustration level:** A comprehension score of less than fifty percent, on factual and inferential questions.

Betts (1946) also identified a Capacity Level which he described as the highest level of readability of material which the learner can comprehend when the material is read to him.

These levels were in part defined in terms of answering factual and inferential comprehension questions. If a different method of assessing comprehension is used, then the criteria would have to be redefined. For example, Bormuth (1967) found a person who correctly answered 38 percent of cloze items also usually obtained a score of 75 percent on multiple choice comprehension questions on the same passage. Similarly, a cloze score of 50 percent was found to be comparable to a multiple choice score of 90 percent.

It is also possible to define readability in terms of characteristics of the passage. For example, the passages used in Experiments One, Two and Three were specifically written to have a very high reading ease score as calculated by the Flesch Formula (Flesch 1948). This formula uses average sentence length and number of syllables per 100 words in the calculation of readability. The formula was specifically related to comprehension in that a score of 100 indicated that a fourth grade child would be able to correctly answer three-quarters of questions asked about a passage so rated. Passages One, Two and Three had reading ease scores of 102, 104 and 103 respectively. For an average Year Four child, this would suggest the passages would be at the instructional level using Betts (1946) criteria.

However, the Flesch Formula was standardised over 40 years ago in the United States. A more modern British reading test, the New MacMillan Reading Analysis (Vincent and de la Mare 1985) can be used to estimate how it transfers to this country and generation in the same way that

Spooncer (1976) used the Neale Analysis of Reading Ability as a standard against which to compare the FOG Index of Readability. The New MacMillan Reading Analysis consists of three parallel sets of short self-contained pieces of prose. Within each set or form of the test, the pieces of prose are of increasing difficulty. According to the norms of the test, a student reading only the first two passages correctly would be reading at approximately the standard of an average seven and a half year old. The Flesch reading ease score of the first two passages is 107. If students read only the first three passages correctly, they would be reading at approximately the standard of an average eight and a half year old. The combined Flesch reading ease score of the first three passages is 97. All three of the experimental passages used appear therefore to be of similar difficulty and to be capable of being read by an average reader between seven and half and eight and a half years of age, ie., a Year Three or Year Four pupil. One can therefore be reasonably confident the passages used in Experiments One, Two and Three were at least at the instructional level for most of the pupils who acted as subjects.

It was hypothesised in Chapter Four that the readability of a passage may differentially affect the rate of detection of different types of substitutions into the passage. It was noted that the rate of detection of consonant strings had been significantly higher than other types of substitution. The application of a purely lexical standard would result in the detection of this type of substitution and this should remain relatively uninfluenced by a passage being close to the frustration level of the readers. However, as semantic standards of comprehension need to be employed in the detection of real but contextually inappropriate words, one could predict a marked decrease in their detection in a harder passage. Passages are defined as

harder if they have a longer average sentence length and more polysyllabic words. This was investigated in Experiment Four (A).

EXPERIMENT FOUR (A)

Subjects

Eighty Year Four pupils from an urban junior school ranging in age from 8 years 1 month to 9 years with a mean age of 8 years 7 months and a standard deviation of 3.6 months took part.

Materials

Passage Three was used in two forms. One was as it was used in Experiment Three. In this form, it had a Flesch reading ease of score of 103. It was also prepared in a modified form so that it would have longer sentences and use more polysyllabic words whilst retaining the same story line and number of words. In this modified form, it had a Flesch reading ease score of 83. Flesch (1948) estimated 10 points on his scale corresponded to a grade level. So the passage with the lower reading ease score of 83 corresponds to approximately two grades higher reading material. The two passages were each prepared in two forms: one with six real, but contextually inappropriate words of the same part of speech substituted for the same six words in both passages, and one with six sets of consonant strings, substituted for the same six words in both passages. That is the two passages differed in their readability but had the same meaning and the same substitutions (either real words or consonant strings) in the same places, replacing the same words.

Procedure

Each subject was seen individually. Each subject was randomly assigned one of the four possible combinations of passage and substitution type.

The task was introduced by the experimenter saying, 'I would like your help. I have got a story here called 'My friend from outer space' which I would like to put in a book. I want to make sure the children reading it can understand it. Would you please read the story to yourself and underline any word that does not make sense in the story.'

Results

A preliminary breakdown of the scores obtained by the subjects for each of the four versions of the passage is shown in Table 5.1.

Table 5.1 Mean scores (and standard deviations) for each version of the passage

		Substitution Type	
		Real Words	Consonant Strings
R E A D A B I L I T Y	E A S Y H A R D	3.20 (1.85)	4.80 (1.80)
		1.35 (1.31)	4.35 (2.50)

Initial inspection the results indicated a pattern consistent with the hypothesis under investigation. 53 percent of real word substitutions were underlined in the original (easier) version of the passage. This is in line with the 51 percent detection rate found for the same real word substitutions at

the same point in the same passage in Experiment Three. The detection rate for consonant strings in the original version of the passage at 80 percent is somewhat less than the 97 percent detection rate for the consonant strings substituted at the same point in the same passage in Experiment Two. However, whilst the detection rate for the same real words in the harder version of the passage plummeted to 22.5 percent, the detection rate for consonant strings in the harder version of the passage remained more stable at 72.5 percent.

Further analysis was carried out using a completely randomised two-factor analysis of variance. The results of the analysis are summarised in Table 5.2.

Table 5.2 Summary of analysis of variance for Experiment Four (A)

Source	SS	df	MS	F	P
Treatments	142.05	3			
A	26.45	1	26.45	7.14	0.01
B	105.8	1	105.8	28.56	0.01
A X B	9.8	1	9.8	2.65	NS
Residual	281.5	76	3.70		
Total	423.55	79			

A = Hard/Easy

B = Real words/consonant strings

From Table 5.2 it can be seen there was a significant main effect for both readability and substitution type. There was no significant overall interaction between the readability of the passage and the type of substitution. A Scheffe test carried out on the mean scores obtained for

consonant string substitutions under both passage conditions indicated that the difference was not significant ($F = 0.55$, $df 1,76$). By contrast, a Scheffe test carried out on the mean scores obtained for real word substitutions under both passage conditions indicated that the difference was highly significant ($F = 9.24$, $df 1,76$, $p < 0.01$). In line with previous experiments, the rate of detection of consonant string substitutions was much higher than real words of the same part of speech as the words they replaced, but which made no sense in the passage. As predicted, fewer substitutions were detected in the version of the passage which had longer sentences and more polysyllabic words. However, the rate of detection of the consonant strings remained relatively constant whilst the rate of detection of the real words dropped significantly.

Discussion

The Flesch Formula which was used to assess the readability of the three passages used in this and earlier experiments uses average sentence length and number of syllables per 100 words in the calculation of readability. Readability formulas have been criticised by Davison (1984) as crude approximations of what reflects difficulties in processing and interpreting written language. Nevertheless, rewriting this passage so that the average sentence length and number of polysyllabic words increased, whilst the total number of words remained the same significantly reduced the number of real words substitutions identified. It is of note that the story remained the same in both versions.

Though the number of words was kept constant, the number of letters in the harder passage increased from 335 to 399. Adams (1990) has argued that in

reading, one reads all the letters and so the harder passage may involve greater reading effort for this reason.

The difference in detection rate of consonant strings over real words is in line with the distinction made by Baker (1985a) between three different standards of comprehension monitoring. The consonant strings can be detected by the application of a purely lexical standard whilst the detection of the real word substitutions in this study would require the application of a semantic standard of comprehension. The significant reduction in the detection of real word substitutions when the sentence length and number of polysyllabic words increased as compared to the high and relatively stable rate of detection of consonant strings also provides evidence as to the existence of different standards of comprehension monitoring.

However, a weakness of this experiment was that it relied on only one passage. The effects reported may be attributable to this one passage and not generalisable to other pieces of text. Moreover, the consonant string substitutions were detectable at the rudimentary level of lexical monitoring of whether or not the substitution was decodable. For this reason, it is proposed to replicate this experiment using a second passage with substitutions consisting of pseudowords and real words of the same part of speech as those substituted.

EXPERIMENT FOUR (B)

Subjects

80 Year Four pupils from an urban junior school ranging in age from 8 years 5 months to 9 years 5 months, with a mean age of 8 years 10 months and a standard deviation of three months, took part. These were all the available Year Four pupils over the experimental period.

Materials

Passage One was used in two forms. One was as it was used in Experiment Two. In this form, it had a Flesch reading ease score of 102. It was also prepared in a modified form so that it would have longer sentences and use more polysyllabic words whilst retaining the same story line and number of words. In this modified form, it had a Flesch reading ease score of 82, ie., approximately two grades higher than in the original form. The two passages were each prepared in two forms: one with six real but contextually inappropriate words of the same part of speech, substituted for the same six words in both passages, and one with six pseudowords substituted for the same six words in both passages. The pseudowords were prepared by changing only one letter in the real word substitutions to preserve the similarity between the substitutions. Both passages therefore, differed in their readability but had the same meaning and the same substitutions (either real words or pseudowords) in the same places replacing the same words.

Procedure

The identical procedure to that used in Experiment Four (A) was followed.

Results

A preliminary breakdown of the scores obtained by the subjects for each of the four versions of the passage is shown in Table 5.3.

Table 5.3 Mean scores (and standard deviations) for each version of the passage

		Substitution Type	
		Real words	Pseudowords
R E A D A B I L I T Y	E A S Y H A R D	2.5 (1.15)	4.25 (1.89)
		0.95 (1.00)	3.4 (0.96)

Initial inspection of the results indicated a pattern consistent with the hypothesis under investigation. Forty two percent of the real word substitutions were underlined in the original (easier) version of the passage. The detection rate of seventy one percent of the pseudowords in the original version of the passage is in line with previously obtained detection rates for pseudowords. However, while the detection rates for real words dropped to sixteen percent in the harder version of the passage, that for pseudowords remained more constant at fifty seven percent.

Further analysis was carried out using a completely randomised two-factor analysis of variance. The results of the analysis are summarised in Table 5.4.

Table 5.4 Summary of analysis of variance for Experiment Four (B)

Source	SS	df	MS	F	P
Treatment	119.45	3			
A	28.8	1	28.8	11.85	0.01
B	88.2	1	88.2	36.30	0.01
AXB	2.45	1	2.45	1.01	NS
Residual	184.5	76	2.43		
Total	303.95	79			

A = Hard/Easy B = Real words/Pseudowords

From Table 5.4 it can be seen there was a significant main effect for both readability and substitution type. There was no significant overall interaction between the readability of the passage and the type of substitution. A Scheffe test carried out on the mean scores obtained for pseudoword substitutions under both passage conditions indicated that the difference was not significant ($F = 2.97$, $df 1,76$). In contrast, a Scheffe test carried out on the mean scores obtained for real word substitutions under both passage conditions indicated that the difference was highly significant ($F = 9.89$ $df 1,76$ $p = 0.01$). As predicted, fewer substitutions were detected in the harder version of the passage. However, the rate of detection of pseudowords remained relatively stable, whilst that for real words dropped significantly.

Discussion

The results of both these experiments support the distinction made by Baker (1985a) between lexical and semantic standards of comprehension monitoring. They also illustrate that it is possible to tell the same story in ways which either inhibit or facilitate comprehension. In these experiments, this was done by manipulating the sentence length and number of polysyllabic words. Marks, Doctorow and Wittrock (1974) also found that by changing one in twelve words in a story from an infrequently occurring to a more commonly occurring synonym, they were able to significantly enhance comprehension of the story by sixth grade students as measured by their performance on a multiple choice test. Thorndyke (1977) also obtained enhanced ratings of comprehensibility for stories which followed a simple narrative structure of setting, theme, plot and resolution as compared to when the same content was presented in ways which deviated from this story grammar.

Kletzien (1981) investigated strategy use by 10th and 11th grade high school students when completing cloze passages at three different levels of difficulty: independent, instructional and frustration. The 48 students were divided into good comprehenders and poor comprehenders. As one might expect, the harder the passage the less well the students completed the spaces. Good comprehenders reported using significantly more strategies on the independent level passage but there was no difference in strategy use between instructional level and frustration level. Poor comprehenders however, reported significantly more strategy use at independent level than the two harder levels, but also significantly greater strategy use at instructional level than frustration level. Whilst one must interpret this result conservatively as only one passage represented each level of difficulty and

this might have encouraged particular response patterns, it does suggest exposing students to passages at their frustration level is likely to result in a poor strategic approach as well as poor performance. Presenting students with passages at their frustration level is therefore unlikely to lead to them having a positive learning experience. In line with this, Carver (1987) has argued that though it is when students are presented with passages at their frustration level that apparently large, but non-transferable gains in comprehension scores can be achieved by the application of study skills, this does not justify a practice ordinarily considered poor teaching.

Armbruster (1984) has described text which violates principles of effective learning as inconsiderate. She considers that during their elementary school years, children read many pages of inconsiderate text from poorly written textbooks. Armbruster suggests two ways of dealing with this problem: careful selection of texts so considerate texts are selected and teaching students to deal with inconsiderate texts. Put more globally, one can focus on teaching students to improve their comprehension of text or one can present them with text that is more comprehensible. It would appear more useful to explore enhancing students' comprehension of text as this could be more generally applied to a range of texts including those using technical words for which easier alternatives may not be readily available. In Appendix C, empirical studies of programmes designed to enhance students' comprehension of text are reviewed. One common strategy for enhancing comprehension referred to in the introductory chapter was encouraging readers to spend more time engaged with the text often by re-reading it. This will be discussed further in Chapter Six.

CHAPTER SIX

READING AND RE-READING

In Chapter One, it was noted that Carver (1990a) had predicted that asking pupils to re-read a text before being assessed on their comprehension of it should enhance their performance. In contrast, Yuill and Oakhill (1991) hypothesised that comprehension involved the concurrent production of a mental model whilst reading and that poor comprehenders would not benefit from the opportunity of studying a text.

A distinction should be drawn between a reader looking back to find a particular part of the text in order to answer a question as investigated by Garner, Hare, Alexander, Haynes and Winograd (1984) and Alvermann (1988) amongst others, and systematically re-reading a passage as in Chan and Cole (1986) and Berkowitz (1986). The former strategy in many ways would appear the more efficient since one would not be able to apply the total rereading strategy routinely to lengthy pieces of prose. However, lookbacks may not work so well for longer texts as students may tend to look back to the wrong section of the text, though graphic organisers outlining the gist of the passage may assist in this (Alvermann 1988). However, rereading in total may lead in time to more selective rereading perhaps only used when the reader is aware of failing to understand the passage.

There is a substantial body of evidence supporting the idea that rereading is an effective strategy for enhancing recall and comprehension of text. In the early forties, Arnold (1942) investigated four study techniques: repeat reading, underlining and margin notes, listing and summarising amongst 242

American college students when presented with learning about Latin-American history. Four measures of comprehension were used by Arnold: single word completion, single word answer, single choice recognition and parallel column matching questions though he does not elaborate on what these involved in any detail. Arnold's original analysis is hard to follow as each student was taught each technique but a reanalysis by Anderson (1980) suggested that repeat reading was the most effective of the four techniques examined.

Rothkopf (1968), also found adult subjects performed better on cloze tests after repeated exposure to text. However, though his subjects read the text 0, 1, 2, or 4 times, he found the effects of repeated exposure levelled off after 2 exposures and if the subjects were repeatedly exposed to the text, the average inspection time per page decreased. Similarly, Howe and Singer (1975) found university students performed better on measures of recall when just asked to read or reread a passage than when asked to copy or summarise it. However, they found no difference in a follow-up experiment between the performance of students asked to read a passage themselves and those who listened to it 3 times at normal speaking rate. Barnett and Seefeldt (1989) also found enhanced performance by university students on a measure of immediate recall by those asked to read a text twice as compared with those who only read it once. Krug, Davis and Glover (1990) found distributed repeated reading with a week separating two readings enhanced performance on a test of free recall more than two successive rereadings followed by immediate testing. Krug et al also looked at inspection times and found it was only when students were asked to read a passage verbatim for a second time there was a marked fall off in inspection times. This fall off was not evident when the second reading was of a paraphrased version of the passage or the readings were separated by a

week. This might suggest possible confounding between total inspection time of the passage and the condition under which it is presented.

These investigations of the effects of rereading have tended to use only a single measure of outcome usually of recall. They have also tended to use adult subjects. Exceptions to this are Chan and Cole (1986) who found that 11 year old mainstream children showed enhanced performance on a multiple choice measure of comprehension following rereading and O'Shea, Sindelar and O'Shea (1985) who found third graders could recall more story propositions after three or seven readings than after one reading. An important question is therefore, does rereading enhance junior aged pupils' ability to comprehend text as judged by their performance on more than one measure of comprehension.

Also of interest is the mechanism by which rereading may improve comprehension. It is possible the first reading may act as an advance organiser (Ausubel 1960) or preview. Evidence exists that previews increase both literal and inferential comprehension (Graves, Cooke and Laberge 1983, McCormick 1989). Mayer (1983) compared the effect of repeated listening to a passage up to three times to the effect of an advance organiser using female college students as subjects. He found similar effects for both in that they both enhanced recall of detail and problem solving transfer.

Mayer (1983), proposed that repetition allows the learner to add more overall information to knowledge previously obtained so influencing how much is learned. This he called his quantitative hypothesis. Mayer (1983) also proposed the qualitative hypothesis which is that repetition allows the learner to use successively more sophisticated encoding strategies based on

knowledge obtained in previous contacts with the material affecting what is learnt. Amlud, Kardash and Kulhavy (1986) looked at this using one, two or three readings of a passage by university students who then took an immediate and delayed recall test. In common with the findings of Rothkopf (1968), a second reading enhanced overall performance but a third reading added little or nothing to the students' recall of either main ideas or details. This would suggest Mayer's hypothesis only holds over two readings.

Another factor that may be influence outcome is that a child's ability to decode the passage may be enhanced by an initial reading. Experimental support for this comes from the work of Gonzales and Elijah (1975) who found that amongst third graders, oral reading errors decreased between the first and second reading of a passage. However, it does not follow that improving decoding improves comprehension, and indeed, Smith (1973) argued that often readers have to understand a passage before they can successfully decode many words. Students have been trained in rapid decoding by Fleisher, Jenkins and Pany (1979), and also by Yuill and Oakhill (1988) and in both studies, the training in decoding was not associated with an improvement in comprehension. Indeed, Yuill and Oakhill found skilled comprehenders' reading times for a word list were slightly, although not significantly slower than less skilled comprehenders. The results of the four experiments on error detection reported in Chapters Two, Three, Four and Five also suggest that though junior school children are quite skilled at detecting words which cannot be sounded out, they are much less skilled at detecting words whose meaning does not fit in with the rest of the passage. The experimental evidence does not suggest that any enhancement of decoding that may result from reading a passage twice is likely to enhance comprehension of the passage.

A possibility is that a second reading might enable a student to organise the content of the text in a more meaningful way by enabling an organising framework to be formed by the first reading. From this would follow the prediction that there would be no difference in the performance of children given an organising framework in advance of reading the passage once and those who read the passage twice before completing a comprehension exercise.

It was felt that enhanced performance on two measures of comprehension should be obtained to indicate generally improved comprehension. As well as an error detection task similar to that used in Experiment Three with real words of the same part of speech substituted into a passage a cloze version of the same passage was also prepared. Cloze which involves the elimination of words from a passage which the reader has to reinsert was a technique pioneered by Ebbinghaus (Burt 1921) and was both named and developed by Taylor (1953). Cloze completion was described by Weaver (1965) as a search procedure involving matching possible words to the semantic and syntactic clues provided by the rest of the sentence, resulting if successful, in the generation of a word which is syntactically and semantically consistent with the remaining words in the sentence. Cloze procedure scores correlate well (0.71 to 0.77) with paragraph meaning scores of standardised tests (Ruddell 1965) and also with multiple choice comprehension tests for which Bormuth (1967) found a correlation of 0.95. Rankin and Culhane (1969) found a lower correlation of 0.68 in a replication of Bormuth's (1967) study. Bormuth (1968) also found correlations ranging from 0.90 to 0.96 between passage difficulties determined using cloze tests and passage difficulties determined using comprehension and word recognition tests.

Cloze has however, been challenged as not measuring readers' ability to integrate information across sentence boundaries. This followed from the findings of Shanahan, Kamil and Tobin (1982) that the group means of readers for cloze tests on original as compared with scrambled sentence order did not significantly differ. However, McKenna and Layton (1990) found fifth grade students' ability to answer comprehension questions on scrambled passages was significantly correlated with their cloze performance on the passage. The comprehension questions required the integration of information from at least two sentences. They concluded that cloze scores may be sufficiently reflective of intersentential comprehension to warrant their use. Cloze procedures have also been incorporated into a number of widely used tests, e.g., GAPADOL, Wide-span, London Reading Test and so their use as an experimental measure of comprehension would seem legitimate.

EXPERIMENT FIVE

Subjects

144 subjects made up of 72 from each of two suburban junior schools were used. All the children came from Year Four and both schools contributed all their available pupils in that year group. The subjects ranged in age from 8 years 9 months to 9 years 9 months with a mean age of 9 years 2 months and a standard deviation of 3 months.

Materials

Two measures of comprehension were used. One was a cloze version of Passage Three from Experiment Two and one was an error detection version of Passage Three. Every tenth word (plus or minus one so as to make the distribution of substitutions unsystematic) was chosen for omission or substitution. No word was chosen whose meaning could not be determined from the surrounding text. From the possibilities that fulfilled these criteria, those for which it was possible to make a substitution that was the same part of speech but did not make sense in the passage were chosen. Of these, the one most likely to be easily spelt was selected. Where a space was left, a line equivalent to a word five letters in length was inserted on which the child could write. The passages were untitled.

Procedure

Each child was seen individually. The subjects each completed one of six possible tasks to which they were randomly assigned.

Group 1:

Completed cloze

Group 2:

Read passage before completing cloze

Group 3:

Given advance organiser, asked to complete cloze

Group 4:

Read passage and underlined words which didn't make sense

Group 5:

Read passage once, given pencil and then asked to read through and underline words which didn't make sense

Group 6:

Given advance organiser, asked to read through passage and underline words which didn't make sense

The control groups (Groups 1 and 4) were simply asked to either fill in any blanks so that the story made sense or to underline any words that did not make sense in the story.

Those who read the passage twice (Groups 2 and 5) were first asked to read it to themselves. They were then given a pencil and asked to read it again filling in any blanks so that the story made sense or underlining any words that did not make sense in the story.

Those subjects given an advance organiser (Groups 3 and 6) were told the story was about a small, space creature who is found by a little boy/girl who makes friends with him. One day, the space creature suddenly returns to space. They were then asked to read the passage and underline any words

that did not make sense in the story or fill in any blanks so the story made sense.

Scoring

The cloze exercises were independently marked by the experimenter and 2 colleagues. Where there were any discrepancies in the scores a simple majority determined whether the item was accepted or not.

The error detection exercises were marked by the experimenter and a colleague. There were no discrepancies in the scores awarded.

Results

A preliminary breakdown of the scores obtained by the subjects is given in Table 6.1.

Table 6.1 Mean scores out of ten (and standard deviations) by condition and task

Task	Condition		
	Control	Read twice	Advance Organiser
Cloze	5.67 (2.62)	6.92 (1.28)	7.21 (1.74)
Error detection	2.75 (2.4)	2.83 (2.28)	3.21 (2.43)

Differences between treatments within each task were small, but the mean scores for those asked to read the passage twice and those given an advance organiser were larger than that obtained by the control groups. Further analysis was carried out using a completely randomised two-factor analysis of variance. The results are summarised in Table 6.2.

Table 6.2 Summary of analysis of variance for Experiment Five

Source	df	SS	MS	F	P
Treatment	5	519.06	103.81		
A	1	484	484	102.33	0.01
B	2	24.89	12.45	2.63	NS
AXB	2	10.17	5.09	1.08	NS
Residual	138	652.92	4.73		
Total	143	1171.97			

A = cloze/error detection B = control/read twice/advance organiser

From Table 6.2 it can be seen there is a significant main effect for the comprehension task used. Examination of Table 6.1 indicated this was because the readers obtained over twice as high a score on average on the cloze task than they did on the error detection task. On the cloze task over all three conditions, the readers averaged 66% correct responses, whilst on the error detection task, they averaged a 29% detection rate. There was no significant effect for how the task was presented, nor was there any significant interaction effect.

The number of correct responses made for each of the ten items in each task is shown in Table 6.3.

Table 6.3 Correct responses made under each task condition

Comprehension Task	Item									
	1	2	3	4	5	6	7	8	9	10
Cloze	62	28	24	8	51	57	68	64	62	51
Error detection	9	20	7	23	41	30	31	23	23	4

To determine whether the order of difficulty of cloze items and detection of substitutions was consistent, a Spearman Rank Correlation was calculated. A coefficient of 0.36 was obtained between cloze scores and error detection scores for each item which is not significant at $p = 0.05$. The total scores obtained and the distribution of scores therefore both appear to differ in the two comprehension tasks.

Discussion

The results of this experiment do not support the hypothesis advanced that rereading would improve junior aged pupils' performance on comprehension tasks. It was noted during the experiment that a number of control pupils spent some time on their task and a number of those given the cloze exercise spontaneously read it first before starting to complete the task. This combined with the very low and limited range of scores obtained on the error detection task, may have combined to minimise any effect. No significant effect was found either on providing an advance organiser. This stands in contrast to the results of other experimental work such as that of Dole, Valencia, Greer and Wardrob (1991), who found two types of prereading instruction; either the teacher reading a script which provided students in Grade Five information necessary for understanding a text, or the

teacher discussing the students' prior knowledge about the text improved students' comprehension test score over those who received no prereading instruction. However, that no difference was found between the performance of those who read the passage twice and those given an advance organiser is in line with the original hypothesis and earlier experimental work (Mayer 1983).

In this experiment, the average score of 29% obtained on the error detection task was much lower than the 51% average obtained in Experiment Three on the same passage. In Experiment Three however, only 6 substitutions were made in the passage, and it may be easier to detect a few substitutions in a generally meaningful passage than a larger number of substitutions which may disrupt the whole meaning of the passage for some readers. In this experiment, no readers identified all the substitutions and only one reader identified nine substitutions whilst 14 readers failed to identify any. However, the significant difference obtained between the cloze results as compared with the error detection results at the same points in the passage indicate it may not be failure to comprehend which is the problem but differences in the demands made by the two tasks. The cloze space prompts the reader that something is missing whilst the error detection task requires the readers to realise themselves that something does not make sense. If little comprehension monitoring is going on, then it would be quite possible to read the passage and be unaware that it fails to make sense at particular points. What this may indicate is not that the readers are not capable of comprehension of age appropriate text but that they do not always actively monitor their comprehension unless prompted to do so. This will be examined further in Chapter Seven.

CHAPTER SEVEN

DO READERS KNOW WHEN THEY DON'T UNDERSTAND WHAT THEY ARE READING?

The difference in results obtained in Experiment Five between the cloze form of a story and one with substitutions, made at the same point in the same story may indicate a failure on the part of the readers to actively monitor their comprehension. This requires the reader to know he is meant to understand what he is reading, i.e., that it is a meaningful text and the purpose of reading it is to understand it.

It can be argued that the use of the word meaningful requires further clarification. Mosenthal (1987) discussed it in terms of either representational meaning the main linguistic idea, world meaning the real phenomenon referred to or cognitive meaning our personal knowledge of the subject of the text. For the purpose of this experiment, meaningful is restricted to representational meaning that is that the text consists of words which combine to form an internally consistent theme. This is not to imply that either world meaning or cognitive meaning will not inform the representational meaning given the passage by the reader but that for our purposes, meaningful refers to the reader at least making a representational meaning from the text.

An early study which asked children what they thought about reading was that of Reid (1966). She found that five year olds did not know the goals of reading. In a replication of Reid's study, Downing (1969) also concluded that young beginning readers have serious difficulty in understanding the purpose of written language. Wixson, Bosky, Yochum and Alvermann

(1984), found that some junior aged children believed that the purpose of reading was to pronounce all the words without making mistakes. Canney and Winograd (1979) as quoted in Jacobs and Paris (1979) found in a study of 8, 10, 12 and 14 year old children that younger and poorer readers attended more to decoding, whereas proficient readers knew that making sense was the goal of reading. Cairney (1988) asked 178 primary aged Australian children for the most important reason for reading a basal reader. Eighty percent of the responses recognised either the need to learn to read or to learn a specific aspect of reading. However, of those who referred to a specific aspect of reading, 26% made reference to decoding and only 2% to gaining meaning. Cairney concluded that the children did not see meaning as important when reading basal readers. By contrast, Weiss and Hagan (1988) found 41% of 100 kindergarten children in Nevada understood the connection between reading and acquiring knowledge or information. However, Weiss and Hagan had predominantly been using common reference material, e.g., menus, newspapers, calendars or letters in their survey. This was a point they themselves emphasised and they concluded by stating 'Understanding why and for what purposes people read is a logical prerequisite for learning how to read.'

There appears to be a large body of evidence that children up to and including those in Year Six do not always appreciate that reading is about conveying meaning. There is also evidence both from both the United States (Durkin 1978-79) and the UK (HMI 1989), that there is not always much systematic instruction in comprehension strategies in schools. This would not encourage an awareness of reading being a meaningful activity. The first pre-requisite for the application of a successful comprehension monitoring strategy that the child knows he should be reading for meaning cannot be assumed to exist and needs to be made explicit. That is the child needs to

be aware that the standard against which he should monitor his comprehension is a semantic standard.

Having established that the child's task is to understand the meaning of the text, how does he know when he's achieved this? He can apply a standard of internal consistency, i.e., that the text is not self-contradictory or a standard of external consistency i.e. that the text does not contradict or conflict with the child's knowledge of the world. Ultimately however, the reader either has to make a judgement that he understands the meaning the writer meant to convey or there needs to be some external criterion against which he can judge the completeness of his comprehension. For example, the author could give questions (and answers he considers acceptable) at the end of the passage against which the reader could judge his comprehension. Alternatively, a full understanding of the text may enable the reader to complete some tasks, e.g., prepare a meal or construct an item or select one out of a number of options.

In the absence of a clear task against which the reader can measure his comprehension, he has to make a judgement. Piaget (1932) has demonstrated that young children often believe they understand what others say to them though their understanding may not be that intended by the writer or speaker. For example, when five year olds were asked in a study by Papandropoulou and Sinclair (1974) to say a long word and a short word, they usually named a large object or small object. Older children up to ten may say 'a cat that is taking a walk' for a long word in which what is long is not so much the object but what is said about it. While a word is a highly abstract concept and children's understanding of its meaning has a literature of its own (Downing, Ayers and Schaefer 1978) this research does illustrate

mis-matches between intended and actual comprehension can occur at a fairly fundamental level.

If the reader is asked to make a judgement as to when he has understood, then besides motivational factors, the most likely factor influencing the accuracy of this judgement is practice. It would therefore be important that in the absence of comprehension assessment tasks, readers who were expected to comprehend text knew that they were meant to understand what was written and had practice at judging their comprehension of text. However, even for mature adult readers this may be a difficult task. Parker (1962) failed to find convincing evidence that college students were aware of organisational features of passages which would be useful to them in their study of the passages. Brown and Smiley (1977) did find that 18 year old students were able to reliably rate the structural units of a 400 word story and their recall of the units increased with their importance. However, the stories they used were short and were written at fifth grade level and so were probably very easy for college students. In a follow-up experiment, Brown, Smiley and Lawton (1978) found that college students prior to having experience recalling a text selected mainly the most important elements to serve as retrieval cues, but after experience of recall, they selected elements of intermediate importance. Brown et al found students attributed this change in strategy to themselves realising they would remember the main ideas without further effort and so concentrated on the intermediate material they had found harder to remember.

One problem with the Brown and Smiley experiments was the material used was short and very easy. Maki and Berry (1984) also investigated college students' prediction of recall, but this time of half or full chapters of an introductory text on psychology. They found only very small correlations

averaging between 0.15 and 0.23 between predictions and performance on a multiple choice test. They concluded that college students were not particularly good at assessing their levels of comprehension and future memory for test material. In related work on the detection of contradictions in passages, Glenberg, Wilkinson and Epstein (1982) found that despite drawing the college students' attention to the existence of contradictions in a text, and asking them to detect them, only 51% of them were detected. However, many students also judged they understood the passage well and 40% of them displayed what Glenberg et al called the illusion of knowing in which high ratings of understanding were linked with failure to detect the contradiction. They claimed the illusion of knowing is inconsistent with an assumption of active, accurate on-line comprehension. They proposed a default method of monitoring in which it is assumed comprehension has occurred until the reader is alerted otherwise by an error signal, e.g., encountering an unfamiliar technical term. This would be consistent with Miller and Yochum's (1991) finding that elementary grade level students with comprehension problems were generally unaware of their reading difficulties.

Hunter-Blanks, Ghatala, Pressley and Levin (1988) investigated university students' ability to monitor the ease with which sentences they were studying could be learned. The subjects who were all shown examples of each sentence type estimated their recall for different sentences either; (a) before studying sentences; (b) after studying but before being tested or (c) after being tested. Only subjects who estimated after the test accurately gauged the different rate of recall for the different sentence types. This Pressley and his colleagues called the 'testing effect'. Despite this, about 80% of the subjects reported being aware of the relative difficulty of learning different sentences and the great majority also reported using this judgement to

regulate their study. It appears that the subjects may have based their estimates of performance on study effort and it was not until a test of actual recall that they adjusted their recall estimates. Pressley, Snyder, Levin, Murray and Ghatala (1987), also found a 'testing effect' amongst students asked to read a chapter from an introductory psychology text before taking a 50 item multiple-choice test. Performance predictions were more accurate after testing than before reading. In another experiment recorded in the same paper, Pressley et al (1987) found that monitoring of overall learning during reading was improved by embedding questions in the text that were similar in format and difficulty to criterion test items. This is consistent with the default model of comprehension proposed by Glenberg in that failure to comprehend is not recognised until an error signal, e.g. failure to answer a question, is received.

Clearly monitoring their comprehension of text is not an easy task for adults and without an external criterion against which to judge their comprehension, they often have delusions of understanding. Similar problems were found in child readers. For example, Ghatala, Levin, Foorman and Pressley (1989) gave fourth grade students a short passage of about 300 words to read and the goal of getting 100% correct on a multiple choice test. Children were either given the opportunity to read the passage as many times as they liked before taking the test, or they read the passage and were tested but could repeat this as many times as they liked, or as in the read-test condition but being asked how many items they had got right in the test or in the final condition the children were given feedback as to how many items they had got right after each test. Only children who received feedback on their test performance were able to regulate their study to achieve mastery. In the other conditions, the children overestimated their mastery of the material and understudied.

In a follow-up experiment reported on the same paper on third and fourth grade students, Ghatala et al obtained evidence that subjects often gave high confidence ratings to incorrect responses. This led them to conclude that distractor items in the multiple choice test may have inflated students' confidence in them as correct responses. This led them to repeat their original experiment, but this time using a short answer recall test and omitting the feedback condition. Children who were able to study the text as many times as they liked before testing understudied. In contrast, children who were tested on the content whether or not they had to estimate their performance studied longer and scored higher. Seretny and Dean (1986) also found written inserted questions facilitated comprehension amongst second grade students as measured by a norm-referenced test of reading comprehension.

Pressley, Ghatala, Woloshyn and Pirie (1990) found similar effects of short answer test questions as compared to multiple choice with Canadian undergraduates in that short answer questions generally promoted appropriate re-reading more than multiple choice questions. The exception was for thematic questions for which subjects rarely reread following an error regardless of whether they took a short answer or multiple choice test. A follow-up experiment demonstrated that this was because readers were often quite confident their answer was correct whatever format the test was in. Roughly 60% of incorrect answers in either format received confidence ratings ranging from somewhat certain to absolutely certain. Moreover, in neither experiment was a significant correlation obtained between main idea monitoring and verbal ability measures though in both experiments the more able students got more answers correct. It seems that even good adult

readers appear not to be aware when they have missed the main point of the text.

In the discussion following Experiment Five, it was felt that the readers may have performed less well on the error detection task because they were not actively monitoring their comprehension. This could be investigated by comparing the pupils' performance on the tasks with or without prior knowledge of the comprehension questions they are to be asked on completing the task. Additionally, by introducing a third version, but without the gap in which the reader has to indicate where there is a word missing would give information as to how precise the prompt has to be to stimulate a correct response.

EXPERIMENT SIX

Subjects

120 Year Four pupils, 60 from each of the 2 suburban junior schools. The pupils ranged in age from 8 years 2 months to 9 years 4 months with a mean age of 8 years 9 months and a standard deviation of 4 months.

Materials

Passage Three was used in three forms: one was the cloze form as used in Experiment Five except for one space, one was identical to this but without any spaces where words were omitted and one was the error detection form used in Experiment Five except for one change to accommodate the omission task. Each was presented in one of two ways, either with no accompanying questions or with five accompanying questions printed beneath the passage.

Procedure

Each child was randomly assigned to one of six groups.

Group 1 completed cloze version and was asked the five questions.

Group 2 completed the omission version and was asked the five questions.

Group 3 completed the error detection task and was asked the five questions.

Group 4 completed the cloze task with questions and was then asked the questions.

Group 5 completed the omission task with questions and was then asked the questions.

Group 6 completed the error detection task with questions and was then asked the questions.

Each child was seen individually. They were told the story was about a small space creature who is found by a little boy who makes friends with him.

Those given the cloze task were asked to complete the story so that it made sense. An example was given: **The cat sat on the mat.** They were told they would be asked some questions on the passage when they had finished to see how well they understood the story. For those in Group Four, the questions at the bottom of the passage were indicated one at a time. On completing the task, the page was removed and they were asked the questions and their answers recorded.

Those given the insertion task were told a number of words were missing and were asked to put a cross where a word was missing. An example: **The cat satxthe mat** was shown. They were told they would be asked some questions on the passage to see how well they understood the story. For those in Group Five the questions at the bottom of the page were indicated one at a time. On completing the task, the page was removed and they were asked the questions and their answers recorded.

Those given the error detection task were told a number of words did not make sense in the passage and they were asked to underline them. An example was given: **The cat read on the mat.** They were told they would be asked some questions on the passage to see how well they understood the

story. For those in Group Six, the questions at the bottom of the page were indicated one at a time. On completing the error detection task, the page was removed and they were asked the questions and their answers were recorded.

Scoring

The cloze exercises were independently marked by the experimenter and two colleagues. Where there were any discrepancies in the scores, a simple majority determined whether the item was accepted or not. The error detection and omission tasks were marked by the experimenter and a colleague. There were no discrepancies in the scores awarded.

Results

A preliminary breakdown of the scores obtained by the subjects is given in Table 7.1.

Table 7.1 Mean scores out of ten (and standard deviations) by condition and task

Condition	Task		
	Cloze	Omission	Error Detection
Questions available	6.05 (2.06)	5.4 (1.6)	1.8 (1.64)
Questions not available	6.35 (2.58)	5.35 (1.79)	2.85 (2.23)

Having the questions available for the children to read prior to carrying out the comprehension task has little effect on the children's performance. There does seem to be a difference between the children's performance on the

cloze and the omission tasks as compared to the error detection task. Further analysis was carried out using a completely randomised two-factor analysis of variance. The results are summarised in Table 7.2.

Table 7.2 Summary of analysis of variance for Experiment Six

Source	df	SS	MS	F	P
Treatment	5	345.267	69.053		
A	1	5.633	5.633	1.388	NS
B	2	333.317	166.659	41.069	0.01
A x B	2	6.317	3.158	0.778	NS
Residual	114	462.6	4.058		
Total	119	807.867			

A = questions available/not available

B = cloze/omission/error detection

From Table 7.2 it can be seen there is a significant main effect for the comprehension task undertaken by the reader. There was no significant effect for whether the task was presented with or without the questions available, nor was there any significant interaction. Scheffe tests showed that there was no significant difference in scores between the cloze and omission tasks ($F = 1.68$, $df 1,117$) but there was a significant difference between the cloze and error detection tasks ($F = 37.02$, $df 1,117$, $p < 0.01$) and the omission and error detection tasks ($F = 22.94$, $df 1,117$, $p < 0.01$).

In this experiment 62 percent of the cloze items and 23 percent of the error detection tasks were completed correctly. This is comparable with

Experiment Five in which 66% and 29% of responses were correct in the cloze and error detection task, respectively.

It can be argued that the omission task could be carried out principally through reliance on syntactic comprehension monitoring. In contrast, the error detection task was designed so that semantic comprehension monitoring would need to be employed to detect the errors. In only one case was an item substituted in detected more often than an omission at that point was identified. This was the fourth item. In the omission version, the story reads at that point as follows '... they all live under the ground so I hid (him) under a rose bush in my garden'. The syntax of the version without the omitted 'him' gives little or no clue that something is missing. It would require semantic monitoring that the story refers to the boy still being seen to identify that this did not make sense.

Discussion

In Experiment Five, a significant difference in frequency of correct responses in favour of the cloze task over the error detection task was obtained. Experiment Six explored whether it may be the cloze line signalling the necessity to make a response which contributed to the difference by also providing a set of questions against which the pupils might judge their comprehension. This made no difference to the performance of the pupils on either the cloze, the omission or the error detection task.

It is plausible as Marton and Saljo (1976) have suggested that the effects of a one off experimental condition may not generalise in that the students will mainly learn the particular responses needed for that condition. They refer to Marton (1975) who carried out a study to induce deep processing by inserting in text content neutral questions (e.g., asking them to summarise a section in one or two sentences). The control group who received no treatment achieved better results on both recall and answering questions. Marton considered this happened because the subjects made answering questions an end in itself, causing them to narrow the task and make it trivial. In this case they perhaps, made answering the questions the focus of their comprehension monitoring so minimally influencing the other tasks.

Though this experiment differed from Experiment Five in also providing an example of the experimental task prior to completing the task and used a different sample of pupils the rate of detection of substitutions and correct completion of cloze were both very similar to those obtained in Experiment Five. Also, as in Experiment Five, a significant difference in favour of the

number of correct cloze responses over the number of substitutions detected was also obtained.

A third task was also introduced in this experiment. In this task, the reader had to identify where a word was omitted. The readers were significantly better at this than detecting when an inappropriate word was substituted even though these were at the same points in the same story. For only one item was the omission task harder than the error detection task and in this instance, the omission did not produce a syntactically unlawful arrangement of words. If the other omissions were detected predominantly through syntactic monitoring one could predict that in situations where omission preserved a lawful syntactic structure so that the omission could only be detected by semantic monitoring then there would be no difference in the rate of detection of real word substitutions and these omissions. This will be investigated in Experiment Seven.

In this experiment as in Experiment Five, the rate of detection of real word substitutions at 23% and 29% respectively was much less than the 53% average for Experiments Three, FourA and FourB. However, in these experiments, only six substitutions were made in each 100 word passage as compared to the 10 substitutions made in Experiments Five and Six. It may be that increasing the density of substitutions decreases the overall intelligibility of the passage so decreasing the rate of detection. This will be investigated in Experiment Eight.

A weakness of Experiment Six was that like Experiment Five, it relied upon only one passage, and the effects may be passage specific. Some information relating to the specificity of the outcomes will be derived from Experiments Seven and Eight as described in Chapters Eight and Nine

respectively. Criticism might also be levelled that the instructions to the error detection task focus the reader on single words which do not make sense. However, though this may influence the outcome, the same instructions were used in Experiments Three and Four and so this cannot explain the less frequent identification of substitutions in this experiment and Experiment Five.

CHAPTER EIGHT

THE WORD OMISSION TASK AND THE SYNTACTIC STANDARD OF COMPREHENSION

Baker (1985a) defined the syntactic standard of comprehension monitoring as involving 'judgements as to the grammaticality of a string or words or the recognition that a particular word is not syntactically acceptable given the surrounding context' (Baker 1985a p.163). A multiple regression analysis on information collected by Vogel (1975) on the reading skills of 40 boys aged between seven and eight had suggested that syntactic ability accounted for approximately half of the variance in their reading comprehension as measured by their performance on the Gates-MacGinitie Reading Tests. However, Baker (1985a) noted that researchers had devoted relatively little attention to readers' use of syntactic standards of evaluation. She nevertheless concluded that although there had been few investigations directly indicating syntactic standards it appeared readers did evaluate text for syntactic appropriateness.

Since Baker's (1985a) review, further work has also implicated children's use of syntactic standards in comprehension monitoring. Schwantes (1991) asked third, sixth and college-grade students to monitor sentences either for words/non-words or meaningfulness/non-meaningfulness. The sentences were either semantically coherent, syntactically intact but meaningless or ungrammatical and meaningless. The sentences were presented in two ways. In the lexical decision task, the subjects indicated by pressing a button whether all of the items in the sentence were words. In the semantic decision task, the subjects indicated by pressing a button whether the sentence was meaningful. All students particularly third-graders, identified

non-words quicker in otherwise semantically coherent sentences as compared to syntactically intact but non-meaningful sentences. All students and particularly third-graders, identified non-words quicker in syntactically intact (but non-meaningful) sentences as compared to ungrammatical and non-meaningful sentences. The difference in response times in making lexical or semantic decisions for semantically coherent sentences was significant for third graders. So for third graders, the decision that all the words in a sentence were words was made significantly faster than that the sentence was meaningful. The finding of interest to this review is that for children of eight to nine years of age, syntactic information facilitates word recognition speed.

Some information as to a possible mechanism that may be involved is provided by Rauenbusch and Bereiter (1991). They carried out a protocol analysis of the strategies revealed by the thinking aloud of seventh grade pupils when asked to read text degraded by the blanking out of every third letter for which a dash was substituted. They identified four strategies:

1. re-reading the text in an effort to clarify meaning
2. reading ahead to see if further information would be helpful
3. summarising what is known so far
4. determining the word type.

Rauenbusch and Bereiter (1991) observed that while the first three strategies are commonly cited in the literature, the fourth strategy is not. This strategy can be largely syntactic. They also noted that studies on oral reading mistakes often show readers substituting syntactically appropriate words for the words actually presented (Clay 1968, Kolars 1970, Simons and Leu 1987, Weber 1970). The work of Willows and Ryan (1986) on assessing the relation between grammatical sensitivity and reading level also led to the

conclusion that children aged from six to nine years are sensitive to syntactic constraints during reading. They also demonstrated that grammatical sensitivity shows substantial growth in these early elementary school years. Earlier work of theirs (Willows and Ryan 1981) had also suggested that there was little growth in syntactic sensitivity in nine to 12 year old readers.

In Experiment Three significantly more inappropriate words of a different part of speech were identified than inappropriate words of the same part of speech. This suggested that eight and nine year old children may be more sensitive to violations of meaning when combined with violations of grammar than violations of meaning alone.

In Experiment Six, no difference was found between the frequency of identification of the location of a missing word and a correct word being offered to fill the gap. It may be that rules of grammar e.g. of the type of word required to fill the gap may be the dominant influence in determining a cloze response rather than meaning per se. It is proposed to investigate this in Experiment Seven by asking children to either identify where a word is missing or identify a real word of inappropriate meaning but the same part of speech at the same point in the passage. However, these passages have been written in such a way as the omission of the target words does not violate grammatical constraints, but does violate the sense of the passage. By biasing the identification of the omission towards requiring the use of a semantic standard of comprehension, one would anticipate finding no difference in the rate of detection of the real word substitutions and omissions.

EXPERIMENT SEVEN

Subjects

Forty eight Year Five pupils were randomly selected from the available pool of subjects in a single junior school. The sample of 24 boys and 24 girls ranged in age from nine years six months to ten years six months with a mean age of ten years no months and a standard deviation of 3.5 months.

Materials

Two passages of 100 words in length and capable of being read by an average reader of between seven and a half and eight and a half years of age were prepared. Both passages were modified in two ways. In one form, five words were omitted so as to leave the passage syntactically acceptable, but not meaningful at the point of the omission. In the second form, five words of the same part of speech were substituted for the same five words omitted in the other form. These substitutions made no sense within the passage.

Procedure

Each subject was seen individually. Each subject completed one form of each passage in counter-balanced order.

	First	Second
Group One	B4 S	B5 O
Group Two	B4 O	B5 S
Group Three	B5 S	B4 O
Group Four	B5 O	B4 S

B4 = Passage Four

O = omission

B5 = Passage Five

S = substitution

When given the omission task, they were told a number of words were missing and were asked to put a cross where a word was missing. An example was shown: The cat satxthe mat. When given the error detection task, they were told a number of words did not make sense in the story and were asked to underline them. An example was given: The cat read on the mat.

Results

A breakdown of the results by passage and task is given in Table 8.1

Table 8.1 Mean Scores (and standard deviations) by passage and task

Passage	Task		
	Omission	Substitution	Combined
Four	3.71 (0.81)	1.75 (1.39)	2.71 (1.49)
Five	1.50 (1.35)	2.75(0.94)	2.13(1.31)
Combined	2.6 (1.57)	2.25(1.28)	

There are indications from these means that there may be a group effect. The group means for each task are given in Table 8.2.

Table 8.2 Group Means (and standard deviations) for each task

Group	Task		
	Omission	Substitution	Overall
1	1.50(1.45)	1.75(1.60)	1.63(1.50)
2	3.67(0.78)	2.92 (1.16)	3.29 (1.04)
3	3.67(0.89)	2.58(0.67)	3.13(0.95)
4	1.50(1.31)	1.75(1.22)	1.63(1.24)

The means for Groups One and Four are identical, while those of Groups Two and Three are very similar.

Further analysis using a Type V analysis of variance (Lindquist 1956), was conducted with three within subject factors: task, passage and session. The results are summarised in Table 8.3.

Table 8.3 Summary of Analysis of Variance for Experiment Seven

Source	df	SS	MS	F	P
Between subjects	47	139.33			
AB	1	60.17	60.17	33.61	0.01
AC	1	0.17	0.17	0.09	N.S
BC	1	0.17	0.17	0.09	N.S
Error	44	78.82	1.79		
Within subjects	48	84.00			
A	1	2.67	2.67	1.61	N.S
B	1	8.17	8.17	4.92	0.05
C	1	0.17	0.17	0.10	N.S
ABC	1	0.17	0.17	0.10	N.S
Error	44	72.82	1.66		

A = Task B = Passage C = Session

From Table 8.3 it can be seen there is no significant main effect for task. Though this is in line with the hypothesis, it is a weak test of the hypothesis as the significance tested for is that of there being a difference which is not likely to have occurred by chance.

Nevertheless, the F value obtained is small, and the overall task means are quite similar. However, there is a highly significant interaction between passage and task.

This was investigated using tests of simple main effects (Winer 1962). These indicated Groups One and Four showed no difference between tasks

($F = 0.45$, $df 1,44$ N.S.) Groups Two and Three showed a significant difference in performance between tasks ($F = 6.09$, $df 1,44$ $p < 0.05$) which examination of Table 8.2 indicated was in favour of the omission task. Groups Two and Three did better than Groups One and Four on the substitution task ($F = 6.70$, $df 1,44$ $p < 0.05$) and on the omission task ($F = 31.47$, $df 1,44$ $p < 0.01$). It is unclear from these results whether Groups Two and Three were more able on these tasks or whether the versions of the tasks they undertook were easier. This reinforces the need for particular caution in interpreting the results of Experiments Five and Six in which only one passage was used as it is possible differences in performances may have reflected a passage effect. This suggests the need to use a minimum of two passages in any further experiments.

An average of 45 per cent of the real word substitutions were detected in this experiment which compares with an average of 53 per cent in Experiments Three and Four and an average of 28 per cent in Experiment Five and Six under comparable conditions to those in Experiments Three and Four except that there were ten substitutions in Experiment Five and Six and compared to five or six in Experiments Three, Four and Seven.

Discussion

The results of this experiment were in line with the hypothesis that there would be no difference in the performance of children on the omission or error detection task when both required the application of a semantic standard of comprehension monitoring for successful completion. However, an unambiguous interpretation cannot be made as there was a significant interaction between the two tasks and the two passages. This suggests that the results of Experiments Five and Six cannot be generalised beyond the

one passage used in these experiments. It also indicates a need to use at least two passages in any further experiments.

In this experiment in which five errors were substituted into the passages, there was a similar rate of detection to Experiment Three and Four in which six errors were substituted. This stands in contrast to the much lower rate of detection in Experiments Five and Six when ten errors were substituted into the 100 word passage. A plausible hypothesis is that the increased number of substitutions degrades the meaning of the passage so impairing the detection of substitutions which require semantic comprehension monitoring to be identified. This will be investigated in Experiment Eight.

CHAPTER NINE

THE INFLUENCE OF THE DENSITY OF SUBSTITUTIONS ON THEIR DETECTION

Two indices of comprehension have been used in the experiments reported on so far in this series of studies; error detection and cloze. Cloze although not originally known by that name has a long history. Ballard (1920) described a reading test involving the completion of a story of 540 words in length from which 67 words were omitted which had to be completed within three minutes by readers who had just read the original story. This had some of the features of the modern cloze test but differed in others. In particular, Ballard removed words which might aid the readers to complete the omissions by inference and substituted an X for the missing word. Burt (1921) devised as a supplementary test, two passages with words missing which were to be replaced. Burt felt this completion task was a test of reading comprehension and he credited Ebbinghaus with having originally devised this type of test.

In 1953, Taylor used the completion method to measure readability and in his article in the *Journalism Quarterly* referred to it as the cloze procedure. This was derived from the Gestalt notion of closure. By 1957, Taylor was advocating that cloze was a valid index of comprehension. In 1982, Shanahan, Kamil and Tobin were able to describe cloze as a conventional measure of reading comprehension. They noted that in the United States, it was used in informal tests, standardised tests and state-wide competency tests. In the United Kingdom, the cloze procedure is also widely used e.g. in the London Reading Test.

Cloze was used as a second index of comprehension together with an error detection task in Experiments Five and Six. Because they were used as parallel measures, they had to share common features including the same number of items. In these experiments the number of deletions determined the number of substitutions.

Taylor (1957) advised the deletion of every *n*th word. However, a convention of deleting every fifth word (Bormuth 1967, McKenna and Layton 1990, Shanahan, Kamil and Tobin 1982) or every tenth word (Schneyer 1965) arose. Alderson (1979) investigated the effect of deleting either every 6th, 8th, 10th or 12th word from one of three texts one of which was difficult, one of medium difficulty and one easy. Though Alderson refers to difficulty effects because only one passage represented each level of difficulty, such effects may have been attributable to the passage. He found significant differences between cloze test scores but these were unpredictable. Deleting every 12th word did not necessarily result in an easier test than deleting every 6th, 8th or 10th word. However, when paired comparison tests were made when only items identical to both cloze tests were compared, no significant differences were found. It is probable this is the most accurate representation of his results as it does not appear from Alderson's account that he controlled for those taking the tests with more deletions being automatically able to score more than those with fewer deletions. For example, a good comprehender scoring 100% would be able to obtain twice as high a score on the every 6th word deleted version as on the every 12th word deleted version.

In the cloze procedure used in Experiments Five and Six, every 10th word (on average) was deleted. In consequence, ten substitutions were made in the alternate test of comprehension through error detection. Not only were

significant differences found between cloze test scores and error detection scores, but a far smaller percentage of substitutions were detected than in previous experiments using the same passage. In Experiments Three and Four when six substitutions were made into Passage Three, 52 per cent were detected while in Experiments Five and Six when ten substitutions were made, only 28 per cent were detected. It is possible the increase in density of substitutions made the task more difficult however different substitutions were made in different places in the different experiments and it is possible this may also have influenced detection rates. In Experiment Eight, it is proposed to investigate the effect of different densities of substitutions controlling for substitution and their immediate textual context.

EXPERIMENT EIGHT

Subjects

All 54 Year Five pupils in one junior school were used. They ranged in age from nine years eight months to ten years eight months with a mean age of ten years one month and a standard deviation of three months.

Materials

Three passages (Passage One, Three and Four) each of 100 words and capable of being read by an average reader of between seven and a half and eight and a half years of age were used. Each passage was modified so it could be presented in one of three ways. One version had ten substitutions, one made every tenth word plus or minus two to avoid any obvious pattern and to facilitate the substitution of a word of the same part of speech as the one it replaced but with a meaning inappropriate to the story. Each passage was also produced in a form where only the even numbered substitutions were retained, the passage otherwise being as the original and a form where only the odd numbered substitutions were retained.

Procedure

Each subject was seen individually. The task was introduced by the experimenter saying "I would like your help. I have got some stories here which I would like to put in a book. I want to make sure the children reading them can understand them. Would you please read this story to yourself and underline any word that doesn't make sense in the story."

Each subject was then presented in succession with all three passages so

that each of the subjects was exposed to all three methods of presentation of substitutions. The order of presentation and type of substitution were grouped into nine different sequences.

	First	Second	Third
Group 1	A ₁ B ₁	A ₂ B ₃	A ₃ B ₄
Group 2	A ₃ B ₃	A ₁ B ₄	A ₂ B ₁
Group 3	A ₂ B ₄	A ₃ B ₁	A ₁ B ₃
Group 4	A ₁ B ₃	A ₂ B ₄	A ₃ B ₁
Group 5	A ₃ B ₄	A ₁ B ₁	A ₂ B ₃
Group 6	A ₂ B ₁	A ₃ B ₃	A ₁ B ₄
Group 7	A ₁ B ₄	A ₂ B ₁	A ₃ B ₃
Group 8	A ₃ B ₁	A ₁ B ₃	A ₂ B ₄
Group 9	A ₂ B ₃	A ₃ B ₄	A ₁ B ₁

A₁ = all 10 substitutions

A₂ = even numbered substitutions only

A₃ = odd numbered substitutions only

B₁ = Passage One

B₃ = Passage Three

B₄ = Passage Four

Results

A preliminary breakdown of the results by passage and task is given in Table 9.1.

Table 9.1 Total scores for Each Item for Each Passage:

Passage One												
No of substitutions	Item										Total Ident -ified	% Identi -fied
	1	2	3	4	5	6	7	8	9	10		
10	9	8	4	8	3	8	3	4	9	3	59	32.8
5	9	15	8	14	4	12	7	6	11	4	90	50
Passage Three												
No of substitutions	Item										Total Ident -ified	% Identi -fied
	1	2	3	4	5	6	7	8	9	10		
10	2	7	2	4	11	11	5	8	10	0	60	33.3
5	1	5	3	3	12	9	10	3	10	2	58	32.2
Passage Four												
No of substitutions	Item										Total Ident -ified	% Ident- ified
	1	2	3	4	5	6	7	8	9	10		
10	10	9	5	6	9	17	9	5	3	7	90	50
5	9	10	7	7	9	16	7	4	11	8	88	48.9

The overall detection rate of 41% is in line with previous studies. Further analysis of the results was carried out using a repeated measures one factor analysis of variance with the two scores out of five for each subject being combined to give a single score out of ten. Table 9.2 gives a summary of the results.

Table 9.2 Summary of Analysis of Variance for Experiment Eight

Source	df	SS	MS	F	P
Between Subjects	1	593.94			
Within Subjects	54	95.5			
Between Treatments	1	6.75	6.75	4.03	0.05
Residual	53	88.75	1.68		

Though this analysis indicates a significant main effect for treatment which examination of Table 9.1 indicates is in favour of the presentation of the items in groups of 5, this is only the case for Passage One. For Passages Three and Four, the total scores under each condition of presentation are only marginally different and that in the favour of presentation of all ten substitutions.

Passage Three is of particular interest as it was this passage which was used in Experiments Five and Six in which much lower rates of detection of errors were found. A Mann-Whitney U Test carried out on Passage Three failed to show a significant difference ($U = 49$, N.S. at $p = 0.05$) for substitution frequency. Passage Three was however, the passage with the lowest overall detection rate of 33 per cent as compared to 41 per cent for Passage One and 49 per cent for Passage Four.

Discussion

This study addressed the question as to whether the use of ten substitutions in the 100 word passage used in Experiments Five and Six as compared to the six substitutions in earlier experiments may have caused the drop in

detection rate form an average of 52 per cent to the 27 per cent found in Experiments Five and Six on the same passage. Though an overall higher rate of detection of substitutions was found in this study when only five substitutions were made, this only occurred in Passage One. No difference in detection rates was found in Passage Three which was the passage used in Experiments Five and Six. The increased number of substitutions in these latter two experiments is therefore unlikely to have brought about the lower rate of detection but it is impossible to be unequivocal about this. However, in this experiment, Passage Three was the passage in which fewest substitutions were found with an overall detection rate of 33 per cent. Table 9.3 shows this to be the median detection rate for real word substitutions in this passage over the five experiments in which it has been used in that way.

Table 9.3 Percentage Detection Rate of Real Word Substitutions made in Passage Three over Five Experiments

Experiment	3	4	5	6	8
Detection Rate	51	53	29	23	33

It is possible the low detection rates found in Experiments Five and Six are just one extreme of the range of detection rates found within a large experimental population. The difference in performance in the cloze as compared to the error detection task in Experiments Five and Six may therefore represent a non-artifactual difference between performance on unprompted and externally prompted comprehension tasks. The convergence of performance on the omission and error detection tasks in Experiment Seven when both could only be detected by semantic

comprehension monitoring would be consistent with this. However, for a comparison of externally prompted comprehension and cognitive monitoring to be made, tasks which require very similar responses would have to be used. This will be examined in Chapter Ten.

CHAPTER TEN

THE RELATIONSHIP OF COMPREHENSION LEVEL TO PERFORMANCE ON MEASURES OF PROMPTED AND UNPROMPTED COMPREHENSION

Yuill and Oakhill (1991) claimed that 'Not only do skilled readers make efforts to understand what they are reading, they also check whether or not they have understood....' (p125). They went on to assert that 'Monitoring of progress is a necessary prerequisite to comprehension repair' (p125). They stated that 'Less skilled comprehenders do not completely fail to perform comprehension monitoring: they are just less likely to detect and resolve anomalies and are influenced to a greater extent than skilled children by increased processing demands' (p137).

Garner (1987) cited error detection as being the dominant research technique used to investigate comprehension monitoring. This technique assumes the reader by monitoring his comprehension of what he reads is able to detect when a word's meaning is inappropriate within that story. There is no external prompt to the reader that they have to make a response at any point. The error detection task may therefore be an indication of the child's ability to monitor their comprehension of books, stories and articles where there may be no external check as to whether or not they have understood. One could predict from Yuill and Oakhill's (1991) model that less good comprehenders would do less well on an error detection task.

In Experiment Five where both a comprehension monitoring task and a cloze task were used, the readers performed much better in the cloze task. This difference persisted in Experiment Six, despite the presence of

comprehension questions which readers saw in advance and knew would be asked. However, in Experiment Six, there was no significant difference between the readers' ability to indicate where words were omitted and to complete the cloze form.

There are a number of problems in interpreting these results. Only one passage was used so it is possible the results may be passage dependent. Also, the responses required; underlining, writing a word or indicating an omission with a cross are very different and it cannot be assumed they do not moderate performance. It is also possible the selection of every tenth word plus or minus one for deletion for the cloze form (so also determining which words would be substituted to make the two tasks comparable), may have enabled readers to use syntactic clues to identify if a word was missing and what it might be. For example, the third sentence in the passage read 'He told me (his) rocket had crashed.' Ninety per cent of readers identified where a word was missing when **his** was omitted. By contrast, only one person identified where a word was missing when **him** was omitted from the sentence, 'On Mars, they all live under the ground so I hid (him) under a big bush in my garden.' In the latter example, the sentence remains syntactically intact without the **him** and it is only in the context of the story that its omission is apparent. In Experiment Seven when the omission and error detection tasks were compared with only five errors or omissions per passage and these required semantic comprehension monitoring for identification, there was no significant difference in performance on the two tasks. This provides some evidence of equivalence of these two comprehension monitoring tasks.

A comparison of prompted and unprompted comprehension could be made by comparing performance on a cloze task with performance on an

insertion task where the reader has to realise a word was omitted and then insert the missing word. As in Yuill and Oakhill's (1991) model, the comprehension monitoring task involves the extra step of identifying the missing element before attempting comprehension repair one would predict less good performance generally on the insertion task as compared to the cloze task. However, they claim better comprehenders are better at comprehension monitoring as they build a mental model of the situation described in the text by making inferences from the text and integrating information as they go along. One would therefore anticipate the difference between prompted and unprompted comprehension would be less for better than for less good comprehenders. This will be investigated in Experiment Nine.

EXPERIMENT NINE

Subjects

Forty eight Year Five pupils aged between nine years ten months to ten years ten months with a mean age of ten years three months and a standard deviation of four months were used. These were selected from 52 possible subjects who were split into better and less good comprehenders on the basis of their score on the Primary Reading Test (France 1981) administered three months before the experiment. The 52 subjects were divided about the modal raw score of 33.5 on the test. Two subjects who scored 33 and two subjects who scored 34 were eliminated. The mean raw score of the less good comprehenders was 28.2 with a standard deviation of 4.1, while the mean score of the better comprehenders was 37.0 with a standard deviation of 2.9.

Materials

Two passages of 100 words in length and capable of being read by an average reader of between seven and a half and eight and a half years of age were used. Both passages were modified in two ways. In one form, five words were omitted so as to leave that passage syntactically acceptable but not meaningful at the point of omission. In the second form, a standard gap was left where each word was omitted.

Procedure

Each subjects was seen individually. Each subject competed one form of each passage in counter-balanced order.

	First	Second
Group One	B4C	B5O
Group Two	B4O	B5C
Group Three	B5C	B4O
Group Four	B5O	B4C

B4 = Passage Four
B5 = Passage Five

O = omission
C = cloze

When given the insertion task, they were shown an example: 'The cat sat on the mat'. They were asked to complete a practice sentence: 'The mouse the cheese' and were given sufficient prompts to ensure they successfully completed this. These were then asked to read the story and to insert any words they felt were needed for the story to make sense.

When given the cloze task, they were also shown an example: 'The cat sat on the mat'. They were asked to complete a practice sentence 'The _____ chased the mouse'. All pupils did so successfully. They were then asked to read the story and to fill in any blanks so that the story made sense.

Scoring

Both exercises were independently marked by the experimenter and two colleagues. Where there were any discrepancies in the scores, a simple majority determined whether the item was accepted or not.

Results

A Preliminary breakdown of the scores obtained by the subjects is given in Table 10.1.

Table 10.1 Mean scores out of five (and standard deviations) by task and comprehension level

Comprehension level	Task		Total
	Cloze	Insertion	
Better	4.33 (0.92)	3.13 (1.30)	7.46 (1.59)
Less good	3.58 (0.97)	1.96 (1.37)	5.54 (1.59)
Total	3.96 (1.01)	2.54 (1.44)	

Further analysis was carried out by analysis of variance the results of which are summarised in Table 10.2.

Table 10.2 Summary of Analysis of Variance for Experiment Nine

Source	df	SS	MS	F	P
Between subjects	47	80			
Between A	1	22.042	22.042	17.494	0.01
Subjects within A	46	57.958	1.260		
Within subjects	48	114			
Between B	1	48.167	48.167	34.185	0.01
Subjects x B	47	65.833			
A x B	1	1.041	1.041	0.739	NS
Subjects x B within A	46	64.792	1.409		
Total	95	194			

A = level of comprehension

B = task

There is a significant difference in performance between the subjects comprising the two different comprehension bands which inspection of Table 10.1 indicates is in favour of the better comprehenders as identified by their score on the Primary Reading Test. There is also a significant difference between the two tasks with the readers performing better on the

cloze task. There was no significant interaction between initial comprehension level and performance on the tasks.

It could be argued that performance on the insertion task is not a measure of comprehension monitoring as the reader automatically repairs any comprehension breakdown and so fails to indicate the omission. If this were the case, one would expect a negative correlation between cloze performance and performance on the insertion task as those words most easily replaced would presumably be under the above argument those omissions most easily overlooked. The total scores for each item for each passage is shown in Table 10.3.

Table 10.3 Total Scores for each item for both tasks and passages

	Passage Four					Passage Five				
Item	1	2	3	4	5	1	2	3	4	5
Cloze	21	24	17	20	22	15	11	13	24	23
Insertion	16	15	10	20	14	4	4	7	18	14
Total	37	39	27	40	26	19	15	20	42	37

The Pearson Product Moment correlation coefficient of the cloze and insertion scores for each item is 0.85 which for $df = 8$ is significant at the one percent level. This would not support the argument that those omissions most easily overlooked are those for which it is easiest to supply the missing word.

A further Pearson Product Moment correlation coefficient was calculated between each subject's raw score on the Primary Reading Test and their score on the insertion task. The correlation coefficient obtained of $r = 0.487$ is significant at $p = 0.001$ for $df = 46$; indicating a small but

significant correlation between performance on a standardised test of reading comprehension and a measure of comprehension monitoring.

Discussion

Following Harris, Kruthof, Meerum Terwogt and Visser (1981), Yuill and Oakhill (1991) noted that 'as readers develop their skills, they may increasingly modify their reading strategy according to the results of comprehension monitoring activity, and monitoring can be seen as a fundamental part of comprehension rather than an epiphenomenon of more basic processes' (p139). Additionally, they claimed that 'poor comprehenders tended not to notice inconsistencies in text, and more generally they did not realise that they had failed to understand a text, or know how to remedy such a failure if it did become apparent to them' (p216). The results of this experiment are not entirely consistent with this. The results of this experiment together with the strong positive correlation obtained between performance on the Primary Reading Test and error detection reported in Appendix D are consistent with better comprehension monitoring being associated with better performance on measures of comprehension. However, the results give no indication as to whether better comprehension monitoring enhances comprehension or better comprehension enables readers to monitor their understanding with more facility.

Yuill and Oakhill (1991) carried out a series of three experiments comparing the effects of training either on inference making, inference making together with comprehension monitoring or comprehension monitoring with training in rapid decoding or comprehension exercises. Those less skilled comprehenders given either inference training by itself or

combined with training in comprehension monitoring did better than controls or those given training in rapid decoding. They did not do better than those given training just on comprehension exercise or those just given training in comprehension monitoring. Those just given training in comprehension monitoring did not do better than those given training in rapid decoding. This would suggest that the training in comprehension monitoring given by Yuill and Oakhill either did not improve comprehension monitoring, or that improved comprehension monitoring did not enhance performance on the Neale Analysis.

In developing their model of reading comprehension, Yuill and Oakhill (1991) made reference to the results of their own investigations. Their investigations generally involved comparing two groups; one skilled and the other less skilled comprehenders. This could result in one group appearing qualitatively different in some respect e.g., metacognitive ability which may only represent one extreme of a continuum. The results of this experiment whilst demonstrating that the comprehension monitoring task of insertion is harder than cloze does not support there being a different relationship between performance on unprompted and prompted comprehension tasks for better than for less good comprehenders.

CHAPTER ELEVEN

READING COMPREHENSION AND COMPREHENSION OF WHAT IS READ TO ONE

The evidence obtained from experimental work described earlier in this study is consistent with there being multiple standards against which readers can monitor their comprehension. These standards can be lexical, syntactic or semantic. The semantic standard subsumes standards relating to the consistency of propositions within a sentence, between sentences, within the whole passage and of the passage against an external standard i.e., prior knowledge.

A determinant of which standard is given precedence may be the instructions given the subjects. For example, Tikhomirov and Klochko (1981) presented high school students, university physics students and graduate and teaching members of physics departments with a passage that contained three propositions that violated a law of physics that water runs downhill. All the subjects were asked to check the passage's grammar and after doing so, were asked to recall the passage. Those that did not notice the contradictions were then asked to read the passage aloud to prepare to retell it. After recall, they were asked if they noticed any problems. If they did not notice the three propositions that referred to rivers flowing up mountains, they were given the specific task of searching for the violations. Throughout the tasks, the subjects galvanic skin responses (GSR) were monitored. Only one out of the 45 subjects reported the prior knowledge violations while checking for grammaticality. More surprisingly, only two people reported the problems after the second task which presumably required deeper processing. However, with specific instructions to find the

problems, an additional 35 people reported them. The GSR data suggested that some of the subjects who did not report the problems initially may have noticed them as their GSR responses fluctuated sharply at the moment of reading the contradiction in the text. All these subjects subsequently went on to report the problems when directed to search for them, whilst those whose GSR response remained stable did not report them.

Given that a reader's attention is directed to the salient features of a task, a significant question is what is their optimum level of comprehension?

Betts (1946) drew a distinction between a student's comprehension of material he reads as compared to material which is read to him. Betts applied the term capacity level to describe the highest level of readability of material which the learner can comprehend with 75% accuracy when the material is read to him. This accuracy of comprehension was also the minimum he specified for reading material to be at the instructional level. Betts felt that when a substantial difference existed between instructional level and capacity level, it usually indicated the possibility of rapid gains. Sticht, Beck, Hauke, Kleiman and James (1974) also felt a mature reader was one who could read as well as he could listen. Oakhill and Garnham (1988) suggested that listening comprehension ability is related to that part of reading comprehension skill that cannot be explained by poor decoding. However, Oakhill and Garnham were careful to point out that they were not arguing that they believed that reading comprehension was no more than decoding plus oral comprehension skills, noting that there were many differences between written and oral language with which the child has to learn to deal. Indeed, Yuill and Oakhill (1991) claimed that additional prosodic cues probably explained why young children of seven to eight did better in their Experiment 3.5 when asked questions on sentences that were

read to them rather than those which they read. It is of note that in this experiment, Yuill and Oakhill's results indicated that their less skilled comprehenders performed at a virtually identical overall level in the auditory mode of presentation as their skilled comprehenders did in the visual mode. Whilst prosodic cues may contribute to this, it nevertheless emphasises the role access to the material can play in restricting comprehension performance.

Another issue as Curtis (1980) pointed out is the competition among skills for limited processing capacity in early readers. Poor decoding skills would require more attention to be given to this aspect allowing less for comprehension of what is read. Nevertheless, children who have trouble understanding written language also often have trouble understanding spoken language.

Early work by Young (1936) suggested that children who do poorly in comprehending through reading, also do poorly in comprehending through hearing. This was also the conclusion of Rubin (1980). Berger (1978) found that poor fifth grade readers were deficient in listening as well as reading comprehension. The measures of reading and listening comprehension obtained by Curtis (1980) for third and fifth grade students were also consistent with listening comprehension being better than reading comprehension and that a deficit in one was associated with a deficit in the other. Smiley, Oakley, Worthen, Campione and Brown (1977) also found a large listening comprehension deficit for seventh grade poor readers and a high correlation between reading and listening comprehension. In their review of listening and reading, Sticht and James (1984) concluded that reading and listening build upon a common language and knowledge base, and that a person's listening comprehension skills establish their reading

potential.

Closely related to Sticht and James' (1984) idea of reading potential is that in learning to read one closes the gap between listening comprehension and reading comprehension. One would anticipate that with beginning readers listening comprehension would exceed reading comprehension. Young (1936) though noting children in grades four, five and six got very little from a single exposure to material demonstrated that they got more from a teacher oral presentation than from a silent reading themselves. This was little changed if during the teacher's oral presentation, the pupils simultaneously read the passage. Sticht et al (1974) found in a review of studies that in grades one to six almost all comparisons favoured listening comprehension. As one moved from grade seven to adults, the proportion of studies showing no difference or a superiority for reading increased. As would be predicted from this analysis, Vosniadou, Pearson and Rogers (1988) found that third graders detected a greater number of inconsistencies embedded in text in a listening task than in a reading task. However, the Flesch Reading Ease scores of those parts of the two example stories read by everyone given by Vosniadou et al lie between 80 and 88. That is they are roughly equivalent to grade five to six in difficulty. In Experiment Four in this study, increasing the reading difficulty of the passage from about 100 to 80 to 85 for children of very similar age, resulted in an overall decrease of 64% in the detection of semantically inappropriate words. Though Vosniadou et al claim to have only selected competent readers, it nevertheless raises the possibility that problems of access may have reduced the reader's performance. Syne (1982), while using passages with an average Flesch Reading Ease score of 84 i.e., of roughly the same readability as those used by Vosniadou et al used them with both eight and twelve year old subjects. Passages of this level of readability would have

been quite accessible for average 12 year olds. Amongst this group, Syne found no difference in detecting inconsistencies whether the passages were read to or read by the subjects. However the argument that differences in performance between listening and reading comprehension are solely the result of problems of access to text is weakened as in Syne's experiment the younger children also showed no difference in performance between listening and reading comprehension.

One can be reasonably certain that for younger pupils their listening comprehension is likely to be their optimum level of comprehension. It has been assumed that for the students in Experiments One to Eight, their failure to identify virtually all the substitutions reflected either the application of a standard of comprehension, other than a semantic standard or a true deficit in reading comprehension. If as Townsend, Carrithers and Bever (1987) argued, modality has little effect on comprehension processes this suggests that difference in performance are 'a matter of how the individual deploys basic knowledge of language in comprehension situations.' (p238) It would be necessary to ensure however, that the pupils had equal access to the materials when reading as when listening.

If a passage is of appropriate readability, then in line with the results of Syne (1982), one would not anticipate a higher level of detection of semantically inappropriate substitutions when the passage is read to as compared to read by the subjects. Moreover, from Yuill and Oakhill's (1991) model of reading comprehension and in line with the results of Experiment Five one would not anticipate repeat reading or listening to the passage to enhance performance. This stands in contrast to Carver's (1987) views that 'students can improve the degree to which they comprehend a passage by spending more time reading the passage'. This will be

investigated in Experiment Ten.

EXPERIMENT TEN

Subjects

Twenty four girls and 24 boys from Year Four of a single junior school acted as subjects. They ranged in age from eight years two months to nine years one month with a mean age of eight years nine months and a standard deviation of four months.

Materials

Twelve three sentence stories were written in the form of an argument allowing an inference to be drawn. Following Tunmer, Nesdale and Pratt (1983), the first sentence which stated a general principal was omitted and was replaced by a neutral introductory sentence. In their work, they referred to this as the implicit condition as contrasted with the explicit condition when the first sentence corresponded to the initial premise of an argument. The second sentence was written in such a way so that by changing one word it could be consistent or inconsistent with the argument. No one sentence was internally inconsistent. This was done to ensure children could only detect inconsistencies by integrating information across sentences.

Procedure

Each child was seen individually and was randomly assigned to one of 16 presentation conditions (see Fig. 11.1). Modality of presentation, sequence of stories and form of story were counterbalanced.

Figure 11.1 The Sixteen Combinations of Modality, Version and Story
Used in Experiment 10.

Group	Listen Once						Read Once					
	1	+1	-2	-3	+4	+5	-6	+7	+8	-9	+10	-11
2	-1	+2	+3	-4	-5	+6	-7	-8	+9	-10	+11	+12
3	+7	+8	-9	+10	-11	-12	+1	-2	-3	+4	+5	-6
4	-7	-8	+9	-10	+11	+12	-1	+2	+3	-4	-5	+6
	Read Once						Listen Once					
5	+1	-2	-3	+4	+5	-6	+7	+8	-9	+10	-11	-12
6	-1	+2	+3	-4	-5	+6	-7	-8	+9	-10	+11	+12
7	+7	+8	-9	+10	-11	-12	+1	-2	-3	+4	+5	-6
8	-7	-8	+9	-10	+11	+12	-1	+2	+3	-4	-5	+6
	Listen Twice						Read Twice					
9	+1	-2	-3	+4	+5	-6	+7	+8	-9	+10	-11	-12
10	-1	+2	+3	-4	-5	+6	-7	-8	+9	-10	+11	+12
11	+7	+8	-9	+10	-11	-12	+1	-2	-3	+4	+5	-6
12	-7	-8	+9	-10	+11	+12	-1	+2	+3	-4	-5	+6
	Read Twice						Listen Twice					
13	+1	-2	-3	+4	+5	-6	+7	+8	-9	+10	-11	-12
14	-1	+2	+3	-4	-5	+6	-7	-8	+9	-10	+11	+12
15	+7	+8	-9	+10	-11	-12	+1	-2	-3	+4	+5	-6
16	-7	-8	+9	-10	+11	+12	-1	+2	+3	-4	-5	+6
<p style="text-align: center;">- Inconsistent version + Consistent version</p>												

The appropriate version of the following instructions were used. 'You are going to hear (read out loud), some stories. Some stories make sense but some stories may have in them one word that does not make sense with the rest of the story. After you have heard (read) each story (twice), I will ask you if there is a word that does not make sense in the story and if so, what it is. I will also ask why the story does not make sense'.

All children were given two practice items, one consistent and one inconsistent in the modality in which their first story was presented. Following completion of the first series of stories, the instructions were repeated for the second series prefixed by 'Now'.

The children were given a word if they failed to read it after four seconds or misread it. The children who failed to read or misread more than two words in each story were eliminated. This procedure was designed to ensure all pupils had adequate access to all the material.

Scoring

All subjects' responses were initially scored as to whether they had correctly identified whether there was a word that did not make sense. A second set of scores was obtained for each subject using the stricter criterion of whether they correctly identified which word failed to make sense in the story.

Results

The initial yes or no responses are summarised in Table 11.1.

Table 11.1 Mean (and standard deviation) of responses correctly identifying whether a story contained an inappropriate word.

Number of Times Presented	Modality of Presentation				Total
	Visual		Aural		
	Error Present	Error not Present	Error Present	Error not Present	
Once	2.3 ^a (0.7)	2.3 (0.8)	2.0 (0.9)	2.4 (0.7)	9.0 ^c (1.5)
Twice	2.1 (1.1)	2.2 (0.9)	2.1 (1.1)	2.5 (0.7)	8.9 (2.0)
Combined	2.2 (0.9)	2.3 (0.8)	2.1 (1.0)	2.5 (0.7)	
	4.4 ^b (1.1)		4.5 (1.1)		

a out of 3 b out of 6 c out of 12

A related samples t-test comparing performance on consistent and inconsistent stories for overall correct identification of whether the story was or was not internally consistent indicated there was no significant difference in performance ($t = 1.71, df = 47$). These results also indicated a very consistent response pattern across presentation frequency and modality.

The results obtained by using the stricter criterion of correctly identifying the inappropriate word were cross tabulated as shown in Table 11.2. The results obtained by this method of scoring were similar to those obtained by only using the yes/no response. Twenty-eight subject's scores remained unchanged and of the 20 which changed, all but one were by one point only. A Pearson Product Moment correlation of 0.94 ($P < 0.001, d.f. 46$) was

obtained between the scores under the two scoring criteria when comparing the results for those stories in which an inappropriate word was substituted. Nevertheless, there was a significant difference between the means of the scores obtained under the two scoring criteria for these stories ($t = 5.9$, $df 47$, $p < 0.01$) indicating that identifying the inappropriate word was harder than realising the story did not make sense.

Table 11.2 Means (and standard deviations) of scores when subjects had to correctly identify an inappropriate word.

Number of Times Presented	Modality of presentation		Mean of Means
	Read	Listen	
Once	4.5 (1.1) ^a	4.3 (1.0)	4.4 (1.1)
Twice	4.0 (1.3)	4.3 (1.2)	4.1 (1.2)
Total	4.2 (1.2)	4.3 (1.1)	

a = out of 6

Overall subjects were scoring 71.5% correct when listening to the stories and 70.5% correct when reading the stories. The scores obtained using the stricter marking criterion were subjected to analysis of variance the results of which are summarised in Table 11.3.

Table 11.3 Summary of analysis of variance for Experiment Ten

Source	df	SS	MS	F	P
Between subjects	3	75.99			
Between A	1	1.26	1.26	0.03	N.S
Subjects within A	2	74.73	37.37		
Within subjects	48	50.5			
Between B	1	0.09	0.09	0.09	N.S
Subjects X B	47				
A X B	1	1.26	1.26	1.18	N.S
Subjects X B					
Within A	46	49.2	1.07		
Total	95	126.49			
A = Exposure			B = Modality		

The analysis indicated no significant effects for frequency of exposure to the stories, modality of presentation or any interaction between modality and frequency of exposure. The correlation between individuals scores in the two modalities failed to reach significance ($r = 0.2$, d.f.46).

Discussion

Before interpreting the results of this experiment, the possibility that the results arose simply because the task was intrinsically non-discriminatory should be addressed. Three grounds can be advanced for arguing that the results are non-artifactual. The first is that the task used is an established

one which has discriminated between the performance of younger and older children (Tunmer et al 1983). The second is that floor and ceiling effects did not appear to present a significant problem in that the range of scores was from five to 12 with only two subjects achieving the maximum score. Finally the finding of no significant difference on reading the text twice is consistent with the results of Experiment Five which used different materials.

The finding of no significant enhancement of comprehension on reading the material twice is not consistent with Carver's (1987) view that 'students can improve the degree to which they comprehend a passage by spending more time reading the passage'.

These findings are however consistent with the concept of comprehension as building mental models 'by the concurrent construction of a mental model of a text while reading or listening' as advanced by Yuill and Oakhill (1991 p.111). They went on to note that they had found allowing poor comprehenders to study a text had not improved their performance on inferential questions and speculated that poor comprehenders did not know how to make use of extra reading time.

Yuill and Oakhill's model of reading comprehension contains more elements than just the concurrent integration of text by drawing inferences to construct a mental model. In particular, they also stress the need for an adequate working memory. The relationship between working memory and reading comprehension will be examined in Chapter 12.

CHAPTER TWELVE

READING COMPREHENSION AND MEMORY

Yuill and Oakhill (1991) have developed a theory of reading comprehension involving the reader in the inferential construction of a mental model of the text as they read it. They distinguished a role for memory in this process. However the results reported by Yuill and Oakhill did not support there being a difference in short term memory between good and less good comprehenders. For example when some children were asked to recall lists of four words or pictures whose names were one, two or three syllables in length no significant difference was found between good and less good comprehenders.

Earlier work carried out by Oakhill (1982) also produced findings inconsistent with differences in verbatim memory playing a major role in achieving different levels of comprehension of discourse. In this experiment children were read eight stories and subsequently had to identify whether they had heard some sentences in those stories. Half the sentences presented to them occurred in the stories, a quarter of them could have been logically inferred from the stories and a quarter were semantically incongruent with the stories. There was no difference between skilled and unskilled comprehenders in their recognition of the original sentences. This suggests there was no difference between the groups in their verbatim memory. However the skilled comprehenders were more likely to falsely identify the semantically congruent sentences as having been heard suggesting they had a better memory for gist. A follow up-study by Oakhill, Yuill and Parkin (1986) replicated this result even though in this experiment the children were only asked to listen and understand the stories rather than

remember them as in Oakhill's original experiment.

Other studies using multiple correlations between a range of measures and reading comprehension have also failed to identify a significant role for verbatim memory. For example Saarnio, Oka and Paris (1990) obtained scores from 213 third-grade and 216 fifth-grade children on measures of reading comprehension, decoding skill, cloze, awareness of reading, self-perception about reading and performance on a test of free-recall of a list of 20 words. All correlations between these measures and reading comprehension except recall by fifth-graders were 0.39 or greater and all were significant. However the correlations among all the measures were also significant. To determine the relative importance of each predictor for reading comprehension step-wise multiple-regression analyses were conducted separately for each grade. For third-graders recall only accounted for 3% of the variance while for fifth-graders recall did not emerge as a significant predictor.

In the light of this evidence of limited difference between good and poor comprehenders in verbatim memory, Yuill and Oakhill (1991) suggested good and less good comprehenders may differ in their capacity to store and process information concurrently. That is they differ in their working memory. In terms of the model of working memory proposed by Baddeley (1990) the role they assigned the working memory used in reading comprehension placed it in the central executive rather than the phonological loop or the visuo-spatial sketchpad.

The major experimental evidence provided by Yuill and Oakhill for the relationship between working memory and reading comprehension came from an experiment in which good and less good comprehenders were

shown lists of two, three or four three digit numbers one group of digits at a time. The students had to recall the last digit of each group in the list in order of presentation. They found a substantial correlation between this measure of working memory and reading comprehension score on the Neale Analysis of 0.51. However the Neale Analysis requires subjects to remember the story they have read rather than refer to it when being asked comprehension questions on it. This may tend to inflate the relationship between performance on the Neale and performance on this measure of working memory. A more stringent test would be if the relationship still existed if the measure of comprehension gave the readers continuous access to the text during assessment. Some indication that there may still be some differences between the two groups identified on the basis of their performance on the Neale when the passage is available is that Yuill and Oakhill even with the text available still found better comprehenders on the Neale were better at answering inferential questions but there was no significant difference between them and less good comprehenders on literal questions. However not all their manipulations of memory loading produced differences between skilled and less skilled comprehenders on inferential reasoning. In their Experiment 5.4 in which an additional filler sentence was inserted in some versions of a short story from which readers were meant to draw inferences the presence or absence of the filler sentence did not differentially affect the performance of more and less skilled comprehenders. One might have anticipated more skilled comprehenders would have been less influenced by the insertion of the filler sentence if they had better working memories.

Though the experimental evidence for skilled comprehenders having better working memories in the experiments carried out by Yuill and Oakhill is mixed it is common in the literature to find a correlation between numeric or

verbal measures of working memory and reading comprehension of between 0.26 and 0.9 (Daneman 1987). However even strong correlations do not indicate causality and Daneman who with Carpenter (Daneman and Carpenter 1980) devised the reading span test of working memory which has produced often large correlations with reading comprehension argued in her 1987 paper against the utility of employing the notion of a single general working memory system being responsible for reading and all information processing. Instead she argued in favour of domain-specific systems with reading being under the control of a system specialised for manipulating verbal or symbolic information only. The evidence to which she referred was the pattern of correlations between different measures of working memory: verbal, mathematics and spatial and measures of reading comprehension, vocabulary knowledge and verbal ability. The strongest correlations were between measures of verbal working memory and the three verbal abilities with no significant correlations being reported between spatial working memory and any of the verbal abilities.

There are a number of problems in interpreting these correlational studies. One is that the size of the correlation between working memory and reading comprehension varies depending on the measures used. The correlation appears to be largest for those measures of working memory closest to reading itself. This has led some people to argue for reading involving a domain specific working memory only used for verbal or symbolic information. Alternatively the correlations obtained between measures of working memory and reading comprehension may be mediated by a third variable e.g. a general language skill and working memory may make little unique and direct contribution on its own.

There is however another technique which has been used to investigate

comprehension of discourse in which experimenters have discerned a role for working memory. This approach uses randomised text to investigate the influence of referential continuity in discourse. The rationale is that the construction of a mental model of the text is facilitated by continuity of reference between sentences and that creating discontinuities in the text increases the load on working memory as readers would need to hold information in memory longer in order to construct a model of the world described in the text.

Ehrlich and Johnson-Laird (1982) found that undergraduates were better able to draw diagrams of spatial arrays from spoken or written descriptions when the descriptions were referentially continuous ie. every sentence referred back to an item in the preceding sentence, rather than when there was a referential discontinuity between the first two sentences of the three sentence descriptions. They argued that their subjects were trying to build up a mental model of the spatial layout described. However when the subjects were unable to integrate a premise because it had no item in common with their then current mental model they were forced to keep the information in working memory so placing a greater load on working memory.

Subsequently Oakhill and Garnham (1985) also found adults reading three sentence referentially discontinuous descriptions of spatial and non-spatial relations took longer than those reading referentially continuous descriptions. They proposed this could be explained if the memory load was considered. They argued that discontinuous descriptions may make a description harder to remember as elements have to be held in memory while the model is being constructed. However they also noted a description may take longer to process because its content is intrinsically

harder to represent or hold in memory.

Garnham, Oakhill and Johnson-Laird (1982) examined the recall of eight year old readers for short stories that were intact or had been randomised or modified from their random state to enhance continuity. The readers had previously been divided into skilled and unskilled comprehenders on the basis of their performance on the Neale Analysis of Reading Ability. The skilled comprehenders recalled best the passages in their original order, followed by the modified randomised passages and performed least well with the randomised passages. The less skilled comprehenders recalled more of the passages in their original order than when randomised but there was no difference between their performance on the randomised and modified randomised versions. The skilled comprehenders also recalled more of the original and revised random passages than the less skilled comprehenders but there was no difference in their performance on the randomised versions. Garnham et al believed this was because readers who were poor at making the inferences required to establish co-reference between sentences gained relatively little from the re-establishment of coherence in the modified randomised stories.

Yuill and Oakhill (1991) in their commentary on this experiment observed that 'the less-skilled group were less aware of or less able to use, the referential features of texts to facilitate their understanding and memory' (p175). They also went on to later conclude that 'less skilled children showed evidence of poorer working-memory efficiency' (p216). Whilst these two statements are not incompatible the experimental work reviewed has not suggested whether skilled comprehenders remember coherent text better because they have a better memory or that they remember it better because they are better able to use the referential features of the text to

facilitate their memory or a combination of both.

It is proposed to investigate this by comparing good and less good comprehenders' abilities to re-establish continuity in randomised text with continuous access to the text, to recognise the presence of sentences in text with the text available and to remember original and partially discontinuous text. It is hypothesised that if ability to establish referential continuity is the prime factor in facilitating comprehension of text then better comprehenders would perform better than less good comprehenders in sequencing text. They should also be better at remembering partially discontinuous text as their ability to re-establish continuity would facilitate recall. If better comprehenders have better working memories they should better recognise the presence of sentences given once orally in text available to them. If good working memory, memory for text and ability to integrate text are important the better comprehenders would produce superior performance compared to the less good comprehenders on all these measures.

In addition to comparing the influences of memory and constructive processing in comprehension the results of the experiment may also have some implications for the model of working memory employed in any theory of reading comprehension. Though Yuill and Oakhill(1991) favour the model of working memory advanced by Baddeley the results reviewed above are compatible with the fuzzy-trace model of working memory advanced by Brainerd and Kingma (1985). This model proposes two types of trace are encoded into working memory. One type which usually serves as the basis for responding to short-term memory tasks retains verbatim information while the other contains gist. An important characteristic of the fuzzy-trace model is that reasoning by preference relies on gist traces and that successful reasoning can be achieved independently of verbatim

memory for the problem information. This model would predict no difference between good and less good comprehenders on verbatim recall.

EXPERIMENT ELEVEN A

Subjects

Forty eight Year Four pupils of whom 27 were girls ranging in age from eight years seven months to nine years six months with a mean age of nine years one month and a standard deviation of four months all attending the same junior school.

Materials

Five stories each of 100 words in length, made up of ten sentences and with Flesch Reading Ease scores of between 97 and 104 were written.

Passage Six which lent itself to a chronological sequence was divided into its 10 sentences with each printed on a separate strip of card of the same width.

Passage Three was used in the same form as in Experiment Five with ten words of inappropriate meaning but the same part of speech substituted for ten words in the passage.

Passages Two, Four and Seven were also prepared in discontinuous versions in which the introductory first sentence was re-inserted at a natural break in the story. This created the discontinuity of there being no referent for the pronouns used in the story until the original first sentence was subsequently reinserted. Twenty recognition sentences were prepared for each passage. Ten were the original sentences of which the stories were comprised. Five were semantically congruent foils and five were semantically incongruent

foils.

Design

Each child completed the error detection task followed by either the sequencing task or the recognition task for a version of each passage. The boys and girls scoring five or over on the error detection task made up the group of better comprehenders and each completed one of six combinations of order of presentation of the three memory and the sequencing tasks shown in Figure 12.1. Those scoring under five on the error detection task made up the less good comprehenders, who also each completed one of the six combinations of memory and sequencing tasks.

The three memory tasks were designed to place differing demands on working memory. In one version the subjects had the passage available so that they could refer to it. In a second version the passage was not available so increasing the demand on working memory. In the third version the story was written so initial pronominal references were not made explicit until part way through the story so further increasing the memory load.

Figure 12.1 The six combinations of tasks

Group 1 ED	S	M ₁ 2	M ₂ 4	M ₃ 7
Group 2 ED	S	M ₃ 4	M ₁ 7	M ₂ 2
Group 3 ED	S	M ₂ 7	M ₃ 2	M ₁ 4
Group 4 ED	M ₁ 2	M ₂ 4	M ₃ 7	S
Group 5 ED	M ₃ 4	M ₁ 7	M ₂ 2	S
Group 6 ED	M ₂ 7	M ₃ 2	M ₁ 4	S

S = Sequencing task

ED= Error detection task

2 = Passage Two

4 = Passage Four

7 = Passage Seven

M₁ = Original text available

M₂ = Original text unavailable

M₃ = Modified text unavailable

After each memory passage was read out loud by the children they were told they were going to be read some sentences. If they had just read a sentence with exactly the same words they were to say yes. If they had not just read a sentence with those exact words they were to say no. They were asked if they understood before the sentences were read. This procedure was designed to eliminate any recency effect.

Procedure

Each child was seen individually by the experimenter in a quiet room. Following administration of a practice item each child was asked to read Passage Three and if they found any word in the passage which did not make sense in the story they were asked to underline it. They were then presented with either the sequencing or the memory tasks. For the sequencing task they were given the ten sentences making up Passage Six in random order and were asked to arrange the sentences in order so they told a story. For the memory tasks they were asked to read out loud a story. They were told that the experimenter would then read out some sentences

and they were to say whether or not they had just read exactly the same sentence. In line with the conventions of individual reading tests the children were allowed to make up to 16 errors which were corrected. Those making more than 16 errors (two subjects) were eliminated from the experiment. They were then read the 20 sentences comprising the recognition set in an individually randomised order having been instructed to respond yes if they had just read exactly that sentence or no if they had not read exactly that sentence in the story they had just read. The child's response was noted at the time of testing.

Scoring

The sequencing task was scored by giving one point for every sentence following the one that was being marked that in the original sequence would follow it e.g. a perfect sequence scored a maximum of 45 points while the sequence 1 2 8 9 5 4 6 7 3 10 scored 30 points. This was made up as follows:

Sentence 1 scored 9 as it was followed by the 9 sentences it should be.

Sentence 2 scored 8 as it was followed by the 8 sentences it should be.

Sentence 8 scored 2 as it was followed by sentences 9 and 10.

Sentence 9 scored 1 as it was followed by sentence 10.

Sentence 5 scored 3 as it was followed by sentences 6, 7 and 10.

Sentence 4 scored 3 as it was followed by sentences 6, 7 and 10.

Sentence 6 scored 2 as it was followed by sentences 7 and 10.
and so on.

Results

Table 12:1 summarises the performance of the better and less good comprehenders on the sequencing task.

Table 12.1 Mean score (and standard deviation) on the sequencing task for good and less good comprehenders

	Mean score on sequencing task
Good Comprehenders	32.5 (5.8)
Less Good Comprehenders	28.0 (4.5)

An independent t-test showed the better comprehenders were significantly better on the sequencing task ($t = 4.044$, $df = 46$, $p = 0.01$).

Tables 12.2 and 12.3 summarise the performance of the better and less good comprehenders on the memory tasks.

Table 12.2 Mean score (and standard deviation) on the memory tasks by passage for good and less good comprehenders shown for the 10 verbatim and total recognition set

	Passage 2		Passage 4		Passage 7		Total %	
	10	20	10	20	10	20	for verbatim	for all
Better Comprehenders	8.5 (1.4)	16.3 (2.1)	8.3 (1.2)	13.9 (1.5)	8.1 (2.1)	15.8 (2.2)	83.2	76.7
Less Good Comprehenders	8.8 (1.4)	15.1 (2.5)	8.9 (1.0)	14.1 (1.9)	8.6 (1.3)	15.7 (1.6)	87.8	75.6

With respect to the set of ten verbatim recognition sentences there is a small but consistent difference across the three passages in favour of the less good comprehenders. A summary of the responses to all the recognition items broken down by text presentation is shown in Table 12.3 in which the raw scores have been converted to proportions to facilitate comparison.

Table 12.3 **Mean proportional score (and standard deviation) for the three recognition types by text presentation**

	Original text available			Original text unavailable			Modified text unavailable		
	V	Si	Sc	V	Si	Sc	V	Si	Sc
Better Comprehenders	0.87	0.82	0.61	0.81	0.82	0.55	0.81	0.83	0.58
Less Good Comprehenders	0.10	0.17	0.34	0.21	0.14	0.32	0.15	0.11	0.24
	0.89	0.83	0.52	0.90	0.74	0.44	0.84	0.73	0.53
	(0.12)	(0.19)	(0.28)	(0.10)	(0.19)	(0.28)	(0.14)	(0.23)	(0.32)

V = verbatim Si = semantically incongruent Sc = semantically congruent

A three factor analysis of variance (Winer 1962) with comprehension level (A) as a between subjects factor and passage presentation (B) and recognition type (C) as within subject factors was carried out. The results are summarised in Table 12.4.

Table 12.4 Summary of analysis of variance of performance of better and less good comprehenders on recognition tasks

Source	df	SS	MS	F	P
Between Subjects	47	2.2699			
A	1	0.1039	0.1039	2.206	NS
Subject W Groups	46	2.166	0.0471		
Within Subjects	384	25.9067			
B	2	0.1489	0.0745	1.966	NS
AB	2	0.0078	0.0039	0.103	NS
B x Subject W Groups	92	3.49	0.0379		
C	2	8.1022	4.0511	74.771	0.01
AC	2	0.3329	0.1665	3.073	NS
C x Subject W Groups	92	4.9849	0.0542		
BC	4	0.1124	0.0281	0.602	NS
ABC	4	0.1289	0.0322	0.690	NS
BC x Subject W Groups	184	8.5987	0.0467		

It can be seen there was only one significant variable recognition type. Inspection of Table 12.3 indicated that both better and less good comprehenders did least well on the semantically congruent recognition items.

Further investigation of the relationship between the major variables in the experiment was carried out by calculating the Pearson Product Moment correlations between them. The coefficients obtained are shown in Table 12.5.

Table 12.5 Correlation coefficients between error detection (ED), sequencing (SE), total recognition score (R), and score on verbatim (V), semantically incongruent (SI) and semantically congruent (SC) items

	SE	R	V	SI	SC
ED	0.361*	0.222	-0.099	0.243	0.215
SE		0.070	-0.195	0.196	0.069
R			0.242	0.697**	0.617**
V				0.567**	0.048
SI					0.084

*p = 0.05

**p = 0.01

No significant relationships between performance on either the error detection or the sequencing task and any of the types of recognition items were found.

Discussion

The results of this experiment support the position of Yuill and Oakhill (1991) that comprehension of text is related to ability to integrate it into a meaningful whole.

The performance of good and less good comprehenders did not vary significantly across presentation modes. The same instructions were used across all three presentation modes so that the children were not specifically advised they could refer back to the text when it was available. Often the children pointed out it had not been removed and even when reassured that this was all right made a point of not referring to it. The lack of variation in performance between the original and modified text is consistent with the

application of a matching strategy where meaning need not be invoked but a judgement made as to the familiarity of a particular string of words. The application of only a matching strategy should result in there being no difference in performance between the types of recognition item. However readers performed significantly better on the semantically incongruent as compared to the semantically congruent items. This suggests both good and less good comprehenders may also be able to apply the additional strategy of assessing the semantic congruency of the sentences to the model of the meaning they had constructed from the text.

The finding of no significant difference in performance between the better and less good comprehenders on the semantically congruent foil items in this experiment is in contrast to the less good performance of the better comprehenders in the experiments reported by Oakhill (1982) and Oakhill et al (1986). There were two differences between the design of these two experiments and the current experiment which may have influenced this.

1. In the Oakhill experiments the children were read all eight stories slowly with a pause between each sentence before the recognition task.
2. Additionally in the Oakhill experiments there was a three minute card sorting task before the children were read the recognition sentences.

Bartlett (1932) demonstrated that after an interval of about 15 minutes recall of a prose passage is characterised by loss of detail. Gomulicki (1956) also showed that omissions of words which contribute least to the general meaning of the passage characterise immediate attempted verbatim recall of prose passages. In both these studies the key parts of the passage or gist

were preserved. Hearing all eight passages before the recognition test combined with the delay before the recognition test was administered may have biased performance in favour of gist over verbatim memory. This may have resulted in equivalence of performance on the verbatim items as the meaning of these clearly fits the passage but better performance on the semantically congruent items by those relying on verbatim memory.

This argument would also lead one to anticipate better performance under the conditions in this experiment as compared to that obtained by Oakhill (1982) and Oakhill, Yuill and Parkin (1986). Examination of Table 12:6 indicates this is the case.

Table 12.6 Percentage correct for each recognition type in this experiment compared with Oakhill (1982) and Oakhill, Yuill and Parkin (1986) for skilled s skilled comprehenders

Recognition Type	This experiment		Oakhill (1982)		Oakhill et al (1986)	
	Skilled	Less Skilled	Skilled	Less Skilled	Skilled	Less Skilled
Verbatim	83	88	71	60	80	74
Semantically Congruent	58	50	36	50	37	47
Semantically Incongruent	82	77	81	70	79	69

However Oakhill used generally younger i.e. seven to eight year old pupils in her experiments as compared to the eight to nine year olds used in this experiment and this too would lead one to anticipate a better standard of

performance in the experiment reported here. Experiment 11B was carried out to investigate the effect of slightly delaying administration of the recognition set for approximately three minutes which was the duration of the delay built into Oakhill's experiments by the card sorting task.

EXPERIMENT 11B

Subjects

Forty Year Four pupils all attending the same junior school aged between eight years ten months and nine years nine months with a mean age of nine years four months and a standard deviation of four months made up the final experimental group.

Materials

Three stories of a hundred words in length made up of ten sentences and with Flesch Reading Ease scores of between 97 and 102 were used. Passage Six was used as the sequencing task as in Experiment 11A. Passages Four and Seven were used as the memory passages in their unmodified form.

Design

All children completed the sequencing task individually. On the basis of their performance on this task they were assigned to either the good or less good comprehender group. Each subject then completed the recognition sets for two passages. Passage order was counterbalanced within the groups. The delay was introduced by the recognition set for the first passage read not being administered until after the recognition set for the second passage which introduced a delay of about three minutes. The text was unavailable to the subjects during the presentation of the recognition set.

Procedure

All 53 available pupils were seen individually to complete the sequencing task as in Experiment 11A. Three pupils were eliminated at this stage: two were non-readers and one was dyspraxic. The 20 best and 20 least good performers were in a second individual session presented with Passages Four and Seven to read out loud. Two further subjects in the less good group were eliminated at this point as they were unable to read a passage without making over 16 errors. The next lowest scorers were used as substitutes. The subjects were then presented with the recognition sets for the passages in the opposite order to which they read them.

Results

The better comprehenders obtained an average score of 35.65 with a standard deviation of 2.11 on the sequencing task while the final experimental group of less good comprehenders obtained an average score of 21.45 with a standard deviation of 4.66. The total pool of scores on the sequencing task ranged from 12 to 40.

A summary of the performances of the two groups on the recognition sets is shown in Table 12:7.

Table 12.7 **Means (and standard deviations) of performances of more and less skilled comprehenders on the three recognition types under both conditions**

Comprehension Level	Immediate			Delayed		
	Verbatim	SI	SC	Verbatim	SI	SC
More Skilled	9.1 (0.91)	3.95 (0.89)	1.75 (1.29)	8.9 (1.25)	3.55 (0.89)	1.55 (1.64)
Less Skilled	9.15 (0.99)	3.65 (0.99)	1.4 (1.47)	9.05 (1.00)	3.25 (1.12)	1.4 (1.39)

The raw scores were converted to proportions to enable a three factor analysis of variance (Winer 1962) with comprehension level (A) as a between subjects factor and passage presentation (B) and recognition type (C) as within subject factors to be carried out. The results are summarised in Table 12:8.

Table 12.8 **Summary of analysis of variance of better and less good comprehenders on recognition tasks under delayed and immediate presentation conditions.**

Source	df	SS	MS	F	P
Between Subjects	39	2.466			
A	1	0.067	0.067	1.063	NS
Subjects W Groups	38	2.399	0.063		
Within Subjects	200	23.043			
B	1	0.088	0.088	2.095	NS
AB	1	0.004	0.004	0.095	NS
B x Subjs W Groups	38	1.591	0.042		
C	2	15.105	7.553	215.8	0.01
AC	2	0.057	0.029	0.829	NS
C x Subjs W Groups	76	2.651	0.035		
BC	2	0.052	0.026	0.565	NS
ABC	2	0.004	0.002	0.043	NS
BC x Subjs W Grps	76	3.490	0.046		

The analysis of variance revealed only one significant variable which was type of recognition item. Inspection of Table 12:7 indicated that both better and less good comprehenders did least well on the semantically congruent recognition items.

Discussion

The results of this experiment are in line with those of Experiment 11A which it partially replicated in that no difference in performance on the recognition items was found between better and less good comprehenders.

This contrasts with the results of Oakhill (1982) and Oakhill et al (1986)

who found better comprehenders did less well on the semantically congruent recognition items than less good comprehenders. No significant difference in performance occurred when recognition items were presented after a short delay as compared to immediately the passage had been read. This would not support the hypothesis that the brief delay which occurred in Oakhill's experiments may cause better comprehenders to rely more on gist memory.

In this experiment a sequencing task was used to assess comprehension level as compared to the error detection task used in Experiment 11A. Both of these measures differ from those used by Oakhill who used the Neale Analysis in that they gave the reader continuous access to the text while completing the task. In contrast in the Neale Analysis the passage is not available for inspection when questions are asked on it. It may be good performance on the Neale comprehension questions is consistent with having good gist memory for text.

A further difference between these experiments and the earlier work of Oakhill and her colleagues is that while Oakhill read both the passages and the recognition set in these experiments the subjects read the text and listened to the recognition set. Modality of presentation effects have been found for sentence recognition memory (Flagg and Reynolds 1977) which would argue for a further replication contrasting all four combinations of aural or visual presentation of the text and recognition set to investigate whether modality of presentation influences outcome .

In conclusion in these experiments no difference was found between the performance of good and less good comprehenders on the different types of recognition item. However the results are consistent with those of Oakhill (1982) and Oakhill et al (1986) in that in all four experiments both classes of

comprehenders misrecognised more semantically congruent sentences. This is also consistent with the earlier work of Blachowicz (1977-78). She employed a very similar design to that used by Oakhill (1982) except she used as subjects 40 seven year olds, 40 ten year olds, 40 twelve year olds and 30 adult students. The school age subjects were selected randomly from their grades. The youngest subjects made most misrecognitions overall, the middle graders formed a homogenous subset and the adults made the fewest misrecognitions. There was a strong tendency for all subjects to recognise semantically congruent inferences as having been present in the original paragraphs. This tendency evident across experiments covering readers aged from seven years to adult would be consistent with most readers typically engaging in constructive processing of text.

CHAPTER THIRTEEN

SUMMARY, CONCLUSIONS, IMPLICATIONS FOR TEACHING COMPREHENSION OF TEXT AND SUGGESTIONS FOR FURTHER RESEARCH

Summary and Conclusions

There is a long tradition stretching back at least as far as Huey (1908) and continuing to the present (Daneman 1987, La Berge and Samuels 1974, Stauffer 1969, Yuill and Oakhill 1991) of viewing reading as a problem solving activity. A recent development of this tradition has been the metacognitive approach to reading (Garner 1987). The metacognitive approach emphasises the active control of reading by the selection of an appropriate reading process or strategy and the monitoring of the outcome when it is applied. A major research tool within the metacognitive approach has been error detection. This involves the reader in detecting errors in a text. The reader has to identify for himself a breakdown in comprehension. This distinguishes error detection from most conventional measures of comprehension in which the need for a response is signalled. For this reason error detection lends itself to the investigation of comprehension in everyday reading during which one is not normally asked to respond to questions or fill in missing words.

Within the metacognitive approach emphasis is placed on readers using a strategic approach to reading. Carver (1990a) distinguished five basic reading processes: scanning, skimming, rauding, learning and memorizing. He stressed the need to identify which process is being used by readers as different performances can be expected by the application of different

processes to achieve different goals. This study was concerned with reading for meaning; a process called rauding by Carver.

Carver regarded the rauding process as the most important of all the reading processes as it is the process individuals use most often when reading. It was assumed pupils were using the rauding process in the research reported upon in this study. During rauding individuals were considered by Carver to be comprehending about 75% or more of the complete thoughts encountered. This comprehension level was also that set by Betts (1946) as the standard to be reached for reading material to be at the Instructional Level of difficulty. This was also the standard set by Flesch (1948) as that achieved by Grade Four pupils reading material with a reading ease score of 100. The three passages used in the first four experiments all had reading ease scores slightly higher than 100; the higher the score the easier being the passage.

The first four experiments investigated the standards of comprehension used by pupils in an error detection task. Different classes of substitutions were made into the three short passages. Generally consonant strings when substituted into the passages were easily detected. Fewer pseudoword substitutions were detected and even fewer real but contextually inappropriate words were detected. The detection rate for real words declined when they were the same part of speech as the word they replaced. In line with the results of Baker (1984a) nearly all children were identifying items in all three classes of substitution. This contrasts with the results of Baker (1984b) who found a significant minority of 35 percent of nine year olds who only detected nonsense words when presented with stories which contained either nonsense words, prior knowledge violations or internal inconsistencies. In Experiment Two only 15 percent of nine year olds only

detected nonsense words or consonant strings, though 5.5 percent detected all consonant strings and nonsense words but no real word substitutions. However in both Baker (1984a) and Experiment Two the detection rates for nonsense words were much higher amongst nine year olds at 86 percent and 77 percent respectively as compared to the 46 percent in Baker (1984b). It may be the passages used by Baker (1984b) were more difficult. In Experiment Four rewriting some of the passages so that they consisted of longer sentences containing a greater proportion of polysyllabic words reduced the rate of detection of real word substitutions whilst the rate of detection of consonant string or pseudoword substitutions remained high. The results of these experiments are summarised in Table 13:1 and are consistent with the application of different standards of comprehension of text: lexical, syntactic and semantic.

Table 13.1 Combined detection rates in percentages for different types of substitution for children aged eight to ten in Experiments 1, 2, 3 and 4

Substitution Type	Flesch Reading Ease Score of Passage	
	100 - 105	80 - 85
Consonant strings	86	73
Pseudowords	71	57
Real words of different part of speech	69	
Real words of same part of speech	53	19

The results are also consistent with the Easiness Principle advanced by Carver (1987) that one can increase the degree to which students will comprehend passages by using passages easier than those at their frustration level and the conclusions reached by Zirinsky (1984) that text factors must

be taken into account when assessing readers' level of comprehension. Zirinsky went on to state that both textual and contextual factors exert a powerful influence on qualitative aspects of comprehension.

If manipulations of text readability can significantly influence comprehension then this reinforces the advice proffered by Carver (1987) that schools should 'adopt textbooks for content area reading that were not written at a difficulty level higher than the ability level of the students. Or, if the given textbooks were indeed at a higher level, teachers would not ask students to read these materials but instead someone would read the books to the students'. Not only is the readability of the text in its narrow sense important but also in the wider sense of the assumptions it makes about its readers. Though it is sometimes argued that teachers compensate for omissions McKeown, Beck, Sinatra and Loxterman (1992) found that for fifth grade pupils even extensive preparation to fill in background did not compensate for inadequacies in the original text.

The results of Experiments One and Two have some implications for Smiths' (1973) assertion that in reading silently readers move from the surface structure of the writing to the deep structure. He states that 'In such a conceptualization there is no room to hypothesise decoding to sound at all.' (Smith 1973 p82). If this were the case one would predict that the rate of detection of consonant strings and pseudowords would be the same as both have no conventionally assigned meaning. However the rate of detection was different. This could be because readers can apply different standards of comprehension and the application of one standard e.g. decoding rather than another would result in the detection of different classes of error. However no direct evidence was obtained as to which standard the reader applied to identify the substitution. Asking the readers why they underlined

a substitution would have given some information as to what standard they applied in that case.

Though the evidence from Experiments One through Four is consistent with the application of different standards of comprehension and the literature contains descriptions of a number of possible taxonomies of comprehension (Baker 1985a, Clymer 1972, Collins and Smith 1982, Spache 1963) commercially available assessment material does not generally distinguish between different levels of comprehension. Indeed some tests which distinguish between reading accuracy and reading comprehension e.g. the Neale Analysis of Reading Ability, MacMillan Individual Reading Analysis and New MacMillan Reading Analysis confound the two measures as the child's accuracy of reading controls their exposure to the comprehension questions which can only be asked on passages they have read.

The focus of this study was on semantic comprehension. This was examined in light of the model of comprehension advanced by Yuill and Oakhill (1991). They believed readers construct a mental model of the text they are reading. McNamara, Miller and Bransford (1991) describe how the construction of mental models by the reader can account for results which cannot be explained by schema theory (Anderson and Pearson 1984). Schema theory involves the instantiation of an existing schema with text specific information. However it does not explain how readers understand texts about unfamiliar objects.

Mental models differ from propositional representations. Propositions which are the smallest units of knowledge that can stand as separate units. For example sentences 11 and 12 below differ by a single proposition that

specifies whether the fish swam under the turtles or under the log.

11. Three turtles rested on a floating log, and a fish swam beneath them.
12. Three turtles rested on a floating log, and a fish swam beneath it.

Bransford, Barclay and Franks (1972) found subjects who memorized sentence 11 later had difficulty deciding whether they had learned that sentence or sentence 12. However subjects who memorized sentence 13 did not confuse it with sentence 14.

13. Three turtles rested beside a floating log, and a fish swam beneath them.
14. Three turtles rested beside a floating log, and a fish swam beneath it.

These latter two sentences describe different events as distinct from the former sentences which describe the same event. Mental models of the latter would be different and so not easily confused whilst mental models of the former would be the same and so easily confused.

Yuill and Oakhill's (1991) version of a mental model theory of reading comprehension hypothesised that readers construct their mental model by drawing inferences and monitoring their comprehension. This they postulated requires a good working memory. Unlike Carver (1987) they did not consider that increased study time would increase comprehension of text. This was tested in Experiment Five using both an error detection and a cloze task. This involved increasing the number of substitutions from six in the earlier experiments to ten to parallel the cloze version of the passage. The amount of engaged time was manipulated by asking some readers to reread the passage before attempting the comprehension tasks. No

significant enhancement of performance was found on rereading the passage through significantly more cloze spaces were correctly completed than inappropriate substitutions at the same points in the passage were identified.

The results described above were consistent with Yuill and Oakhill's position. However the readers' performance on the error detection task on the one passage used in this experiment was very much poorer than that obtained in previous experiments. This remained the case even when subjects were informed they would be asked questions on the passage to encourage them to monitor their comprehension (Experiment Six). Though in this follow up experiment a significant difference in performance was found between when subjects had to indicate where a word was missing and where a word with an inappropriate meaning was substituted this difference in favour of the omission task was shown in later work to be possibly related to the omission of words creating syntactical anomalies which could be detected by syntactic comprehension monitoring (Experiment Seven). The poorer performance on the error detection task also appeared not to be related to there being more substitutions in the passage used in Experiments Five and Six than in earlier experimental work, (Experiment Eight).

The results of Experiment Nine suggested that tasks that required readers to identify for themselves a breakdown in comprehension were harder than similar tasks where the potential difficulty was signalled for the reader. This difference was the same for both good and less good comprehenders. A significant correlation between an error detection task which required readers to identify comprehension breakdown and their performance on a sentence completion test of reading comprehension was also found. This suggests a relationship exists between comprehension performance and comprehension monitoring, akin to that between reading accuracy and

reading comprehension in that the former sets limits on the latter. Whilst better comprehenders were better at identifying comprehension breakdown the relationship between their performance on prompted and unprompted comprehension tasks was no different to that of less good comprehenders. It may be more parsimonious to distinguish between tasks in which comprehension is prompted and unprompted tasks where the reader has to identify and repair a breakdown in comprehension.

One weakness of Experiment Five in which no difference was found between comprehension performance on reading the material once or reading it twice was that it relied on only one passage. Experiment Ten examined this further using 12 short stories which were presented once or twice either aurally or visually. No significant difference in performance was found either for frequency of presentation or for modality of presentation. This is consistent with comprehension being seen as the construction of a representation or model of the text whilst one is reading or listening to it. If the concurrent construction of a mental model takes place while reading one would predict that reading comprehension would be unimpaired by rapid serial presentation of individual words. Juola, Ward and McNamara (1982) compared performance on multiple choice comprehension questions between material presented in normal paragraph format with rapid serial presentation of chunks of either 5, 10 or 15 characters (approximately one, two or three words) on a computer display amongst undergraduates. No difference in performance was found between the two presentation formats.

Yuill and Oakhill (1991) claimed that drawing appropriate inferences spontaneously while reading or listening to a story makes demands on working memory. While this is at one level self-evidently true the results of

Experiments Eleven A and B did not support there being differences between good and less good comprehenders in recognition memory. Experiment Eleven A did however provide some support for there being a positive relationship between performance on an error detection task and sequencing the component sentences of a story to make it coherent. The ability to integrate the components of a text would be necessary for the construction of a mental model of what it represents. That better comprehenders are better at this is in line with the central claim of the theory of reading comprehension advanced by Garnham (1987) and Yuill and Oakhill (1991) which is that determinate texts (and discourse) are encoded in mental models and that these representations are the psychologically important ones. In some circumstances such as when subjects are presented with indeterminate descriptions, for example:

The bookshelf is to the right of the chair
The chair is in front of the table
The bed is behind the chair

the information presented is likely to be held in memory in a verbal form as contrasted with determinate descriptions such as:

The spoon is behind the knife
The knife is to the right of the plate
The fork is to the left of the plate

which are more likely to be retained as a mental model (Mani and Johnson-Laird 1982).

However the building of a mental model of what is represented in text is dependent on the reader's world knowledge (Garnham 1987). It goes beyond the idea that a semantic representation of a sentence contains all the information required for a complete understanding of that sentence. That is it differentiates between the meaning and the significance of a sentence (Johnson-Laird 1977). 'The significance of a sentence is very closely related to the particular situation in the world that it describes' (Garnham 1987 p19). For example the sentence; 'The light is red' has a meaning in the abstract but its significance might depend critically on whether one was referring to the colour of traffic lights one was approaching in a car or the light on a breathalyser one had blown into. For someone from a culture unfamiliar with the car their representation of the meaning of the text of which that sentence could be part may be characterised as propositions related by textual signals. 'In that sense their representation is closer to the surface form of the text than the representation of experts!' (Noordman and Vonk 1992 p338).

For this reason comprehension of text cannot be divorced from the world knowledge of the reader. The problem for those reading text for meaning is to integrate the meaning of the words in the text with their world knowledge of the situation on which the text is based, i.e. both the situation described in the text and the context in which it is presented to build a model of the meaning conveyed by the text. The emphasis is on integrating prior knowledge and text as McCormick (1992) has demonstrated that too heavy reliance on background knowledge can result in comprehension errors caused by dismissing text information in favour of prior knowledge. As McNamara, Miller and Bransford (1991) concluded 'In general, reading needs to be seen as an engineering problem: using available resources to build a model of meaning that is well suited for the job, cost effective, and structurally sound.'

Implications for Teaching Comprehension of Text

The results of this series of experiments suggest that children listening to or reading narrative passages are not particularly good at identifying at what points what they hear or read does not make sense in the story. The work reviewed in this study suggests four ways in which teachers could address this issue.

i) Identifying comprehension as a goal of reading.

Durkin (1979) and HMI (1989) have identified a lack of emphasis on teaching reading comprehension during the primary years. It has been suggested (Frederiksen 1979) this may lead to some children emphasising word pronunciation at the expense of passage comprehension. This could be addressed by:

- a) Teachers always using meaningful reading material ie. material which does not sacrifice meaning to phonic regularity or repeated exposure to a set of words. Ideally the reading material should encourage synthesis of sentences (Blachowicz 1977-78).
- b) Teachers either always asking children whom they have heard read about the meaning of what they have read or getting the children to ask the teacher about the story. This latter strategy has the potential advantage of encouraging the child's active construction of meaning whilst reducing their anxiety at being 'tested' on what they had read.

ii) Using listening comprehension tasks with younger children.

If reading and listening tasks share common components then practice at listening comprehension would have these advantages:

- a) It would allow the earlier detection of those who may later experience problems with reading comprehension (Yuill and Oakhill 1991).

- b) It would allow the child who has poor access skills the opportunity to practice comprehension at a level unavailable through reading (Curtis 1980).
- c) It would encourage the continuous construction of meaning on the part of the listener as one could not refer back to the passage on being asked a question.

iii) Using unprompted comprehension tasks

Using unprompted comprehension tasks would not only enrich the reading curriculum but also help focus readers on the importance of understanding what they were reading. The extensive use of unprompted comprehension exercises with appropriate feedback (Winne, Graham and Prock 1993) may also help close the gap between readers' performance on unprompted as compared to prompted comprehension tasks.

iv) Training readers to integrate text

If as has been argued semantic comprehension involves the integration of text to form a coherent model then training children in skills which promote this should facilitate comprehension. In particular poor comprehenders find difficulty in determining pronominal reference and this could be targeted by teachers both for their discussions with pupils of what the pupils have read and for specific comprehension exercises e.g. selectively eliminating pronouns for later insertion by pupils as an unprompted task.

Suggestions for Further Research

This will be considered under two headings:

i) Refinements of experiments carried out in this series of studies.

In Experiments One to Three the children were only asked about those errors which they failed to identify. The results provided only indirect evidence on how children identified errors. This could have been addressed by also asking the readers why they underlined the errors they identified.

If a repeated measures design had been used in Experiment Four to investigate the influence of changes of readability on error detection then it would have enabled the question to be addressed as to whether when individuals are given text of too hard a readability for them they employ a restricted range of standards of comprehension.

Further investigations of the influence of rereading could exploit the application of the different standards of comprehension which emerged in Experiments One to Four. Comparing rates of detection of errors which can be detected by the application of lexical or syntactic standards as contrasted to those errors which can only be detected by the application of a semantic standard on rereading a passage could provide empirical support for the practise of encouraging students to proof-read their own stories whilst providing evidence pertaining to the utility of distinguishing between standards of reading comprehension.

Experiment Eleven employed mixed modalities in that the readers read the stories and then listened to the sentences they were asked to identify as having been or not been just read. Flagg and Reynolds (1977) found

modality of presentation effects for sentence recognition memory. This would argue for a replication contrasting all four combinations of aural or visual presentation of the text and recognition sentences to investigate whether modality of presentation influences outcome.

ii) Further Research

The processes involved in the integration of prior knowledge and text in the construction of a mental model require further investigation. Brewer (1987) has argued that the term mental model is too general when used to describe specific knowledge structures constructed to represent new situations and has advanced the term episodic model. A potentially fruitful area for enquiry into the construction of episodic models is those devices which confer coherence on text which are known as anaphors. In the simplest case anaphors have the same meaning as a preceding portion of text, e.g. in the sentences:

George went to the library.
He chose a book.

the pronoun He has the same meaning as George. However some anaphors require the application of knowledge external to the text for resolution. In the sentences:

It had been raining when Hazel walked across the park to visit Heather.
She took her coat off before she sat down.

She is more likely to refer to Hazel than Heather as it would be normal to wear a coat when walking in the rain as Hazel did but not in ones own home into which it is inferred Heather invited Hazel.

Garnham and Oakhill (1989) have collected examples of anaphors used in everyday writing and in advertising (Garnham and Oakhill 1992). Oakhill (1993a) has summarised their experimental work on these. Oakhill concluded that despite extra work being required to interpret linguistically deviant pronouns this was not a significant obstacle to comprehension. She considered that readers used a range of cues to help interpret these anaphors but that the initial mental model of a text was tied fairly closely to the way ideas in the text were expressed. Further work needs to be carried out to establish if training students to identify pronominal referents significantly enhances their comprehension of text. Factors worthy of consideration in designing an empirical evaluation of the instructional impact of this training are identified in Appendix C.

APPENDICES

READABILITY

Many factors determine how well a child understands text. Some of these factors are related to the reader and what he brings to the text. Other factors are related to the text. Readability formulae attempt to quantify the influence of textual factors on comprehension.

Readability formulae are predictors of the comprehensibility of text. As such they use easily identified aspects of the text or of samples from it e.g. number of polysyllabic words, words in a sentence, number of frequently used words. To that extent they open themselves to the charge that they would produce the same readability score if the words were written in random order (Oakhill and Garnham 1988). However as predictors of readability rather than as guides to the production of readable text one would anticipate them only being applied retrospectively to coherent text. Applied in this way the ten formulae considered by Lunzer and Gardner (1979) as part of the School's Council - Effective Use of Reading Project produced correlations of about 0.7 with pooled teacher judgements. This research was carried out in the United Kingdom but the results are in line with American research (Harrison 1980).

Harrison (1977) following a review of 10 readability formulae concluded that the Flesch formulae was reliable as it produced a close correlation to both pooled teacher's estimates of difficulty and also childrens' performance on passages. Lunzer and Gardner (1979) following their review of the 10 readability formulae also selected the Flesch Formula as the one they wished to use in a readability level survey. Details of the formula and its derivation are given by Flesch (1948).

To calculate the Flesch Reading Ease Score a 100 word excerpt is selected and the number of syllables and the average number of words per sentence are calculated. The Flesch Reading Ease Score can then be calculated by substituting into the formula below:

$$206.835 - (0.846 \times \text{No.of syllables}) - (1.015 \times \text{average No.of words per sentence})$$

This gives a Reading Ease Score ranging from a theoretical maximum of over 110 (very easy) to 0 (very difficult).

Using this formula the passages used in this study had the following Reading Ease Scores.

Passage One	102
Passage Two	104
Passage Three	103
Passage Four	102
Passage Five	99
Passage Six	97
Passage Seven	99

A score of 100 originally indicated an average fourth grade child would be able to correctly answer three-quarters of the questions about the passage. However the Flesch Formula was devised over 40 years ago in the United States. A more modern British reading test, the New Macmillan Reading Analysis (Vincent and de la Mare 1985) can be used to estimate how it transfers to this country and generation in the same way that Spooner (1976) used the Neale Analysis of Reading Ability as a standard against which to compare the FOG Index of Readability. The New Macmillan

Reading Analysis consists of three parallel sets of short self-contained pieces of prose. Within each set or form of the test, the pieces of prose are of increasing difficulty. According to the norms of the text, a student reading only the first two passages correctly would be reading at approximately the standard of an average seven and a half year old. The combined Flesch reading ease score of the first two passages is 107. If the student read only the first three passages correctly they would be reading at approximately the standard of an average eight and a half year old. The combined Flesch reading ease score of the first three passages is 97. All the experimental passages used appear therefore to be of similar difficulty and to be capable of being read by an average reader of between seven and a half and eight and a half years of age ie. a Year Three or Year Four pupil. One can therefore be reasonably confident the passages used in this study were at least at the instructional level for most of the pupils who acted as subjects. Some indirect support for this was provided by the elimination of fewer than six percent of subjects in Experiments 11A and B who failed to read passages 2, 4 and 7 to the criterion of accuracy used in the Neale Analysis of Reading Ability.

MATERIALS

Alternatives are shown in parenthesis.

EXPERIMENT ONE**Passage One : Ben's First Day**

As Ben's blmb (blad) walked home, she felt sad. She had talked to Ben about how one day he would go to trck (sant) and that all children had to go some time. She had talked about the good bits and not the mrst (hist) bits. Still Ben had seemed tsng (stug) happy when she had left him with his cnlf (bing). She had said she would come and get him at lunch. Ben talked dwlt (fest) about his morning on the way home. But his face fell when after lunch she told him it was time to go back. But I went this morning, he said.

Passage Two: The Dog

Mr Smith had a big rmng (colk) dog called Sam. It used to fndk (pott) Tom as he walked past it to school. But one day it was whining. Tom went over and glst (gost) it. The dog licked him and brnd (glat) day from then on Tom stopped to play with the dog. On the way to school one morning, a big boy lsck (dost) Tom and hit him. He wanted Tom to give him money. Tom would not give him his money. Suddenly Sam ran down the road barking. The big boy ran off and Tom was bwst (feek) pleased he had made friends with Sam.

EXPERIMENT TWO**Passage One: Ben's First Day**

As Ben's dcrt (dert, dart) walked home she felt sad. She had talked to Ben about how one day he would go to grsb (greb, grub) and that all children had to go some time. She had talked about the good bits and not the snwp (snop, snap) bits. Still Ben had seemed pmst (pust, past) happy when she had left him with his swnt (sant, sent). She had said she would come and get him at lunch. Ben talked wslk (wulk, walk) about his morning on the way home. But his face fell when after lunch she told him it was time to go back. But I went this morning, he said.

Passage Two: The Dog

Mr Smith had a big tmlk (telk, talk) dog called Sam. It used to gld (gald, gold) Tom as he walked past it to school. But one day it was whining. Tom went over and fnst (fost, fist) it. It licked him and flst (flot, flat) day from then on, Tom stopped to play with the dog. On the way to school one morning, a big boy mrst (mest, most) Tom. He wanted Tom to give him money. Tom would not give him his money. Suddenly Sam ran down the road barking. The big boy ran off and Tom was slxw (sluw, slow) pleased he had made friends with Sam.

Passage Three: My Friend

I found my friend one morning. He was lying in a black circle in a field. He told me his strp (stip, stop) had crashed. He was small and swck (seck, sack). On dcgs (degs, dogs) they all live under the ground so I hid him under a big bush in my garden. I used to play with him sxng (seng, sing) day. My mum used to wonder who I was playing with. One day there was a black circle on our lawn. I looked for my friend but he was gone. I was clmp (clup, clap) sad and I got told off for making a brlt (bult, belt) on the lawn.

EXPERIMENT THREE**Passage Two: The Dog**

Mr Smith had a big deep (talk) dog called Sam. It used to read (gold) Tom as he walked past it to school. But one day the dog was whining. Tom went over and sent (door) the dog. It licked him and good (fist) day from then on, Tom stopped to play with the dog. On the way to school one morning, a big boy rang (most) Tom. He wanted Tom to give him money. Tom would not give him his money. Suddenly Sam ran down the road barking. The big boy ran off and Tom was neatly (sister) pleased he had made friends with Sam.

Passage Three: My Friend

I found my friend one morning. He was lying in a black circle in a field. He told me his sock (shut) had crashed. He was small and long (sack). On dogs (dull) they all live under the ground so I hid him under a big bush in my garden. I used to play with him hard (sing) day. My mum used to wonder who I was playing with. One day there was a black circle on our lawn. I looked for my friend but he was gone. I was fast (bell) sad and I got told off for making a belt (glad) on the lawn.

EXPERIMENT FOUR A**Passage Three: My Friend - Easier version**

I found my friend one morning. He was lying in a black circle in a field. He told me his sock (srck) had crashed. He was small and long (lsng). On dogs (dmgs) they all live under the ground so I hid him under a big bush in my garden. I used to play with him hard (hwrđ) day. My mum used to wonder who I was playing with. One day there was a black circle on our lawn. I looked for my friend but he was gone. I was fast (fnst) sad and I got told off for making a belt (bnlt) on the lawn.

Passage Three: My Friend - Harder version

I discovered my friend one morning lying in a black circle in a furrowed field. He informed me his personal sock (srck) had crashed. He was small and long (lsng). On dogs (dmgs) they all reside underneath the surface so I hid him underneath a big bush, in my garden. I used to play with him hard (hwrđ) day. My mother wondered whom I was playing with until one day I found a black circle on our lawn and though I searched for my friend he had disappeared. I was fast (fnst) upset and moreover I was punished for making a belt (bnlt) on the lawn.

EXPERIMENT FOUR B**Passage One: Ben's First Day - Easier Version**

As Ben's dad (ded) walked home she felt sad. She had talked to Ben about how one day he would go to tree (trem) and that all children had to go some time. She had talked about the good bits and not the pink (ponk) bits. Still Ben had seemed slowly (skowly) happy when she had left him with his grass (gress). She had said she would come and get him at lunch. Ben talked greedily (greekily) about his morning on the way home. But his face fell when after lunch she told him it was time to go back. But I went this morning, he said.

Passage One: Ben's First Day - Harder version

As Ben's dad (ded) walked homewards she had felt unhappy. She had discussed with Ben about how one day he would attend tree (trem) and that all children had to go sometime. She discussed the pleasant aspects but avoided the pink (ponk) bits. Ben appeared slowly (skowly) happy when she had left him with his grass (gress), though she had informed him she would collect him at lunchtime. When she got him Ben talked quite greedily (greekily) about his morning as they returned home, but his face fell when after lunch she informed him it was time to return. But I went this morning, he said.

EXPERIMENT FIVE**Passage Three**

I found my friend one morning. He was lying (along) a black circle in a field. He told me (its) rocket had crashed. He was small and round. (Above) Mars they all live under the ground so I hid (those) under a big bush in my garden. I used to (ring) with him each day. My mum used to wonder (which) I was playing with. One day there (felt) a black circle on our lawn. I looked for (their) friend but he was gone. I was very sad and (it) got told off for making a fire on the (tree).

EXPERIMENT SIX**Passage Three**

I found my friend one morning. He was lying (along) a black circle in a field. He told me (its) rocket had crashed. He was small and round. (Above) Mars they all live under the ground so I hid (those) under a big bush in my garden. I used to (ring) with him each day. My mum used to wonder (which) I was playing with. One day there (felt) a black circle on our lawn. I looked for (their) friend but he was gone. I was very sad and I (did) told off for making a fire on the (tree).

- Questions:**
- 1) Where did he find his friend?
 - 2) Was his friend male or female?
 - 3) From where did his friend come?
 - 4) What did he and his friend do each day?
 - 5) What made the black circle on the lawn?

EXPERIMENT SEVEN**Passage Four: The Crash**

Jim rode his bike to school each day. He had to pedal hard to get (down) the hill. Then he took his feet off the pedals and cruised all the way to the school gates. On the way back, once he got (for) the top of the hill he could lift his (hat) up and cruise home. One day a teacher was driving his new car out of the gate. Jim came tearing (up) the hill. He turned (on) the school gates and smashed into the teacher's car. He was angry and Jim had to pay for all the damage.

Passage Five: The Bird's Nest

Ken was not happy. He had climbed to the top of the tree to look in the bird's nest but now he was (painted) up the tree. He had walked past (her) to school each day but had only spotted the nest (tomorrow). He waited till Sunday morning when he did not have to go to school and there were (more) people about. He had climbed up with no problem. Now he was stuck. Suddenly he saw a man. Ken called out and the man saw him. The man went and got a ladder and helped Ken climb (up).

EXPERIMENT EIGHT**Passage One: Ben's First Day**

As Ben's mum walked home she felt sad. She had (walked) to Ben about how one day he would go to (tree) and that all children had to go some (thing). She had talked about the good bits and not the (pink) bits. Still Ben had seemed quite happy when (he) left him with his teacher. She had (sunk) she would come and get him for lunch. Ben talked (greedily) about his morning on the way home. But (its) face fell when after lunch she told him it was time to (swim) back. But I went this morning, (she) said.

Passage Three: My Friend

I found my friend one morning. He was lying (along) a black circle in a field. He told me (its) rocket had crashed. He was small and round. (Above) Mars they all live under the ground so I hid (those) under a big bush in my garden. I used to (ring) with him each day. My mum used to wonder (which) I was playing with. One day there (felt) a black circle on our lawn. I looked for (their) friend but he was gone. I was very sad and I (did) told off for making a fire on the (tree).

Passage Four: The Crash

Jim rode his bike to school each day. (She) had to pedal hard to get up the hill. Then (it) took his feet off the pedals and cruised all (a) way to the school gates. On the way back, (twice) he got to the top of the (roof) he could lift his feet up and cruise home. One (book) a teacher was driving his new car out of the (mine). Jim came tearing down the hill. He turned (over) the school gates and smashed into the teacher's (horse). He was angry and Jim had to pay for all (a) damage.

EXPERIMENT NINE**Passage Four: The Crash**

Jim rode his bike to school each day. He had to pedal hard to get (_____) the hill. Then he took his feet off the pedals and cruised all the way to the school gates. On the way back, once he got (_____) the top of the hill he could lift his (_____) up and cruise home. One day a teacher was driving his new car out of the gate. Jim came tearing (_____) the hill. He turned (_____) the school gates and smashed into the teacher's car. He was angry and Jim had to pay for all the damage.

Passage Five: The Birds Nest

Ken was not happy. He had climbed to the top of the tree to look in the bird's nest but now he was (_____) up the tree. He had walked past (_____) to school each day but had only spotted the nest yesterday. He waited till Sunday morning when he did not have to go to school and there were (_____) people about. He had climbed up with no problem. Now he was stuck. Suddenly he saw a man. Ken called out and the man saw (_____). The man went and got a ladder and helped Ken climb (_____).

EXPERIMENT TEN**Practice Items**

Fish live in the ocean. It is very dark at the bottom of the ocean. Fish that live at the bottom of the ocean often find food by its smell (colour).

Fishermen carefully choose where they fish. Fishermen need to keep their hooks under the water. That is why they put corks (leads) in their lines.

Story One

People like to stick things on the fridge in their kitchen. Fridges are made of steel. People often use magnets (pins) to hold things on them.

Story Two

Bob liked cold drinks. Bob's can of drink was too warm. He put it in the fridge (oven) before he drank it.

Story Three

Mrs Smith uses her car for shopping. One day Mrs Smith's car did not have any petrol at all. She walked (drove) to the shop to get some bread.

Story Four

Ann was sleepy so she went to bed. In the middle of the night Ann got out of bed and looked out of the window. She saw the stars (sun) up in the sky.

Story Five

Andrew liked to fly his kite in the summer. One day last summer Andrew flew his kite all day long. It was windy (calm) all that day.

Story Six

Tom liked to race sticks in the stream. Tom put two sticks in the stream. Then he ran to the bottom (top) of the hill to see which came first at the finish.

Story Seven

The little baby girl was asleep in her pram. When the baby girl woke up she was very hungry. She started to cry (laugh).

Story Eight

Cats are very pretty animals. Kim has a pet cat. Kim gives it some fresh fish (fruit) to eat each day.

Story Nine

Peter was a boy who hardly ever cleaned his teeth. He woke up one morning with a bad tooth. His mother took him to the dentist (vet) straight away.

Story Ten

Shops are full of things that cost money. Shops have to be able to put out little fires. That is why shops keep buckets of sand (sawdust) to put them out.

Story Eleven

John has a new red sports car. John fitted his car with an alarm. He always switches it on (off) when he parks his car.

Story Twelve

We like to live in warm houses. Fires give off lots of heat. All houses with a coal fire have a chimney (garage).

EXPERIMENT ELEVEN

Passage Six

Sam and Bob had wanted to make this trip for years.

They wanted to row all of the river to the port of Troon.

They were now tired as they had been rowing for six hours.

But they had reached half way.

The river began to run fast down the hill.

They pulled the oars in and rested while the fast current carried them.

They were getting closer to the river's mouth.

They drifted into the river's mouth and started to row to Troon.

After half an hour they reached port.

Sam and Bob had ended their twelve hour journey.

QUESTIONS ASKED IN EXPERIMENTS ELEVEN A & B**Passage Two**

	Yes	No	Cat
1. Mr Smith has a big black dog called Sam.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2. It used to scare Tom as he walked past it to school.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3. But one day it was whining.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4. Tom went over and stroked it.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5. It licked him and each day from then on Tom stopped to play with it.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
6. On the way to school one morning, a big boy grabbed Tom and hit him.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7. He wanted Tom to give him money.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
8. Tom would not give him money.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
9. Suddenly Sam ran down the road barking.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
10. The big boy ran off and Tom was really pleased he has made friends with Sam.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
11. Mr Smith had a black cat called Sam.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> SI
12. Tom ran off and the big boy was really pleased he had made friends with Sam.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> SI
13. On the way to school one morning, a big boy grabbed Sam and hit him.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> SI

- | | | | |
|---|--------------------------|--------------------------|-------------------------------------|
| 14. Tom wanted to give him money. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 15. Tom went over and grabbed it. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 16. He demanded money from Tom. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 17. Suddenly Sam rushed at the big boy. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 18. It used to bark at Tom as he walked past it to school. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 19. The dog licked him and each day from then on Tom
stopped to pat the dog. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 20. But one day the dog was sad. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

Passage Four

	Yes	No	Cat
1. Jim rode his bike to school each day.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2. He had to work hard to get up the hill.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3. Then he took his feet off and cruised all the way to the gates.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4. On the way back, once he got to the top of the hill he could lift his feet up and cruise home.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5. One day a teacher was driving his new car out of the school.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
6. Jim came tearing down the hill.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7. He turned into the school gates.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
8. Jim smashed into the teacher's car.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
9. The teacher was angry.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
10. Jim had to pay for all the damage he had caused.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
11. The teacher rode his bike to school each day.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> SI
12. Jim came tearing up the hill.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> SI
13. Jim smashed into the teacher's bike.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> SI
14. Then he put his feet on and pedalled all the way to the gates.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> SI

- | | | | |
|--|--------------------------|--------------------------|-------------------------------------|
| 15. The teacher was driving his new car into the school. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox" value="SI"/> |
| 16. Jim always rode his bike to school. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox" value="SC"/> |
| 17. It was hard work getting up the hill. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox" value="SC"/> |
| 18. Jim's bike smashed against the teacher's car. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox" value="SC"/> |
| 19. Jim had to pay for the damage to the car. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox" value="SC"/> |
| 20. The teacher was angry at Jim. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox" value="SC"/> |

Passage Seven

	Yes	No	Cat
1. Fred was very fond of ice lollies.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2. He loved to eat them when he was out for the day.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3. At first he used to rip the wrapper off.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4. However he did not like the way small bits of paper stuck to it.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5. Then his father showed him how to blow into the wrapper to loosen it.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
6. It then came out with no paper stuck to it.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7. But the lollies he liked the most were the ones in a paper cone.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
8. He always left the last bit.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
9. He held the cone in his hands till the ice melted.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
10. Then he drank it.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
11. Fred was very fond of sweets.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> S1
12. Then his father told him to blow on the wrapper to loosen it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> S1
13. The lollies he liked the most were strawberry splits.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> S1
14. He did not mind if the last bit melted on the stick.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> S1

- | | | | |
|--|--------------------------|--------------------------|-------------------------------------|
| 15. Fred's mother was fond of ice lollies. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 16. Fred loved eating ice lollies at the seaside. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 17. Fred's dad blew into the wrapper to loosen his
ice lolly. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 18. Fred most liked the cones of frozen juice. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 19. Fred didn't like paper stuck to his lolly. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 20. He liked to drink the juice from the cones. | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

**A REVIEW OF EMPIRICAL STUDIES OF TRAINING TO
ENHANCE COMPREHENSION OF TEXT:**

Introduction

There have been a large number of studies of training programmes which have aimed to enhance children's comprehension of prose. These training programmes have taken many forms, but in order to help structure this review, the approaches used will be considered in three groupings: indirect techniques, direct training in reading comprehension and direct training in reading comprehension with a metacognitive component. This classification is primarily for organisational convenience and no great claims are made that the division of training studies is entirely clear and objective. The classification was made using the following criteria. Indirect techniques were considered to be those which did not primarily involve explicit instruction in comprehension of text but instead, emphasised language training, accessing prior knowledge or organisational strategies. That is they emphasised a process that was perceived to underpin comprehension or which if applied would enhance comprehension. Direct training in techniques of reading comprehension emphasised ways of accessing the meaning of the passage(s) presented often involving focusing on a product of comprehension e.g. a summary. Those direct training techniques which made explicit reference to or gave emphasis to the metacognitive process of comprehension monitoring were included in the third category.

Again, a large number of measures have been used to assess subjects comprehension of text. This review will concentrate only on these

established measures:

- (a) Responses to oral or written questions on the text
- (b) Cloze exercises
- (c) Adequacy of a summary made of the passage
- (d) Error detection

This is not to suggest that these four measures are somehow pure measures of comprehension. Far from it. All these measures, and likely all possible measures, are in some way confounded by motivational and attentional factors and general language competence. However, these four measures are often used in formal exam or experimental assessments of comprehension of text and so for practical purposes, performance on these measures is a measure of reading comprehension. One common measure of comprehension; single sentence completion has not been included. This is because the measures of comprehension of interest are measures of comprehension of text comprising a number of linked sentences. Completion measures related to text are subsumed under either cloze procedure or responses to oral or written questions which could be multiple choice.

It could be argued that by limiting this review to training studies which measured effectiveness by enhanced performance on any one or combination of four measures, this may lead to programmes having an effect because the students were trained in the technique used to assess comprehension. Carver (1987) referred to this as the Practice Principle. This can of course, be taken into account when assessing studies and the advantage of using multiple measures of outcome will be touched on in the discussion of the conclusions which can be drawn from the training studies reviewed.

One problem to be immediately acknowledged is that any review of the literature is at some level subjective. One way in which the biases an individual reviewer brings to the literature can be minimised is by making the selection of studies as objective and as wide ranging as possible, and by both establishing clear criteria for including a study and by making clear the pool of studies from which those selected come.

The pool of studies from which those meeting the criteria specified were selected comprised:

1. All the studies already referred to in preceding chapters.
2. All the studies evaluated in the reviews of Carver (1987), Lysynchuk, d'Ailly, Smith and Cake (1989), and Paris, Wasik and Van der Westhuizen (1988), who have carried out similar exercises.
3. Any studies published from mid-1988 (the latest studies covered by Lysynchuk et al 1989) to 1990 in these journals:

Journal of Reading Behaviour
 Reading Research Quarterly
 Educational Psychologist
 Journal of Educational Psychology
 Cognition and Instruction
 American Educational Research Journal

which were the core journals on which Lysynchuk et al based their review. From this pool of studies, all of those training studies which used at least one of the four measures of comprehension of text outlined above were reviewed.

The subjects who participated in these studies ranged in age from early

readers to mature college students. It is not being argued that a technique taught to pupils of age seven would have the same value when taught to college students or vice versa. However some research findings may have a general application and as this review is general rather than focused age of subjects was not used to restrict those studies considered.

Indirect Techniques:

Weaver (1979) investigated the possibility of improving reading comprehension by training “sentence organisation skills” which were felt to be those skills which enabled the reader to process sentence information in units larger than the single word. Thirty-one third grade experimental students individually received training in sentence anagrams for approximately 15 minutes, three times a week until they could solve sentence anagrams five words longer than those with which training began. Generally, students completed between six and ten sentence anagrams in a session. The number of sessions ranged from 14 to 21. Students were taught to use a word-grouping strategy that was designed to induce them to arrange words systematically into phrases and then to arrange the phrases into sentences. Whilst the experimental students were receiving their training, the control students remained in their classroom and received no treatment. The experimental students did better than the controls on the post test cloze exercise, but not on the comprehension subtest of the Metropolitan Achievement Tests. However, as the control group received no training other than regular reading instruction, it was not possible to tell whether the effects were due to general syntax sensitisation or the specific training. White, Pascarella and Pflaum (1981) carried out a replication using 9 to 12 year old learning disabled students. One group received training on materials like Weaver's for 21 half hour lessons in groups of three or four

over eight weeks. Another group received training in sentence patterning tasks e.g., identifying sentences as statements, questions or commands, matching noun and verb phrases or replacing nouns with pronouns. The children who received instructions using the anagram materials developed by Weaver did significantly better than the students receiving the more traditional sentence patterning tasks on the post-test cloze exercise.

Winograd and Johnston (1982) investigated whether giving readers assistance in selecting the appropriate schema for understanding a passage would improve comprehension. Their subjects were 20 sixth grade students who had to read two passages and detect errors in them. Half the students on retesting a week later were given assistance to select the appropriate schema by being shown a picture of a circus or church and asked to imagine everything two children might see at the circus/church and tell it to the experimenter. They were then asked to read and detect by oral response, errors in two passages about a circus or church. Schema training was found to have no significant effect on error detection.

Prater and Terry (1988), looked at the effect of mapping strategies on the reading comprehension of fifth graders in a series of experiments. Semantic mapping is a teaching strategy involving activating, assessing and embellishing students' prior knowledge of a topic before reading about it. In their first experiment, thirty fifth-grade students in two intact average reading ability groups were used. One group mapped their prior knowledge of the subject area of the first story before silently reading the story and then the next day, confirming accurate information on the map, eliminating inaccurate information and adding additional information. On the third day, these students wrote a short summary of the story and on the fourth day, took a ten item comprehension test on the story. This was repeated six

times for six separate stories. The control group had new vocabulary introduced and read the story on the first day, discussed the story on the second day, wrote a summary on the third day and took the comprehension test on the fourth day. On each of the six comprehension tests, the treatment group did significantly better.

In their second experiment, 80 students were randomly assigned to treatment or control conditions. The treatment groups followed the same process as previously described except that there were only three stories, nine questions in each comprehension test and they were asked to write a journal entry about the story rather than summarise it. Again, statistically significant differences in performance on the comprehension tests were found favouring the treatment group.

In their third experiment carried out two weeks after experiment two and using the same subjects and teachers, mapping was only carried out after reading three stories. Again, there was a nine item comprehension test but this time, students were asked to summarise the stories. This time, no significant difference was found between the treatment and control groups. This may suggest either that post-reading mapping is not effective in enhancing comprehension and/or that three sessions are not sufficient for students to internalise this strategy and utilise it without prompting.

Garner et al (1984) randomly assigned 24 children aged between 9½ to 13½ to either a treatment group where they were taught to look back at text to answer questions or to a control group. The 12 experimental subjects were taught in three 20 minute lessons why they should look back, when they should and where they should with accompanying questions. The controls each received three 20 minute lessons individually or in pairs on text-

processing strategies other than look backs. On a post training comprehension test there was no difference between trained students and controls on correct answers provided by recall alone, but the trained students correctly answered significantly more questions with the use of look backs. Garner at al felt this simple strategy was not used more often because students felt it was “illegal”.

Raphael and Wonnacott (1985) and Raphael and Pearson (1985) conducted training studies on fourth and sixth grade students to investigate whether training in categorizing questions with respect to whether the answers were textually implicit, textually explicit or scripturally implicit (required background knowledge on the part of the reader) improved comprehension. In both studies, significant effects were only found for their low ability groups. Training did not however help the lower ability students in scripturally implicit questions the ability to answer which is related to the students knowledge base.

McCormick (1989) using four classes of fifth grade students got them to read two passages without a preview and two passages with a preview consisting of the teacher asking questions, the reading of a synopsis and vocabulary instruction. The order of presentation of the passages and the condition under which they were administered was counterbalanced across the classes. Following the reading of each passage, a 12 item multiple choice comprehension test was administered. Students scored significantly better on the comprehension test under the preview condition. This may indicate a positive effect on comprehension of previewing a passage, but as there was no control for amount of time spent across conditions a plausible alternative explanation could be that the students in the preview condition performed better as they spent in total more time on the passage.

Table C1. Studies using indirect training of reading comprehension which have resulted in positive outcomes

Outcome Measure				
Training Method	Comprehension Questions	Cloze	Summary	Error Detection
Sentence anagrams		Weaver (1979) White et al (1981)		
Mapping	Prater & Terry (1988)			
Look backs	Garner et al (1984)			
Question categorizing	Raphael & Wonnacott (1985) Raphael & Pearson (1985)			
Preview	McCormick (1989)			

Direct training in techniques of reading comprehension:

Schunk and Rice (1987) investigated how providing 40 remedial readers ranging in age from nine years seven months to 13 years two months information that strategy use improves performance influenced their comprehension skill. The children were pre and post tested on 20 orally administered comprehension questions on eight passages selected from 'Scoring High in Reading'. Following pre-test, the children were randomly assigned within sex, grade level and school to one of four experimental conditions. All the children received instruction in comprehension of the main idea in passages. However, one group received specific guidance as to how useful their instructions were for the purpose in hand, one group guidance as to how generally useful it was, one group both types of guidance and one group no guidance. Each child received a 35 minute training session on each of 15 consecutive school days in groups of five to six with one or two adult female trainers. Significant differences in post-test scores were found between those children who received both types of guidance and those in the three other conditions. Schunk and Rice then went on to look at the effects of giving 30 students of mean age 11 years either performance feedback e.g., that's correct, specific feedback as to how effectively they were using the strategy being trained (comprehension of main ideas), or both on the comprehension measure outlined above. Those children receiving either performance feedback or both performance and specific feedback did significantly better on the post test with those receiving both doing the best on the comprehension test items.

Gambrell, Pfeiffer and Wilson (1985), investigated the effects of retelling by free recall upon comprehension of the text. Their subjects were 93 fourth grade students attending one of four elementary schools in North Carolina.

The subjects were assigned randomly to one of two treatment conditions: retelling or illustrating. Subjects participated in four training sessions and one test session. For each of the four training sessions, subjects silently read a passage of about 240 words in small groups of six to eight. Those in the retelling condition retold the important parts of the passage by completing a blank outline asking them for the important idea and some supporting details. Though initially assisted with this by the teacher, the amount of assistance was reduced over the four training sessions. After the silent reading and completing the outline, each of the subjects retold the story. The subjects in the illustrating group read silently, completed the outline and were then asked to illustrate all the important ideas from the story. For the text passage, all students were individually tested two days later answering ten literal and ten inferential questions about the test passage. The students who received practice in retelling, answered the comprehension questions significantly better.

Taylor (1984) investigated the relative effects of reading or writing a prose or diagrammatic summary upon the comprehension of expository prose on 125 first and second year college students. There were 94 females and 31 males ranging in age from 17 to 51, with 48% being under 21. All subjects were tested on the 'Degrees of Reading Power Test'. Their average score was 64 and the reading materials used were chosen to be at that reading level. Five passages of about 400 words in length were selected from text books. Prose and diagrammatic studies were prepared for each passage.

Each student had to:

1. Read a passage and write a diagrammatic summary
2. Read a passage and write a prose summary of no more than 180 words
3. Read a passage and its diagrammatic summary
4. Read a passage and its prose summary
5. Read a passage only.

So each student read each passage under different conditions. Order of presentation was counterbalanced. The students had to answer ten multiple choice comprehension questions on each passage. A significant task/passage interaction was found which makes it hard to evaluate the results, however in general writing a summary was more effective than reading a summary and prose summaries were more effective than diagrammatic summaries, but it was content dependent.

Armbruster, Anderson and Osterlag (1987) investigated how giving fifth-grade students instruction in conventional expository text structure including instruction on summarising would affect their ability to comprehend expository text of conventional structure. Forty-one students received this structure training for 45 minutes a day over 11 consecutive days. The forty-one control students were given conventional instruction. It has to be noted the experimental and control students were in pre-existing classes. The control and experimental students were asked to answer an essay question on a passage, to summarise a 200 word passage in 50 or less words, to give answers to a ten item comprehension test and to write six days after instruction, another essay. In line with the criteria specified, our concern is with the experimental groups performance on the summary and short answer

comprehension test. There was no significant difference between subjects and controls on their performance on the short-answer comprehension test. There were too many interaction effects for there to be any unambiguous interpretation of the results of the analysis of the summarizing task.

Baumann (1984) investigated the effectiveness of a direct instruction model for teaching children the comprehension skill of identifying the main idea. Sixty-six, sixth grade students blocked into high, middle and low achievement level using their performance on the Metropolitan Achievement Test were randomly assigned from within these levels to one of three experimental groups. These consisted of a strategy group in which subjects received intensive main idea instruction by direct instruction; a basal group in which subjects had lessons on main idea comprehension from a popular basal series; or a control group in which subjects engaged in unrelated vocabulary development exercises. All the groups had eight lessons of 30 minutes duration. The outcome was evaluated using three multiple choice tests, a summary task and a written free recall task. The group given direct instruction in strategies for identifying the main idea did significantly better than the basal reader or control group in all the outcome measures except the free-recall task.

Bean and Steenwyk (1984) compared the effectiveness of three forms of instruction in writing summaries using sixty sixth-grade students in three classes from a suburban district in southern California. The students were randomly assigned in their class groups to one of three training groups. Each group met for 12 instructional periods of 25 to 30 minutes over five weeks. They were all taught by Fern Steenwyk. One group received direct

instruction in six summarisation rules:

1. Delete unnecessary material
2. Delete redundant material
3. Compose a word to replace a list of items
4. Compose a word to replace the individual parts of an action
5. Select a topic sentence
6. Invent a topic sentence if one is not available.

The second group received training in a more intuitive strategy devised by Cunningham (1982) and called GIST (Generating Interactions between Schemata and Text). In this group, students were taught to compose 15 word summaries first of one sentence then adding sentences until they could compose a single 15 word summary of the whole passage.

The control group received the same length of tuition, but were simply advised to write summaries by finding the main idea of the paragraph. No explicit step-by-step instruction was offered.

The post test was asking the students to summarise a paragraph in fifteen or fewer words and completing Form B of the reading comprehension Nelson Reading Test. No difference was found between the groups receiving direct instruction in summarisation on either the summary or reading comprehension measure. But both these groups did significantly better than the control group on both measures. Some caution must be used in interpreting this result as whole classes were used as the unit of randomisation which leaves open the possibility of a class/instruction interaction and the instructor was also the experimenter and so was not 'blind' to the comparisons being made.

Rinehart, Stahl and Erickson (1986) gave sixth grade students 45 to 50 minutes training in summarising on each of five consecutive days. The students were taught to: identify the main information, delete trivial information, delete redundant information and to relate main information and important supporting information. The students given the training, produced better summaries than controls who carried out basal reader work. There was however, a significant interaction between the passage used and the performance of the subjects. It also emerged from this study that training in summarising enhanced recall of major information but not minor information.

In contrast, Taylor (1982) initially found with fifth grade students given one hour a week training in summarising that they were no better than controls at giving short answers to questions given one day after summarising a passage. In a replication of this study, Taylor then found fifth grade students given conventional instruction which involved them independently completing short answer questions on a passage rather than summarising it actually did significantly better on the post-test short answer questions than those given training in making summaries. In a further similar experiment with seventh grade students, Taylor and Beach (1984) did not find any significant difference between students taught to summarise and those given conventional instruction in terms of their ability to give short answers to questions a day after having studied a passage. It could well be argued that this is as much a measure of recall as comprehension of text and so can only throw a tangential light on the impact of training in summarising on ability to give correct short answers on a passage.

Bereiter and Bird (1985) randomly assigned 20 students from grades seven

and eight to one of four groups consisting of three treatment and one control group. The three treatment groups received nine 40 minute lessons over three weeks. Group One's instruction consisted of explanation and modelling of various reading strategies, practice at identifying the use of different strategies and practise in using the strategies. Group Two's instruction consisted of the instructor modelling the strategies and the students practising their use. Group Three's instruction consisted of the experimenter directly asking the student to use a particular strategy at a particular juncture, both orally and in writing. Pre and post-testing took place using the Spache Reading Scales and the Nelson Reading Skills Test (Form C). On the post-tests, all the groups including the control group, showed large gains on the Spache Scales. However, the students in Group One where the teachers both modelled and gave explicit instruction in strategies did significantly better. Rauenbusch and Bereiter (1991) also trained seventh grade students on reading strategies using text degraded by deleting every third letter and replacing it with a dash. Both the strategy training group and their controls received six 40 minute training sessions. The strategy training group did significantly better on the Nelson-Denny Reading Rest.

Carr, Dewitz and Patberg (1983) randomly assigned classes of sixth grade pupils to one of three groups. One treatment group received instruction only in the cloze technique which included advice on a self-modelling of responses. The other treatment group received a combination of a structured overview to activate background knowledge and the cloze technique. There was also a control group. Instruction took place over eight weeks in forty minute social studies class periods. Analysis indicated that cloze groups did significantly better on the literal comprehension questions used as a post-test than the control group and they also did significantly better on the textually

implicit comprehension questions also used as a post-test.

Dewitz, Carr and Patberg (1987) elaborated on their earlier work by assigning randomly three of four classes of fifth grade pupils to one of three treatment conditions. The fourth class where due to illness in the family the teacher expressed a desire for a less demanding task was assigned the structured overview instruction. The remaining three groups were the control group, a group instructed in cloze procedure and a group that received both cloze instruction and a structured overview. Instruction took place over six weeks (the abstract) or eight weeks (page 107) during 40 minute social studies class periods. On the post-tests of reading comprehension and metacognitive awareness, the groups receiving cloze instruction did significantly better. This was also the case on tests of inferential and literal comprehension taken some six months after instruction ceased (page 108) or four months (page 110) or 20 weeks (the abstract). The results then are in line with their previous study, but confidence in them has to be undermined by the lack of precision in the detail of the reporting, the use of assignment by existing class and the failure to control for the plausible hypothesis that the teacher of the structured overview condition may well have been concerned more about family matters than teaching so potentially depressing the results of the pupils in this group.

Linden and Wittrock (1981) investigated whether teaching ten year old Hispanic children to attend to text and to generate metaphors, summaries or illustrations of the text enhanced their comprehension. The children were taught to do this in three one hour sessions over three consecutive days. Post-testing using multiple -choice and completion tests of reading comprehension provided significant differences in favour of the treatment groups over controls on the comprehension measures but not on the fact

retention measures.

Carnine and Kinder (1985) compared the effects of training 27 fourth to sixth grade students to apply generative or schema strategies when reading. The generative strategy training involved instructing students to close their eyes and make a picture after each meaningful unit of prose. Individual students were asked to describe their image. At the end of the story, each subject gave a summary. The teacher modelled correct answers where necessary. In schema training on narrative stories, the students listened to or read passages at various points in which four questions were asked by the teacher: What is the story about? What does he or she want to do? What happens when he or she tries to do it? What happens in the end? In the nine expository passages, the students were told to look for the rule in the story. All students were exposed to 39 passages of 500 to 600 words in length of about third to fourth grade reading level. They were trained on these by a teacher working with the same three to five students three to four days a week for 20 to 30 minutes each day. Pre-testing, daily testing and follow up testing two weeks after the training, took place using 100 word passages on each of which five questions were asked. No significant differences were found between the two interventions. Not surprisingly however, the students performed better after the training.

Table C2: Studies using direct training of reading comprehension which produced positive outcomes

Outcome Measure					
Training Method	Comprehension Questions	Cloze	Summary	Error Detection	
Guidance in strategy use	Schunk & Rice (1987)				
Feedback about results	Schunk & Rice (1987)				
Re-telling	Gambrell et al (1985)				
Identifying main idea	Baumann (1984)		Baumann (1984)		
Writing Summaries			Bean & Steenwyk (1984) Rinehart et al (1986)		
Cloze	Carr et al (1983)				
Schema training	Carnine & Kinder (1985)				
Strategy training	Rauenbusch & Bereiter (1991)				

Direct training with a metacognitive component:

Paris and his co-workers (Paris, Lipson and Wixson 1983; Paris, Cross and Lipson 1984; Paris and Jacobs 1984; Paris 1986; Paris and Oka 1986; Cross and Paris 1988), have developed and evaluated an instructional programme called 'Informed Strategies for Learning', designed to increase young children's understanding of reading tasks, goals and strategies. The emphasis is both on providing reading strategies and increasing children's metacognitive knowledge about reading. Paris (1986) also emphasises that strategic readers combine knowledge about the task with the motivation to act accordingly.

The programme consisted of 20 modules designed for Grades 3, 4 and 5 (Paris 1986). Each module emphasises one comprehension strategy and includes three half-hour lessons. Paris and his co-workers created metaphors to help make each reading strategy more concrete. The modules were arranged in four groups and are listed below:

(i) Awareness of reading goals, plans and strategies

1. Goals and purposes of reading
'hunting for reading treasure'
2. Evaluating the reading task
'be a reading detective'
3. Comprehension strategies
'a bag full of tricks for reading'
4. Forming plans
'planning to build meaning'
5. Review

(ii) Components of meaning in text

6. Kinds of meaning and text content
'turn on the meaning'
7. Ambiguity and multiple meanings
'hidden meaning'
8. Temporal and casual sequences
'links in the chain of events'
9. Clues to meaning
'tracking down the main idea'
10. Review
- (iii) Constructive comprehension skills**
11. Making inferences
'weaving ideas'
12. Preview and review of goals and task
'surveying the land of reading'
13. Integrating ideas and using context
'bridges to meaning'
14. Critical reading
'judge your reading'
15. Review
- (iv) Strategies for monitoring and improving comprehension**
16. Comprehension monitoring
'road signs for reading'
17. Detecting comprehension failures
'road to reading disaster'
18. Self-correction
'road repair'
19. Text schemas and summaries
'round-up your ideas'
20. Review
'Plan your reading trip'

Informed Strategies for Learning is not linked to any one set of curricular materials or any one method of reading instruction or reading scheme. Paris (1986) suggested that many of the skills addressed might be best taught during social studies or science lessons.

Paris, Cross and Lipson (1984) evaluated Informed Strategies for Learning (ISL) on 87 subjects in the third grade (mean age eight years five months) and 83 subjects in the fifth grade (mean age ten years five months). The subjects made up eight intact classes. Two third grade and two fifth grade classes received two 30 minute lessons a week over four months of ISL while the remaining four classes acted as controls. This design does not eliminate there being a group-instruction interaction that may invalidate the outcome. This weakness was compounded by experimental and control classes being in separate schools so that it is not possible to assume there are no school effects. Pre and post-testing was carried out using comprehension tests (the Gates-MacGinitie and Paragraph Reading sub-test of the Tests of Reading Comprehension), a cloze procedure and an error-detection task in which the children had to read two passages and underline those words or sentences which did not make sense. The versions of ISL used for instruction was a less well developed version than that described previously and consisted of 14 modules. No significant differences between treatment and control groups were found on post-testing for either of the two comprehension tests. However, significant differences favouring the treatment group were found on post-testing on the cloze and error detection tasks.

Paris and Jacobs (1984) evaluated ISL on 91 third grade subjects (mean age eight years five months) and 92 fifth grade subjects (mean age ten years five months). All the children were members of eight intact classes in four

schools. Separate schools were chosen for experimental and control classes so again, there was a confounding of school and class with treatment. Pre and post-testing was carried out on a standardised comprehension test (the Gates-MacGinitie Reading Test) a cloze task and an error detection task in which children read two passages and underlined those words or sentences that did not make sense. The experimental group received two short lessons of about half an hour each week for 14 weeks of ISL. The post-test results were analysed using ANCOVA. It is inappropriate to use this with pre-existing groups. That aside, no significant differences were found between control and treatment groups on the Gates-MacGinitie comprehension test. However, the treatment group did significantly better than the control group on the post test cloze and error detection tasks.

Paris and Oka (1986) evaluated the implementation of the ISL programme with 800 third and 800 fifth grade children from 28 schools in Detroit. Pre and post-testing was carried out using the Gates-MacGinitie Reading Test comprehension sub-test, a cloze task and an error detection task. The programme involved the experimental groups receiving three half-hour class lessons on each of the 20 modules listed earlier. As for the previous two studies, there was no significant enhancement of the treatment groups performance on the Gates-MacGinitie Test but the treatment group did significantly better on the cloze and error-detection tasks on post-testing.

Cross and Paris (1988), evaluated ISL using 87 third graders (mean age eight years five months) and 84 fifth graders (mean age ten years five months) from eight intact classes. Two third and two fifth grade classes received ISL, the remaining four classes acted as controls. Separate schools were chosen for experimental and control classes so all the caveats made previously apply. Pre-testing was again with the Gates-MacGinitie

comprehension subtest; a cloze task and an error detection task. The experimental students received two thirty minute lessons a week in three phases of five to six weeks covering the 14 modules making up the cut-down version of ISL. The results of the post-testing was analysed by multivariate cluster analysis. Amongst 29 third graders who received ISL twelve achieved a good combined score on the three pre-test measures noted above but 21 achieved good scores after ISL. These group sizes remained unchanged for the controls. Amongst fifth graders, 20 achieved good or average combined scores on the three measures quoted on pre-testing but this rose to 28 on post-testing after they had received ISL. Amongst fifth grade controls, 18 achieved good or average combined reading comprehension scores on pre-testing and this rose to 19 on post testing. Jacobs and Paris (1987) evaluated ISL on 783 third grade and 801 fifth grade students from intact classrooms. Amongst pre-treatment measures was the Gates-MacGinitie Reading Test. The treatment group received all 20 modules each module consisting of three half-hour lessons presented to the entire class. The post-test included Form 2 of the Gates-MacGinitie Reading Test. No analysis of the pre and post-test results are made and one can only assume that no significant differences were found. This would be consistent with other evaluations of ISL. In this paper, they claim that standardised reading achievement scores “are so highly correlated with intelligence and test-taking skills that they may not be sensitive to specific knowledge about reading nor specific reading strategies.” Forrest-Pressley, Waller and Pressley (1989) reject this argument. First they note that skilled readers differ in both general intelligence and other skill factors so 'such confoundings with reading ability are natural.' Secondly, they argue Paris and his associates have provided 'no data that illuminate whether the measures they prefer are less loaded on general intelligence and non reading skill factors.' They also note that the two measures preferred by Paris and

his group, cloze and error detection, may well be mediated by short-term memory and general language competence. They conclude by suggesting one interpretation of Jacobs and Paris's 1987, motivation for favouring these non-standard measures of reading comprehension is that informed Strategies for Learning has proven effective at enhancing error detection and performance on cloze exercises, but not with the comprehension subtest of the Gates-MacGinitie! However it could be argued that error detection and cloze measures the reader's construction of meaning whilst multiple choice comprehension tests such as the Gates-MacGinitie measure more the reader's location of literal information within a text.

Overall, it would appear ISL does enhance students performance on cloze and error detection measures of comprehension after about 28 spaced half-hour lessons. It is unsuccessful in increasing eight to 11 year olds performance on standardised comprehension tests.

Chan and Cole (1986) used a robot to help train 36 learning disabled students of mean age 11 years one month and 36 mainstream pupils of mean age eight years three months in a number of reading strategies. The students were trained in groups of four or five over four half-hour daily sessions to: generate questions about the content of the passages they read, or underline two interesting words in the passage, or both generate questions and underline two interesting words or to read the passage to the robot twice. After each session, the children had to answer multiple choice questions on the passage they had read. For the first and second sessions only, they were given the correct answers after they had completed the questions. There was a differential effect of training on the two groups: The mainstream children did best under the reading twice condition while the learning disabled students did best when asked to underline two interesting

words either with or without self-questioning. That the mainstream students did best under the read twice condition may simply be a reflection of the version of the total time hypothesis advanced by Carver (1987) the Reading Time Principle which is that students can improve the degree to which they comprehend a passage by spending more time reading the passage. This however, did not seem to hold for the learning disabled students in this study.

Short and Ryan (1984) investigated the effects of story grammar training designed to increase comprehension monitoring and attribution training designed to increase awareness of the effort required for efficient reading amongst 42 fourth grade poor readers. These were split into three groups. One received strategy training only. This consisted of prompting them to use and vocalise 'wh' questions:

- i) Who is the main character?
- ii) Where and when did the story take place?
- iii) What did the main character do?
- iv) How did the story end?
- v) How did the main character feel?

One group received attribution training in which they were reminded of the importance of effort in successful reading performance and the third group received both. Pre-testing of the children was carried out in which they were asked 14 short answer questions. Fourteen skilled readers acted as controls. The three groups of poor readers received three half-hour sessions of instruction. The group trained in the five 'wh' questions did significantly better on the comprehension questions than those given attribution training only. Attribution training did not significantly enhance performance on the

comprehension task. The strategy only and combined groups performance on this task was indistinguishable from the skilled readers performance.

Anne Marie Palincsar and Ann Brown have developed a novel method of instructing students in reading for meaning called reciprocal teaching. Palincsar and Brown (1986) felt reciprocal teaching was 'best characterised as a dialogue in which the students and teacher work together to comprehend text' (Palincsar and Brown 1986 p776). The dialogue begins with trying to activate the relevant background knowledge the students already possess. It is 'structured by the use of four comprehension fostering and comprehension monitoring strategies: predicting, questioning, summarising and clarifying . The teacher balances the use of explanation, instruction and modelling with guided practice, so that there is a gradual transfer of responsibility for sustaining the dialogue...' (Palincsar and Brown 1986 p776-777). Eventually, adult and student take turns at being the teacher and so responsible for leading the dialogue. The teaching takes place in small group settings. The evidence for the effectiveness of reciprocal teaching will be considered next.

Palincsar and Brown (1984) assessed the effects of reciprocal teaching using 24 seventh grade students who had problems with reading comprehension in that they scored two years below grade level on a standardised reading comprehension test but were able to read grade appropriate texts at a rate of at least 80 words per minute with two or fewer errors. Initially, six students were assigned to the reciprocal teaching condition and six to an untreated control group. Six months later, six more were assigned an alternative intervention; locating information and six had practice in daily assessment, but no intervention. In total, there were four groups of six receiving; reciprocal teaching, locating information, test only and control. In the

reciprocal teaching groups in pairs under the supervision of a teacher, the students took turns to ask each other questions, summarise and offer predictions. They had approximately 20 sessions, one a day lasting about 30 minutes, centred on 13 passages of about 1500 words. Additionally, 45 shorter (400-475 words) assessment passages on each of which ten questions were set taken from the same sources were produced. Pre and post testing took place using a summarising task where the students deleted sections from a passage they considered trivial or redundant, wrote superordinate terms for any lists, and underlined or wrote in a topic sentence and an error detection task where the students made timed responses to whether a sentence in a passage presented on a VDU made sense. Comprehension was also assessed pre-intervention, in the first half and the second half of training and post-intervention by students responses to ten questions on a new passage.

The reciprocal teaching group achieved about 80% correct levels of response on the comprehension questions post-intervention as compared with about 45% for the other three groups. This enhanced performance was maintained on testing eight weeks after the intervention had ceased. The reciprocal teaching group also did significantly better than the control group on the summarising task on post-testing. The reciprocal teaching group also did significantly better than the control group on post-testing on the error detection task. The Gates-MacGinitie test was also re-administered three months after the termination of the intervention and on average the reciprocal teaching group had improved by 15 months. This is an impressive gain though its importance has to be qualified by the absence of data from the control group.

Palincsar and Brown replicated this study using four pre-existing groups of

sixth, seventh or eighth grade readers taught by their own teachers. The same procedure was followed for these 21 students except they were taught in groups of four, five or seven students. The students started at a baseline of 40% correct on the comprehension questions and ended at a level of 80% correct. They also improved significantly on the summarising task and on the error detection task. The teachers only received three training sessions involving an introduction, then a practice session with Palincsar and finally a practice session with Palincsar and some seventh grade students. Each teacher also received a weekly visit from Palincsar while the intervention was taking place.

Carver (1987) has suggested an alternative explanation for the results obtained by Palincsar and Brown (1984), which is that students who are taught to spend large amounts of time reading difficult material can achieve the same level of comprehension as average students who spend lesser amounts of time reading. Carver also argues that the increase in score of the reciprocal teaching group on the Gates-MacGinitie test of reading comprehension could be explained by regression to the mean as they had been selected at the outset only if they scored at least two years below grade level on the basis of pretest scores on this test.

Gilroy and Moore (1988) reported a replication of Palincsar and Brown's (1984) study. The subjects were three nine and ten year old girls, three 11 year old girls and four 12 to 13 year old girls attending a New Zealand urban school. They were each able to mechanically read up to their chronological age but were about two years behind their age on the 'Progressive Achievement Test in Reading Comprehension.' Nine girls of the same age whose comprehension scores, were above the 85th centile on the test of reading comprehension acted as controls. A multiple testing model was

followed with a baseline period, 21 days of 20-25 minute lessons of reciprocal teaching for the experimental group, a maintenance phase three weeks after completion of the intervention lasting five consecutive days and a follow-up session over three school days held eight weeks after the maintenance phase. All students answered ten questions on a short passage each day of intervention.

The results were analysed separately for each age band. Generally, the experimental group made a significant improvement in their comprehension test performance though at least one of the above average control groups also improved significantly. The improved performance tended to be maintained over the maintenance and follow-up periods so that over the baseline the average score for the experimental group was 26.3% correct, over the maintenance period it was 69% correct and over the follow-up period it was 76.6% correct. Post-intervention testing on the 'Progressive Achievement Test' also suggested that the experimental students generally made measurable gains. As this is a timed comprehension test it addresses some of the concerns of Carver (1987) that reciprocal teaching may influence comprehension performance by encouraging students to study texts for longer.

Palincsar, Brown and Martin (1987) also carried out an exploratory study in which seventh-grade remedial reading students acted as peer tutors using reciprocal reading techniques in dyads or triads. The nine tutors received ten days of training and the 15 tutees 12 days of reciprocal teaching. Significant differences between both the tutors and tutees pre and post test scores on ten comprehension questions were found over the 12 days of reciprocal teaching. Significant differences between both the tutors and tutees pre and post test scores on ten comprehension questions were found

over the 12 days of reciprocal teaching. It is however, not possible to know how much weight to give this as there was no control group.

In a British evaluation McLeod (1993) found no significant enhancement in performance over controls arising from the application of reciprocal teaching with a group of 13 to 14 year olds. However the reciprocal teaching was limited to between 7 to 9 sessions of about 30 minutes duration spread over a 5 week period. Moreover the students were probably less literate than those used by Palincsar and Brown (1984) or Gilroy and Moore (1988).

One aspect of reciprocal teaching of comprehension is questioning, and Davey and McBride (1986), provided some independent confirmation of the value of training in question generating. Using five groups of between 23 and 25 randomly assigned sixth grade students, they compared question training with a no question control, a question generating group, an inference-question practice group and a literal-question practice group. All five groups received five 40 minute lessons over two weeks. The question training group were trained to generate questions linking information across sentences and questions tapping the most important information. The question generating group were asked to generate two good questions for each passage which made them think about what they had read and could not be answered by underlining parts of the passage. Both the inference-question practice and the literal-question practice groups read the same three passages each session as the other groups, but had to answer four inference or four literal questions per passage per session respectively. On the post-test passage, the question-training and the question-generation practice groups out performed the other groups on the literal questions, and on the post-test inferential questions the question training group out performed all the other groups. This result would tend to support the validity of including

question generating activities in reciprocal teaching of comprehension. However, McDonald (1986) who also investigated the effects of training seventh and eight grade students to ask questions within the context of reading being described as a problem solving exercise, did not find any significant difference between control and trained subjects on comprehension measures and these students received about twice as much training as in the Davey and McBride (1986) study.

Nolte and Singer (1985) however, carried out a study which though designed to measure the effects of training question generation on comprehension of text has a number of parallels with the reciprocal teaching studies in that instruction was given initially in a group of 19 by teacher modelling, then in groups of five or six with a student chairing within the class, then in pairs and finally, individually and repeated comprehension tests were carried out daily on both controls and the experimental group. Training of the fourth and fifth graders took place over two successive school days for 40 minutes each day. Consistent differences in favour of the experimental group were found from the ninth day on. It is also interesting that some of the teacher modelling explicitly dealt with questions about what would happen next in the story. This study's instruction does however differ from reciprocal teaching in giving direct instruction in self-questioning.

Seretny and Dean (1986) investigated whether written inserted questions facilitated comprehension as measured by a norm-referenced test of reading comprehension. This involved small group instruction over four weeks in which after the children had orally read a page, the teacher then stopped the reading and asked and discussed the interspersed questions. Both average and below average second grade readers given training on interspersed post-

passage questions did significantly better on the comprehension subtest of the SRA Achievement Test Battery. Above average readers showed no improvement over controls but this may be attributable to both groups scoring near the ceiling of the test. Generally therefore, there appears to be support for the role of both oral and written questions on passages enhancing pupils performance on comprehension measures.

Miller (1985) investigated the degree to which self-instructional training enhanced the ability of students between eight and a half and ten and a half years to detect errors in prose. All the students received three 45 minute individual training sessions over a week. The type of training they received fell into four categories, either specific self-instruction, general self instruction, didactic instruction or control. Immediately, and three weeks after training, the students were asked to read nine short essays, six of which contained one between-sentence inconsistency which the students were asked to underline. Those students given general self instruction training did better than those given didactic training or controls on both the immediate and delayed test. Those given specific self-instruction training did better than those given didactic training or controls only on the delayed training, but there was no significant difference between those given either specific or general self-instruction training in terms of their overall performance. The self-instruction training used in this study has parallels with the clarifying and questioning components of reciprocal teaching.

Wong and Jones (1982), investigated the hypothesis that insufficient metacomprehension is one possible cause underlying learning disabled adolescents comprehension problems and that training them to monitor their understanding of important textual elements fosters metacomprehension and consequently improves their comprehension. They investigated this with 60

learning disabled students who scored averagely on the Wechsler Intelligence Scale for Children - Revised, but were three to four years behind in their reading as measured on the Nelson Reading Skills Test and 60 normally achieving sixth grade students. Half the subjects were randomly assigned to a control condition. The remaining half having been given training on identifying the main idea were taught to ask themselves five questions about the passage they were reading. This training only significantly enhanced the learning disabled students comprehension performance. Interestingly, Wong and Jones also recorded the length of time students took to complete assignments and those who received it took significantly longer, sometimes well over double the time to complete assignments than the untrained controls. There was no difference in study time between the trained learning disabled and the trained normally achieving sixth grade students, nor was there any difference between those two groups performance on the comprehension items. This may well be an example of Carver's Reading Time Principle (Carver 1987) that students can improve the degree to which they comprehend a passage by spending more time reading the passage.

Hansen (1981) investigated whether providing second grade children with a strategy for integrating text information and prior knowledge would enable them to perform better on an oral text comprehension test than students given only practice on inferential questions or basal reader instruction. Her results indicated no difference in results between those given strategy training and those given practice on inferential questions, though both did better than the control group. A similar picture also emerged from post-testing on the reading test of the Stanford Achievement Test, which contains a preponderance of cloze type items.

Hansen and Pearson (1983) investigated the effect of four classroom teachers providing instruction to improve the inferential comprehension of good and poor fourth grade readers. Initially, 40 students who were randomly selected from the available group, were assigned as good or poor readers on the basis of their score on the Stanford Achievement Test and then randomly assigned to control or treatment groups. Instruction of the students took place for 20 minutes, two days a week over ten weeks following a two week training period. The control group followed the suggestions in the teachers manual for the basic readers. The training of the experimental group focused on (a) making students aware of the importance of drawing inferences between new information and existing knowledge structures, (b) getting students to discuss prior to reading something that was similar to the events in the text and to hypothesize what would happen in the text and (c) providing students with six inferential questions to discuss after reading a text. Following discussion of each of these stories, each student completed ten written open-ended questions of which three were literal and three inferential. A 16 question oral comprehension test on a final passage was also given to all students as a post-test. The instruction in inferential thinking only significantly improved the scores of the poor readers on the inferential questions and in fact, the good control readers did better than the good experimental readers on the literal questions. On the common post-test passage, the experimental group did significantly better on the oral inferential questions. Overall, the instruction in inferential comprehension seemed to particularly benefit the initially poorer readers.

Berkowitz (1986) using four existing classes of sixth grade students, randomly assigned one class to each of four instructional procedures which took place for one 45 minute session each week for six weeks. The four

procedures were:

1. Map - construction procedure. In this, students were instructed how to construct a map on a blank sheet of paper outlining in words four to six main ideas from a passage.
2. Map - study group. Students received prepared maps of the material to study and discuss.
3. Question - answering procedure. Students wrote out answers to 20 questions on each weekly passage and were instructed in memorising the answers.
4. Re-reading procedure. Students were instructed to read each passage twice and the instructor modelled a review/study procedure for students to copy and practice.

Clearly the use of pre-existing groups makes comparison between the procedures dubious, though Berkowitz did find there to be no significant differences between the groups on their performance on the Gates-MacGinitie Reading Test administered two weeks before instruction commenced. Berkowitz also rotated instructors between the groups. Post-testing took place using two passages on which students had to answer 20 questions. The students were also tested on the second passage two weeks later as a measure of delayed recall and under a transfer condition where they were just asked to think about the strategies they had studied. Measures of free-recall were also taken. The results are difficult to interpret as the pattern of results was different under different test conditions. As a single passage was used for each measure (except delayed recall), a treatment/passage interaction may render the results spurious. For example there is some evidence that Passage Two was better suited to mapping than the other two passages. It was on Passage Two that the map-construction

treatment group did significantly better than other groups. Overall, the map study group did not do particularly well. The interpretation of the results is made more problematic because the question-answering group had the opportunity to answer half the test questions prior to testing so as not to penalize them “by requiring that they focus on material unrelated to the tests.” Taking into account the possible inflationary effect this would have on the results of the question-answering group, the rereading group appeared to perform satisfactorily in comparison.

Yuill and Oakhill (1988) trained Year Three pupils who were either skilled or less skilled at comprehension to look for clue words that would help them to understand a text and also trained them to generate questions on the text. There were two control groups who either completed comprehension exercises or had practice in rapid decoding. The trained less skilled pupils performed significantly better on post testing on the comprehension component of the Neale Analysis than those given training in rapid decoding. No other comparisons reached significance.

Yuill and Oakhill (1991) also found nine and ten year olds given training in learning to think in pictures as they read stories, performed better on comprehension questions than controls who spent the same amount of time with the experimenters but just answered questions on the training stories. Though there was an overall main treatment effect, only the initially less skilled comprehenders did better overall on imagery training; the difference for the initially more skilled comprehenders failing to reach significance.

Table C3: Studies using direct training with metacognition in reading comprehension which produced positive outcomes

Outcome Measure					
Training Method	Comprehension Questions	Cloze	Summary	Error Detection	
Informed strategies for learning		Paris et al (1984) Paris & Jacobs (1984) Paris & Oka (1986)		Paris et al (1984) Paris & Jacobs (1984) Paris & Oka (1986)	
Underlining	Chan & Cole (1986)				
Rereading	Chan & Cole (1986)				
“Wh” questions	Short & Ryan (1984)				
Reciprocal teaching	Palincsar & Brown (1984) Gilroy & McBride (1987)		Palincsar & Brown (1984)	Palincsar & Brown (1984)	
Question generating	Davey & McBride (1985) Nolte & Singer (1985)				
Inserted questions	Seretny & Dean (1986)				
Self instruction training				Miller (1985)	
Inferential comprehension training	Hansen & Pearson (1983)				
Inferential comprehension training and question generation	Yuill & Oakhill (1988)				
Imagery training	Yuill & Oakhill (1991)				

Conclusions:

At the beginning of this review, measures against which the effectiveness of different comprehension training procedures could be judged were selected. A large number of the training programmes reviewed enhanced the participants' performance on at least one of these four measures of comprehension of text. Whilst it is encouraging that so many different interventions appear to enhance students' performance on these measures of reading comprehension, it would also be surprising were reading comprehension to involve so many variables that over 20 different approaches could positively influence it each through a unique mechanism.

It may be that some of the results can be explained relatively simply through Carver's Practice Principle which is that 'students ordinarily improve on any reading-related task simply by practising on that task'. (Carver 1987 p 116). Thus, it is not surprising that Rinehart et al (1986) found students trained in summarising produced better summaries than controls. It would follow from this that one should use different measures of outcome in assessing the effectiveness of any training than those on which the subjects are trained. Confidence in a positive effect would be enhanced by also showing an effect on more than one measure.

It is also of note that, even relatively brief interventions such as the ten hours of reciprocal teaching in the study by Palincsar and Brown (1984) sometimes had a significant impact on performance. Also, where studies compared a number of interventions often no difference was detected in their outcomes. For example, Taylor and Beach (1984) found no difference between those taught to summarise and those given conventional instruction

in their ability to give short answers to questions. Carnine and Kinder (1985), found no statistically significant difference between generative and schema strategy training and Bean and Steenwyk (1984) found no statistically significant difference in outcome between two different types of training in writing a summary. Conversely, Berkowitz (1986) did find some differences between the interventions she investigated but the differences were not consistent and some may have been due to spurious factors which were uncontrolled for in her experiment.

If the conclusion that there is often no difference when direct interventions are compared drawn from this review is correct, one would anticipate a similar result emerging from other reviews of reading comprehension studies. Another substantive review is presented in the book edited by Fitzgerald (1990). This book contains a review of 369 studies of reading comprehension in the United States of America between 1910 and 1987. Table C4 summarises the results. Not all the studies were included because some of them provided no results, some results were unclear or mixed and some focused on the changing nature of reading comprehension under treatment rather than overall comprehension. The results are broadly supportive of the conclusion drawn from the studies reviewed earlier in that the great bulk of interventions had a positive outcome but when interventions were compared, just less than half significantly differed in their outcomes.

Table C4: Summary of Review of Studies in Reading Comprehension Instruction in Fitzgerald (1990)

Method	No. of studies reviewed in which outcomes are given	No. of positive outcomes	No. of studies comparing interventions	No. of comparative studies showing a significant difference
Exercises & drills	59	48	6	2
Teacher directed skill lessons	29	25	5	5
Rereading	3	3	9	2
Reading practice	10	9	3	1
Study skills	27	26	7	1
Instrument presentations	11	7	2	1
Word study	16	11	2	1
Oral presentations	12	9	1	1
Schema activation	74	60	41	19
Text structure	23	19	4	4
Monitoring	10	9		
Total	274	226	80	37

As there is evidence relatively brief interventions have significant effects, that where different interventions are compared with each other then often no intervention seems significantly better and that large numbers of different interventions have significant effects on comprehension one possible explanation is that these different interventions operate through a common mechanism.

There are a number of possible candidates for such a mechanism. For example, Tobias proposed that 'any arrangement that increases student's macro processing of instructional impact is likely to improve achievement' (Tobias 1982 p6) where macro processing refers to the frequency and intensity with which students cognitively process instructional input. Another possibility has been advanced by Carver (1987); the Reading Time Principle. Carver asserted that students can improve the degree to which they comprehend a passage by spending more time reading the passage and has developed (Carver 1977, 1990a) a complex series of prediction equations by which the percentage increase in comprehension can be estimated. If a relatively straightforward factor such as spending more time reading a passage does underlie a number of the interventions reviewed in this Appendix then this would be consistent both with relatively brief interventions having significant impact and it often being found that apparently dissimilar interventions have similar outcomes on students performance on measures of comprehension. One of the simplest ways to increase the amount of time a student spent reading a passage would be to ask them to reread it. However, the experimental work reported in Chapters Six and Eleven did not support rereading as enhancing comprehension, when the materials used were accessible to the readers.

Another possibility is that the training improves the subject's mechanical reading skills. There is at least one area in which increases in reading accuracy has led to potentially artifactual increases in reading comprehension and that is in paired reading. Topping (1985) claimed that gains of three times 'normal' in reading accuracy and five times 'normal' in comprehension seem typical in paired reading projects. However, in the section on paired reading in the book edited by Topping and Woldendale (1985), all eight studies described in detail used the Neale Analysis as their measure of comprehension. On this test, the questions used to assess comprehension can only be asked if the passage has been read to a criterion of accuracy (usually 16 or fewer errors). Any increase in reading accuracy would give the opportunity for more questions to be asked and so by itself provide the potential for there to be an increase in reading comprehension score. Joscelyne (1991) has provided evidence that when independent measures of reading accuracy and reading comprehension are used, no significant increases in reading comprehension are obtained through paired reading as compared to when children read aloud to a peer tutor for the same length of time. While there is no doubt accessing the written word is an extremely important skill, there would seem to be some virtues in experimental situations in separating access and comprehension as far as possible by using material of a strictly controlled readability or by using listening comprehension tasks as well as or instead of reading comprehension tasks so that the influence of access on comprehension performance is controlled for.

Caution should also be exercised in interpreting a finding of no difference in training studies using a single pool of skilled and less skilled comprehenders.

For example, Yuill and Oakhill (1988), found no overall effect for training in inference making and question generation but further analysis revealed the less skilled comprehenders made significant improvements in comprehension score over controls given training in rapid decoding. In making comparisons at the lower end of the ability range, one would also want to ensure one had build in sufficient control for regression effects not to pose a problem when interpreting the results. Despite this hazard studies which employ both less and more skilled comprehenders to demonstrate whether training less skilled comprehenders to use a specific skill closes the gap between different classes of comprehenders may identify which specific skills underlie differences in performance. In interpreting the outcomes one would have to bear in mind Oakhill's (1993b) caution that poor comprehenders may not have acquired the target strategy because they may be unable to do so. One therefore cannot assume 'that poor comprehenders can be turned into good ones by trying to teach them the skills that good comprehenders possess.'

Finally the structure and content of this review may indicate how subsequent reviews could be structured. For example the categorization of studies covered in this review was acknowledged to be crude; in part because it was an initial inspection of a large body of data. Alternative categories could have been employed. The most obvious would have been comparing studies employing specific training techniques as in Fitzgerald (1990). However the overwhelmingly positive outcome to most interventions reviewed here and in Fitzgerald (1990) would not suggest this structure would necessarily have led to any clearer indications as to the processes which may underlie comprehension. However though most published interventions appear to be successful often only short term outcomes were

measured and it may be instructive to review studies which carried out long term follow up assessments some time after the intervention had ended.

**THE RELATIONSHIP BETWEEN ERROR
DETECTION AND PERFORMANCE ON A STANDARDISED
COMPREHENSION TEST**

The 54 subjects who took part in Experiment Eight of this study had as part of a screening programme also a month previously taken the Primary Reading Test Level Two (France 1981). This is a 48 item reading test which involves the testee in selecting the appropriate word out of five options to match a picture or complete a sentence. It is an untimed test. It was judged by France to be a test of comprehension. It is widely used for screening populations and was used by Vincent and de la Mare (1985) as the basis for selecting children for the standardisation of the New MacMillan Reading Analysis. The availability of this information enabled a comparison to be made between the pupils performance on the three error detection tasks they completed as part of Experiment Eight and their performance on a widely used measure of comprehension that was standardised and validated on 20,000 children in 1977 and 1978. The Pearson Product Moment Correlation between the children's raw scores out of 20 on the error detection tasks and their raw scores on the Primary Reading Test was 0.612 which for $df = 52$ is highly significant ($p = 0.001$). This indicates there is a strong relationship between performance on an established test of reading comprehension and the error detection tasks used in experimental work reported in this study and supports the use of the error detection task as a measure of comprehension of text.

CONTROLLING INDIVIDUAL DIFFERENCES IN PSYCHOLOGICAL EXPERIMENTS

Experiments involve measuring the effects of the manipulation of variables. They are designed to minimise error variation so the experiment gives a clear and unbiased measure of the experimental effect. The individual subjects participating in the experiment differ in many ways and this is a major potential source of error variation. Some experimental designs hold individual differences constant through all subjects receiving all treatments. However this is sometimes neither possible or practicable. For example subject fatigue may preclude the administration of all treatments or exposing a subject to one treatment e.g. a passage containing consonant strings which the subject has to identify and underline would sensitize them to another treatment e.g. the same passage in which the subjects have to identify and underline nonsense words substituted at the same point in the passage as the consonant strings. In these circumstances it may be possible to control some individual differences by counterbalancing. In a counterbalanced design in comparisons involving one factor the effects of the other factors are equalized so the comparisons are unbiased. Experimental designs of this type in which each subject takes more than one but not all combinations of factors are referred to by Lindquist (1956) as 'mixed' designs. A feature of these designs is that some of the treatment comparisons are inter-subject and some are intra-subject.

Lindquist (1956) separately identified 7 out of this family of designs. The design employed in the first three in this series of experiments was a Lindquist Type V Design. It can be used when there are three factors which occur at the same number of levels. In the cases of Experiments One and Three there were 2 levels while in Experiment Two there were 3 levels of

the 3 factors: A (type of substitution), B (passage), and C (session). The Type V design provided a means of counterbalancing both the effects of order and passage. For example in Experiment Two the 3 levels of each factor were combined in a Latin square to give 9 combinations of treatments so that each subject contributed one and only one score at each level of each factor. Further discussion of this design will refer to the interpretation of the summary table of Experiment Two which is reproduced here for the reader's convenience.

Table 3.2 Summary of analysis of variance for Experiment Two

Source	df	SS	MS	F	P
Between Subjects	53	259.59			
AB	2	40.01	20	4.88	0.05
AC	2	13.20	6.6	1.61	
BC	2	2.09	1.05	0.26	
ABC	2	19.79	9.89	2.41	
error	45	184.50	4.10		
Within subjects	108	369.33			
A	2	178.01	89	50.59	0.01
B	2	15.64	7.82	4.45	0.05
C	2	1.94	0.97	0.55	
AB	2	2.09	1.05	0.60	
AC	2	3.71	1.86	1.06	
BC	2	4.08	2.04	1.16	
ABC	6	5.53	0.92	0.52	
error	90	158.33	1.76		

A = type of substitution B = passage C = session part

The significance of the main effects of A, B and C are judged relative to the within - subject error term. However as noted above some interactions are between subjects and some are within subjects. The initial significance level for each interaction is judged relative to the appropriate between - subjects or within - subjects error term. If both comparisons are non-significant then it can be accepted there is no interaction. However if as in the case of the example shown one of the interactions is significant then one must accept that there may be an interaction. However there is an increased risk of falsely accepting there is an interaction. This arises because the two interactions are mutually independent. The chances of one being significant at the same time as the other is very small; the square of the appropriate significance level. This increases the chances of the null hypothesis being rejected to almost 10% if one accepts a 5% significance level i.e. it almost doubles the table value of probability of making a type one error. In the case illustrated the F value of the between subjects AB interaction is significant at the 2¹/₂% level which has been doubled and shown as 5% to compensate for the increased risk of making a type one error.

REFERENCES

REFERENCES

- Adams M J** (1989). Thinking Skills Curricula: their promise and progress. *Educational Psychologist* 24, 25-77.
- Adams M J** (1990). *Beginning to Read: thinking and learning about print.* Cambridge Massachusetts: M.I.T. Press.
- Alderson J C** (1979). The effect on the cloze test of changes in deletion frequency. *Journal of Research in Reading* 2, 108-119.
- Alvermann D E** (1988). Effects of spontaneous and induced lookbacks on self-perceived high and low ability comprehenders. *Journal of Educational Research* 81, 325-331.
- Amlund J T, Kardash C A M, and Kulhary R W** (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly* 21, 49-58.
- Anderson R C and Pearson P D** (1984). A schema-theoretic view of basic processes in reading comprehension. In P D Pearson (ed) *Handbook of Reading Research.* New York: Longman.
- Anderson T H** (1980). Study strategies and adjunct aids. In R J Spiro, B C Bruce and W F Brewer (eds), *Theoretical issues in Comprehension: perspectives from cognitive psychology, artificial intelligence, linguistics and education.* Hillsdale N T: Erlbaum.
- Armbruster B B** (1984). The problem of 'Inconsiderate Text.' In G G Duffy, L R Roehler and J Mason (eds), *Comprehension instruction: perspectives and suggestions.* New York: Longman.
- Armbruster B B, Anderson T H and Ostertag J** (1987). Does text structure / summarization instruction facilitate learning from expository text ? *Reading Research Quarterly* 1987 22, 331-346.

- Arnold H F** (1942). The comparative effectiveness of certain study techniques in the field of history. *Journal of Educational Psychology* 33, 449-457.
- Asher S R** (1980). Topic interest and children's reading comprehension. In R J Spiro, B C Bruce and W F Brewer (eds), *Theoretical Issues in Reading Comprehension: perspectives from cognitive psychology, linguistics, artificial intelligence and education*. Hillsdale N J: Lawrence Erlbaum.
- Ausubel D P** (1960). The use of advance organizers in learning and retention of meaningful material. *Journal of Educational Psychology* 51, 267-272.
- Baddeley A** (1990). *Human Memory: theory and practice*. Hove: Lawrence Erlbaum.
- Baker L** (1984a). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology* 76, 588-597.
- Baker L** (1984b). Spontaneous versus instructed use of multiple standards for evaluating comprehension: effects of age reading proficiency and type of standard. *Journal of Experimental Child Psychology* 38, 289-311.
- Baker L** (1985a). How do we know when we don't understand? Standards for evaluating text comprehension. In D L Forrest-Pressley, G E MacKinnon and G T Waller (eds), *Metacognition, Cognition and Human Performance, Volume One, Theoretical Perspectives*. London: Academic Press.
- Baker L** (1985b). When will children acknowledge failures of word comprehension ? Paper presented at the meeting of the Society for Research in Child Development, Toronto.
- Ballard P B** (1920). *Mental Tests*. London: Hodder and Stoughton.

- Barnett J E and Seedeldt R W** (1989). Read something once, why read it again ? : repetitive reading and recall. *Journal of Reading Behaviour* xxi, 351-360.
- Bartlett F C** (1932). *Remembering: a study in experimental and social psychology*. Cambridge. Cambridge University Press.
- Baumann J F** (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. *Reading Research Quarterly* 20, 93-108.
- Bean J W and Steenwyk F L** (1984). The effect of three forms of summarization instruction of sixth grader's summary writing and comprehension. *Journal of Reading Behaviour* xvi, 297-306.
- Beck I L, McKeown M G, Omanson R C and Pope M T** (1984). Improving the comprehensibility of stories: the effects of revisions that improve coherence, *Reading Research Quarterly* 19, 263-277.
- Beebe M J** (1980). The effect of different types of substitution miscues on reading. *Reading Research Quarterly* 15, 324-336.
- Bereiter C and Bird M** (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction* 2, 131-156.
- Berger N S** (1978). Why can't John Read ? Perhaps he's not a good listener. *Journal of Learning Disabilities* 11, 633-638.
- Berkowitz S J** (1986). Effects of instruction in text organisation on sixth-grade students' memory for expository reading. *Reading Research Quarterly* 21, 161-178.
- Betts E A** (1946). *The Foundations of Reading Instruction*. New York: American Book Company.
- Blachowicz C L** (1977-78). Semantic constructivity in children's comprehension. *Reading Research Quarterly* 13, 188-199.

- Bormuth J R** (1967). Comparable cloze and multiple choice comprehension test scores. *Journal of Reading* x, 291-299.
- Bormuth J R** (1968). Cloze test readability: criterion reference scores. *Journal of Educational Measurement* 5, 189-196.
- Brainerd C J and Kingma J** (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology* 17, 210-247.
- Bransford J D, Barclay J R and Franks J J** (1972). Sentence memory: a constructive versus interpretative approach. *Cognitive Psychology* 3, 193-209.
- Brewer W F** (1987). Schemas versus mental models in human memory. In P Morris (ed) *Modelling Cognition*. Chichester: John Wiley and Sons Ltd.
- Brown A L** (1978). Knowing when, where and how to remember:- a problem of metacognition. In R Glaser (ed), *Advances in Instructional Psychology, Volume One*. Hillsdale N J: Lawrence Erlbaum.
- Brown A L and Smiley S S** (1977). Rating the importance of structural units of prose passages: a problem of metacognitive development. *Child Development* 48, 1-8.
- Brown A L, Smiley S S and Lawton S Q C** (1978). The effects of experience on the selection of suitable retrieval cues for studying texts. *Child Development* 49, 829-835.
- Bryant P and Bradley L** (1985). *Children's Reading Problems: psychology and education*. Oxford: Basil Blackwell Ltd.
- Burt C** (1921). *Mental and Scholastic Tests*. London County Council.

- Cairney T H** (1988). The purpose of basals: what children think. *The Reading Teacher* 41, 420-428.
- Canney G and Winograd P** (1979). Schemata for reading and reading comprehension performance (Tech Rep No 120). Urbana - Champaign: University of Illinois, Centre for the Study of Reading.
- Carnine L, Carnine D and Gersten R** (1984). Analysis of oral reading errors made by economically disadvantaged students taught with a synthetic - phonics approach. *Reading Research Quarterly* 19, 343-356.
- Carnine D and Kinder D** (1985). Teaching low-performing students to apply generative and schema strategies to narrative and expository material. *Remedial and Special Education* 6, 20-30.
- Carpenter P A and Daneman M** (1981). Lexical retrieval and error recovery in reading: a model based on eye fixations. *Journal of Verbal Learning and Verbal Behaviour* 20, 137-160.
- Carr E M, Dewitz P and Patberg J P** (1983). The effect of inference training on children's comprehension of expository text. *Journal of Reading Behaviour* xv, 1-18.
- Carroll J B, Davies P and Richman B** (1971). *Word Frequency Book*. Boston: Houghton Mifflin Company.
- Carver R P** (1977). Toward a theory of reading comprehension and rauding. *Reading Research Quarterly* 13, 8-63.
- Carver R P** (1987). Should reading comprehension skills be taught. In J E Readance, S R Baldwin, J P Konopak and H Newton, *Research in Literacy: merging perspectives*. Thirty-sixth Yearbook - National Reading Conference. New York: National Reading Conference.
- Carver R P** (1990a). *Reading Rate: a review of research and theory*. New York: Academic Press.

- Carver R P** (1990b). Predicting accuracy of comprehension from the relative difficulty of the material. *Learning and Individual Differences* 2, 405-422.
- Carver R P** (1992). Effect of prediction activities, prior knowledge and text type upon amount comprehended: Using rauding theory to critique schema theory research. *Reading Research Quarterly* 27, 165-174.
- Ceci S J, Caves R D and Howe M J A** (1981). Children's long-term memory for information that is incongruous with their prior knowledge. *British Journal of Psychology* 72, 443-450.
- Chan L K S and Cole P G** (1986). The effects of comprehension monitoring training on the reading competence of learning disabled and regular class students. *Remedial and Special Education* 7, 33-40.
- Cipielewski J and Stanovich K E** (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Psychology* 54, 74-89.
- Clay M M** (1968). A syntactic analysis of reading errors. *Journal of Verbal Learning and Verbal Behaviour* 7, 434-438.
- Clay M M** (1985). *The Early Detection of Reading Difficulties*. (Third Edition) Auckland: Heinemann.
- Clymer T** (1972). What is 'reading' ? : some current concepts. In A Melnick and J Merritt (eds), *Reading Today and Tomorrow*. London: University of London Press.
- Colman E B** (1968). Experimental studies of readability. *Elementary English* 45, 166-178.

- Collins A, Brown J S and Larkin K M** (1980). Inference in text understanding. In R J Spiro, B C Bruce and W F Brewer (eds), *Theoretical Issues in Reading Comprehension*. Hillsdale N J: Lawrence Erlbaum Associates.
- Collins A and Smith E E** (1982). Teaching the process of reading comprehension. In D K Detterman and R J Sternberg (eds), *How and How Much Can Intelligence be Increased*. New Jersey: Ablex.
- Cromer W** (1970). The difference model: a new explanation for some reading difficulties. *Journal of Educational Psychology* 61, 471-483.
- Cross D R and Paris S G** (1988). Developmental and instructional analysis of children's metacognition and reading comprehension. *Journal of Educational Psychology* 80, 131-142.
- Cunningham J W** (1982). Generating interactions between schemata and text. In J A Niles and L A Harris (eds), *New Inquiries in Reading Research and Instruction*. Rochester New York: National Reading Conference.
- Curtis M E** (1980). Development of components of reading skills. *Journal of Educational Psychology* 72, 656-669.
- Daneman M** (1987). Reading and working memory. In J R Beech and A M Colley (eds), *Cognitive Approaches to Reading*. Chichester: Wiley.
- Daneman M and Carpenter P A** (1980). Individual differences in working memory and reading, *Journal of Verbal Learning and Verbal Behaviour* 19,450-466.
- Davey B and McBride S** (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology* 78, 256-262.

- Davison A** (1984). Readability formulas and comprehension. In G G Duffy, L R Roehler and J Mason (eds). *Comprehension instruction: perspectives and suggestions*. New York: Longman.
- Dewitz P, Carr E M and Patberg J** (1987). Effects of inference training on comprehension and comprehension monitoring. *Reading Research Quarterly* 22, 99-121.
- Department of Education and Science** (1991). *Testing 7 year olds in 1991: results of the National Curriculum assessments in England*. London: D. E. S.
- Diack H** (1965). *In Spite of the Alphabet*. London: Chatto and Windus.
- Doctor E A and Coltheart M** (1980). Children's use of phonological encoding when reading for meaning. *Memory and Cognition* 8, 195-209.
- Dole J A, Valencia S W, Greer E A and Wardrop J L** (1991). Effects of two types of prereading instruction on the comprehension of narrative and expository text. *Reading Research Quarterly* xxvi, 142-159.
- Downing J** (1969). How children think about reading. *Reading Teacher* 23, 217-230.
- Downing J, Ayers D and Schaefer B** (1978). Conceptual and perceptual factors in learning to read. *Educational Research* 21, 11-17.
- Durkin D** (1978-79). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly* 14, 481-533.
- Ehrlich K and Johnson-Laird P N** (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behaviour* 21, 296-306.
- Evans M A and Carr T H** (1985). Cognitive abilities, conditions of learning and the early development of reading skill. *Reading Research Quarterly* 20, 327-350.

- Fitzgerald J** (1990). Reading Comprehension Instruction 1783-1987: a review of trends and research. Newark: International Reading Association.
- Flagg P W and Reynolds A G** (1977). Modality of presentation and blocking in sentence recognition memory. *Memory and Cognition* 5, 111-115.
- Flavell J H** (1971). First discussants comments: what is memory development the development of ? *Human Development* 14, 272-278.
- Flavell J H** (1985). *Cognitive Development*. Second edition. Englewood Cliffs New Jersey: Prentice Hall Inc.
- Flavell J H, Friedrichs A G and Hoyt J D** (1970). Developmental changes in memorization processes. *Cognitive Psychology* 1, 324-340.
- Flavell J H and Wellman H M** (1977). Metamemory. In R V Kail and J W Hagen (eds), *Perspectives on the Development of Memory and Cognition*. Hillsdale: Lawrence Erlbaum.
- Fleischer L S, Jenkins J R and Pany D** (1979). Effects on poor reader's comprehension of training in rapid decoding. *Reading Research Quarterly* 15, 30-48.
- Flesch R** (1948). A new readability yardstick. *Journal of Applied Psychology* 32, 221-233.
- Forrest-Pressley D and Waller T G** (1984). *Cognition, Metacognition and Reading*. New York: Springer-Verlag.
- Forrest-Pressley D, Waller T G and Pressley M** (1989). Forrest-Pressley and Waller (1984) concluded that metacognition about reading is related to reading performance: a comment about Jacob and Paris (1987). *Educational Psychologist* 24, 207-210.

- Frederiksen C H** (1979). Discourse comprehension and early reading. In L B Renick and P A Weaver (eds). *Theory and Practice of Early Reading Vol. 1*. Hillsdale N J: Lawrence Erlbaum Associates.
- France N** (1981). *The Primary Reading Test: Teacher's Guide* (revised edition). Windsor: NFER - Nelson.
- Francis H** (1992). Patterns of reading development in the first school. *British Journal of Educational Psychology* 62, 225-232.
- Gambrell L B, Pfeiffer W R and Wilson R M** (1985). The effects of retelling upon reading comprehension and recall of text information. *Journal of Educational Research* 78, 216-220.
- Garner R** (1980). Monitoring of understanding: an investigation of good and poor reader's awareness of induced miscomprehension of text. *Journal of Reading Behaviour* xii, 55-63.
- Garner R** (1981). Monitoring of passage inconsistency among poor comprehenders: a preliminary test of the 'piecemeal processing' explanation. *Journal of Educational Research* 74, 159-162.
- Garner R** (1987). *Metacognition and Reading Comprehension*. Norwood N J: Ablex.
- Garner R, Hare V C, Alexander P, Haynes J and Winograd P** (1984). Inducing use of a text lookback strategy among unsuccessful readers. *American Educational Research Journal* 21, 789-798.
- Garnham A** (1987). *Mental Models as Representations of Discourse and Text*. Chichester: Ellis Horwood Ltd.
- Garnham A and Oakhill J V** (1989). The everyday use of anaphoric expressions: Implications for the 'mental models' theory of text comprehension. In N E Sharkey (ed) *Models of Cognition: a review of cognitive science Volume One*. Norwood N. J: Ablex.

- Garnham A and Oakhill J V** (1992). Aberrant ellipsis: advertisers do, but why. *English Today* 8, 37-40.
- Garnham A, Oakhill J V and Johnson-Laird P N** (1982). Referential continuity and the coherence of discourse. *Cognition* 11, 29-46.
- Gilroy A and Moore D W** (1988). Reciprocal teaching of comprehension - fostering and comprehension - monitoring activities with ten primary school girls. *Educational Psychology* 8, 41-49.
- Glass G V, Peckham P D and Sanders J R** (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research* 42, 237-288.
- Ghatala E S, Levin J R, Foorman B R and Pressley M** (1989). Improving children's regulation of their reading PREP. time. *Contemporary Educational Psychology* 14, 49-66.
- Glenberg A M, Wilkinson A C and Epstein W** (1982). The illusion of knowing: failure in the self-assessment of comprehension. *Memory and Cognition* 10, 597-602.
- Gomulicki B R** (1956). Recall as an abstractive process. *Acta Psychologica* 12, 77-94.
- Gonzales P C and Elijah D V** (1975). Rereading: effect on error patterns and performance levels on the I.R.I. *Reading Teacher* 28, 647-652.
- Goodacre E J** (1972). Methods of teaching reading. In A Melnik and J Merritt (eds), *The Reading Curriculum*. London: University of London Press.
- Goodman K S** (1982). Reading: a psycholinguistic guessing game. In F V Gollasch (ed), *Language and Literacy: the selected writings of Kenneth S Goodman*. Volume One: Process, Theory, Research. Boston: Routledge and Kegan Paul Ltd.

- Goodman K S and Goodman Y M** (1979). Learning to read is natural. In L B Resnick and P A Weaver (eds), *Theory and Practice of Early Reading Volume 1*. Hillsdale N J: Lawrence Erlbaum Associates.
- Goswami U and Bryant P** (1990). *Phonological Skills and Learning to Read*. Hove: Lawrence Erlbaum Associates.
- Graves M F, Cooke C L and Laberge M J** (1983). Effects of previewing difficult short stories on low ability junior high school students' comprehension, recall and attitudes. *Reading Research Quarterly* 18, 262-276.
- Guthrie J T** (1973). Reading comprehension and syntactic responses in good and poor readers. *Journal of Educational Psychology* 65, 294-299.
- Hansen J** (1981). The effect of inference training and practice on young children's reading comprehension. *Reading Research Quarterly* 16, 391-417.
- Hansen J and Pearson P D** (1983). An instructional study: improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology* 75, 821-829.
- Hare C V, Rabinowitz M and Schieble M K** (1989). Text effects on main idea comprehension. *Reading Research Quarterly* 24, 72-88.
- Harris P L, Kruithof A, Meerum Terwogt M and Visser T** (1981). Children's detection and awareness of textual anomaly. *Journal of Experimental Child Psychology* 31, 212-230.
- Harrison C** (1977). Assessing the readability of school texts. In Gilliland J (ed), *Reading: research and classroom practice*. London: Ward Lock Educational.
- Harrison C** (1980). *Readability in the Classroom*. Cambridge: Cambridge University Press.

- Henderson L and Chard J** (1980). The reader's implicit knowledge of orthographic structure. In U Frith (ed) *Cognitive Processes in Spelling*. New York: Academic Press.
- H. M. I.** (1989). *Reading Policy and Practice at Ages 5-14*. D. E. S.
- Hoover W A and Tunmer W E** (1993) The components of reading. In G B Thompson, W E Tunmer and T Nicholson (eds) *Reading Acquisition Processes*. Clevedon: Multilingual Matters.
- Howe M J A and Singer L** (1975). Presentation variables and Students' activities in meaningful learning. *British Journal of Educational Psychology* 45, 52-61.
- Huey E B** (1908). *The Psychology and Pedagogy of Reading*. New York: Macmillan.
- Hunter-Blanks P, Ghatala E S, Pressley M and Levin J R** (1988). Comparison of monitoring during study and during testing on a sentence - learning task. *Journal of Educational Psychology* 80, 279-283.
- Isakson R L and Miller J W** (1976). Sensitivity to syntactic and semantic cues in good and poor comprehenders. *Journal of Educational Psychology* 68, 787-792.
- Jacobs J E and Paris S G** (1987). Children's metacognition about reading: issues in definition, measurement and instruction. *Educational Psychologist* 22, 255-278.
- Johnson-Laird P N** (1977). Psycholinguistics without linguistics. In N S Sutherland (ed) *Tutorial Essays in Psychology*. Volume One. Hillsdale N.J. : Lawrence Erlbaum Associates.
- Johnson-Laird P N** (1983). *Mental Models: towards a cognitive science of language, inference and consciousness*. Cambridge: Cambridge University Press.

- Joscelyne T** (1991). Peer tutored Paired Reading. Unpublished D. Phil. Thesis. University of Sussex.
- Juola J F, Ward N J and McNamara T** (1982). Visual search and reading of rapid serial presentations of letter strings, words and text. *Journal of Experimental Child Psychology: General* 111,208-227.
- Kieras D E** (1981). Component processes in the comprehension of simple prose. *Journal of Verbal Learning and Verbal Behaviour* 20, 1-23.
- Kintsch W and van Dijk T A** (1978). Toward a model of text comprehension and production. *Psychological Review* 85, 363-394.
- Kletzien S B** (1991). Strategy use by good and poor comprehenders reading expository text of differing levels. *Reading Research Quarterly* xxvi, 67-86.
- Kolers P A** (1970). Three stages of reading. In Levin H and Williams J P (eds), *Basic Studies on Reading*. New York: Basic Books.
- Krug D, Davis T B and Glover J A** (1990). Massed versus distributed repeated reading: a case of forgetting helping recall ? *Journal of Educational Psychology* 82, 366-371.
- La Berge D and Samuels S J** (1974). Towards a theory of automatic processing in reading. *Cognitive Psychology* 6, 293-323.
- Levy B A and Carr T H** (1990). Component process analysis: conclusions and challenges. In T H Carr and B A Levy (eds) *Reading and its Development: component skills approaches*. London: Academic Press.
- Linden M and Wittrock M C** (1981). The teaching of reading comprehension according to the model of generative learning. *Reading Research Quarterly* 17, 44-57.
- Lindquist E F** (1956). *Design and Analysis of Experiments in Psychology and Education*. Boston: Houghton Mifflin Co.

- Lunzer E A and Gardner K** (1979). *The Effective Use of Reading*. London: Heinemann Educational Books.
- Lysynchuk L M, Pressley M, d'Ailly H, Smith M and Cake H** (1989). A methodological analysis of experimental studies of comprehension strategy instruction. *Reading Research Quarterly* xxiv, 458-470.
- McCormick S** (1989). Effects of previews on more skilled and less skilled reader's comprehension of expository text. *Journal of Reading Behaviour* xxi, 219-239.
- McCormick S** (1992). Disabled readers' erroneous responses to inferential comprehension questions: description and analysis. *Reading Research Quarterly* 27, 55-77.
- MacDonald J D** (1986). Self-generated questions and reading recall: does training help. *Contemporary Educational Psychology* 11, 290-304.
- McKenna M C and Layton K** (1990). Concurrent validity of cloze as a measure of intersentential comprehension. *Journal of Educational Psychology* 82, 372-377.
- McKeown M G, Beck I L, Sinatra G M and Loxterman J A** (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research quarterly* 27, 79-93.
- McLeod D** (1993). *Reciprocal Teaching for the Comprehension of Text*. Dissertation in part fulfilment of the M Sc in Psychology of Education; University of London Institute of Education.
- McNamara T P, Miller D L and Bransford J D** (1991). Mental models and reading comprehension. In R Barr, M L Kamil, P Mosenthal and P D Pearson, *Handbook of Reading Research* vol II. New York: Longman.

- Maki R H and Berry S L** (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10, 663-679.
- Mani K and Johnson-Laird P N** (1982). The mental representation of spatial descriptions. *Memory and Cognition* 10, 181-187.
- Markman E and Gorin L** (1991). Children's ability to adjust their standards for evaluating comprehension. *Journal of Educational Psychology* 73, 320-325.
- Marks C B, Doctorow M J and Wittrock M C** (1974). Word frequency and reading comprehension. *Journal of Educational Research* 67, 259-262.
- Marton F** (1975). On non-verbation learning: II - the erosion effect of a task induced learning algorithm. Reports from the Institute of Education; University of Goteborg, No.40.
- Marton F and Saljo R** (1976). On qualitative differences in learning - II. Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology* 46, 115-127.
- Massaro D W, Taylor G A, Venezky R L, Jastrembski J E and Lucus P A** (1980). Letter and Word Perception. Amsterdam: North-Holland Publishing Company.
- Masson M E J** (1982). Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory and Cognition* 8, 400-417.
- Mayer R E** (1983). Can you repeat that ? Qualitative effects of repetition and advance organizers on learning from science prose. *Journal of Educational Psychology* 75, 40-49.
- Miller G E** (1985). The effects of general and specific self-instruction training on children's comprehension monitoring performances during reading. *Reading Research Quarterly* 20, 616-628.

- Miller J W and Isakson R L** (1978). Contextual sensitivity in beginning readers. *Elementary School Journal* 78, 325-331.
- Miller S D and Yochum N** (1991). Asking students about the nature of their reading difficulties. *Journal of Reading Behaviour* xxiii., 465-485.
- Morris D** (1983). Concept of word and phoneme awareness in the beginning reader. *Research in the Teaching of English* 17, 359-373.
- Mosenthal P B** (1987). Understanding meaning in reading. *The Reading Teacher* 41, 206-209.
- Myers M and Paris S G** (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology* 70, 680-690.
- Neale M D** (1966). *Neale Analysis of Reading Ability: second edition.* London: Macmillan Education Ltd.
- Neale M D** (1989). *Neale Analysis of Reading Ability: revised British edition.* Windsor: N.F.E.R.-Nelson.
- Niemi P and Virjano M** (1986). Proofreading: visual, phonological, syntactic or all of these ? *Journal of Research in Reading* 9, 31-38.
- Nolte R Y and Singer H** (1985). Active comprehension: teaching a process of reading comprehension and its effects on reading achievement. *The Reading Teacher* 39, 24-31.
- Noordman L G M and Vonk W** (1992). Readers' knowledge and the control of inferences in reading. *Language and Cognitive Processes* 7, 373-391.
- Oakhill J** (1982). Constructive processes in skilled and less skilled comprehenders' memory for sentences. *British Journal of Psychology* 73, 13-20.
- Oakhill J** (1993a). How do we understand anaphors ? *The Psychologist: Bulletin of the British Psychological Society* 6, 204-208.

- Oakhill J** (1993b). Children's difficulties in reading comprehension. *Educational Psychology Review* 5, 223-237.
- Oakhill J and Garnham A** (1985). Referential continuity, transitivity, and the retention of relational descriptions. *Language and Cognitive Processes* 1, 149-162.
- Oakhill J and Garnham A** (1988). *Becoming a Skilled Reader*. Oxford: Basil Blackwell Ltd.
- Oakhill J, Yuill N and Parkin A** (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading* 9, 80-91.
- O'Shea L J, Sindelar P T and O'Shea D J** (1985). The effects of repeated readings and attentional cues on reading fluency and comprehension. *Journal of Reading Behaviour* 17, 129-142.
- Owings R A, Petersen G A, Bransford J D, Morris C D and Stein B S** (1980). Spontaneous monitoring and regulation of learning, a comparison of successful and less successful fifth graders. *Journal of Educational Psychology* 72, 250-256.
- Palincsar A S and Brown A L** (1984). Reciprocal teaching of comprehension fostering and comprehension monitoring activities. *Cognition and Instruction* 1, 117-175.
- Palincsar A S and Brown A L** (1986). Interactive teaching to promote independent learning from text. *The Reading Teacher* 771-777.
- Palincsar A S, Brown A L and Martin S M** (1987). Peer interaction in reading comprehension instruction. *Educational Psychologist* 22, 231-253.
- Papandropoulou I and Sinclair H** (1974). 'What is a word?' Experimental study of children's ideas in grammar. *Human Development* 17, 241-258.

- Paris S G** (1986). Teaching children to guide their reading and learning. In Raphael T (ed), *The Content of School-based Literacy*. New York: Random House.
- Paris S G, Cross D R and Lipson M Y** (1984). Informed strategies for learning: a program to improve childrens reading awareness and comprehension. *Journal of Educational Psychology* 76, 1239-1252.
- Paris S G and Jacobs J E** (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development* 55, 2083-2093.
- Paris S G, Lipson M Y and Wixson K K** (1983). Becoming a strategic reader. *Contemporary Educational Psychology* 8, 293-316.
- Paris S G and Myers M** (1981). Comprehension monitoring, memory and study strategies of good and poor readers. *Journal of Reading Behaviour* 13, 5-22.
- Paris S G and Oka E R** (1986). Children's reading strategies, metacognition and motivation: *Developmental Review* 6, 25-56.
- Paris S G, Wasik B A and Van der Westhuizen** (1988). Meta-metacognition: a review of research on metacognition and reading. In J E Readance, S R Baldwin, J P Konopak and P R O'Keefe (eds), *Dialogues in Literacy Research: Thirty Seventh Yearbook National Reading Conferences*. Chicago: National Reading Conference.
- Parker J P** (1962). Some organizational variables and their effect on comprehension. *Journal of Communication* 12, 27-32.
- Perfetti C A** (1985). *Reading Ability*. Oxford: Oxford University Press.
- Piaget J** (1932). *The Language and Thought of the Child*. London: Kegan Paul, Trench, Trubner and Co. Ltd.
- Prater D L and Terry C A** (1988). Effects of mapping strategies on reading comprehension and writing performance. *Reading Psychology* 9, 101-120.

- Pressley M, Ghatala E S, Woloshyn V and Pirie J** (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in response to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly* xxv, 232-249.
- Pressley M, Snyder B L, Levin J R, Murray H G and Ghatala E S** (1987). Perceived readiness for examination performance (PREP) produced by initial reading of text and text containing adjunct questions. *Reading Research Quarterly* xxii, 219-236.
- Rankin E F and Culhane J W** (1969). Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading* 13, 193-198.
- Raphael T E and Pearson D P** (1985). Increasing students' awareness of sources of information for answering questions. *American Educational Research Journal* 22, 217-235.
- Raphael T E and Wonnacott C A** (1985). Heightening fourth-grade students' sensitivity to sources of information for answering comprehension questions. *Reading Research Quarterly* 20, 282-296.
- Rauenbusch F and Bereiter C** (1991). Making reading more difficult: a degraded test microword for teaching reading comprehension strategies. *Cognition and Instruction* 8, 181-206.
- Recht D R and Leslie L** (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology* 80, 16-20.
- Reid J F** (1966). Learning to think about reading. *Educational Research* 9, 56-62.
- Rinehart S D, Stahl S A and Erickson L G** (1986). Some effects of summarisation training on reading and studying. *Reading Research Quarterly* 21, 422-436.

- Roberts B** (1992). The evolution of the young child's concept of word as a unit of spoken and written language. *Reading Research Quarterly* 27, 125-138.
- Rothkopf E Z** (1968). Textual constraint as function of repeated inspection. *Journal of Educational Psychology* 59, 20-25.
- Rubin A D** (1980). A theoretical taxonomy of the difference between oral and written language. In R J Spiro, B C Bruce and W F Brewer (eds), *Theoretical Issues in Reading Comprehension*: Hillsdale N.J. : Lawrence Erlbaum Associates.
- Ruddell R B** (1965). The effect of the similarity of oral and written patterns of language structure on reading comprehension. *Elementary English* 42, 403-410.
- Saarnio D A, Oka E R and Paris S G** (1990). Developmental predictors of children's reading comprehension. In T H Carr and B A Levy, *Reading and its Development*. London: Academic Press.
- Schank R and Abelson R** (1977). *Scripts, plans, goals and understanding*. Hillsdale N.J. : Lawrence Erlbaum Associates.
- Schonell F J** (1961). *The Psychology and Teaching of Reading*. Edinburgh: Oliver and Boyd.
- Schneyer J W** (1965). Use of the cloze procedure for improving reading comprehension. *The Reading Teacher* 19, 174-179.
- Schunk D H and Rice J M** (1987). Enhancing comprehension skill and self-efficacy with strategy value information. *Journal of Reading Behaviour* xix, 285-302.
- Schwantes F M** (1991). Children's use of semantic and syntactic information for word recognition and determination of sentence meaningfulness. *Journal of Reading Behaviour* xxiii, 335-350.

- Seretny M L and Dean R S** (1986). Interspersed post-passage questions and reading comprehension achievement. *Journal of Educational Psychology* 78, 228-229.
- Shanahan T, Kamil M L and Tobin A W** (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly* 17, 229-255.
- Shnayer S W** (1969). Relationships between reading interest and reading comprehension. In J R Figurel (ed), *Reading and Realism Proceedings Volume 13*. Newark, Delaware: International Reading Association.
- Short E J and Ryan E B** (1984). Metacognitive differences between skilled and less skilled readers: remediating deficits through story grammar and attribution training. *Journal of Educational Psychology* 76, 225-235.
- Simons H D and Leu D J** (1987). The use of contextual and graphic information in word recognition by second, fourth and sixth grade readers. *Journal of Reading Behaviour* 19, 33-47.
- Smiley S, Oakley D, Worthen D, Campione J and Brown A** (1977). Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. *Journal of Educational Psychology* 69, 381-387.
- Smith F** (1969). The use of featural dependencies across letters in the visual identification of words. *Journal of Verbal Learning and Verbal Behaviour* 8, 215-218.
- Smith F** (1973). *Psycholinguistics and Reading*. New York: Holt, Reinhart and Wilson.
- Smith F and Holmes D L** (1971). The independence of letter, word, and meaning identification in reading. *Reading Research Quarterly* 6, 394-415.

- Smith R J and Barrett T C** (1974). *Teaching Reading in the Middle Grades*. Reading, Massachusetts: Addison-Wesley.
- Spache G D** (1963). *Toward Better Reading*. Champaign. Illinois: Garrard Publishing Co.
- Spooncer F A** (1976). Reading ability and readability. *Reading* 10, 2-10.
- Stauffer R G** (1969). *Teaching Reading as a Thinking Process*. New York: Harper and Row.
- Sticht T G, Beck L J, Hauke R N, Kleiman G M and James J H** (1974). *Auding and Reading: a developmental model*. Alexander V A: Human Research Organisation.
- Sticht T G and James J H** (1984). Listening and Reading. In P D Pearson (ed), *Handbook of Reading Research*. New York: Longman.
- Syne J** (1982). *Metacognition and Text Comprehension*. Dissertation submitted in part fulfilment of the M.Sc in Psychology of Education; University of London Institute of Education.
- Taylor B M** (1982). Text structure and children's comprehension and memory for expository material. *Journal of Educational Psychology* 74, 323-340.
- Taylor B M and Beach R W** (1984). The effects of text structure instruction on middle-grade student's comprehension and production of expository text. *Reading Research Quarterly* 19, 134-146.
- Taylor C A** (1984). *The relative effects of reading or writing a prose or diagrammatic summary upon the comprehension of expository prose*. Ed. D. Dissertation: University of Kentucky.
- Taylor W L** (1953). 'Cloze Procedure' : a new tool for measuring readability. *Journalism Quarterly* 30, 415-433.

- Taylor W L** (1957). Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology* 41, 19-26.
- Thorndike E L** (1917). Reading as reasoning: a study of mistakes in paragraph reading. *Journal of Educational Psychology* 8, 323-332.
- Thorndyke P W** (1977). Cognitive structures in comprehension and memory for narrative discourse. *Cognitive Psychology* 9, 77-110.
- Tikhomirov O K and Klochko V E** (1981). The detection of a contradiction as the initial stage of problem formulation. In J V Wertsch (ed). *The Concept of Activity in Soviet Psychology*. Armonk N.Y. : M.E Sharpe.
- Tobias S** (1982). When do instructional methods make a difference. *Educational Researcher* 11, No 4, 4-9.
- Topping K** (1985). An introduction to paired reading. In K Topping and S Wolfendale (eds), *Parental Involvement in Children's Reading*. London: Croom Helm.
- Topping K and Wolfendale S** (1985). *Parental Involvement in Children's Reading*. London: Croom Helm.
- Townsend D J, Carrithers C and Bever T G** (1987). Listening and reading processes in college-and middle school-age readers. In R Horowitz and S J Samuels (eds). *Comprehending Oral and Written Language*. New York: Academic Press.
- Tunmer W E, Nesdale A R and Pratt C** (1983). The development of young children's awareness of logical inconsistencies. *Journal of Experimental Child Psychology* 36, 97-108.
- Turner M** (1990). *Sponsored Reading Failure: an object lesson*. Education Unit, Warlingham Park School, Chelsam Common Warlington, Surrey.

- Turner M** (1991). Squaring the diamond. *Educational and Child Psychology* 8, 67-74.
- Upward C** (1992). Is traditional english spelng mor diffictl than jermn ? *Journal of Research in Reading* 15, 82-94.
- Vincent D and de la Mare M** (1985). *Manual for the New Macmillan Reading Analysis*. Basingstoke: Macmillan Education Ltd.
- Vogel S A** (1975). *Syntactic abilities in normal and dyslexic children*. Baltimore: University Park Press.
- Vosniadou S, Pearson D P and Rogers T** (1988). What causes children's failures to detect inconsistencies in text ? Representation versus comparison difficulties. *Journal of Educational Psychology* 80, 27-39.
- Weaver P A** (1979). Improving reading comprehension: effects of sentence organization structure. *Reading Research Quarterly* 19, 134-146.
- Weaver W W** (1965). Theoretical aspects of the cloze procedure. In E L Thurstone and E H Hafner (eds), *The Fourteenth Yearbook of the National Reading Conference*. Milwaukee, Wisconsin: The National Reading Conference.
- Weber R M** (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly* 5, 427-451.
- Wechsler D** (1974). *Manual for the Wechsler Intelligence Scale for Children - Revised*. British Edition. Windsor: NFER-Nelson.
- Weiss M J and Hagen R** (1988). A key to literacy: kindergartners' awareness of the functions of print. *Reading Teacher* 41, 574-578.
- White C V, Pascarella E T and Pflaum S W** (1981). Effects of training in sentence construction on the comprehension of learning disabled children. *Journal of Educational Psychology* 73, 697-704.

- Willows D M and Ryan E B** (1981). Differential utilization of syntactic and semantic information by skilled and less skilled readers in the intermediate grades. *Journal of Educational Psychology* 73, 607-615.
- Willows D M and Ryan E B** (1986). The development of grammatical sensitivity and its relationship to early reading achievement. *Reading Research Quarterly* 21, 253-266.
- Winer B J** (1962). *Statistical Principles in Experimental Design*. New York: McGraw-Hill Book Co.
- Winne P H, Graham L and Prock L** (1993). A model of poor readers' text-based inferencing: effects of explanatory feedback. *Reading Research Quarterly* 28, 53-66.
- Winograd P and Johnson P** (1982). Comprehension monitoring and the error detection paradigm. *Journal of Reading Behaviour* 14, 61-74.
- Wixson K K, Bosky A B, Yochum M N and Alvermann D E** (1984). An interview for assessing students' perceptions of classroom reading tasks. *Reading Teacher* 37, 346-352.
- Wong B Y L and Jones W** (1982). Increasing metacomprehension in learning disabled and normally achieving students through self-questioning training. *Learning Disability Quarterly* 5, 228-240.
- Young W E** (1936). The relation of reading comprehension and retention to hearing comprehension and retention. *Journal of Experimental Education* v, 30-39.
- Yuill N and Oakhill J** (1988). Effects of inference awareness training in poor reading comprehension. *Applied Cognitive Psychology* 2, 33-45.
- Yuill N and Oakhill J** (1991). *Children's Problems in Text Comprehension: an experimental investigation*. Cambridge: Cambridge University Press.

Zirinsky S (1984). Qualitative Developments and Text Effects in Children's Reading Comprehension. Unpublished D Phil Thesis. University of London Institute of Education.

