

**The Relationship Between Reading Comprehension and On-line
Inference Generation in Children.**

Claudine Bowyer-Crane, B.Sc.

**A thesis submitted for the degree of Doctor of Philosophy in the
Department of Psychology, University of York.**

September, 2002.

Abstract

This thesis investigated the relationship between on-line inference generation and reading comprehension in children. An initial investigation was conducted in order to assess the degree of inferential ability required to meet the demands of standardised tests of reading comprehension. While all of the tests required some degree of inference generation, the tests varied in the extent to which they required inferences based in real-world knowledge and those that required the use of cohesive devices. Moreover, the tests also varied in the extent to which they required inferences that were necessary for maintaining textual coherence. Children's ability to answer questions that required inferences was related to their reading comprehension ability and those with less well developed comprehension skills performed particularly poorly on inferences requiring the application of real-world knowledge to the information provided in the text.

The remainder of this thesis described a series of studies that were conducted using an on-line passage reading and sentence verification task in which children read a series of passages presented on the screen. At the end of each passage, children were required to judge whether a target sentence was true or false, using a key-press response on the computer keyboard. Each target sentence related to specific information embedded in the passage. Three types of test sentence were used; sentences that relayed information explicitly stated in the text, sentences that represented the mediating idea underlying a causal inference required for textual coherence, or sentences that represented a possible elaboration from the text. Off-line comprehension questions relating to the experimental passages and additional measures of cognitive ability were also used.

The results presented in Chapters 4, 5, 6, and 7 indicate that children generate causal inferences but not elaborative inferences on-line during reading. In Chapter 4, evidence is presented to suggest that the same pattern of inference generation is found in texts of both a narrative and expository nature, and thus that text genre does not affect inference generation. Similarly, the results presented in Chapters 5, 6, and 7 suggest that children are able to generate inferences from texts containing information that is inconsistent with real world knowledge. These results are discussed in terms of previous

research into inference generation, the development of logical reasoning and the belief bias effect.

Correlational analyses reveal a consistent relationship between inference generation and reading comprehension throughout this thesis. Moreover, a developmental trend revealed in Chapter 6 suggests that reading skill and not chronological age is important in the development of inferential skills. However, in contrast to previous research, the results presented in Chapter 7 reveal that children with specific comprehension difficulties were able to generate causal inferences on-line. These results suggest that, while performance on the inference generation task clearly depended on reading comprehension skill, other factors also played a role. These results are discussed in terms of possible differences in working memory capacity and levels of strategic processing.

Table of Contents

Abstract	ii
Table of Contents	iv
List of Tables and Figures	ix
Acknowledgements	xii
Introduction	xiii
CHAPTER 1	
An Introduction to Reading Comprehension and Inference Generation	
1.1 Reading Comprehension: Typical and Atypical Development	1
1.1.1 Components of Reading Comprehension	1
a) Decoding	1
b) Listening Comprehension	3
c) Vocabulary	4
d) Suppression	5
e) Working Memory	9
1.1.2 The Simple View of Reading	12
1.1.3 Children with Specific Comrehension Difficulties	14
a) General Cognitive Ability	14
b) Inferential Skills in Poor Comprehenders	15
c) Semantic Skills in Poor Comprehenders	20
d) General Language Skills in Poor Comprehenders	22
1.1.4 Summary	26
1.2 Inference Types and Models of Inference Processing	26
1.2.1 Types of Inference	28
1.2.2 Two Accounts of Inference Generation	30
1.2.3 The Constructionist Theory	32
1.2.4 The Minimalist Hypothesis	35
1.2.5 Evidence for the Constructionist Theory and the Minimalist Hypothesis	38
1.2.6 Summary	39

CHAPTER 2

Methodological Considerations In The Study Of Inference Generation

2.1 Introduction	41
2.2 Off-Line Measures of Inference Generation	41
2.3 On-Line Measures of Inference Generation	42
2.3.1 Reading Time	43
2.3.2 Activation Measures	44
a) Lexical Decision	44
b) Naming Speed	45
c) Sentence Verification	47
2.3.4 Eye-Movements	49
2.4 The Three-Pronged Approach	49
2.5 Conclusions	51

CHAPTER 3

An Investigation into the Relationship between Reading Comprehension and Inference Generation

3.1 Introduction	52
3.2 Experiment 1	53
3.2.1 Method	53
a) Participants	53
b) Design	53
c) Materials	53
d) Procedure	57
3.2.2 Results	58
3.2.3 Discussion	66

CHAPTER 4

The Effect of Text Format on Inference Generation

4.1 Introduction	69
4.1.1 Structural Differences	69

4.1.2 Familiarity of Content	70
4.1.3 Readers Purpose	70
4.2 Research into the Effect of Text Format on Inference Generation	71
4.2.1 Inference Generation in Expository Text	71
4.2.2 Comparisons of Inference Generation in Narrative and Expository Text	74
4.2.3 Other Factors Affecting Inference Generation	75
4.3 Material Appropriate Processing	78
4.4 Methodological Considerations Surrounding Research into the Effects of Text Format on Inference Generation	80
4.5 Experiment 2	82
4.5.1 Method	82
a) Participants	82
b) Design	82
c) Materials	83
d) Procedure	86
4.5.2 Results	88
4.5.3 Discussion	94
CHAPTER 5	
On-Line Inference Generation During Children's Reading of Fairy Tales	
5.1 Introduction	99
5.2 Reasoning from False Premises: Evidence from Deductive Reasoning	99
5.3 Generating Text-Based Inferences from False Premises	105
5.4 Experiment 3	107
5.4.1 Method	107
a) Participants	107
b) Design	108
c) Materials	108
d) Procedure	113
5.4.2 Results	115
5.4.3 Discussion	122

CHAPTER 6

The Development of Children's On-Line Inference Generation During Reading

6.1 The Development of Inferential Skills	127
6.2 Experiment 4	133
6.2.1 Method	133
a) Participants	133
b) Design	134
c) Materials	134
d) Procedure	140
6.2.2 Results	140
6.2.3 Discussion	150

CHAPTER 7

Inferential Skills of Children with Specific Comprehension Difficulties

7.1 Introduction	155
7.2 Experiment 5	157
7.2.1 Method	157
a) Design	157
b) Participants	157
c) Materials	158
d) Procedure	159
7.2.2 Results	159
7.2.3 Discussion	164

CHAPTER 8

General Discussion

8.1 Summary of Findings	168
8.2 Discussion of Findings	171

8.2.1 The Effect of Text Genre, Content and Reading Skill on Inference Generation during Reading	171
8.2.2 The Time-Course of Inference Generation	177
8.2.3 The Relationship between Reading Comprehension and Inference Generation	181
8.3 Conclusions	186
8.4 Educational Implications	186
REFERENCES	189
 APPENDICES	
Appendix I Experiment 2 Test Materials	202
Appendix II Experiment 2b	224
Appendix III Experiment 3 Test Materials	227
Appendix IV Experiment 4 and Experiment 5 Test Materials	248
Appendix V Experiment 4 Off-Line Validation Materials	263

List of Tables and Figures

Table 1.1 Summary of different inference types and predictions of constructionist theory and minimalist hypothesis regarding inference generation.	37
Table 3.1 Mean chronological age of children in each year group.	53
Table 3.2 Overall mean scores of each year group on each standardised test.	58
Table 3.3 Zero-order and partial correlations (below diagonal) between raw scores obtained on standardised tests.	59
Table 3.4 Multiple regression with NARA II Comprehension as the dependent variable and Chronological Age, NARA II Acc, GNWRT, Mathematical Reasoning and WORD Comprehension as the predictor variables.	61
Table 3.5 Multiple regression with Mathematical Reasoning as the dependent variable and Chronological Age, Numerical Operations, NARA II Comprehension and WORD Comprehension as the predictor variables.	62
Table 3.6 Percentage of each category of question identified in each standardised test.	63
Table 3.7 Mean standard scores of Skilled and Less-Skilled Comprehenders on NARA II and WORD tests of reading comprehension.	63
Table 4.1 Mean chronological age and Suffolk Reading Scale (SRS) standard score for each of the two participant groups.	82
Table 4.2 Example of expository and narrative test passages, target sentences and comprehension questions.	85
Table 4.3 Mean response time and number of correct responses across subjects for each Test Sentence by Content condition at each Format Level.	89
Table 4.4 Mean coherence break reading times for passages of each format in each condition.	90
Table 4.5 Mean scores on standardised measures of reading, vocabulary and comprehension.	92
Table 4.6 Zero-order correlations between target sentence response times and performance on standardised measures, and partial correlations (below diagonal) between these variables when reading accuracy is controlled for.	93
Table 4.7 Multiple regression with NARA II comprehension as the dependent variable and response times to target sentences, chronological age and WISC	94

vocabulary as the predictor variables.

Table 5.1 Examples of valid arguments with sound and unsound premises.	100
Table 5.2 Examples of Real and Fairy story passage types, test sentences and comprehension questions.	110
Table 5.3 Mean no of syllables and book grades for experimental passages.	112
Table 5.4 Mean reading times (secs) for critical sentences in Early and Late sections of Real and Fairy passages.	115
Table 5.5 Mean response times and number of correct responses for target statements in each experimental condition.	116
Table 5.6 Mean standard scores on the standardised assessment measures.	118
Table 5.7 Zero order correlations between standardised measures, response times, error rates and experimental comprehension scores and partial correlations (<i>in italics</i>) between these variables when age and TOWRE are controlled for.	120
Table 5.8 Multiple regression with NARA II comprehension as the dependent variable and TOWRE, WISC Vocabulary, target sentence response times and target sentence accuracy as the predictor variables.	122
Table 6.1 Mean standard Suffolk Reading Scale (SRS) score, chronological age (months), NARA II reading accuracy age (months) and NARA II comprehension standard score for each experimental group.	134
Table 6.2 Examples of Real and Fairy story types, test sentences and comprehension questions.	136
Table 6.3 Mean number of syllables and book grades for experimental passages.	137
Table 6.4 Example of story and questions presented to raters in the off-line validation task.	138
Table 6.5 Mean test sentence response times and number of correct responses for each Reading Group in each condition.	143
Table 6.6 Mean performance on experimental comprehension questions for each Reading Group.	146
Table 6.7 Standard scores of each Reading Group on the standardised assessment measures.	147
Table 6.8 Zero-order correlations between response times, response accuracy, experimental comprehension scores and performance on the standardised	148

measures and partial correlations (below diagonal) between these variables when age is controlled for.

Table 6.9 Multiple regression analyses with NARA II Reading Comprehension as the dependent variable and Chronological Age, WISC vocabulary, RT Causal, RT Static, Causal Accuracy and Static Accuracy as the predictor variables. 150

Table 7.1 Mean performance of Less-Skilled Comprehenders and RA Controls on tests of reading and language ability 158

Table 7.2 Mean response time and number of correct responses to target sentences for Less-Skilled Comprehenders and RA controls 159

Table 7.3 Mean performance of Less-Skilled Comprehenders and RA controls on experimental comprehension questions. 163

Table 7.4 Multiple regressions with Causal Sentence RT as the dependent variable and Wisc III Vocabulary, Static Sentence RT and Group as the predictor variables (N=20). 163

Table A2.1 Mean response time and mean number of correct responses to each test sentence. 226

Figure 3.1 Performance of Skilled and Less-Skilled Comprehenders on each question type identified in the standardised measures. 65

Figure 6.1 Mean response time of each experimental group in each test sentence condition. 141

Figure 6.2 Mean number of correct responses of each experimental group in each condition. 142

Figure 6.3 The relationship between speed and accuracy of responding to Causal and Static test sentences for the Skilled group. 144

Figure 6.4 The relationship between speed and accuracy of responding to Causal and Static test sentences for the Less-Skilled group. 145

Figure 7.1 Mean response time of Less-Skilled comprehenders and RA Controls to Causal and Static test sentences. 161

Acknowledgements.

I have been fortunate in receiving a great deal of support over the past three years and I would like to thank everyone who has helped me in the completion of this thesis. However, there are a number of people who deserve special thanks.

Firstly, I would like to thank my supervisor, Maggie Snowling, for her support, advice and patience over the past three years. Without her help this thesis would not have been completed.

I am grateful to all of the children who took part in this research and would like to thank them for all their hard work. I would also like to thank the staff of all the schools I worked in over the past three years for their patience and co-operation.

A special note of thanks must go to all the members of 'the Zoo', both past and present for companionship and empathy. In particular I would like to thank Drew for his continued support and friendship over the past 6 years. I would also like to thank Jo for sharing with me the frustrations of studying whilst bringing up a toddler.

I will always be indebted to my Mum and Dad who have never stopped believing in me and who have supported me in ways too numerous to mention. Finally I must thank Michael, for his infinite patience, without whom I would never have got this far, and my daughter, Anaïs, to whom I dedicate this thesis, for reminding me that there is more to life than gaining a PhD.

Introduction.

The motivation for this thesis evolved from an interest in the reasoning ability of children with specific comprehension difficulties. The initial research question asked whether children with specific comprehension difficulties could perform reasoning tasks as well as children with normally developing reading skills. The point of departure for this research was to define more clearly the kind of reasoning ability that was to be investigated. An initial review of the literature revealed that while deductive reasoning ability had been looked at extensively in children and adults, research into inference generation during reading had largely been conducted with adult participants. Moreover, while a comprehensive body of research into text-based inference generation did exist in the developmental literature, it did not fully address the question of how inference generation relates to reading comprehension. Arguably, the kind of reasoning tasks children will be faced with on a regular basis are those embedded in their everyday reading experience i.e. explaining the actions of a character or following a sequence of events using information that is implied by, but not explicitly stated in the story. The ability of children to conduct this kind of reasoning task and how that skill relates to reading comprehension is therefore an interesting question. With this in mind, I chose to address the initial research question within the context of inference generation during reading rather than deductive reasoning.

However, it became apparent that there were some unanswered questions that needed to be considered before the original research question could be tackled. First, it was important to establish a pattern of inference generation in children with normally developing reading skills. Children are presented with written material in many different forms. For example, as well as reading stories, children are also presented with text written in a more informative, expository manner, as in a textbook. Also, the stories children read might contain information that is not based in real world knowledge, as in fairy tales. It was not clear from the developmental literature whether children show a consistent pattern of inference generation across these different types of text or whether inference generation is affected by the type of text a child is presented with. Moreover, much of the adult literature demonstrates a well-established pattern of inference generation such that causal inferences are generated on-line during reading and elaborative inferences are generated off-line. This pattern of inference generation had not been fully established in children and

it was not clear whether children would show this on-line/off-line distinction between different inference types. These questions needed to be addressed before one could compare the inference generation of normally developing readers with that of children with specific comprehension ability.

To this end, the focus of this thesis moved from the original research question and concentrated on establishing a consistent pattern of inference generation in normally developing readers, thus adding to the developmental literature in this area. Furthermore, this thesis aimed to expand the literature by using a systematic on-line measure of inference generation, resulting in a clearer picture regarding the time-course of inference generation in children. In line with the initial research question posed at the outset of this research, the relationship between inference generation and reading comprehension is considered throughout this thesis although it is not the main focus of the research. However, this relationship is looked at in more detail in a final study comparing the inference generation of children with normally developing reading skills to those with specific comprehension difficulties. Thus, this thesis has two main aims; to establish a pattern of inference generation in children with normally developing reading skills and to explore the relationship between inference generation and reading comprehension.

Chapter 1

An Introduction to Reading Comprehension and Inference Generation.

1.1 Reading Comprehension: Typical and Atypical Development

The ability to read is one of the most important skills for a child to develop. Indeed, the importance of this skill is reflected in the emphasis the National Curriculum currently gives to reading instruction. However, becoming a skilled reader requires the development of a number of different skills. While the ability to decode words is clearly important to reading, children must also be able to access the meanings of words. Moreover, they must be able to integrate the individual word meanings in order to understanding the meaning of the sentence or paragraph they are reading as a whole. In other words, as well as developing basic decoding skills, children must also develop skills that allow them to understand what they are reading. The exact nature and complexity of these skills has been the subject of much research. While the aim of this thesis is to investigate the importance of inference generation to the development of reading comprehension skills, the following review provides an overview of a number of other skills thought to be important to reading comprehension.

1.1.1 Components of Reading Comprehension

Reading comprehension has traditionally been considered a complex process tapping a number of sub-components including decoding, listening comprehension, working memory, vocabulary and suppression. This review considers each of these individual components in turn.

a) Decoding

Decoding is defined by Hoover and Gough (1990) as "the ability to derive a representation from printed input that allows access to the appropriate entry in the mental lexicon, and thus, the retrieval of semantic information at the word level" (Hoover & Gough, 1990, p130). In other words, decoding is the ability to recognise individual printed words and to retrieve their meanings from the mental lexicon. Decoding has consistently been shown to be an important component of reading comprehension. For example, in a recent study of 96 American children, Zinar (2000) investigated the relative contributions of word identification and comprehension monitoring behaviour to reading comprehension using a series of standardised tests and a comprehension monitoring task. Results of a zero-order correlation between all the

tasks used in the study showed that scores on a test of single word reading, in this case the Word Identification subtest of the *Woodcock-Johnson Psycho-Educational Battery – Revised* (Woodcock & Johnson, 1989, 1990), correlated more highly with reading comprehension ($r = 0.68, p < .001$) than any of the other tasks used. De Jong and Van der Leij (2002) extended these findings to a transparent orthography by administering a series of tests to Dutch children in Grade 1 and then again at the end of Grade 3. The battery of tests included measures of receptive and expressive vocabulary, phonological awareness and word decoding, and listening and reading comprehension. Results of a hierarchical regression analysis revealed that reading comprehension at Grade 3 was significantly predicted by decoding skills at Grade 1 after reading comprehension performance at Grade 1 was controlled. This finding supports previous research which asserts that decoding is a reliable predictor of reading comprehension (Curtis, 1980; Hoover & Gough, 1990; Stanovich, Cunningham, & Feeman, 1984; Stanovich, Nathan, & Vala-Rossi, 1986)

However, research has also demonstrated that the importance of decoding to reading comprehension alters over time (Curtis, 1980; Willson & Rupley, 1997). For example, Willson and Rupley (1997) investigated the relationship between background knowledge, strategy knowledge, phonemic knowledge and reading comprehension in approximately 1200 children in grades 1 to 6. Background knowledge was defined as familiarity with the specific topic of the text being read and strategy knowledge was defined as knowledge of how to engage with a text. Using the structural equation modelling technique, Willson and Rupley found that while decoding skill directly affected reading comprehension performance at grades 2 and 3, its contribution to reading comprehension was outweighed by that of strategy knowledge by grades 5 and 6. This suggests that, while decoding is important in the early stages of reading comprehension development, more advanced reading comprehension is driven by top-down strategic processes. In contrast, Stanovich et al. (1986) found that while decoding skill was a reliable predictor of reading ability in fifth-grade children, reading ability in the third-grade was related to vocabulary knowledge and word-recognition speed. Thus, the change in the contribution of decoding skill to reading ability over time is not clear.

b) Listening Comprehension

Listening comprehension has consistently been identified as a reliable predictor of reading comprehension (Curtis, 1980; de Jong & van der Leij, 2002; Sears & Keogh, 1993). For example, in the study by de Jong and Van der Leij (2002) described above, listening comprehension at grade 1 was a significant predictor of reading comprehension at grade 3 after reading comprehension at grade 1 was controlled. Curtis (1980) gave a battery of standardised tests to 100 children in grades 2, 3 and 4. The tests included measures of reading and listening comprehension, memory span, word matching and vocalization. In the word matching task, two words or pseudowords were presented together and the children were required to press a 'yes' key if both items were the same or a 'no' key if the items were different. Thus, this task involved the recognition of orthographic patterns and as such can be considered a measure of decoding. The vocalization task was administered to measure phonological decoding. In this task, children were shown letters, words and pseudowords one at a time on a computer screen. They were asked to say the name of the item out loud as soon as they had identified it or, in a second condition, after a 3-second delay.

Correlational analysis showed that, while second-grader's reading comprehension did not significantly correlate with listening comprehension, the correlation between these two variables was highly significant for third and fifth graders (third grade $r = 0.66$, $p < .001$; fifth grade $r = 0.74$, $p < .01$). In addition, multiple regression analyses showed that listening comprehension was a significant predictor of reading comprehension in the third and fifth grades. 23% of the unique variance in reading comprehension scores was accounted for by listening comprehension in third grade children, a further 6% by word vocalization, 3% by letter matching and 6% by a combination of other variables. 35% of the unique variance in reading comprehension was accounted for by listening comprehension in the fifth grade, a further 3% by pseudoword vocalization, and 5% by other variables. Listening comprehension did not predict reading comprehension in the second grade children.

Curtis also looked at the relationship between these variables in three groups of skilled and less-skilled readers. For less-skilled readers and the younger skilled readers, listening comprehension accounted for only a small proportion of the variance in reading comprehension (3% and 2% respectively). However, for older skilled readers, listening comprehension accounted for 28% of the variance in reading comprehension

ability. Stanovich, Cunningham and Feeman (1984) reported similar findings such that for second grade children, listening comprehension made no significant contribution to reading comprehension after general intelligence was controlled. In contrast, for third and fifth grade children, listening comprehension accounted for 14% ($p < .05$) and 13% (ns) respectively of the variance in reading comprehension after non-verbal ability had been controlled. However, when decoding skill was controlled, the contribution of listening comprehension to reading comprehension dramatically decreased. These findings suggest that, while listening comprehension clearly makes a significant contribution to reading comprehension, it is mediated by the development of basic decoding skills.

c) Vocabulary

Previous research has found a consistent relationship between word-recognition and reading comprehension. For example, Seignuriac, Ehrlich, Oakhill and Yuill (2000) investigated the relative contribution of working memory, vocabulary knowledge and decoding to reading comprehension in a sample of fourth grade children using a series of regression analyses. In each of five simultaneous regression analyses, vocabulary was shown to be the best predictor of reading comprehension. Further evidence for the contribution of vocabulary knowledge to reading comprehension comes from the finding that children with comprehension difficulties also demonstrate poor vocabulary knowledge. For example, Nation and Snowling (1998a) measured the receptive and expressive vocabulary knowledge of children with specific comprehension difficulties using the *Test of Word Knowledge* (TOWK; Wiig & Secord, 1992). Children with comprehension difficulties showed poorer receptive and expressive vocabulary than children with good comprehension who were matched for decoding skill and non-verbal ability.

In addition to the evidence discussed above, the contribution of vocabulary to reading comprehension has been shown by a number of instructional studies. For example, Beck, Perfetti and McKeown (1982) implemented a long-term, in-depth instructional programme aimed at improving the vocabulary knowledge of 27 children aged approximately 9 years. Over a five-month period, the children taking part in the instructional programme were given a variety of tasks including games and group exercises during which they were taught 104 new words. Following training, the children were given three tests of comprehension; a semantic decision task, a sentence

verification task and a story recall task. Performance on all three tests of comprehension was better for the instructional group than for a control group who were given no vocabulary training.

This study was replicated and refined by McKeown, Beck, Omanson and Perfetti (1982). In addition to the semantic decision and story recall tasks used in the previous study, McKeown et al. (1982) included a multiple-choice comprehension task in which children were asked questions about a passage they had previously read. Results showed that children who received long-term vocabulary instruction were more accurate on the multi-choice task than the control group when the passages that the questions related to contained examples of the newly learned vocabulary. Thus, instruction aimed at improving vocabulary knowledge also enhanced reading comprehension, supporting the assumption that vocabulary knowledge is an important component of the reading comprehension process. Similar results have been found in other studies implementing vocabulary instruction programmes (Kameenui, Carnine, & Freschi, 1982). Taken together, these findings suggest that vocabulary knowledge is a fundamental component of reading comprehension ability.

d) Suppression

A further component thought to be important to reading comprehension is the suppression of irrelevant information. Gernsbacher and Faust (1991) suggest that during comprehension, readers activate many potential meanings of the words they encounter. They propose that the activation of these competing meanings is suppressed or enhanced in accordance with the contextual, semantic and syntactic constraints of the surrounding passage by means of a "structure building framework" in which modulation of the different levels of activation eventually results in the sole activation of the contextually appropriate meaning. They further propose that less-skilled comprehenders may have less efficient suppression mechanisms, resulting in an inability to reject the inappropriate meanings of words.

In support of this hypothesis, Gernsbacher and Faust (1991) conducted a series of experiments investigating the ability of skilled and less-skilled comprehenders to suppress linguistic and non-linguistic information. For example, participants were presented with short sentences followed by a test word and were asked to state whether the test word fitted the meaning of the sentence. On some of the trials, the final word of

the sentence was a homophone, that is one of two or more words with the same pronunciation but different meanings and written forms (e.g. 'patients' and 'patience'). On trials containing homophones, the test word presented after the sentence was related to an alternative form of the sentence-final homophone e.g. following the sentence 'he had lots of patients', the test word 'calm' would be presented. Gernsbacher and Faust (1991) predicted that if the incorrect form of the homophone was activated, there would be a large difference between the response latencies to test words following homophones and those to test words following non-homophones. The time taken to reject test words after reading sentences containing homophones was compared with sentences containing non-homophones (e.g. he had lots of students). The results demonstrated that when subjects were asked to respond immediately after the test word was presented, both skilled and less-skilled comprehenders showed a large difference in response times to homophone versus non-homophone conditions indicating that both groups of readers had activated the incorrect form of the homophone. However, when asked to respond after a delay of 100 msecs, the skilled comprehenders no longer showed a difference in response times to homophone versus non-homophone conditions indicating that they had successfully suppressed the irrelevant form of the homophone. In contrast, less-skilled comprehenders continued to demonstrate slower response latencies in the homophone condition, indicating that the incorrect form of the homophone was still active after a 1-second delay.

In another experiment, Gernsbacher and Faust (1991) investigated the abilities of skilled and less-skilled comprehenders to suppress information across modalities. They first presented participants with a series of context displays, each containing a picture with a word superimposed onto it and participants were informed whether they were being given a picture trial or a word trial. For the picture trials, participants were asked to ignore the word and concentrate on the picture in the context display. In the word trials, participants were asked to concentrate on the word while ignoring the picture. Participants were then presented with a test display. In the picture trials, participants were shown another picture and were asked to state whether the picture in the test display was related to the picture in the context display. Conversely, for the word trials, participants were shown another word and were asked to state whether the word in the test display was related to the word seen in the context display. On some of the trials of each type, the test display was related to the item they were instructed to ignore in the context display. Less-skilled comprehenders were slower to reject test

items on these trials than skilled comprehenders, adding further support to the hypothesis that less-skilled comprehenders have less efficient suppression mechanisms.

Similar results were found when subjects were asked to reject named objects typical of, but not present in, previously presented arrays of objects. For example, participants were presented with pictures of objects typical of a farm scene. They were then presented with the word 'tractor', an object commonly found in a farm scene but one that was not present in the previously presented array. Gernsbacher and Faust (1991) found that, after a delay of 100 msec, less-skilled comprehenders were slower than skilled comprehenders to reject objects that were typical of but absent from previous presented arrays of objects. Once again, this finding provides support for the hypothesis of a less efficient suppression mechanism in less-skilled comprehenders. However, Gernsbacher and Faust showed that less-skilled comprehenders were equally as aware as skilled comprehenders of what was contextually appropriate. For example, on some of the trials in experiment 2, the test object presented was typical of and present in the previous array (e.g. a tractor presented with an array of objects from a farm scene), while on other trials, the test object was atypical of but present in the previous array (e.g. a tractor presented with an array of objects from a kitchen scene). On these trials, both skilled and less-skilled comprehenders were quicker to accept the test word as being present in the previous array of objects if it was typical of the scene suggested by those objects. Therefore, the problems demonstrated by less-skilled comprehenders in rejecting inappropriate information could not be attributed to a lack of awareness of what is contextually appropriate and as such were more likely the result of a less-efficient suppression mechanism.

These findings are supported by an earlier study by Merrill, Sperber and McCauley (1981) in which the performance of good and poor comprehenders was measured on a stroop task. Merrill et al. presented participants with target words presented in coloured ink that were immediately preceded by a sentence context. Participants were asked to name the colour of the target as quickly as possible. The target words varied in their relatedness to the preceding context. For example, following the sentence 'the girl fought the cat', a related target word would be 'claw', while an unrelated word would be 'fur'. However, both target words were semantic properties of the object in the preceding sentence (e.g. cat).

Analysis of response latencies showed that, for good comprehenders, response latencies were faster when the target word was related to the preceding sentence context, than when it was unrelated. However, poor comprehenders showed no difference in response times to related or unrelated target words. This finding suggests that good comprehenders had activated only those semantic properties of the object that were consistent with the context of the sentence. In contrast, the poor comprehenders did not suppress the semantic properties of the object that were irrelevant to the sentence context, thus experiencing an equal amount of interference from related and unrelated target words.

Recent work by De Beni and Palladino (2000) analysed the recall errors made by skilled and less-skilled comprehenders on a version of the listening span test (Daneman & Carpenter, 1980) in which they were asked to listen to sequences of two, three, four and five sentences and make true/false judgements about them. They were also asked to remember the sentence-final word in each sentence. Less-skilled comprehenders made more 'intrusion' errors than skilled comprehenders. That is to say, less-skilled comprehenders were more likely to recall a word that was present in the previous sequence of sentences but not in the final position. Thus, less-skilled comprehenders were less able to suppress the irrelevant information in the sentences in order to recall the sentence-final word. Furthermore, regression analysis revealed that number of intrusion errors significantly predicted reading comprehension performance one year later, explaining 43% of the variance. In a second task, when asked to read through a passage and identify idea units related to the title, skilled comprehenders identified more relevant sentences than poor comprehenders. When asked to recall the passage, skilled comprehenders recalled more relevant parts of the passage than poor comprehenders. Furthermore, although both groups of children recalled an equal amount of irrelevant information, poor comprehenders recalled more irrelevant than relevant information.

Palladino, Cornoldi, De Beni and Pazzaglia (2001) presented skilled and less-skilled comprehenders with an updating task in which children heard sequences of 4, 6, 8 or 10 words and were instructed to remember the last four words in the sequence. They were not told prior to the presentation of each sequence how many words they would hear. Thus, they had to update their memory representation of the sequence with the presentation of each new word. They predicted that less-skilled comprehenders

would find this task more difficult than skilled comprehenders because of their reported difficulties with suppression. In contrast to this prediction, both groups of children performed at a comparable level on this task.

In a second experiment, Palladino et al. (2001) presented children with lists of twelve words, half of which were "sizeable" concrete words (e.g. names of objects or animals that were measurable by size) and half of which were abstract words. Participants were instructed to listen to the words and remember the three smallest objects or animals presented in the list. For example, participants presented with the following list '*meeting, sense, woodpecker, passion, law, cow, happiness, amount, caterpillar, lamb, feast and frog*' would have to recall the words '*woodpecker, frog and caterpillar*'. Analysis of the recall errors made by the less-skilled comprehenders showed that they made more intrusion errors than skilled comprehenders.

In a third experiment, Palladino et al. manipulated the suppression element of the task by including more concrete and less abstract words in the list, thereby enlarging the pool of relevant information from which the target words had to be extracted. Analysis of the recall errors again showed that less-skilled comprehenders made more intrusion errors than skilled comprehenders. Furthermore, while the number of intrusion errors made by skilled comprehenders was not affected by the increased suppression load, less-skilled comprehenders made twice as many intrusion errors when a higher degree of suppression was required. Together with the findings of Gernsbacher and Faust (1991), Merrill et al. (1981), and De Beni and Palladino (2000), these results suggest that the suppression of irrelevant information is an important component of comprehension.

e) Working Memory

Working memory is thought to be a fundamental component of reading comprehension because of the need to store information during reading and integrate that stored information with incoming information (Daneman & Carpenter, 1980). Daneman and Carpenter (1980) devised a reading span task in which subjects read aloud a list of sentences and then recalled the last word of each sentence. Thus, the task comprised both a processing and a storage component and was therefore considered a reliable measure of working memory. Having administered this task to a group of college students, Daneman and Carpenter (1980) found that performance on the reading

span task correlated with three separate measures of reading comprehension. Moreover, performance on a simple recall task in which participants had to recall lists of single words did not correlate significantly with any of the reading comprehension tests administered. These results suggest that reading comprehension is strongly mediated by working memory.

Yuill, Oakhill & Parkin (1989) compared the performance of skilled and less-skilled comprehenders on a task in which they were asked to read lists of groups of three digits aloud and then recall the last digit in each of the groups. Children were given lists of two, three and four groups of digits. Results indicated that for lists containing two groups of digits, there was no significant difference between groups in mean number of final digits recalled. However, the less-skilled comprehender group performed significantly worse on tasks requiring recall of three or four digits.

Yuill et al. also gave participants a text-based task requiring the resolution of text anomalies using information placed elsewhere in the passage. For example, the anomaly of a child being praised for not sharing his sweets with his brother was resolved by the information that his brother was very fat and was on a diet. In the passages presented to participants, the anomaly and resolution information was either contained in adjacent sentences or separated by additional sentences, thus increasing the working memory load. Following the presentation of each passage, the children were asked a number of comprehension questions. The results of this task indicated that when the anomaly and resolution information were separated in the text, less-skilled comprehenders performed significantly worse on the comprehension questions than skilled comprehenders. Taken together the findings of Yuill et al. suggest that children's text comprehension is at least partially reliant on efficient working memory.

More recently, Leather and Henry (1994) administered a battery of 11 tests to 71 children with a mean age of 7 years 5 months. The battery included measures of receptive vocabulary, arithmetic skill, reading accuracy and comprehension, memory span and phonological awareness. Four memory span measures were used; two simple span tasks and two complex span tasks. In both of the simple span tasks, children were required to listen to lists of words and repeat them back in the correct order. The words were either rhyming words (phonemic span task) or non-rhyming words (control span task). Thus, these tasks can be considered to test storage capacity. In contrast, the

complex span tasks included a processing component as well as a storage component and thus can be considered measures of working memory. In the complex counting span task, children were presented with a series of cards on which were a number of dots to count. After all the cards in the series had been counted, the children were asked to recall the number of dots on each card in the order in which they were presented. In the complex listening memory span test, children were required to listen to a series of sentences and fill in the missing final word in each sentence. Having listened to a group of two or more sentences, the children were asked to recall the sentence-final words in the correct order.

Correlational analyses revealed a significant relationship between reading comprehension and both complex span tasks (counting span $r = 0.36$, $p < .01$; listening span $r = 0.61$, $p < .01$) which was not reflected in the relationship between reading comprehension and the simple span tasks (phonemic span $r = -.05$, ns; control span $r = 0.30$, ns). Moreover, a series of fixed order regression analyses revealed that complex listening span scores accounted for between 9% and 33% of the variance in reading comprehension performance ($p < .01$). In contrast, complex counting span accounted for a maximum of 11% of the variance in reading comprehension ($p < .01$) and was only a significant predictor if listening span and phonemic awareness were not controlled. These results suggest that verbal working memory is a significant predictor of reading comprehension, a finding replicated in more recent research (Seigneuric et al., 2000). In addition to this, Nation, Adams, Bowyer-Crane & Snowling (1999) found that children with poor comprehension had normal working memory for spatial tasks, but their verbal working memory was impaired, findings similar to an earlier study by Swanson and Berninger (1995). These findings are supported by Cataldo and Oakhill (2000) who demonstrated that a test of spatial memory did not differentiate between groups of good and poor comprehenders.

In the study by De Beni and Palladino (2000) described previously, skilled and less-skilled comprehenders were given two tests of verbal short-term memory; recall of digits forwards and recall of digits backwards. Neither of these memory tests differentiated between the two groups of children. However, a significant difference was found in the performance of good and poor comprehenders on a listening span test. Good comprehenders performed better than poor comprehenders when presented with four sentences, although their performance did not differ when given two, three and five

sentences. Similarly, on the tasks presented by Palladino et al. (2001), less-skilled comprehenders recalled fewer target words overall than skilled comprehenders, even when the memory load was increased to recall of the five smallest items in the list instead of the original three. Furthermore, at the end of experiments 2 and 3, participants were asked to recall all the words heard in each list. Results from these final recall tasks showed that skilled comprehenders consistently recalled more of the target items than less-skilled comprehenders. The results of experiment 2 also showed that both groups of children recalled a similar number of non-target concrete items, although this finding was not replicated in experiment 3. These results suggest that, while the two groups differed in terms of working memory, their memory span was similar in other respects. This finding reflects that of De Beni and Palladino (2000). Taken together the findings described above suggest that verbal working memory is a vital element of reading comprehension.

The preceding discussion highlights some of the typical processes thought to be important to reading comprehension. Moreover, it is clear from this discussion that previous research has failed to highlight the relative contribution of each of these components to the reading comprehension process, with researchers focusing on specific numbers of components in each study. However, Hoover and Gough's (1990) *Simple View of Reading* aims to overcome this difficulty and provide a comprehensive framework for considering the many different processes involved in reading comprehension. It is to this model of reading that we now turn.

1.1.2 The Simple View of Reading

The *Simple View of Reading* (Hoover & Gough, 1990) states that reading comprehension is reliant on the co-operation of two components; *decoding* and *linguistic comprehension*, and that the use of one component in the absence of the other does not constitute reading. *Decoding* can be defined simply as the ability to recognise words while *linguistic comprehension* can be defined as the ability to form higher level discourse representations using the information derived from word level decoding. In Hoover and Gough's model (1990), linguistic comprehension incorporates vocabulary knowledge and listening comprehension at the sentence, paragraph and text levels. Importantly, the model also states that the contribution of each of these components to efficient reading comprehension is equal, both skills being necessary for reading but

neither being sufficient alone. Importantly, the simple view of reading does not deny the importance of factors such as working memory or suppression to reading comprehension. Indeed, an important assumption of the simple view is that, while decoding and linguistic comprehension are the only components of reading comprehension that are specific to reading, other non-specific components may be utilised during reading comprehension. For example, Gernsbacher and Faust (1991) state that suppression is not specific to language processing but is a general cognitive mechanism. Therefore, in accordance with the simple view of reading, suppression and working memory would still play a role in reading comprehension.

The model further states that decoding and linguistic comprehension are separable components. This view is supported by a number of investigations looking at the relationship between linguistic comprehension and decoding in the typically developing reader. Such studies have revealed significant independent contributions of both decoding and linguistic comprehension to reading comprehension (Curtis, 1980; de Jong & van der Leij, 2002; Singer & Crouse, 1981; Stanovich et al., 1984). Indeed, Singer and Crouse (1981) used a measure of vocabulary as their linguistic comprehension measure thus adding support to the notion of a multi-faceted linguistic comprehension component. The separability of these two components is also evidenced by their dissociation in people with specific reading difficulties. For example, children with dyslexia have typically been demonstrated to have poor decoding skills in conjunction with intact comprehension ability. Conversely, a large percentage of children have comprehension skills which are impaired relative to both their chronological age and their reading age while their decoding skills are in line with, or even ahead of those expected for their age (Nation & Snowling, 1997; Yuill & Oakhill, 1991). Research has shown that this deficit in comprehension is not restricted to reading comprehension but is also evident in listening comprehension (Stothard, 1992), a finding consistent with the simple view of reading (Hoover & Gough, 1990).

The evidence cited above lends support to the simple view of reading and indeed to the notion that decoding and linguistic comprehension are separable constructs. However, for children with specific comprehension deficits, there is debate about the underlying cause of their difficulties. By investigating the comprehension deficits of such children in more detail, we can learn more about the multi-faceted nature of the linguistic comprehension component suggested by Hoover and Gough (1991).

1.1.3 Children with Specific Comprehension Difficulties

a) General Cognitive Ability

Children with specific comprehension difficulties in reading have been shown to have difficulties with a number of different tasks. In terms of general cognitive capabilities, an extensive study by Stothard (1992) revealed a distinct profile of cognitive skills in children with comprehension difficulties. Stothard compared a group of children with specific comprehension difficulties to two control groups; one matched for chronological age and one for comprehension age. Comparison of the performance of these groups on a standard IQ assessment battery revealed a specific deficit in verbal skills for the poor comprehender group. On two sub-tests of verbal IQ, both the chronological-age match control group and the comprehension-age match control group demonstrated age-appropriate verbal skills. However, the poor comprehender group performed at a level comparable to the younger comprehension-age match group, exhibiting weaker verbal skills than expected for their age. In contrast, on measures of performance IQ, all three groups of children performed at a level commensurate with their age. Moreover, unlike the two control groups, the poor comprehenders showed a significant discrepancy between verbal IQ and performance IQ. In contrast to the findings of Nation and Snowling (1998a) the less skilled comprehenders had normal receptive vocabulary indicating that they could access the meanings of words. However their expressive vocabulary was impaired suggesting that while they could find the correct meaning for a word they could not use that meaning in a wider context. Reading speed was comparable to that of chronological age-match controls indicating that their comprehension difficulties could not be due to inefficient decoding. Nor could their problems be attributed to phonological deficits since phonology was intact in the poor comprehender group, a finding since replicated in a study by Cain, Oakhill and Bryant (2000). Finally, it was established that the deficits seen in this group could not be attributed to problems with working memory since their working memory skills were normal.

In a recent study, Nation, Clarke and Snowling (in press) administered the *British Ability Scales II* (Elliot, Smith, & McUlloch, 1996) to 25 poor comprehenders and found a similar cognitive profile to that described by Stothard (1992) in the majority of children tested. 20 of the 25 children who participated in the study exhibited general cognitive ability (GCA) in the normal range (mean GCA 98.5). However, verbal IQ was impaired compared to non-verbal and spatial IQ (mean

standard scores 91, 101.1, & 104.5 respectively). Performance on the achievement scales administered revealed single word reading and number skills higher than expected on the basis of GCA. Furthermore, reading accuracy as measured by the NARA II was in line with their GCA while NARA II reading comprehension performance was significantly below that predicted by their GCA.

Taken together the findings of Stothard (1992) and Nation et al. (in press) indicate that the problems poor comprehenders experience may be due to generally weak verbal skills rather than specific comprehension skills. However, Nation et al. (in press) also discovered a sub-group of children with a different cognitive profile to the average poor comprehender. Five of the 20 children seen in this study showed generally low GCA (mean 76.4) with below average performance on verbal, non-verbal and spatial scales (mean standard scores 80.4, 74.8, & 85.8 respectively). Achievement scale performance showed single word reading and number skills markedly higher than GCA. Furthermore, in contrast to the majority of poor comprehenders, reading comprehension performance measured by the NARA II was in line with GCA while reading accuracy was significantly higher than expected on the basis of GCA. These findings suggest that caution needs to be taken when identifying poor comprehenders since they clearly form a heterogeneous group. While most poor comprehenders appear to be identified due to surprisingly low comprehension skills, a small sub-group may be erroneously classified as poor comprehenders because of unexpectedly high reading accuracy.

b) Inferential Skills in Poor Comprehenders

Poor comprehenders have been consistently shown to have weak inferential skills. Early work by Oakhill (1982) indicated that less-skilled comprehenders do not engage in active constructive processing in the same way as skilled comprehenders. Oakhill presented children with short 'stories' followed by a recognition test in which children were read two original sentences from the story, a valid inference foil which contained information inferred from the passage, or an invalid inference foil, containing information that could not be inferred by the passage. The task of the children was to say 'yes' if they had heard exactly the same sentence before, and 'no' if they had not. Skilled comprehenders made fewer errors than less-skilled comprehenders when presented with the original sentences and the invalid inference foils. However, skilled comprehenders erroneously recognised more valid inference foils as sentences they had

heard before than less-skilled comprehenders, indicating that skilled comprehenders were actively integrating information across sentences while less-skilled comprehenders were not.

In a later study, Oakhill (1983) presented skilled and less-skilled comprehenders with a series of 16 simple sentences. The children were then given a cued recall task in which they were presented with a 'clue' word and asked to recall the sentence that the word reminded them of. The 'clue' word was either an original noun taken from the sentence or an instantiation of that noun. For example, following the sentence 'the animal chased the mouse', the original cue word would be 'animal' and the instantiated cue word would be 'cat'. The results of the study showed that for skilled comprehenders, the use of an instantiated 'clue' word as a cue to recall was more effective than using the original noun. Recall in the instantiated condition was 71.9% while recall in the original condition was 45.8% for skilled comprehenders. In contrast, less-skilled comprehenders recalled significantly fewer sentences overall, their performance being poorest when given instantiated cue words: 35.4% in the instantiated condition and 38.5% in the original condition. In addition, less-skilled comprehenders were slower to respond to instantiated cue words than to original cue words, while skilled comprehenders showed no difference in response time to either cue type. Finally, when asked to choose two possible specific meanings of a noun, less-skilled comprehenders performed at a level comparable with the skilled comprehenders, indicating that their difficulties were not related to poorer general knowledge. For example, when the sentence 'the animal chased the mouse' was re-read, and participants were asked whether the animal in the sentence was more likely to be a cat or a rabbit, less-skilled comprehenders were just as likely to correctly choose the word 'cat' as skilled comprehenders. These results reflect the notion that, unlike skilled comprehenders, less-skilled comprehenders do not actively construct a mental representation of the text as they are reading.

In a study designed to look more closely at the inferential and memory skills of poor comprehenders, Oakhill (1984) presented skilled and less-skilled comprehenders with short stories to read, and subsequently asked them a series of questions about each story. Some of the questions tapped information that was explicitly stated in the story, while other questions required some degree of inference to provide a correct answer. When required to answer the questions from memory, the results showed that less-

skilled comprehenders were worse than skilled comprehenders on literal and inference questions. However, when the text was available to refer to, the advantage of the skilled comprehenders was only maintained for the inference questions.

Cain and Oakhill (1999) conducted a similar study and found that even when the stories were removed from sight before asking the questions, less-skilled comprehenders performed at a comparable level to chronological age-match peers on questions tapping literal information. However, their ability to answer questions tapping text-connecting inferences was inferior to children in both a chronological age-match group and a comprehension age-match group. Furthermore, the less-skilled comprehenders were significantly worse at answering questions that required a gap-filling inference than the chronological age-match group, although their performance was comparable to the comprehension age-match group. Interestingly, when prompted to refer to the text, the performance of the less-skilled comprehenders improved for the text-connecting inferences to a level comparable with both control groups. However, they continued to have difficulty with the gap-filling inferences, performing at a level comparable to the comprehension age-match controls.

Garnham, Oakhill and Johnson-Laird (1982 experiment 2) investigated the effect of referential continuity on text comprehension of skilled and less-skilled comprehenders aged between 7 and 8 years. Children were given examples of three types of text; the original version, a random version and a revised random version. In the original version of the story, the main character and object were referred to by name in the first sentence. In the second and third sentence, pronouns were used to refer to both character and object. The object and character were then referred to by name once more in the fourth and fifth sentence respectively. For example, the following is the original version of one of the passages presented:

'David was playing with his big, coloured ball in the garden. He bounced it so hard that it went right over the fence. The people next door were out so he climbed over to get it. He found his ball and threw it back. David carried on with his game.'

In the random version, the sentences of the text occurred in random order. In the revised random version, the sentences were presented in random order but some of the noun phrases were changed to re-establish referential continuity. For example, the

sentence 'he found his ball and threw it back' was changed to 'David found his big coloured ball and threw it back'. The children were asked to read and recall two stories in each of the three versions. The results showed that overall, skilled comprehenders recalled more of the stories from all three versions than poor comprehenders. However, both groups of children recalled more of the stories presented in the original version than in either the randomised or revised random version. Moreover, the skilled comprehenders showed a marked recall advantage for the revised random over the random stories while the less-skilled comprehenders showed no such advantage. Recall for the less-skilled comprehenders was comparable for both the revised and the revised random stories indicating that the additional referential continuity present in the revised random version did not benefit less-skilled comprehenders. Garnham et al. (1982) interpreted this finding as evidence for weak inferential skills in poor comprehenders.

Yuill and Oakhill (1986) looked at pronoun resolution in skilled and less-skilled comprehenders. Their results showed that poor comprehenders were less able to use gender cues to resolve pronouns than skilled comprehenders. In a later study, Yuill and Oakhill (1988a) examined poor comprehender's understanding of anaphoric relations in more detail. They looked at four particular types of anaphor; reference, ellipsis, substitution and lexical. Reference items were items where a pronoun referred to a previously mentioned object e.g. 'I have bought some sandwiches' Bill said, 'I hope you remembered yours', where *yours* refers to the aforementioned sandwiches. Ellipsis and substitution both refer to items where one thing is substituted for another. However, in the case of substitution, the replacement is explicit e.g. 'Who gave you your new fishing rod?' she asked. 'Mum did' said Bill', where *did* replaces *gave me the fishing rod*. In contrast, in the case of ellipsis, the replacement is implicit e.g. 'Are you going fishing?', 'Yes, I am', where *am* should be followed by *going fishing*. Finally, lexical items were items where an inference should be made to maintain cohesion e.g. 'The Captain was busy loading up his fishing boat. He fixed the motor in place...' where *motor* refers to the motor of the fishing boat.

Yuill and Oakhill (1988a) gave skilled and less-skilled comprehenders pre-training in these four types of anaphor. They then presented the children with a story containing examples of each of these anaphors. In some cases the anaphor and its referent were contained in adjacent sentences, while in other cases a number of intervening sentences separated the anaphor from its referent. The experimenter read

the story to the child twice, and on the second occasion, whenever an anaphor was encountered, the experimenter asked the child what each anaphor stood for or referred to. During the second reading, the text was available for the child to refer to. The results showed that for both groups of children, increasing the distance between anaphor and referent adversely affected performance on the task. However, for all four anaphor types, less-skilled comprehenders performed significantly worse than skilled comprehenders at one or all distance manipulations. These results indicate that less-skilled comprehenders are poorer at using lexical cues to maintain textual cohesion than skilled comprehenders, a finding later replicated by Yuill and Oakhill (1991).

Taken together, the studies described above indicate that children with specific comprehension difficulties have difficulty with inference processing. However, Cain and Oakhill (1999) found that less-skilled comprehenders could answer literal questions about a text equally as well as skilled comprehenders even when the text was covered, indicating that the problem is not one of memory for text. Similarly, the finding that the performance of less-skilled comprehenders improved for text-connecting inferences when prompted to refer to the text, indicates that it is not the case that poor comprehenders cannot make inferences. Rather, it would seem that the difference in performance between skilled and less-skilled comprehenders lies in the strategies used during reading and in particular, in the ability to use general knowledge to interpret a text.

This finding is supported by a training study conducted by Yuill and Oakhill (1988b) in which skilled and less-skilled comprehenders took part in three independent training programmes; inference training, comprehension exercises and rapid decoding training. The inference training programme comprised activities designed to encourage inference generation. For example, children were given a sentence such as 'Sleepy Tom was late for school again'. They then had to pick a word from the sentence and work out what information that word gave them about the sentence i.e. the name 'Tom' implies that subject of the sentence is male. Children were also encouraged to generate questions about sentences and use the clues in the stories to predict the content of hidden sentences. The comprehension exercises involved the experimenter reading the children a story and asking a variety of literal and inferential comprehension questions. Finally, the rapid decoding training involved the children being given a list of words taken from a story and were asked to practice reading them out loud accurately. The

children were then given the story to read. After reading the story the children were asked to re-read the list of words and the experimenter timed them with a stopwatch. At the beginning of the next session, the children re-read the previous word list again and the experimenter recorded the time.

Results demonstrated that the skilled comprehenders did not significantly benefit from any of the training programmes, showing little improvement in their comprehension skills. In contrast, children with comprehension difficulties who were given inference training showed a mean improvement in their comprehension age of 17.38 months over the course of the training programme. Similarly, less-skilled comprehenders given comprehension exercises showed a mean increase in their comprehension age of 13.71 months. However, less-skilled comprehenders given rapid decoding training showed little improvement in their comprehension skills at the end of the training programme.

c) Semantic Skills in Poor Comprehenders

Nation and colleagues (Nation, Marshall, & Snowling, 2001; Nation & Snowling, 1998a, 1999) investigated the semantic processing abilities of children with reading comprehension difficulties. Nation and Snowling (1998a) matched 16 poor comprehenders aged approximately 9 years with 16 normal readers of the same chronological age. Performance of the two groups of children was compared on tasks requiring access to semantic information e.g. synonym judgement and semantic fluency, and tasks requiring phonological awareness e.g. rhyme judgement and rhyme fluency. The results from these tasks showed that poor comprehenders performed at a level comparable to the normal readers on phonological tasks but were significantly poorer than normal readers on both synonym judgement and semantic fluency. The children were also given a word recognition task in which they were presented with a series of individual words and were asked to read each word as quickly as possible. The words presented varied in terms of regularity (i.e. regular and irregular spellings) and frequency (i.e. high or low frequency words). Results from the word recognition task showed that poor comprehenders performed as well as normal readers for regular and high frequency words. However, for low frequency and exception words, poor comprehenders performed significantly worse than normal readers suggesting that when word recognition requires semantic support, children with comprehension difficulties are at a disadvantage. Furthermore, the poor comprehender group exhibited lower

expressive and receptive vocabulary on a standardised vocabulary test than the normal readers. These findings indicate that the poor comprehenders in this study had good phonological awareness but poor semantic processing skills.

Similarly, Nation and Snowling (1999) compared the performance of 16 poor comprehenders aged approximately 10 years with 16 normal readers of the same chronological age on a semantic priming task. Children were presented with pairs of words constituting a prime and a target word and were asked to make a lexical decision to both the prime and the target. The words were presented one at a time on a computer screen. The experimental items were comprised of semantically related pairs (e.g. DOG-CAT) or functionally related pairs (e.g. BEACH-SAND). Furthermore, the items in each pair were either highly associated (e.g. BROTHER-SISTER; BOW-ARROW) or non-associated (COW-GOAT; FRIDGE-CHEESE) according to word association frequency norms. The words in each pair were also presented with an unrelated prime (e.g. DOG-BROTHER; BOW-CHEESE). The results showed that children with normal reading skills demonstrated strong priming effects for both semantically and functionally related words. Furthermore, while normal readers showed enhanced priming effects for semantically related words when they were highly associated, priming was still present for non-associated words in both the functionally and semantically related conditions. In other words, normal readers made faster lexical decisions about the second word of each pair when the preceding word was related to it than when it was not. Similarly, poor comprehenders showed priming effects for pairs of words that were functionally related, whether highly associated or non-associated. However, for semantically related items, children with comprehension difficulties showed no evidence of priming for non-associated words. Priming effects were only present for highly associated words in this condition suggesting that poor comprehenders are able to access concrete semantic relations but have difficulty with more abstract semantic processing. Finally, Nation, Marshall and Snowling (2001) found that poor comprehenders were slower and less accurate than chronological age-match controls at naming pictures, a task requiring access to semantic information, particularly when the pictures had low frequency names. Taken together the results of these three studies suggest that children with specific comprehension difficulties have problems with semantic processing.

d) General Language Skills in Poor Comprehenders

Other areas of language processing have been investigated in children with specific comprehension difficulties. Donaldson (1986) investigated knowledge of causal connectives in normally developing children. Children were presented with two tasks; a question answering task and a sentence completion task. In the question answering task, participants were shown a pair of pictures depicting a series of events that were described by the experimenter e.g. 'Mary gets soaked. Mary is sneezing.' Participants were then asked a question to ensure that they had understood the state of the protagonist in the picture sequence e.g. 'Mary has got a cold, hasn't she?' They were then asked the experimental question, which took one of two forms; an empirical form or a deductive form. Empirical items were defined as those which required the participant to show some knowledge of cause and effect e.g. to answer the question 'Why does Mary have a cold?', the participant would have to infer that Mary had a cold because she got soaked. In contrast, the deductive items required participants to provide some evidence for their previous conclusion e.g. in order to answer the question 'How do you know Mary has a cold?', participants would have to provide evidence that Mary exhibited the symptoms of a cold - she was sneezing. Results indicated that children aged between 5 and 10 years of age found the deductive version of this task more difficult than the empirical version and tended to answer deductive questions with an empirical answer, although this tendency decreased with age.

In a sentence completion task, participants were again presented with the pictures and accompanying description of events. They were then presented with an incomplete sentence in either the empirical or deductive form e.g. Mary has a cold because _____ (*empirical*), or We can tell Mary has a cold because _____ (*deductive*) and they were asked to complete the sentence. As in the question answering task, the younger children found the empirical sentence completion task easier than deductive sentence completion, and tended to provide empirical completions e.g. 'she got soaked' for both types of sentence (Donaldson, 1986).

Overall, the findings of Donaldson (1986) showed that young children failed to understand that the word 'because' can introduce not only the cause of an event in an empirical sense but also the evidence for a conclusion in the deductive sense. However, Donaldson also found evidence of individual differences such that some children in each age group tested performed well on both types of task, indicating that the lack of

knowledge about the use of the causal connective 'because' cannot be attributed solely to age.

Oakhill, Yuill and Donaldson (1990) extended these findings by presenting skilled and less-skilled comprehenders with the tasks described above. They found that both groups of children performed equally well on the empirical items in both the question answering and sentence completion tasks. However, while both groups of children performed at a lower level on the deductive items in both tasks than on the empirical items, the difference in performance was more extreme for the less-skilled comprehenders. The less-skilled comprehenders achieved a significantly lower score on the deductive items than the skilled comprehenders. Thus, it would seem that it is comprehension skill, and not age, that is related to the understanding of causal connectives. Furthermore, the results showed that less-skilled comprehenders have a less well-developed understanding of 'because', although only when asked to use it in the deductive mode. Thus, they appear to be able to make simple cause and effect links but find complex deductive reasoning processes more difficult. However, it is important to note that, when prompted to attend to the information in the pictures, the performance of both groups improved on the deductive reasoning items in both tasks. Therefore, as in the previous studies described above (Cain and Oakhill, 1999; Yuill and Oakhill, 1988a), it would seem that the problems of less-skilled comprehenders lie in their use of strategies rather than their ability to perform the more complex operations per se. This hypothesis is supported by a recent study from Cataldo and Oakhill (2000) investigating strategy use in less-skilled and skilled comprehenders. Cataldo and Oakhill (2000) found that when children needed to search for specific pieces of information in a coherent text, either for recall or to answer questions about the text, skilled comprehenders consistently used more efficient search strategies than less-skilled comprehenders.

Research has shown that children with comprehension difficulties are less able to use context to aid in word recognition (Nation & Snowling, 1998b). Nation and Snowling (1998b) demonstrated that children with comprehension difficulties do not benefit from context as much as normal readers or dyslexic children. Children were asked to read target words which were either presented individually or were preceded by a contextually appropriate sentence prime (e.g. I went shopping with my mother and my....*aunt*). Results indicated that dyslexic children gained the most benefit from the

sentence context, exhibiting greater improvements in both response time and accuracy than either normal readers or poor comprehenders when presented with words preceded by a sentence prime. In contrast, poor comprehenders gained the least benefit from sentence context. This group showed less improvement in response accuracy than normal readers when presented with words preceded by a sentence prime, although the two groups did not differ in terms of response times in the context condition. Thus it would seem that poor comprehenders are less efficient in their use of context to guide their reading.

Nation and Snowling (2000) showed that the language difficulties of less skilled comprehenders also included poor syntactic awareness. Normal readers and poor comprehenders were presented with sentences in which the word order had been scrambled. The children were asked to recall the sentences, putting the words in the correct order. The results of the first experiment showed that both groups of children were adversely affected by sentences of increasing difficulty such that overall, performance was better for active than for passive sentences, with groups performing near ceiling for active sentences. However, poor comprehenders performed significantly worse than normal readers on both active and passive sentences. The results of a second experiment showed that the performance of both groups of children suffered as a result of increasing ambiguity. However, poor comprehenders performed at a significantly lower level than normal readers at all three levels of semantic ambiguity included in this study while skilled comprehenders only had problems with the most ambiguous sentences. To summarise, the results of both experiments indicate that poor comprehenders have weak syntactic awareness skills.

Finally, Cain and Oakhill (1996) investigated the ability of less-skilled comprehenders to tell a story. They tested three groups of children; a less-skilled comprehender group (mean age 7 years 9 months), a chronological age-match skilled comprehender group (mean age 7 years 8 months) and a comprehension age-match control group (CAM; mean age 6 years 10 months). They provided all three groups of children with two types of prompt; a title prompt and a picture prompt. In the title prompt condition, children were given five titles from which they had to choose three to plan their stories around. In the picture prompt condition, children were presented with three picture sequences representing the events of a story. The sequence included the setting of the story, the main event and the resolution of the story. The child's task in

both conditions was to tell a story out loud to the experimenter based on the prompt presented to them. Children were given practice items in both conditions and were allowed time to plan their stories before responding. The stories were scored in terms of the presence of a number of story conventions; beginnings and endings, and scene and character setting information. They were also scored for story structure, word length and use of connectives.

The results of the study showed that all groups produced shorter stories in the picture condition than in the title condition, although this reduction in story length was most marked for the less-skilled comprehender and CAM groups. There was no difference between groups in the use of beginnings and endings, or character and scene setting information, indicating that less-skilled comprehenders had a comparable degree of knowledge about the conventions of a story as both control groups. Given the findings of Yuill and Oakhill (1991, 1990) described above one might have predicted that the less-skilled comprehenders would use fewer connectives in their stories than either of the other two groups. Surprisingly no such difference was found indicating that all three groups were equally able to use linguistic devices to express their ideas. It should be noted, however, the connectives used in these stories were simple connectives e.g. and/then, temporal connectives e.g. the next day, and causal connectives such as the empirical form of 'because', with which less-skilled comprehenders are familiar (Oakhill, Yuill & Donaldson, 1990). The use of more complex connectives may have differentiated between groups.

In terms of story structure, the less-skilled comprehenders told poorer stories in the title condition than either the chronological age-match or comprehension age-match control groups. However, the less-skilled comprehenders also received more benefit from the picture prompt condition than the other two groups. In the picture prompt condition the less-skilled comprehenders produced stories that were significantly poorer in story structure than the skilled comprehender group but comparable in story structure to the CAM group. The performance of the skilled comprehender and CAM groups did not differ from each other in either condition. The poor structure of the stories produced by the less-skilled comprehender group in the title prompt condition cannot be a result of their comprehension ability as the comprehension age-match controls produced stories superior in structure. Furthermore, the fact that the picture prompt condition proved to be of more benefit to the less-skilled comprehenders than to either of the other

two groups indicates that this condition provided information that the less-skilled comprehenders could not provide themselves about the structure of events. To summarise, the results of this study suggest that children with less-well developed comprehension skills have a poorer knowledge of story structure which may be causally related to their comprehension difficulties and which seems to be related to a difficulty in building up a sequence of events rather than a general lack of knowledge about stories.

1.1.4 Summary

Clearly, children with reading comprehension difficulties have difficulties that extend into other areas of language processing. However, the underlying cause of these difficulties is unclear. The purpose of this thesis is to investigate the importance of inferential skills to reading comprehension in general and to examine the inferential skills of poor comprehenders in more detail. While the work of Oakhill and colleagues discussed above (Cain & Oakhill, 1999; Oakhill, 1982, 1984; Yuill & Oakhill, 1988a) has indicated that poor comprehenders have difficulty making inferences, a number of questions have yet to be addressed. For example, in many of the aforementioned studies, it is not always clear what types of inferences are being examined and it is therefore not clear whether poor comprehenders have difficulties with all classes of inference, or whether their difficulties are limited to inferences requiring attention to certain types of information. Furthermore, the question answering procedure used by Oakhill (1984) and Cain and Oakhill (1999) does not allow for any conclusions to be drawn concerning the time-course of inference generation in poor comprehenders. Primarily, the focus of this investigation will be to establish a pattern of on-line inference generation in children with normally developing reading skills. In addition to this, a study will be conducted looking specifically at inferential processing in children with comprehension difficulties.

1.2 Inference Types and Models of Inference Processing.

The work discussed in section 1.1 suggests that comprehension of text or speech involves much more than a simple understanding of the individual words presented and their meaning in any given sentence. Essential to comprehension is the ability to form a representation of the information presented as a whole. This involves integrating information across sentences and formulating representations that are only implicitly mentioned in the text being read. In other words, successful comprehension is highly

related to the ability to draw inferences. This relationship has been consistently shown in studies that have investigated the representation of discourse in memory. For example, Bransford, Barclay and Franks (1972) found that people tend to recall discourse in terms of the general overall meaning rather than as a verbatim account of the input. Furthermore, they presented subjects with a number of sentence frames and consequently with a recognition task whereby subjects read test sentences that were either inferred from the original sentence frame, included in the original sentence frame or not inferred from the original sentence frame. Their results demonstrated that people erroneously recognised as familiar a sentence that was inferred by a sentence frame while correctly recognising as novel a sentence that was not inferred by the original input. Anderson et al. (1976) found that a word inferred by a sentence could be as good a cue for recall as a word actually present in the original sentence. For example, subjects presented with the phrase 'the fish attacked the swimmer' found the word 'shark' was a better cue for recall of the sentence than the general term 'fish' that was used in the original sentence. Similar results have been found in number of other studies (Anderson & Ortony, 1975; Paris & Lindauer, 1976). Thus it would appear that people's representation of discourse in memory contains a high proportion of information obtained by the drawing of inferences.

Van Dijk and Kintsch (1983) propose that there are three levels to the discourse representation constructed during reading; the surface code, the textbase, and the situation model. The surface code represents the precise wording and syntax of each clause. The textbase is a representation of the explicit propositions in the text in terms of their meaning rather than their surface code. The situation model is a representation of the actions, events, characters and meaning of the text as a whole. The situation model is thought to be constructed by relating the explicit information in the passage with ideas left implicit or to be inferred. Thus it is in the construction of the situation model that inference generation is thought to be most important. However, there are many different types of inferences that could potentially be generated when constructing a situation model. In the following discussion I will first consider the different inference types that have been commonly identified in studies of discourse processes, and secondly I will consider two of the leading models of inference generation.

1.2.1 Types of Inference

Two superordinate categories of inferences that have been consistently identified in the literature are coherence inferences and elaborative inferences. Coherence inferences are those that are necessary to form a consistent and intelligible mental representation of a text and can be classified either as those that use *cohesive devices* or as those that are *knowledge-based*.

Cohesive devices are commonly used to maintain textual integrity. Such devices include anaphora and pronoun resolution. For example, consider the following text:

"The ship left the harbour in a storm. The vessel lurched as a wave hit the bow."

In order to form a meaningful and conjoined representation of the two sentences the reader would need to infer that 'the ship' referred to in the first sentence was the same object as 'the vessel' referred to in the second sentence. As such, anaphora is used to aid in the construction of a situation model. Similarly, for the following sentence: "John asked Pam if he could borrow some money." the reader would need to determine that the pronoun 'he' refers to "John" in order to realise a complete representation of the sentence. Thus, cohesive devices can be used to resolve incongruities or ambiguities in a text, helping to construct a coherent representation and therefore can be viewed as inferences. However, some researchers are undecided as to whether this use of cohesive devices can be classed as a type of inference generation (McKoon & Ratcliff, 1992; Singer, Harkness, & Stewart, 1997). Indeed, the distinction between coherence and cohesion is not entirely clear (Fulcher, 1989), and Graesser, Trabasso & Singer (1994) suggest the two concepts overlap such that cohesion is simply a form of coherence and as such the use of cohesive devices to form textual links can be viewed as a type of inference generation. More importantly, researchers are generally agreed that these inferences are encoded into the textbase representation of a passage during reading (Garnham, Oakhill, & Cain, 1997; Graesser et al., 1994; McKoon & Ratcliff, 1992; Swinney & Osterhout, 1990). Therefore, the remainder of this discussion will focus on the distinction between knowledge-based coherence inferences and elaborative inferences.

Knowledge-based inferences are commonly used to bridge a gap in the discourse by applying real-world knowledge to the explicit information in the text in order to form a link between the ideas that are clearly stated and those that are left implicit. They are also frequently termed *bridging inferences*. The most commonly studied bridging inference is that which is generated to explain the cause of an event. For example, consider the following text:

"The campfire started to burn uncontrollably. Tom grabbed a bucket of water."

In order to understand why "Tom grabbed a bucket of water", it is necessary to relate the second sentence to the first, ("The campfire started to burn uncontrollably."), by generating the inference that Tom was trying to put the campfire out. To do this you must first activate the mediating idea that 'water puts out fire' from your own real-world knowledge. Without generating this inference the two sentences remain disjointed and do not form a coherent whole. In this example, therefore, knowledge of the real world is used to forge a link between the two explicit sentences by generating an explanation of Tom's actions; an explanation that is left implicit in the text itself.

Bridging inferences are not only causal however. Arguably, actions and events in a text can also be explained in terms of temporal information, the motivations and intentions of the characters and the spatial location of the objects and agents involved. The relative importance of each of these inference types to the situation model has not been clearly established and causal bridges remain the most widely recognised type of knowledge-based inference consistently included in a reader's discourse representation.

Elaborative inferences are those that do not add anything to the coherence of a passage but serve to enrich its mental representation. They are therefore not necessary for textual coherence. Like bridging inferences, elaborative inferences are generally knowledge-based. For example, inferences about the consequences of an action, predictions about forthcoming events, speculations regarding the instrument used to perform a task, or suppositions about the physical properties of textual characters and objects are all elaborative inferences. By way of illustration, consider the following passage:

"The knight lunged towards the dragon and pierced his shining scales. The dragon turned towards the knight and let out a fiery roar."

It is possible to make many elaborations from this short passage. It could be inferred that the knight attacked the dragon with a sword because that is the usual tool of the knight and is inferred by the verb 'pierced'. However, it is not necessarily the case since the knight could have been using any other sharp object had he not been in possession of his sword. Furthermore, it could be assumed that the knight was wounded by the dragon's fiery breath. However, it may be the case that the knight was able to avoid the dragon's attack. As interesting and text-enriching as these elaborations may be, none of them are necessary in order to form a full and intelligible representation of the original text. As such, a clear distinction can be drawn between these and the more necessary inferences required for coherence.

While the preceding discussion was intended to give an indication of the different types of inferences available from text, it is important to note that these examples are by no means exhaustive. There are many potential inferences available in text and taxonomies of inference types have proved controversial in the past. For the most-part, the distinction between knowledge-based coherence and elaborative inferences outlined above will be adopted by this thesis.

1.2.2 Two Accounts of Inference Generation.

It is clear that many inferences could conceivably be generated while reading in order to construct a rich discourse representation. However, the generation of all of these inferences while reading would result in an overwhelming drain on cognitive resources. One of the most controversial areas of inference research focuses on establishing which inference types are automatically integrated into the situation model during encoding of the text and which require more strategic processing.

It is important at this point to define what is meant by the term 'automatically'. Automatic processes can be defined as those which are fast, unintentional, require little conscious effort, interfere minimally with other tasks and do not make many demands on attentional resources (Thurlow & van den Broek, 1997). The notion that inferences can be drawn in accordance with this definition seems incongruous with the act of comprehension which, as discussed in the previous section of this chapter, is a complex task requiring the co-operation of a number of different cognitive processes. However,

it is important to consider that automaticity is relative, not absolute; that is to say, a task performed automatically will be performed faster and with less conscious effort than if it were to be performed consciously (Thurlow & Van den Broek, 1997). Thus, it may be that certain types of inference generation during comprehension can be deemed automatic in that they require no more effort than would normally be necessary for reading comprehension. Conversely, strategic processing is that which requires intentional effort. With this in mind it is plausible that some inferences can be made automatically while others require more conscious effort and are usually drawn when the reader has a specific reading goal.

A number of theoretical models of inference generation have been proposed in recent years including the Prediction-Substantiation Model (Bower, Black, & Turner, 1979; DeJong, 1979; Dyer, 1983; Schank & Abelson, 1977), the Causal Inference Maker Model (van den Broek, 1994; van den Broek, 1990), the Event Indexing Model (Zwann, 1999), the Structure Building Framework (Gernsbacher, 1997) and the computational Landscape Model (van den Broek, Young, Tzeng, & Linderholm, 1999). However, two models in particular have been influential in providing a framework for many of the contemporary studies of inference generation; the *constructionist theory* (Graesser & Kreuz, 1993; Graesser et al., 1994) and the *minimalist hypothesis* (McKoon & Ratcliff, 1992). Both models agree that any of the potential inferences can be drawn if a reader has a particular goal and is therefore employing a particular reading strategy. For example, Van den Broek, Lorch, Linderholm and Gustafson (2001) recently demonstrated that readers who read for enjoyment were less concerned with the coherence of a text than people reading for study. Consequently, readers with a study goal engaged in more coherence-building inferential activity and showed a higher level of recall for text than readers with an entertainment goal. The difference between the constructionist theory and the minimalist hypothesis lies in the types of inferences the two models believe to be drawn automatically. This difference centres on the issue of whether the reader strives to achieve global or local coherence. Global coherence occurs if a reader can integrate an incoming statement with the overall structure of the text they have previously read. Conversely, local coherence involves connecting the incoming information with that in the preceding one or two sentences. Therefore, if a reader attempts to achieve both global and local coherence, it is conceivable that many more inferences will be required for comprehension than if the reader is simply striving for local coherence.

1.2.3 The Constructionist Theory

The constructionist theory (Graesser et al., 1993, 1994) is based on three critical assumptions which distinguish it from other theories of inference generation; the assumption that the reader will construct a mental representation that fulfils their reading goal; the assumption that the reader attempts to achieve both local and global coherence; and the assumption that the reader looks for explanations of events, actions and states that occur in the text. These assumptions are referred to as the *reader goal assumption*, the *coherence assumption* and the *explanation assumption*. In terms of automatic inference generation, it is the latter two assumptions that are the most important since the reader goal assumption allows any and all inferences to be generated if they fulfil the reader's reading goal and is thus related more to strategic than automatic inference generation.

Based on the explanation assumption and coherence assumption, certain predictions can be made by the constructionist theory regarding the classes of inferences automatically drawn during encoding. In terms of the inference categories identified above, Graesser et al. (1993, 1994) predict that only those inference categories required for text coherence or to explain events, actions and states, will be drawn on-line. Thus, text-connecting, or cohesive inferences and knowledge-based, or bridging inferences will be generated. Both classes of inference can be said to aid local coherence as illustrated in the short passages in section 1.2.1. Cohesive inferences achieve local coherence by resolving text-based ambiguities as in the case of pronoun resolution or anaphora. Furthermore, Graesser et al. (1993, 1994) include case-structure role assignment in this category whereby a noun phrase is assigned to the role of agent, recipient, object, location or time. For example, in the sentence 'the elephant gave his bananas to the monkey', the monkey will be assigned the role of recipient, the elephant will be identified as the agent and the bananas as the object. Knowledge-based inferences however, achieve local coherence by providing explanations for the actions and events described in the text as in the "Tom versus the campfire" example given above.

Knowledge-based inferences can also facilitate global coherence. Graesser et al. (1993, 1994) identify three classes of knowledge-based inference that contribute to

global coherence: *superordinate goal*, *thematic* and *character emotional reaction*. Superordinate goal inferences refer to a goal that motivates the actions of a character; thematic inferences refer to the overall theme or moral of the text; and character emotional reaction inferences refer to the reactions of a character to the actions or events that occur in the text. These inferences can be deemed knowledge-based since in order to generate them it is necessary to apply knowledge of the world to the information presented in the text. They can be identified as global because the information required to generate these inferences will be integrated across the passage as whole rather than across one or two sentences. Graesser et al. (1994) propose that in order to achieve a complete situation model of the discourse, these global inferences are necessary.

In contrast to coherence inferences, the constructionist theory predicts that elaborative inferences will not be drawn on-line due to the "computational explosion" (Graesser, Millis, & Zwann, 1997) that would occur if all possible alternative elaborative inferences were to be generated. Specifically, Graesser et al. (1994) identify five such inference classes; *causal consequence*, *instantiation of noun category*, *instrument*, *subordinate goal-action*, and *state*.

- A causal consequence inference refers to hypothetical predictions regarding characters' actions and future events based on preceding discourse context. The constructionist theory proposes that the only predictions that will be made automatically are those that are made when drawing coherence inferences such as characters' reactions and what they aimed to do.
- Instantiation of a noun category involves elaborating a specific exemplar from a more generic noun. For example, given the sentence, 'in the explosion the man was thrown clear of the blazing vehicle', one may infer that the 'vehicle' in question was a car. However, it could conceivably have been a motorcycle, a bus, or any other exemplar of the category vehicle. The construction theory predicts that such instantiation will not take place on-line.
- An instrument inference involves inferring a particular object used by an agent to complete an action; for instance, inferring that the knight used a sword to pierce the dragons scales in the example in section 1.2.1. As with the instantiation of a noun category, it is often the case that there are many plausible instruments that

an agent may have used. Thus, this inference class is not thought to be generated on-line.

- A sub-ordinate goal action inference is one in which the reader may infer how an agent achieved a goal not crucial to the over-riding plot of the story. For example, given the sentence, 'John was drinking his coffee', you could infer that John raised the coffee cup to his lips in order to fulfil the act of drinking. However, this inference is not essential to the understanding of the text as a whole, and indeed, however likely this inference may seem there are many alternative ways in which John could commit the act of drinking e.g. through a straw, from a flask etc.

- State inferences refer to static properties that are not causally related to the plot of the text. For example, properties of objects, traits of characters etc. While these properties may serve to enrich the representation of a text they are not thought to be essential for local or global coherence and as such are not drawn on-line according to the constructionist standpoint.

It is important to note, however, that these predictions are not absolute. The constructionist model does allow for variance in the classes of inferences to be drawn on-line, but only when certain criteria are met. For example if the reader has a specific reading goal, they may generate particular classes of inference not normally drawn. Similarly, if the context of the discourse is sufficiently constraining such that few alternative consequences are available, certain elaborative inferences may be drawn. However, the predictions outlined above are valid if the reader has no specific reading goal or if the discourse is not constrained. Furthermore, the constructionist theory primarily confines its predictions to narrative text. However, narrative texts are not the only written medium to which people are exposed. Expository texts are also a common, and sometimes more appropriate way of presenting information. Graesser and Wiemer-Hastings (1999) state that they are "confident that most of the predictions would also hold up in investigations of expository texts on history and even scientific mechanisms" but this issue requires further investigation. The effect of text format on inference generation will be considered in more detail in chapter 4. The inference types discussed in this and the following section, and the predictions of each model regarding the generation of these inferences, are presented in Table 1.1

1.2.4 The Minimalist Hypothesis

In contrast to the constructionist theory, the *minimalist hypothesis* (McKoon & Ratcliff, 1992) asserts that, when a specific reading goal is absent, local coherence is the over-riding goal of the reader. Thus, the minimalist hypothesis asserts that the only inferences to be drawn automatically during encoding are those required for local coherence; that is those necessary to relate incoming information with that in the preceding sentence. The main difference, therefore, between the two models is that the constructionist theory predicts the automatic generation of global coherence inferences while the minimalist hypothesis does not. In terms of local coherence inferences, the two theories appear to be in agreement. However, on closer examination clear differences emerge between the two theories in terms of what each classifies as a local coherence inference. In line with the constructionist theory, the minimalist hypothesis classes the use of cohesive devices as an example of generating a local coherence inference. However, in terms of knowledge-based inferences the stance of the minimalist hypothesis is not so clear.

The minimalist hypothesis proposes that knowledge-based inferences will be drawn automatically during reading if they are a) necessary for local coherence and/or b) based on readily available information. Readily available information is defined as that available from explicit text-based information, from short-term memory and from long-term memory and therefore, the source from which the relevant information can be obtained is not constrained in any way. As a result of these criteria, the minimalist hypothesis allows for knowledge based inferences to be drawn on-line whether necessary for local coherence or not. Furthermore, unlike the constructionist theory, the minimalist hypothesis does not make any specific predictions regarding the classes of knowledge-based inferences that will be made on-line.

The minimalist hypothesis is also less consistent than the constructionist theory in predictions regarding global inferences. For example, while McKoon & Ratcliff (1992) purport to deny the automatic generation of global inferences, they do not rule out global inferences completely. If local coherence is not achieved by relating incoming information to the preceding sentence held in working memory, global inference information can be utilised to achieve local coherence. Moreover, while McKoon and Ratcliff (1992; experiment 3) demonstrate that one particular class of global inference, subordinate goal or action, is not drawn automatically, their findings

cannot be interpreted as evidence against the constructionist theory since Graesser et al. (1994) make the same prediction.

In terms of elaborative inferences, the minimalist hypothesis makes the same prediction as for local coherence inferences; that is elaborative inferences will be drawn if they are necessary for local coherence. However, it further suggests that elaborative inferences may be drawn if they are based on readily available information. Thus, the prediction is similar to that of the constructionist theory that elaborative inferences will not be drawn automatically since elaborative inferences are by their nature not required for local coherence, nor are they usually based on easily available information. However, the model also proposes that elaborative inferences can be encoded to greater or lesser extent (McKoon & Ratcliff, 1986) allowing for any unexpected occurrences of elaborative inference generation.

<i>Inference Type</i>	<i>Description</i>	<i>Classification</i> e.g. local	<i>Source</i> e.g. knowledge-based Text-based	Prediction (drawn on-line \mathcal{P} , not drawn on-line \mathcal{P})	
				<i>Constructionist</i> <i>Theory</i>	<i>Minimalist</i> <i>Hypothesis</i>
Anaphora	Forming a link between two terms referring to the same thing e.g. "The car came racing round the corner. Everybody scattered as the vehicle crashed into the wall."	Local	Text-based	\mathcal{P}	\mathcal{P}
Pronoun resolution	Linking a pronoun to its previous referent e.g. <i>John</i> picked up <i>Mary's</i> book. <i>He</i> had wanted to read it for ages.	Local	Text-based	\mathcal{P}	\mathcal{P}
Case-structure role assignment	Assigning the role of agent, object, recipient, time or location to a noun phrase.	Local	Text-based	\mathcal{P}	\mathcal{P}
Causal antecedent	Provides an explanation for the actions and events in a text.	Local	Knowledge-based	\mathcal{P}	\mathcal{P}
Superordinate goal	The overall goal that motivates the characters in the text.	Global	Knowledge-based	\mathcal{P}	\mathcal{P}
Thematic inference	The overall goal or moral of the passage.	Global	Knowledge-based	\mathcal{P}	\mathcal{P}
Character	The reactions of a character to the actions and events in the text.	Global	Knowledge-based	\mathcal{P}	\mathcal{P}
emotional reaction	The predicted consequences of the actions and events in the text.	Elaborative	Knowledge-based	\mathcal{P}	\mathcal{P}
Causal consequence	Elaboration of specific exemplar from a generic noun.	Elaborative	Text-based	\mathcal{P}	\mathcal{P}
Instantiation of noun category	Inferring a particular object used by an agent to complete an action.	Elaborative	Knowledge-based	\mathcal{P}	\mathcal{P}
Instrument inference	Static properties of objects, characters etc, not related to the causal structure of the text.	Elaborative	Knowledge-based	\mathcal{P}	\mathcal{P}
State inference	How an agent achieved a goal not relevant to the superordinate goal of the text.	Elaborative	Knowledge-based	\mathcal{P}	\mathcal{P}
Sub-ordinate goal action	How an agent achieved a goal not relevant to the superordinate goal of the text.	Elaborative	Knowledge-based	\mathcal{P}	\mathcal{P}

Table 1.1 Summary of inference types and predictions of the constructionist theory and the minimalist hypothesis.

1.2.5 Evidence for the Constructionist Theory and the Minimalist Hypothesis

Empirical evidence supports the claims of both the constructionist theory and the minimalist hypothesis that coherence inferences are drawn on-line to support local coherence. Such inferences include the cause of an event (Fincher-Kiefer, 1996; Magliano, Baggett, Johnson, & Graesser, 1993; Potts, Keenan, & Golding, 1988; Valencia-Laver & Light, 2000), the assignment of roles such as agent, object etc (Swinney & Osterhout, 1990), and the encoding of the link between anaphor and referent (e.g. Garnham, Oakhill & Cain, 1997). However, in opposition to the minimalist hypothesis, there is evidence to suggest that some inferences which facilitate global coherence are also generated (Long, Oppy, & Seely, 1997; Trabasso & Magliano, 1996), including the over-riding goal of the character (Long & Golding, 1993) and character emotional reactions (Gernsbacher, Hallada, & Robertson, 1998). Furthermore, McKoon and Ratcliff (1992) cite an earlier study (McKoon & Ratcliff, 1980) as evidence that anaphoric inferences are drawn on-line to maintain local coherence. However, while they state that:

"local coherence is defined for those propositions of a text that are in working memory at the same time; in other words, propositions that are no farther apart in the text than one or two sentences." (McKoon & Ratcliff, 1992, p441)

the materials used by McKoon and Ratcliff (1980) require anaphoric relations to be drawn between the first and last sentence of four sentence texts. Therefore, considering the definition of local coherence given by McKoon and Ratcliff (1992), it could be argued that in this case the anaphoric inference was being drawn to maintain global, not local, coherence since it involves the relation of information from non-adjacent sentences. Thus, in terms of local and global coherence inferences, the empirical data supports the constructionist theory to a greater extent than the minimalist hypothesis.

Empirical evidence concerning elaborative inferences is not so clear-cut. Traditionally research suggests that elaborative inferences are not made until the processing of the incoming information is complete and then, only if the inference is specifically probed (Graesser et al., 1994; Magliano et al., 1993; Potts et al., 1988; Singer & Ferreira, 1983; Valencia-Laver & Light, 2000). This view is consistent with

the constructionist theory of inference generation. However, some research has suggested that elaborative inferences are made automatically during processing but only to a minimal extent (McKoon & Ratcliff, 1986), a finding in line with the minimalist hypothesis. One of the reasons for this lack of encoding of elaborative inferences is purported to be the unconstrained nature of these inferences. In other words, too many plausible alternative inferences would be generated and this would exceed processing limitations. However, researchers do concede that given certain circumstances, e.g. highly accessible knowledge, high contextual constraints, few alternative consequences, or adequate processing time, elaborative inferences could be drawn (Calvo, 2000; Calvo & Castillo, 1996; Calvo, Castillo, & Estevez, 1999; Fincher-Kiefer, 1996; Graesser et al., 1994; McKoon & Ratcliff, 1992; Murray, Klin, & Myers, 1993). Thus, the empirical evidence for elaborative inference generation does not distinguish between the two theories of inference generation considered in this review. Instead the evidence serves to strengthen the assumption made by both theories that elaborative inferences are not made automatically but require strategic processing.

A number of methodological problems can be raised concerning the studies used as empirical evidence for theories of inference generation. First, because the constructionist theory concentrates primarily on narrative texts, studies supporting a constructionist approach have tended to use only narrative materials. Similarly, while the minimalist hypothesis does not explicitly state that it is confined to narrative texts, researchers tend concentrate on this text type. However, there are many different types of text from which to draw inferences. Texts are not always presented in a narrative format and expositions are a more appropriate medium for some forms of textual discourse. In order for a model of inference generation to fully account for all instances where an inference might be made, it needs to incorporate expository text (Millis, Morgan, & Graesser, 1990; Noordman, Vonk, & Kempff, 1992; Singer et al., 1997). Second, many studies have concentrated on only two categories of inference; causal bridging inferences and causal consequences. By limiting the materials in this way, it is not clear how far such studies can be generalised to include other inference categories.

1.2.6 Summary

To summarise, it is not clear whether the constructionist theory or the minimalist hypothesis provides the best account of inference generation. Arguably, the application of a particular model varies depending on the reader's goal and background knowledge.

However, the minimalist hypothesis has attracted much criticism (Glenberg & Mathew, 1992; Singer, Graesser, & Trabasso, 1994; Trabasso & Magliano, 1996). Of most importance is the criticism levelled by Singer et al. (1994) that the minimalist hypothesis is difficult to falsify. Indeed, the outline of the minimalist hypothesis given above illustrates that it contains many exceptions and provisos, which allow it to encompass a broad range of empirical findings. It may be that the simplicity and straightforwardness of the constructionist theory provides the clearest framework for predicting inference generation. Further, contemporary models of inference generation have used the constructionist theory as a foundation. For example, the Scaffold Model proposed by Revlin and Hegarty (1999) views text-based inferences as a form of deductive reasoning akin to solving syllogisms. While this approach is very different from the more traditional models of inference generation described above, Revlin and Hegarty (1999) demonstrate that the constructionist theory provides the basis for their model while the minimalist hypothesis cannot account for their findings. This lends further support to the validity of the constructionist theory as a model of inference generation. It is this theory which the present thesis will adopt when making predictions regarding the generation of inferences.

Chapter 2

Methodological Considerations In The Study Of Inference Generation.

2.1 Introduction

Research into inference generation typically involves the presentation of a series of short passages that have been designed to induce the generation of specific inferences. Experimenters then seek to assess which types of inferences were generated from the experimental passages, and the time course of those inferences i.e. when the inferences were made. Many different methods have been used in the study of inference generation. However, it is still not clear which method yields the most reliable results. The following discussion considers the types of methodology that have been used in the study of inferences and the problems associated with them.

2.2 Off-line Measures of Inference Generation

Early research employed off-line tasks to measure inference generation. For example, many studies used simple question answering tasks as a means to assess the generation of inferences. Participants taking part in studies using this paradigm are usually presented with short passages to read and are asked a series of questions about each passage. The questions can be presented after the passage is completed, or at strategic points during the reading of the passage. Typically a variety of questions are asked; some require attention to the explicit information given in the passage, and others require some form of inference generation. The degree to which inferences are made is measured by the accuracy with which readers answer each type of question.

The use of memory measures such as cued recall, free recall and recognition as measures of inference generation is also reported in the early literature. In a cued recall task, the inference concept being tested is used as a cue for recall (e.g. Anderson et al. 1976; Anderson & Ortony, 1975; Black & Bern, 1981; Paris & Lindauer, 1976). For example, in one such study, the word 'shark' was used as a cue for recall of the sentence 'the fish attacked the swimmer' since the sentence was thought to encourage the inference that the 'fish' in question was likely to be a 'shark' (Anderson et al., 1976). Similarly, in a free recall task, readers are typically asked to read a passage and then to recall it after a short delay (e.g. Black & Bern, 1981). The recalled stories are then analysed for the number of target sentences correctly recalled (e.g. Black & Bern, 1981) or the number of sentences included in the recalled story that must have been inferred

from the original text. In a recognition task, readers are presented with short stories and are then shown a series of test sentences. Some of the test sentences contain information that was explicitly stated in the text while others contain information that was only inferred from the text. The task for the reader is to decide whether or not each of the sentences appeared in the original passage. The logic behind this paradigm is that, if an inference is generated during reading, readers should erroneously recognise as familiar those sentences that were simply inferred by the text. This prediction has been upheld by a number of studies (e.g. Bransford, Barclay and Franks, 1972; Oakhill, 1982).

However, while recognition and other off-line tasks give some indication of the types of inferences readers can generate from text, they provide no information regarding the time-course of inference generation. That is to say, neither the question answering task nor the free and cued recall tasks can distinguish between those inferences that are made automatically during encoding of the text, and those that are made later when the reader is required to answer questions or complete the recall task. Similarly, in the recognition task, readers could be engaging in some kind of reconstructive process when presented with the recognition cue, checking the compatibility of the test sentence with the prior context. In order to address this possibility, a speeded recognition task has been used in which readers are presented with short passages followed by the presentation of test words (McKoon & Ratcliff, 1986; 1989). The task of the reader is to determine within 300-milliseconds whether the test word appeared in the passage. The imposition of the 300-millisecond deadline is thought to prohibit the reader from engaging in reconstructive processes. Therefore, any erroneous recognition of inferred test words must be due to on-line inference generation. However, research suggests that 300-milliseconds may not restrict the reader from engaging in context-checking and therefore the speeded recognition task may still be subject to the same criticism as the original task (Keenan, Golding, Potts, Jennings, & Aman, 1990; Potts, Keenan, & Golding, 1988). The methods described above are now rarely used in isolation from the more sensitive on-line measures of inference generation because of the problems involved.

2.3 On-line Measures of Inference Generation

On-line measures of inference generation attempt to distinguish inferences that are made during reading i.e. on-line, from those made at a later point in time. There are

a variety of on-line measures used in research into inference generation. Outlined below are some of the more frequently used on-line measures.

2.3.1 Reading Time

Some researchers have used reading time as a measure of inference generation. The rationale behind this method is that the reading time for sentences will increase if the generation of an inference is required for coherence. For example, in a study designed to investigate global coherence, Albrecht and O'Brien (1993) presented participants with short passages divided into five sections; introduction, elaboration, filler information, critical sentences, and conclusion. The introduction section introduced the character and setting of the story and was immediately followed by one of three elaboration conditions; consistent, inconsistent or neutral. The information contained in these sections related to the actions or events conveyed by the critical sentences and was either consistent with, inconsistent with or bore no relation to (neutral) action or event. The elaboration sections were immediately followed by filler information and then the critical sentences were presented. The critical sentences related a sequence of events involving the protagonist. The passages were presented using a self-paced reading paradigm and the time taken to read each sentence was recorded. The results of the study showed that reading times for the critical sentences were slower in the inconsistent condition than in either the consistent or neutral condition which did not differ from each other. These findings were interpreted as evidence that readers strive for global coherence during reading, such that when information was presented that was inconsistent with that previously encountered, readers engaged in inferential processing in order to re-establish global coherence.

Other studies have included measures of reading time in their design (Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Millis, Golding, & Barker, 1995; Myers, Shinjo, & Duffy, 1987). However, reading times can be confounded by a number of other factors. For example, long reading times may reflect difficulties with efficient decoding rather than with inference generation. Similarly, the test sentences may vary in difficulty, comprising different syntactic structures, differing levels of ambiguity and varying familiarity of vocabulary. Finally, it may be that some of the words contained in a passage are either repeated in the corresponding test sentence or are highly related semantically to words in the test sentence. Thus, reading times for test sentences may be a result of word-based priming, an issue that will be discussed

further during consideration of activation measures. To summarise, the use of sentence reading times is not a wholly reliable method for investigating the generation of inferences.

2.3.2 Activation Measures

Activation measures enable researchers to assess the level of activation of an inference concept at various points during the comprehension process. The appeal of these methods is that they lend themselves to a number of experimental manipulations. For example, researchers can vary the point at which the level of activation is measured by inserting the target probe at different points in the stimulus sentence. The time constraints placed on inference generation can also be varied by manipulating the Stimulus Onset Asynchrony (SOA) of a target probe. The activation measures typically used in the study of inference generation are Naming Speed and Lexical Decision.

a) Lexical Decision: In a lexical decision task, participants are presented with a letter string and have to decide whether or not that string is a real word. In the case of research into inference generation, the letter string is usually presented after a short story and the words presented in the lexical decision task are representative of a concept evoked by having made an inference from the preceding passage. Making a lexical decision does not require a comparison between the target item and the preceding passage, since the participant is required only to decode the letter string. Thus, the lexical decision task is thought to overcome the context checking confound of the recognition procedure. However, despite the fact that such comparisons are unnecessary, research suggests that the degree of relatedness between the target word and the text prime does affect lexical decision speed and therefore backward context checking does take place (Keenan et al., 1990). In a comparison of lexical decision and naming speed tasks, Keenan et al. (1990) and Potts et al. (1988) report that the results of the lexical decision task owe more to post-comprehension processes than on-line inference generation. Decision latencies to target words were faster and more accurate when the word was related to a causal inference embedded in the preceding text than when it was an unrelated control word. These findings agree with previous research suggesting that causal coherence inferences are generated on-line. However, decision latencies were also faster and more accurate to target words related to predictive inferences than to unrelated control words, suggesting that predictive inferences are also generated on-line, a finding that contrasts with the majority of previous research. This

pattern of results was not replicated in the data from the naming task. Keenan et al. (1990) and Potts et al. (1988) interpret these findings as evidence of backward context checking during presentation of the target word in the lexical decision task. This finding suggests that lexical decision tasks are not free from the confound of post-comprehension processes and are therefore not a reliable measure of on-line inference generation.

b) Naming Speed: Several researchers have used naming speed as a means of assessing inference generation on-line (Murray, Klin, & Myers, 1993; Potts et al., 1988). In a naming speed task, participants are presented with short stories followed by a probe word that they are required to name as quickly as possible. The probe word is usually representative of an inference implied by the passage. Naming latencies are recorded as a measure of inference generation, the hypothesis being that if an inference is generated on-line, the corresponding probe word for that inference will be named faster than a probe word corresponding to an off-line inference. Naming latencies are made up of a combination of lexical access time and articulation speed. It has been suggested that this articulation speed component cannot be affected by the relatedness of the target word with the preceding text prime. Thus, unlike lexical decision, the naming speed task is thought to be free from the effects of context checking and studies have shown that naming speed is not affected by the relatedness of prime and target (see Keenan et al. 1990 for a review). Results from a comparison of lexical decision and naming speed tasks (Keenan et al., 1990; Potts et al., 1988) showed that naming responses were faster for target words relating to a coherence inference than to control words, reinforcing the generally accepted idea that causal coherence inferences are generated on-line. Moreover, in contrast to the lexical decision results described above, there was no difference in the naming latencies for target words relating to predictive inferences and control words. This finding is in line with previous research suggesting that elaborative inferences are not generated on-line.

However, despite the fact that the naming speed task is not subject to the confound of post-comprehension processes, it does carry with it some possible complications. For example, the speed and relative ease of the task may make it insensitive to the detection of inferences during comprehension. Furthermore, it has been proposed that the modularity theory of language processing (Fodor, 1983) undermines the validity of the naming speed task (Keenan et al., 1990; Magliano &

Graesser, 1991). The modularity theory suggests that reading involves a number of different modules (i.e. lexical, semantic and syntactic). These modules are thought to operate independently of each other and as such, the higher level syntactic and semantic modules should have no effect on the lower level lexicon. Therefore, the naming speed task may activate the lexical module at the point at which the target word is presented and furthermore, the lexicon would be activated in isolation from any of the other higher level modules. Thus, the task may actually be assessing a person's ability to activate entries in the mental lexicon rather than higher-level inference processes. However, the naming speed task has been shown to differentiate between target words that relate to inference concepts and those that do not, indicating that the task taps more than just the lower level lexical module (Potts et al., 1988; Keenan et al., 1990). Thus, the naming speed task may still be a more appropriate measure of text-based inference generation than other on-line tasks such as the lexical decision task.

In addition to the problems discussed above, naming speed and lexical decision share a number of common problems. First, both procedures are intrusive and interrupt the normal flow of reading. Thus, they may interfere with normal comprehension processes. Second, both measures involve the representation of an inference concept using one word. This constrains the kinds of inferences that can be tested since it would be difficult to express a complex inference concept using a single word. Finally, both measures share the problem of text-based versus word-based priming. Text-based priming can be defined as the priming that occurs when a previously activated inference concept is tapped by a target probe. This type of priming is based on the integration of the text with the readers' own knowledge of the situation described by the text. However, it may be the case that previous research has confounded text-based priming with word-based priming. Word-based priming occurs when a word contained in the priming stimulus is highly associated with the target probe. Therefore, any increase in the activation of the inference concept measured by the speed with which the target probe is responded to may be due to the association between the target word and the previously presented information rather than to the generation of the inference itself (see Keenan et al. 1990).

Clearly, it is important to overcome this confound in order to gain a true picture of inference generation and there are a number of ways to do this. First, when constructing the experimental materials, it is important to ensure that none of the target

words are semantically related to the words used in the preceding passage or sentence prime. However, elimination of all semantic associations is impossible (Keenan et al., 1990). A second alternative is to control for the effect of word-based priming by creating a control version of the inference priming materials containing exactly the same words but with no inference required. For example, the phrase 'Dick cleared the snow from the driveway' may prime responses to the word 'shovel' because of the association between the words 'snow' and 'shovel' rather than because of the activation of the inference concept that a shovel is a likely instrument for clearing snow. A control version of this phrase would be 'The sky cleared as Dick drove through the snow on the driveway.'¹ If this control version primed responses to 'shovel' to the same degree as the inference version, one could assume it was due to word-based priming i.e. the association between snow and shovel, since the control version does not imply the use of a shovel. However, if responses to the inference version showed more priming than to the control version, one could attribute this effect to inference generation. Unfortunately, there are problems associated with the generation of these materials such that the control sentences can often sound contrived in an attempt to include all relevant content words. Thus, any differences seen in response times to control sentences compared to inference sentences may be due to a lack of coherence in the control sentences making them difficult to interpret (Magliano & Graesser, 1991). However, when using tasks which may succumb to word-based priming, incorporating a control version of the materials may help to overcome this confound.

c) Sentence Verification: An alternative to the lexical decision or naming procedure is the sentence verification task, and it is this task which has been adopted for the experiments reported in this thesis. In this procedure, participants are presented with short stories with inference-inducing sentences embedded within them. Following the presentation of a story, the participant is presented with a Yes/No question or True/False statement representing the inference concept being tested. The task of the participant is to answer the question or, where a statement is presented, to determine whether the statement is true or false. The logic behind this paradigm is that if the inference concept has been generated on-line, response times for the target questions or statements will be significantly faster than if the inference has not been generated on-line. Thus, like lexical decision and naming speed, sentence verification provides an

¹ Example taken from Keenan et al. (1990).

on-line measure of inference generation. The advantage of sentence verification is that, because the target probes contain more than one word, the inference concepts being tested can vary in complexity. Furthermore, it is unlikely that word-based priming due to the association between a word in the priming stimulus and a word in the target statement will affect the response time since the target statements consist of full sentences, not single words. Thus, the likelihood that one of the words in the target statement will be more or less semantically associated with a word in the priming stimulus is greatly reduced. Finally, unlike naming speed, sentence verification involves more than the simple activation of the lexicon and therefore is not undermined by the modularity theory of language processing.

However, the sentence verification procedure does present a problem in that the response time to any given target statement is comprised of the time it takes to produce the correct answer or make a judgement and the time it takes to make the response. It is difficult to separate the decision making process involved in answering a question or making a true or false judgement, from the actual time taken to make the response. Thus, it may be that the required inference concept is accessed off-line during the decision making process as opposed to being generated during comprehension. However, if responses were solely mediated by the decision making process, one might expect that responses made to statements representing inferences thought to be generated on-line would be no different from those made to off-line inference statements or control statements. Using the true/false target statement version of this task, McKoon and Ratcliff (1988) demonstrated that responses were faster and more accurate to test sentences representing concepts that were contextually relevant to previously mentioned target objects and thus necessary for textual coherence than to contextually irrelevant concepts. Moreover, research using the question answering version of this paradigm has shown a response time advantage for causal coherence inferences (i.e. Singer et al., 1992, 1997). In short, while responding to a question or target statement in this way clearly involves making a decision, decisions to questions or target statements pertaining to on-line inferences are made faster than those to off-line inferences because the inference concept has already been accessed on-line.

To summarise, on-line activation measures are clearly preferable to the use of off-line measures when investigating inference generation but they are not entirely devoid of problems. In order to gain a comprehensive measure of inference generation,

it is vital to be aware of these difficulties and attempt to control for them wherever possible. However, one problem that is not easily overcome is the degree to which these measures can be regarded as purely on-line. As outlined above, all the tasks have a metacognitive component and incorporate processing requirements over and above making the required response. The effect of these added processing requirements on response times is unclear. One methodology proposed as an alternative to these behavioural on-line measures is the measurement of eye-movements during reading.

2.3.4 Eye-Movements

Reading is an active process during which we make a series of fixations, saccades and regressions as our eyes move across the text. Fixations are the periods of time in which our eyes rest on a specific location in the text. Saccades are the rapid eye-movements separating the fixations during which our eyes jump from one location in the text to another. Regressions are backward eye-movements during which we revisit the text that has already been processed. Variability in eye-movements have been shown to give a clear indication of cognitive processes during reading. For example, when reading a difficult text, readers tend to fixate longer on specific locations in a text, show shorter saccades and make more regressions, i.e. revisit the text more frequently. By monitoring the eye-movements of readers during comprehension of inference-inducing passages, one can record the levels of fixations, saccades and regressions at particular points in the passage and thus pinpoint the time at which inference generation occurs (see Rayner, Raney & Pollatsek, 1995 for a detailed discussion). Eye-movement procedures have been used to study cohesive devices (e.g. Ehrlich & Rayner, 1983; Duffy & Rayner, 1990; Myers, Cook, Kambe, Mason and O'Brien, 2000) and elaborative inferences (e.g. Calvo, 2001; Calvo, Meseguer & Carreiras, 2001). While the collection of eye-movement data is clearly a desirable way in which to measure on-line inference generation, it is a labour-intensive procedure and is therefore not at present feasible for all research.

2.4 The Three-Pronged Approach

A parsimonious method for use in the study of inference generation is the three-pronged approach (Magliano & Graesser, 1991). This approach combines the use of behavioural measures, such as those outlined above, with verbal protocols and theories of discourse processing. To summarise this approach, verbal protocols such as question generation or think aloud procedures are used initially to reveal potential inferences

within a text. Theories of discourse processing are then applied to predict which of these potential inferences will be generated during comprehension. Finally, behavioural measures are used to assess whether these predictions are upheld. This approach has been used successfully by a number of researchers (e.g. Suh & Trabasso, 1993; see Magliano & Graesser, 1991 for a review).

However, there are a number of potential problems with this approach. Firstly, including all three-stages proposed by the three-pronged approach is time-consuming. Second, while many verbal protocols have been collected from adult participants and have been used to create behavioural measures for use with adults, it is not clear whether these adult verbal protocols would lend themselves to the creation of children's behavioural measures. It may be that the types of inferences adults generate during the collection of these verbal protocols vary greatly from those that children might generate. There are two methods for revealing inferences through verbal protocols; question answering and think-aloud. The question-answering procedure requires the participant to read a passage and answer questions pertaining to specific sentences or parts of that passage. The questions are designed to reveal the inference concepts those parts of the passage might induce. While a relatively simple task, the generation of questions requires some a priori assumptions regarding the types of inferences that might be generated from the text.

In contrast, the think-aloud method makes no such a priori assumptions and as such can be viewed as a less contaminated way of revealing inferences. However, to complete this task readers are asked to read a passage sentence by sentence and tell the experimenter whatever comes to mind at the end of each sentence. As such, the task is demanding and requires a certain degree of meta-cognitive skill. Therefore, the collection of verbal protocols from young children using this method would prove difficult and may render the results unreliable. Trabasso, Suh, Payton and Jain (1995) report the results of a series of studies with adults and children, in which participants were asked to think-aloud during reading. The results from the verbal protocols were similar for adults and children. However, while the adults were asked to produce think-aloud data for eight stories, the children were only given two stories to read. The comparison between adult and child verbal protocols is therefore limited to just two stories and does not fully address the reliability of the think-aloud method for use with

children. Thus, the use of the three-pronged method for investigating inference generation is not employed in this thesis.

2.5 Conclusions

Many procedures have been employed in the study of inference generation. Research has progressed from the use of off-line measures to more valid on-line measures such as lexical decision and naming speed. Moreover, the development of eye-movement procedures provides an unequivocal on-line measure of the processes that occur during comprehension, thus overcoming some of the problems inherent in the behavioural on-line measures. Finally, the three-pronged approach appears to be the most comprehensive means by which to measure inference generation, combining verbal protocols with theoretical predictions and behavioural measures. While this method has been used extensively in the adult literature, its suitability for use with children has yet to be thoroughly investigated. In short while many procedures have been used in the study of inference generation, no procedure is entirely free of problems. The appropriateness of any one procedure will depend on the needs of the individual research project.

In this thesis, the sentence verification task has been used to assess inference generation in children. This task was chosen for a number of reasons. First, as outlined in the discussion above, an on-line measure of inference generation is clearly preferable to the use of off-line measures. Resource limitations prohibited the use of eye-tracking as a means of assessment and it was therefore necessary to use an on-line behavioural measure. Sentence verification was chosen over lexical decision and naming latency measures to make it possible to use counterfactual materials in which the inference concepts to be tapped would be too complex to express in a single word. In addition to this, it was felt that using this task would reduce the risk of word-based priming effects. Taken together, these considerations led to the sentence verification task being the most appropriate for use in this thesis.

Chapter 3

An Investigation into The Relationship between Reading Comprehension and Mathematical Reasoning.

3.1 Introduction

The evidence described in chapter 1.1 clearly suggests that poor comprehenders have difficulty making inferences at both the word and text levels. However, as discussed in chapter 1.2 there are many potential inferences to be made in text comprehension and it is not clear from previous research whether poor comprehenders have generally poor inferential skills, or whether their difficulties are more pronounced for particular types of inferences. Furthermore, the pervasiveness of these inferential skill deficits has not been established. For example, it is not clear whether these difficulties extend to tasks other than reading and listening comprehension and whether poor comprehenders are equally as poor at making inferences in all types of text formats. One way of addressing these issues is to compare the performance of poor comprehenders with a variety of tests that require some degree of inferential ability. For example, mathematical reasoning problems typically require inferences to be made in order to solve them. Thus, looking at the mathematical reasoning ability of poor comprehenders may help to assess the pervasiveness of their inferential skills deficits. However, in terms of different types of inferences, it is unclear whether all of the inferences required to complete mathematical reasoning problems are also important to reading comprehension. Mathematical reasoning may rely more on deductive inferences based on formal logic while reading comprehension may require more inductive, flexible inference generation. Therefore, it is important to establish the types of inferences that are of primary importance to each of these tasks. Furthermore, before we can assess the relative performance of poor comprehenders on tests of mathematical reasoning and reading comprehension, we first need to look at the relationship between these variables in children with normal comprehension skills.

Aim of the present study: The present study was designed to establish the normal pattern of inference generation in children with unimpaired comprehension on a variety of tasks. Furthermore it was intended to establish the overlap between the

inferential skills needed to complete mathematical reasoning and reading comprehension tasks.

3.2 Experiment 1

3.2.1 Method

a) **Participants:** 38 children from a local primary school were recruited for the study. 8 children from each of years 2, 4, 5 and 6, and 6 children from year 3 were tested. The children were randomly selected from each class by their teacher with the only criterion being that they were judged as having average reading ability. Details of the participants can be found in Table 3.1.

Year Group	Number of Children	Mean Age (mths)	Standard Deviation
2	8	81.96	3.48
3	6	97.32	4.92
4	8	106.32	4.56
5	8	118.68	4.44
6	8	130.92	2.64

Table 3.1 Mean chronological age of children in each year group

b) **Design:** This study was a within-subjects correlational design with each child completing each of five standardised tests in the same order.

c) **Materials:** A battery of standardised reading and mathematical ability tests were administered to the children. Three of these tests were also rated according to a set of inference criteria described below.

Standardised Measures of Reading and Language Ability

1. *The Neale Analysis of Reading Ability* (NARA II; Form 1; Neale, 1989) was administered to assess reading accuracy and reading comprehension. The NARA II is a story-based test. The child is presented with a booklet of short stories which they are

required to read aloud to the experimenter. The stories get progressively harder and longer as the child works his way through the book. In order to obtain a reading accuracy score, the experimenter simply records the number of reading errors the child makes during each individual story. These separate error scores are then totalled to give the cumulative reading accuracy score. Reading comprehension is measured in the following way. Following the completion of each story as described above, the child is required to answer a number of questions which assess their understanding of what they have just read. The questions are presented orally by the experimenter and the child is allowed to refer back to the text in order to answer the questions. An overall comprehension score is obtained by summing the total number of comprehension questions answered correctly throughout the test as a whole.

2. *The Wechsler Objective Reading Dimensions* (WORD; Wechsler, 1990) reading comprehension sub-scale was administered as a second test of reading comprehension. Rather than being a story based task, the WORD can be described as a general text comprehension task since the passages used do not all conform to a narrative format. Some items are merely single sentence statements while others are paragraphs that are more analogous to expositions, or commentaries, than to narratives. In this test children are required to read the items in silence and do not receive any assistance with reading accuracy. When they have read the passage the experimenter asks them a single question based on what they have read. The overall score is the total number of questions correct for the whole test.

In both tests of reading comprehension the text was available for reference by the children during questioning.

3. *The Wechsler Objective Numerical Dimensions* (WOND; Wechsler, 1996) was administered to assess mathematical skills. In the Mathematical Reasoning sub-test, children are presented with word problems to which they have to provide an oral response. The problems are presented in printed form but the experimenter also reads the problem aloud. The children are given a pencil and paper which they are permitted to use for “workings out”.

The Numerical Operations test is a pencil and paper test in which participants are given a worksheet containing arithmetical problems of increasing complexity. Problems range from simple addition to more complex operations using fractions and algebra.

4. *The Graded Non-Word Reading Test* (GNWRT; Snowling, Stothard, & McLean, 1996) was administered as a test of decoding ability. This test requires children to read a number of printed non-words (i.e. fot, gug, chamgalp) aloud. They are instructed to simply read the non-words using the pronunciation they think is appropriate. The stimuli range from one to two syllables in length. The test is scored by calculating the absolute number of non-words the child pronounces correctly.

Assessment of Inferential Skill

The questions that comprise the two comprehension measures and the mathematical reasoning test were examined by the author. The questions posed in each test were placed into one of the following 7 categories indicating the type of information required to answer the question correctly.

Categories for the classification of test questions

1. *Elaborative Inferences*: Elaborative inferences are those that are not necessary for comprehension of the text presented but would enrich the mental representation of that text. Such inferences typically require text to be interpreted in terms of the reader's real world knowledge. For example, predicting a possible outcome from a description of events would not be necessary for understanding the text but would be an elaboration of the information given in the text.

2. *Cohesive Inferences*: These inferences rely on linguistic cues present in the text e.g. anaphora, pronoun resolution. For example, in order to form a coherent representation of the sentence 'Tom hit Janet with his tennis racket.' the reader would need to infer that the tennis racket belonged to Tom. This type of inference is aided by the fact that the two people referred to in the sentence are of different gender types and

that the 'tennis racket' is preceded by the pronoun 'his' indicating that the tennis racket belongs to someone of the male gender. Since Tom is the only male mentioned in the sentence the inference that the tennis racket belongs to Tom can be made.

3. *Knowledge Based Inferences*: These inferences rely on the application of the reader's real world knowledge to the text. They differ from elaborative inferences in that they are *necessary* for textual coherence. The information needed is not always explicit in the text but the inference is necessary in order to form a coherent representation of the text.

4. *Evaluative Inferences*: This type of inference relates to the emotional outcomes of events; the consequences of actions etc. They are also necessary to the understanding of the text and, like knowledge based inferences, rely on the readers use of real world knowledge in interpreting textual information. For example, if you read the passage 'Helen fell off her bike and broke her arm. She started to cry.' the reader could infer that Helen was likely to be crying because she was in pain. Based on the information presented in the text no other interpretation would make sense.

5. *Transitive Inferences*: This is a type of deductive inference involving the relating of a group of objects/ numbers/ people etc. to each other in terms of a specific feature. For example, when reading that 'John was older than Mary but Mary was older than Peter.' you would infer that John was also older than Peter.

6. *Literal Information*: This category was used for those questions which required no inference since the answer was literally written in the text. For example, if the sentence 'Sally wrote to her Grandma' was followed by the question 'Who did Sally write to?' this question would be categorised as requiring attention to literal information only since it is clearly stated in the text that Sally wrote to her Grandma.

7. *Vocabulary Dependent*: Items were rated as vocabulary dependent if the respondent considered that the most important factor in providing the correct answer was understanding a key word i.e. if the vocabulary used was particularly difficult. For

example, if having read the sentence 'Mary is a myope' the reader was asked 'Does Mary have good eyesight?' they would need to have some knowledge of the term 'myope' in order to answer the question.

Following this initial classification phase, seven adult raters assessed each test on the basis of the classifications identified by the author. Raters were given details of the seven categories which were identified as classifying the different types of information needed to answer the questions in each of the tests. They were told to remember that these inference types are not mutually exclusive. That is to say, one piece of text may provide a large number of potential inferences. However, the point here was to classify each specific question in terms of the specific type of information needed to provide a correct answer.

d) Procedure: In session 1 the NARA II, WORD, GNWRT and WOND Mathematical Reasoning task were administered to each child individually and were completed according to the instructions laid out in each test manual. At session 2, the WOND test of numerical operations was administered as a group test, the 8 participants from each year group being tested together. The children were told to complete the test in silence and not to discuss the answers with their neighbours. They were instructed to simply complete as much of the workbook as they could, missing out any questions they could not do and stopping when they felt they could go no further. They were not given a time limit but all children completed the test within twenty minutes. When administered individually, this test has a discontinue criteria of four consecutive scores of zero. When scoring these tests, any scores achieved after the discontinue criteria was met were ignored.

3.2.2 Results

The mean standard scores achieved on each test by each age group are shown in Table 3.2. The individual raw scores were entered into a zero-order correlation in order to assess the relationship between each of the variables.

Measure		Year 2 (n=8)	Year 3 (n=6)	Year 4 (n=8)	Year 5 (n=8)	Year 6 (n=8)	Total (n=38)
WORD	Mean	99.5 (7.41)	95.67 (7.97)	99.40 (8.8)	99.50 (10.99)	103.25 (8.96)	99.66 (8.78)
	Range	84 – 106	84 – 105	90 – 117	80 – 117	90 – 115	80 - 117
NARA II Acc	Mean	104.38 (6.76)	102 (8.94)	98.38 (6.93)	100.25 (6.07)	115.63 (10.29)	104.24 (9.76)
	Range	91 – 111	89 -111	90 – 112	93 – 109	94 – 130	89 – 130
NARA II Comp	Mean	101.25 (7.01)	91.67 (7.31)	93.25 (8.96)	96.25 (15.65)	108.63 (12.98)	98.50 (11.68)
	Range	90 – 109	84 – 100	88 – 99	81 - 128	91 – 124	81 – 128
GNWRT	Mean	112.75 (13.99)	113.67 (11.67)	94.13 (12.96)	102.75 (10.61)	107.88 (7.40)	105.84 (13.10)
	Range	98 - 130	90 – 120	77 - 115	90 - 117	98 - 115	77 - 130
WOND MR	Mean	102.13 (7.88)	105.33 (8.04)	104.13 (11.04)	102.38 (10.82)	115.40 (13.10)	105.89 (11.14)
	Range	88 - 113	96 – 116	93 – 129	90 - 107	94 – 128	88 – 129
WOND NO	Mean	103.50 (9.90)	106.33 (14.46)	102.38 (11.59)	104.50 (4.87)	114.50 (9.41)	106.24 (10.66)
	Range	95 - 123	90 - 125	86 - 118	96 - 111	97 - 123	86 - 125

Table 3.2 Overall mean scores of each year group on each standardised test (MR = Mathematical Reasoning; NO = Numerical Operations; standard deviations in parentheses; all tests have an average score of 100 and a standard deviation of 15).

a) **Zero Order and Partial Correlations:** The zero order correlations can be seen above the diagonal in Table 3.3. As expected, there were significant correlations between the NARA II accuracy and comprehension measures, between both NARA II measures and the WORD comprehension measure, and between the measures of mathematical reasoning and numerical operations. However, the significant positive correlations between the comprehension measures and the tests of mathematical ability were not expected and were therefore looked at in more detail.

Reading comprehension is often confounded by reading accuracy such that good reading skills are usually accompanied by good comprehension skills while poor reading skills typically appear alongside poor comprehension skills. Indeed, in the present study, the correlation between NARA II reading accuracy and NARA II comprehension skills reflected this relationship ($r = 0.56, p < .01$). This fact may have influenced the correlation between the comprehension and mathematical measures seen here such that good word level skills may have contributed to the performance on all tests. Therefore, a partial correlation was performed controlling for word level reading skills by partialling out NARA II accuracy and GNWRT scores. Age was also controlled for. The results of this partial correlation are shown below the diagonal in Table 3.3.

	GNWRT	NARA II Acc	NARA II Comp	WORD Comp	Mathematical Reasoning	Numerical Operations
Chron Age	.42***	.83***	.69***	.85**	.88***	.93***
GNWRT		.56***	.32**	.36***	.30	.31**
Nara II Acc			.70***	.80***	.75***	.80***
Nara II Comp				.74***	.71***	.72***
WORD Comp			.35**		.79***	.91***
Math Reasoning			.30*	.14		
Num Ops			.27	.20	.47***	

Table 3.3 Zero order and partial correlations (below diagonal) between raw scores on standardised tests (Chron age = chronological age; NARA II Acc = NARA II Accuracy; NARA II Comp = NARA II Comprehension; WORD Comp = WORD Comprehension; Math Reasoning = Mathematical Reasoning; Num Ops = Numerical Operations). Partial Correlations in italics controlling for chronological age and word-level reading skills (NARA II accuracy and GNWRT). * $p < .10$, ** $p < .05$, *** $p < .01$

Table 3.3 shows that controlling for word level reading skills and age significantly reduces the correlation between the comprehension measures and the mathematical measures. However, the correlation between NARA comprehension and mathematical reasoning remains statistically significant. This relationship was not reflected in that between WORD comprehension and WOND mathematical reasoning and cannot be explained in terms of word-level reading skills.

b) Concurrent predictors of reading comprehension: The results of the zero-order and partial correlations described above suggest that there is a relationship between performance on the NARA II comprehension measure and the WOND test of mathematical reasoning. However, the nature of this relationship is unclear. For example, it may be the case that performance on one of these factors underlies performance on the other, or alternatively that performance on both of these tests is reliant on a third underlying factor. In order to examine this relationship in more detail, multiple regression analyses were conducted to assess the concurrent relationship between reading comprehension and mathematical reasoning skills.

With NARA II comprehension ability as the dependent variable, chronological age accounted for 47% of the variance when entered on the first step. NARA II accuracy and GNWRT together accounted for 6% of the variance in reading comprehension when entered on the second step which is not significant. WOND mathematical reasoning scores were entered on step 3 and accounted for 4% of the variance, which was not significant. The WORD comprehension scores accounted for a further 5% of the variance when entered on step 4, which approached significance. Thus, when the variance due to word level skills was accounted for, mathematical reasoning did not account for a significant amount of the variance in NARA II comprehension skills. When entered on step 3, WORD comprehension skill significantly accounted for a significant 6% of the variance. However, when entered on step 4, mathematical reasoning accounted for just 3% of the variance, which is not significant. These results can be seen in Table 3.4 and indicate that mathematical

reasoning ability is not a strong predictor of performance on the NARA II comprehension measure.

Step	Variable	R ² change	P
1	Chron Age	.47	.00
2	Nara II Acc	.06	.13
	GNWRT		
3	Math Reasoning	.04	.08
4	WORD Comp	.05	.06
3	WORD Comp	.06	.04
4	Math Reasoning	.03	.12

Table 3.4 Multiple regression with NARA II Comprehension as the dependent variable and Chronological Age, NARA II Acc, GNWRT, Mathematical Reasoning and WORD Comprehension as the predictor variables.

c) Concurrent predictors of mathematical reasoning: The results of the multiple regression analyses described above suggest that mathematical reasoning is not a significant predictor of reading comprehension. However, the relationship between the two variables may be inverse, with reading comprehension predicting mathematical reasoning. A second set of multiple regression analyses was conducted to investigate this possibility. With mathematical reasoning as the dependent variable, controlling for chronological age on step 1 accounted for 78% of the variance. Controlling for numerical operations on step 2 accounted for a significant 5% of the variance. However, when entered on steps 3 and 4, neither the NARA II comprehension nor the WORD comprehension measures accounted for a significant proportion of the variance, irrespective of the order in which they were entered. These results indicate that reading comprehension skill is not a significant predictor of mathematical reasoning skill.

Step	Variable	R ² change	P
1	Chron Age	.78	.00
2	Numerical Operations	.05	.00
3	WORD Comp	.00	.78
4	NARA II Comp	.01	.28
3	NARA II Comp	.01	.26
4	WORD Comp	.00	.89

Table 3.5 Multiple regression with Mathematical Reasoning as the dependent variable and Chronological Age, Numerical Operations, NARA II Comprehension and WORD Comprehension as the predictor variables.

d) Inter-rater reliability for qualitative analysis of test questions.

The results of the qualitative analysis of comprehension questions can be seen in Table 3.6, which clearly illustrates that all three tests require a high proportion of inferences to be made in order to correctly answer the questions posed. However, this is particularly true in the case of the NARA II comprehension measure for which the percentage of inference type questions is far higher than in either of the other two tests. The WORD requires a far greater proportion of questions to be answered using literal information than either the WOND or the NARA II. Also illustrated in Table 3.6 are the inter-rater reliability results for each test following classifications by seven independent adult raters.

	WORD	NARA II	WOND
Total Questions	38	44	50
Literal Questions	31.59	13.63	12
Vocabulary Dependent	7.89	13.64	20
Cohesive	18.42	34.10	28
Knowledge-Based	13.16	29.55	28
Elaborative	26.32	4.55	0
Evaluative	2.63	4.55	0
Transitive	0	0	12
Inter-Rater Reliability (mean)	0.84	0.80	0.76
Inter-Rater Reliability (range)	0.44 – 0.80	0.37 – 0.76	0.02 – 0.63

Table 3.6 Percentage of each category of question identified in each standardised test.

e) The relationship between inferential skill and comprehension ability:

In order to examine the relationship between inferential skill and comprehension ability, a skilled and less-skilled comprehension group were selected from the original sample, and their performance on both tests of reading comprehension was compared. The 10 children with the highest NARA II comprehension standard scores were classified as skilled comprehenders and the 10 children with the lowest NARA II comprehension standard scores made up the less-skilled group (see Table 3.7). Performance on both tests of reading comprehension were then compared using a 2x2 ANOVA with the within-subjects factor Test (NARA II and WORD) and the between-subjects factor Group (Skilled and Less-skilled).

	NARA II Comprehension	WORD Comprehension
Skilled Group	113.90 (8.66)	107.00 (7.09)
Less-skilled Group	86.70 (3.56)	97.40 (7.53)

Table 3.7 Mean standard scores of Skilled and Less-skilled Comprehenders on NARA II and WORD tests of reading comprehension (standard deviations in parentheses).

A significant main effect of group revealed that the skilled group performed better on both tests of reading comprehension than the less-skilled group ($F(1,18) = 55.74, p < .01$). However, there was also a significant Group by Test interaction ($F(1,18) = 21.38, p < .01$). Post-hoc Tukeys HSD showed that the less-skilled group performed better on the WORD than on the NARA II test of comprehension with scores on the WORD being in the average range (HSD $p < .05$). Conversely, the skilled group performed better on the NARA II test of comprehension than on the WORD comprehension sub-test (HSD $p < .05$). These results are shown in Table 3.7.

In order to establish whether children who differ in comprehension skill also differ in inferential ability, and whether this difference is specific to a particular type of inference, a detailed analysis was performed on the children's responses to the different question types identified in each test. Performance on the evaluative, vocabulary based and transitive categories were not analysed as none of these categories were highly represented in any of the tests. Performance on the remaining 4 question types (Literal, Cohesive, Knowledge-based and Elaborative) was examined in the following way. In order to control for the fact that not all participants answered the same number of questions in each test, the total number of questions that each individual was presented with in each of the tests of reading comprehension was calculated. For each individual, the number of correct responses to each question in each test was then expressed as a percentage of the number of questions asked. Finally, the average percent correct for each question type across both tests of reading comprehension was calculated (see Figure 3.1).

These data were then analysed using a 2x2 repeated measures ANOVA with the between-subjects factor group and the within-subjects factor Inference Type. The main effect of Group ($F(1,18) = 8.51, p < .01$) was significant such that the skilled comprehenders performed significantly better than the less-skilled comprehenders. The main effect of Inference Type ($F(3,54) = 16.21, p < .05$) was also significant such that there were more correct responses given to literal information and cohesive questions than to any knowledge based and elaborative questions (HSD $p < .01$). Furthermore, the

percentage of correct responses given to literal information questions was significantly.

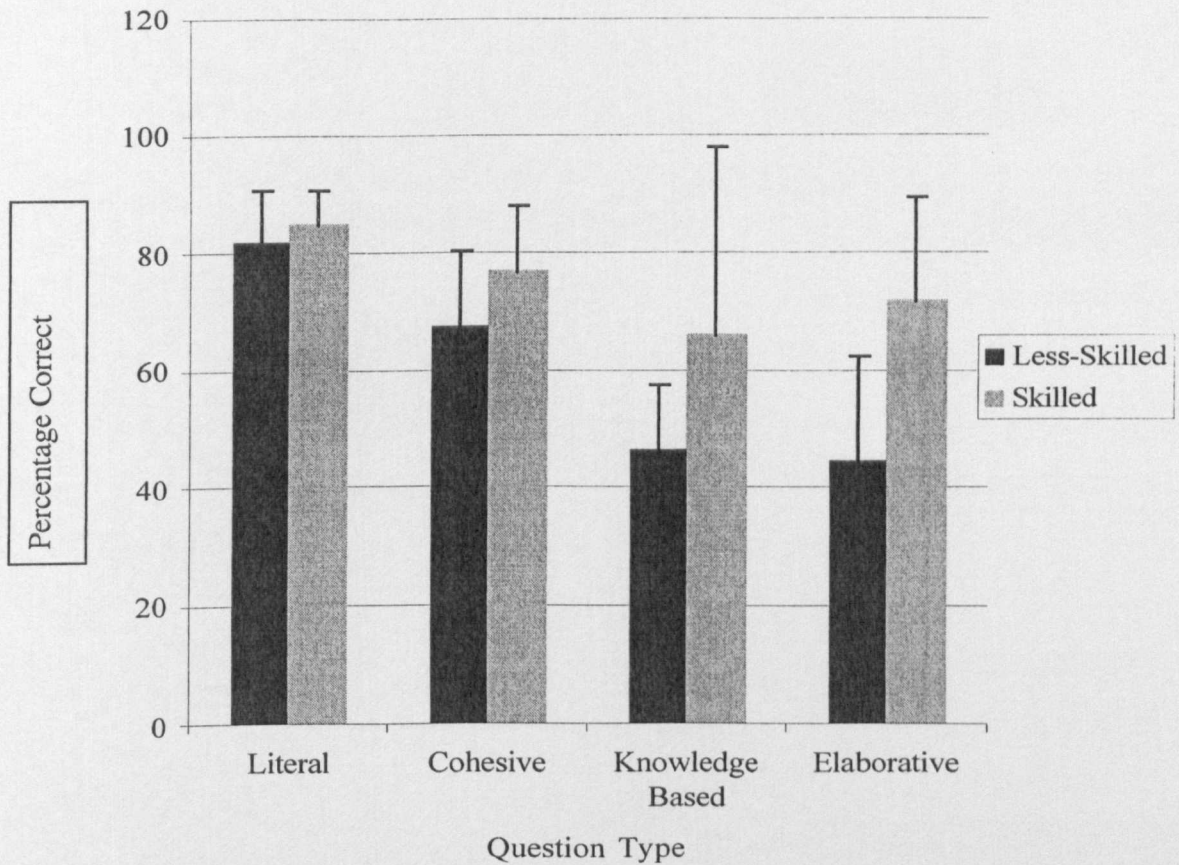


Figure 3.1 Performance of skilled and less-skilled comprehenders on each question type identified in the standardised measures

greater than that given to cohesive questions, while knowledge based and elaborative questions did not differ from each other. A significant Group by Inference Type interaction was also found ($F(3,54) = 2.89, p < .05$). Post-Hoc Tukeys HSD ($p < .01$) revealed that within the less-skilled group, performance on the literal questions was better than on any of the other question types. However, performance on the cohesive questions was better than on the knowledge-based or elaborative questions which did not differ from each other. Within the skilled group, performance on the literal questions was better than performance on the knowledge-based questions. None of the other question types differed from each other. Finally between groups, the less-skilled group's performance on the knowledge-based and elaborative questions was significantly

poorer than the performance of the skilled group on all 4 question types. However, on literal and cohesive questions, the less-skilled group's performance was not significantly different from that of the skilled group on each of the corresponding question types.

3.2.3 Discussion

The results of the multiple regression analysis showed that mathematical reasoning is not a strong predictor of reading comprehension when age and word-level reading skills are taken into account. Similarly, NARA II reading comprehension does not significantly predict mathematical reasoning ability when age and simple number skills are first accounted for. Therefore, there does not seem to be a significant relationship between reading comprehension and mathematical reasoning ability over and above that mediated by basic level skills. No further consideration will be given to mathematical reasoning ability in the remainder of this thesis.

The results of the comparison between skilled and less-skilled comprehenders performance on both the story-based NARA II test of reading comprehension and the more general WORD test of reading comprehension showed that the less-skilled comprehenders performed in the normal range on the WORD while their NARA II comprehension scores were below average. However, while the skilled comprehenders performed above average on both tests of reading comprehension, their performance was significantly better for the NARA II than for the WORD comprehension test. This finding suggests that the varied nature of the WORD test of reading comprehension differs from the story-based NARA II test of reading comprehension in such a way as to facilitate the comprehension of the less-skilled group but inhibit that of the skilled group.

Examination of the qualitative analysis of each of the standardised test measures suggests that the tests of reading comprehension used here vary greatly in their reliance on different inference types. For example, a large proportion of questions in the WORD reading comprehension test rely on attending to the literal information in the text while the NARA II questions rely more on cohesive and knowledge-based inferences than any other question type. If the less-skilled comprehension group are capable of answering

literal questions but have a difficulty specific to inferential questions, this could explain the discrepancy in their performance on the two tests.

In order to examine the inferential skills of the two comprehension groups in more detail, the performance of each comprehension group on 4 of the question types identified in the qualitative analysis was examined. The findings suggested that the less-skilled comprehenders were perfectly capable of answering questions based in literal information or questions requiring a cohesive inference. Their performance on each of these questions types was comparable to that of the skilled comprehension group, and significantly better than the performance of either group on knowledge-based and elaborative inference questions. Furthermore, the skilled group performed significantly worse on the knowledge-based questions than on the literal questions.

These findings indicate that both groups of children found inferences that required the use of real-world knowledge more difficult than literal and cohesive questions. This difficulty was more pronounced for the less-skilled comprehenders and may explain the difference in performance on the two tests of reading comprehension used in this study. This discrepancy in performance can be explained by the finding that less-skilled comprehenders perform at a normal level on tests of verbatim memory (Yuill & Oakhill, 1991). If one considers that literal information questions could be answered by remembering specific propositions in the text and that cohesive inferences can be made using explicit cues from the text, it is not surprising that the less-skilled comprehenders are able to answer these questions correctly.

Taken together the findings of study 1 suggest that the specific reading comprehension difficulties experienced by some children may be mediated by a difficulty with specific inference types. This conclusion is based on the finding that, while the performance of the less-skilled comprehension group on the NARA II was well below average, the performance of this group on the WORD test of reading comprehension was within the normal range. However, the NARA II and the WORD test of reading comprehension also differ in terms of the types of passages included in each test. The majority of the passages in the NARA II are of a narrative structure

while the WORD contains passages of a variety of different text types ranging from single sentences to informative commentary type texts. Research has not yet established whether the pattern of inference generation is the same in texts of different formats. Therefore, a further possible reason for the discrepancy in performance of less-skilled comprehenders on the two tests of reading comprehension used here is that they find the comprehension of commentary style texts easier than that of narrative texts because a lesser degree of inferential skill is required in commentary style texts. In order to address this question, study 2 will examine the on-line inferences of children with normal comprehension skills during the reading of narrative and expository, or commentary style, texts.

Chapter 4

The Effect of Text Format on Inference Generation.

4.1 Introduction

A narrative can be defined as a spoken or written account of a series of connected events in the order in which they occur. In other words, a narrative tells a story. Conversely, an exposition is broadly defined as an explanatory statement or commentary and can take many different forms. Olson, Mack and Duffy (1981) identified two common forms of exposition; the inductive argument where the author aims to convince the reader of a particular proposition, or a comparative essay where the author compares and contrasts two objects, arguments, or phenomena. However, there are many other types of exposition. As discussed in chapter 1, research has shown that certain types of inferences are important to narrative comprehension. However, it is not clear whether there are significant differences in the type of inference processing required in the comprehension of narrative as opposed to expository text. In order to fully consider this question it is necessary first to discuss the basic differences between these text types.

4.1.1 Structural Differences: It is generally agreed that a narrative text follows a specific structure and contains certain fundamental elements that differentiate it from an expository text. Narrative text is considered to be active in that it is centred around a series of actions and events, and is constantly changing and developing. Olson et al. (1981) proposed that a narrative text centres around a pattern of complication and resolution such that a complication or conflict occurs which is resolved by the actions of the central character or characters. This pattern is not thought to be relevant to expository text, which typically does not contain characters, actions or events and can therefore be thought of as having a more passive format. Furthermore, narrative text is generally divided into a series of segments which are temporally or spatially discrete, but which must be causally connected. Conversely, expositions do not have these structural constraints. Given the causal structure of narrative texts, one might expect that the generation of coherence inferences is necessary in order to maintain a clear representation of the text. This assumption however, does not necessarily apply to expositions since they are not constrained in the same way.

4.1.2 Familiarity of content: It has been suggested that one of the major differences between narrative and expository text is that people are usually familiar with the content of narratives whereas the content of expositions is often novel. The familiarity of narrative content is thought to stem from the expectation of the reader that the content will be consistent with real world knowledge such that it will deal with concepts, actions and events which occur in every day life (Olson et al., 1981). Expositions on the other hand are expected to contain information of a more abstract and novel nature. A recent study by Narvaez et al. (Narvaez, van den Broek, & Barron-Ruiz, 1999) found that subjects reported more knowledge-based coherence breaks when reading expository text than when reading narrative text. They classified knowledge-based coherence-breaks as breaks in the coherence of the text due to the reader's lack of knowledge or experience. Thus, it would appear that lack of knowledge is more likely to cause a problem for readers of expositions than readers of narratives. However, this distinction does not seem to account for the fact that some narrative texts can be based entirely in fictional worlds with no real correspondence to everyday life. Nor does it account for the fact that expositions do not always contain unfamiliar content. Britton and Gülgöz (1991; experiment 2) studied expository text comprehension using texts containing information about the Vietnam War. While the subject of the Vietnam War may be unfamiliar to some people, the participants in the experiment were U.S. Air Force recruits and presumably had some knowledge of the content of the passages. Britton and Gülgöz devised a paradigm which revealed participants mental representation of a previously presented passage. They presented participants with a passage describing some of the events of the Vietnam War. They then compared the mental representations of the U.S. Air Force recruits with expert military historians and the passage author's own representation. The results showed that when an expository text was modified such that no inference processing was required during reading, the representation of the U.S. Air Force recruits more closely resembled both the military experts and the author's own representations than when inference processing was required. Therefore, familiarity with the content of an expository text does not necessarily aid in inference generation.

4.1.3 Reader's Purpose: Narrative and expository text are thought to differ in terms of purpose such that a narrative text is generally written to entertain (Olson et al., 1981) while an exposition usually has some informative value (Graesser, Singer, & Trabasso,

1994). Indeed, this difference in reading purpose has been shown to produce differences in inference generation. For example, Narvaez et al. (1999) divided subjects into two groups; a study group who were instructed to read for learning, and an entertainment group who were instructed to read for pleasure. Using a 'think aloud protocol', a test of delayed free-recall and a straightforward question and answer comprehension measure they assessed the effects of reading purpose on inference generation. Overall, they found that those readers told to read for study purposes engaged in less inference generation than those instructed to read for pleasure. Thus, it would appear that reading purpose does affect the generation of inferences.

4.2. Research into the Effect of Text Format on Inference Generation.

It is clear from the research described above that there are differences between narrative and expository text formats. The question of whether inference generation is differentially affected by the format of the text is therefore an important one. However, research into this question has provided inconsistent results. As summarised in chapter 1, most researchers agree that bridging inferences are reliably drawn during narrative text comprehension (Fincher-Kiefer, 1996; Magliano, Baggett, Johnson, & Graesser, 1993; Potts, Keenan, & Golding, 1988; Singer, 1983) and some researchers propose that elaborative inferences can also be drawn in narrative comprehension (Fincher-Kiefer, 1996; Graesser et al., 1994; Murray, Klin, & Myers, 1993). However, the issue of inference generation during the reading of expository text is unclear.

4.2.1 Inference Generation in Expository Text: Millis, Morgan and Graesser (1990) conducted a study which incorporated a comparison between causal bridging inferences and predictive inferences in expository text. Specifically, they calculated the number of inferences that could be generated by each sentence in passages with scientific or technological subject matter. They then related the reading time of each passage to this "inference volume score" (Millis et al., 1990). They found that overall, the number of possible inferences that could be generated by each passage significantly predicted the reading time of that passage, such that more inferences resulted in increased reading times. Furthermore, when they looked at specific inference types, they found that this relationship was the same for both causal bridging inferences and predictive inferences. Their results suggested that both causal bridging and predictive inferences were drawn on-line in

expository texts of this nature. This finding was surprising considering that only causal antecedent inferences have been reliably shown to be drawn on-line in narrative text comprehension (e.g. Potts, Keenan & Golding, 1988). However, a subsequent study produced somewhat different results (Millis & Graesser, 1994). Participants were presented with short scientific texts in both a word-by-word rapid serial visual presentation procedure and a word-by-word self-paced reading paradigm with only the sentence-final word being of fixed duration. They were then required to make lexical decision judgements about inference words and unrelated words. The inference words were related to either a causal antecedent inference present in the text or a causal consequence inference that could potentially be generated from the text. Lexical decision judgements were faster for causal antecedent inference words than for unrelated words. However, lexical decision latencies for causal consequence inference words were comparable with those for unrelated words. The results suggested that only causal antecedent inferences were drawn on-line. Despite the discrepancy with the earlier study, these findings seem to be more in-line with previous research on narrative comprehension.

In contrast, Noordman, Vonk and Kempff (1992) found that people did not draw causal bridging inferences on-line during expository text comprehension. Instead they found that causal bridging inferences were only drawn when subjects were specifically asked to do so. In their first experiment, Noordman et al. (1992) presented participants with short expository passages that they read sentence by sentence using a self-paced reading procedure. There were two versions of each passage; an explicit and an implicit version. A causal inference could be invoked in each version of the passage. However, in the explicit version, the mediating idea that needed to be generated in order to draw the inference was explicitly included before the sentence containing the target inference sentence. Thus, in the explicit condition, no actual inference needed to be made since all the relevant information was available. The mediating sentence was omitted in the implicit version. For example, consider the following passage:

"(1) Spray cans usually contain a gas as a propellant. (2) Propellants serve to disperse the product in the spray can, e.g. paint. (3) Propellants must not combine with the product in the spray can. (4) Chlorine compounds make good propellants, because they react with almost no other substances." (adapted from Noordman et al., 1992, table 1)

Sentence 3 is the explicit information sentence, while sentence 4 is the target inference sentence. In the explicit version of the passage both sentence 3 and 4 were present. However in the implicit version, sentence 3 was not included in the passage. In the absence of sentence 3, the information that "propellants must not combine with the products in the spray can" must be inferred from the information in sentence 4.

If the causal inference was made on-line, the reading time for target sentences in the implicit condition was predicted to be longer than that for target sentences in the explicit condition since more processing time would be required. However, when the reading times for target sentences in each passage were compared, no significant difference was found between versions. Following presentation of each passage, subjects were presented with verification statements which they had to judge to be either true or false. One of these statements described the causal inference that had been included in the text. It was predicted that if the inference had been drawn on-line, the verification times would be the same in both the implicit and the explicit conditions. However, it was found that verification time was longer for sentences in the implicit condition implying that the causal inference was drawn when verifying the statement and not when reading the passage. These findings were replicated in a second experiment conducted to address some methodological issues in experiment 1.

In a third experiment, readers were presented with the same passages but the mediating idea sentences in the explicit condition were altered so that they were no longer consistent with the target inference sentences. For instance, sentence 3 in the example above would be replaced by the sentence "propellants must combine with the products in the spray can." All readers were told to read the passages with the objective of making suggestions for improvement. Furthermore, half of the participants were warned that the passages might contain inconsistencies. It was found that a large proportion of readers in

the no-warning group did not detect the inconsistency between the mediating sentence and the inference sentence, indicating that they did not integrate information across the sentences and thus did not make the inference. Moreover, while a significantly larger proportion of the warning group did detect the inconsistencies, 45% still failed to make the inference.

Finally, Noordman et al. (1992) conducted two further experiments in which they manipulated the task demands of the study. In short they found that, when given a specific task to which the making of the inference was relevant, readers did draw the causal inferences on-line. To summarise, the findings of Noordman et al. (1992) suggest that readers do not make causal inferences in expository text, although given a specific purpose they can be encouraged to do so. While clearly in conflict with the findings of Millis et al. (1990, 1994), these findings are supported by other studies of inference generation during the reading of expositions.

Britton and Gülgöz (1991; experiment 1) modified an expository text such that the gaps in the text that required an inference to be made were explicitly filled so that no inferences were necessary. They presented participants with either this modified version or the original version of the text. In a free recall task, readers of the modified version recalled far more than readers of the original version. Furthermore, participants were given a multiple-choice task that tested whether the information provided by the drawing of an inference in the original version could be accessed. Results showed that participants who were given the modified version of the text where all the inferences had been explicitly drawn, performed far better on this test than readers of the original version of the text although the groups did not differ in their memory for the factual information included in the passages. Thus, these results support the assumption that inferences are not made on-line in expository text.

4.2.2 Comparisons of Inference Generation in Narrative and Expository Text:

All of the studies described above (Britton & Gülgöz, 1991; Millis et al., 1990, 1994; Noordman et al., 1992) focus solely on expository text. None of the studies include a direct comparison between expository and narrative text. In order for a clear distinction to be drawn between the two text types, they must be directly compared.

Olson et al. (1981) performed such a comparison between essays and narrative texts using a similar paradigm to that used by Millis et al. (1990). They gave examples of each of these text types to one group of subjects and recorded their reading times. They also gave these passages to a second group of subjects and asked them to talk aloud while reading the texts. For example, they encouraged subjects to comment on the text and their comprehension of it, on connections they might see between sentences in the text, on predictions they felt they could make on the basis of what they had read, and on any inferences they could draw from the text. From this 'talk-aloud' data they could record the comments people made while reading each type of passage. They compared the talk-aloud data with the reading time data, the assumption being that the more inferences a sentence could generate, the longer the associated reading time for that sentence would be, since generating the inferences would require extended processing time. They found that this relationship was true for the narrative texts. However, for the expository texts, the number of inferences included in a sentence was unrelated to the reading time of that sentence, suggesting that the inferences were not drawn in expository texts.

The findings of Olson et al. (1981) are supported by the more recent study of Narvaez et al. (1999) described above. As well as dividing their subjects into groups with different reading purposes, they presented each group of subjects with both an expository and a narrative text. They found that, irrespective of study purpose, expository text evoked more study-type behaviours than narrative text. In terms of inference generation, this meant that readers of expository text were more likely to repeat words or phrases, evaluate the content of the text or indicate coherence breaks due to a lack of knowledge than were readers of narrative text. Thus, they did not seem to engage in a great deal of inference generation. Readers of narrative text however, were more likely to engage in predictions and to explain the content of the text in terms of background knowledge and text-based inferences than readers of expository text. They therefore seemed to engage in more active inference generation.

4.2.3 Other Factors Affecting Inference Generation: Other research has indicated that text type has no effect on the generation of inferences. Indeed in the Narvaez et al. (1999) study, readers of the expositions did produce a number of associations between the

content of the text and both background and text-based information, indicating that they did engage in some form of integrating activity such that they were relating the information presented to prior knowledge. Research by Singer and colleagues (Singer, Harkness, & Stewart, 1997) has suggested that, at least in the case of causal bridging inferences, text format is not an important factor in inference generation. For example, Singer et al. (1997; experiment 1) looked at the expository texts used by Noordman et al. (1992) and the narrative texts used by Singer, Halldorson, Lear and Andrusiak (1992). Participants were presented with the passages using a sentence by sentence self-paced reading paradigm. Following presentation of the passage, they were given two yes/no questions. One of these questions tapped the reader's memory of the specific detail of the passage. The second question tapped the idea that was needed to draw the causal link between the clauses in the target sentences. Using this paradigm, Singer et al. (1992) had previously demonstrated that causal bridging inferences were drawn on-line in narrative texts since the causal link question was answered more quickly than a similar temporal link question. The results of experiment 1 confirmed the findings of Singer et al. (1992). They also supported the findings of Noordman et al. (1992) such that causal bridging inferences were not made in expository text.

However, in a series of subsequent experiments, Singer et al. (1997; experiments 2 – 4) designed a further set of expository texts and compared them with Noordman et al.'s (1992) materials. They found that while the 'Noordman' materials consistently produced the same results, causal bridging inferences could be generated using the new 'Singer' materials. They proposed a number of reasons to explain this discrepancy. For example, Singer et al. (Singer, 1993; Singer et al., 1992; Singer et al., 1997) found that causal bridging inferences are reliably made when they can be validated against real world knowledge. However, if expository texts contain unfamiliar content (Olson et al., 1981) then the validation hypothesis cannot apply. Instead, it is proposed that the difficulty of the text will influence whether or not inferences are drawn. In support of this, the 'Singer' materials were shown to be less difficult than the 'Noordman' materials such that the target sentences took less time to read despite being equated for length. Furthermore, the error rates for the 'Noordman' materials were higher than for the 'Singer' materials. In addition to text difficulty, Singer et al. (1997) manipulated factors such as discourse cues, reading time and wording of the questions, and found that all of these factors had an inhibitory effect on

inference processing. As such, it was concluded that while causal bridging inferences may be hindered by expository text, they are not completely prevented. Given a relatively easy text containing clear discourse cues and unambiguous wording, and given unlimited reading time, causal bridging inferences can be drawn on line in expository text.

Perrig and Kintsch (1985) presented participants with a text describing the spatial layout of an imaginary town. Two versions of the text were constructed; a route text and a survey text. The route text could be compared to a narrative text because it contained a large number of causal connectives. Conversely, the survey text was written in a less coherent fashion, containing no causal connectives and was therefore more analogous to an expository text. This distinction is supported by the finding that in a free recall task the route text was recalled better than the survey text. Having read the passages, participants completed a sentence verification task in which they were asked to judge whether a sentence was true or false. One category of sentences required participants to judge the accuracy of an inferred spatial location. For example, the following is a section of the route version of the experimental passage:

"To get there drive east along the East-West Highway to the Green River which rushes out of some low hills to your left. Among the hills, before you come to the river, you see the high school which is connected to the highway by a small road."
(taken from Perrig & Kintsch, 1985; experiment 2, table 5.)

An inference verification sentence could take the following form:

(1) "To reach the high school, turn left off the highway ."

In order for participants to judge this sentence as being correct they would have had to infer from the passage that the high school was to the left of the highway, a piece of information not explicitly given.

Sentences were presented in both the route text style and the survey text style so that, irrespective of which style of passage they were given, participants received a mixed selection of verification sentences. Perrig and Kintsch (experiment 2) found that responses to an inference sentence were more accurate if the sentence was written in the style of the

text that the participant had read. However, the type of passage originally presented to participants did not affect their ability to draw inferences. Whether they had read a route or a survey text, they were still able to respond accurately to the inference verification sentence provided it was written in the same style as the text itself. What was important was the level of difficulty of the text. When presented with complex versions of the text in either the route or the survey style, performance on the inference verification task showed large floor effects (experiment 1). Thus, these results suggest that inferences can be made equally well in both text formats.

Surprisingly these results were only true for female participants. Male participants showed a preference towards the survey format such that, irrespective of the original passage read, they were more accurate on the inference verification sentences if they were presented in the survey style. It would appear therefore, that the male participants constructed a particular mental representation of the passage based on the spatial information regardless of the style in which that information was presented. They did not seem to attend to the coherence cues present in the route text and thus performed poorly when presented with inference verification sentences written in this style. This gender difference may be related to the difference in spatial ability consistently found across sexes (Halpern & Le May, 2000). However, it does serve to cast some doubt on these materials as a suitable means of assessing the construction of a discourse model.

4.3 Material Appropriate Processing

The research described above demonstrates the inconsistency that has characterised the question of inference processing in expository text. However, the findings of Singer et al. (1997) that causal bridging inferences are inhibited by expository text but not wholly prevented does not seem entirely incongruous with Noordman et al.'s (1992) findings that causal bridges can be encouraged in expository text comprehension. Furthermore, these findings are in line with the findings of McDaniel and colleagues regarding 'material appropriate processing' (Einstein, McDaniel, Owen, & Cote, 1990; McDaniel & Einstein, 1989; McDaniel, Einstein, Dunay, & Cobb, 1986). McDaniel et al. propose that certain types of materials will induce particular processing strategies to aid recall. In particular they propose that expository text will induce *individual-item processing* while narrative text will induce *relational-item processing*.

Individual item processing refers to the processing of each individual component of the text into the mental representation of that text. Relational-item processing refers to the relating of these components to one another. Relational-item processing therefore is clearly analogous to drawing inferences. As such, the common finding that inferences are reliably drawn during narrative text comprehension but not during expository text comprehension is clearly in line with the theory of material appropriate processing. Contrary to this view, McDaniel and colleagues further propose that for a complete representation of a text to be formed, relational-item processing and individual-item processing need to interact. For example, in order to facilitate construction of a complete representation of a narrative, individual item processing needs to be artificially induced; that is, encouraged by the performance of specific task. Conversely, in the case of expository text where individual item processing is naturally invited, relational item processing needs to be encouraged as a study strategy.

Studies have shown that recall of a text is enhanced by artificially inducing the processing strategy that is not naturally induced by the text format. However, if the naturally induced processing type is further encouraged, no improvement in recall is observed. Einstein et al. (1990; experiment 1) presented subjects with short descriptive texts and short narrative texts. While reading the passages, subjects engaged in one of four extra processing tasks; letter deletion, embedded questions, sentence reordering and outlining. Participants in the letter deletion condition received their passages with a number of letters omitted and had to fill these in as they read through the text. In the embedded questions condition, the paragraphs of the text were interspersed with questions regarding the explicit information in the text. These two tasks were thought to encourage individual-item processing. The sentence reordering group received their passages with each sentence printed on a separate slip of paper. Their task was to arrange the slips of paper in the correct order. Finally, in the outlining condition, subjects were encouraged to summarise the text as they read. These final two tasks were intended to invoke relational-item processing.

Following completion of these tasks, subjects were given a free-recall task in which they had to recall as much information as they could about the passage. The results of this

recall task demonstrated that the individual item processing tasks improved recall for the narrative texts but not for the expository texts, while the relational item tasks showed the opposite pattern of results. Further experiments demonstrated that the benefit of a complementary processing task is maintained even when recall is delayed (Einstein et al. 1990; experiment 2 and 3). It was concluded that complementary processing aids construction of a full representation of the discourse, while supplementary processing does not. The important point to note here though, is that despite being unnatural to an expository text, relational item-processing was achieved. This finding clearly supports the findings of some researchers that inference making can be observed in expository text given the correct conditions and motivation (Noordman et al., 1992; Perrig and Kintsch, 1985; Singer et al. 1997). In summary, research into the effects of text type on inference generation has produced inconsistent results. While it is clear that narrative texts are conducive to drawing inferences, the extent to which expository texts can encourage this type of processing is unclear.

4.4 Methodological Considerations Surrounding Research Into The Effects Of Text Format On Inference Generation.

a) Indirect comparisons of narrative and expository texts: In order to develop a fuller understanding of the difference between narrative and expository texts, it is necessary to directly compare participant's performance on the two types of test passage. As shown in the previous discussion, this has seldom been done in practice and while some attempts have been made, the passages that have been used have differed in terms of content as well as structure. For example, Narvaez et al. (1999) compared narrative texts that relayed fictional stories with expository texts that conveyed information about scientific phenomena. Thus, the subject matter of the two text types was vastly different and therefore the extent to which the passages could be compared is questionable. In order to fully examine the effects of text type on inference generation, it would be more desirable to control for content and simply create two versions of the same text.

b) Empirical truth of passage content: The content of the passages used in studies of inference generation generally agrees with real world knowledge. The inferences to be drawn from the passages are therefore also in agreement with real world knowledge.

However, narratives often contain fictional information that is based on facts that are true only in the context of the story. Research has shown that drawing inferences from false premises can be difficult even with narrative context as an aid (e.g. Franks, 1996). Thus, in order to gain a clearer picture of the difference between narrative and expository effects on inference generation it would be interesting to include passages based on false premises and assess the relative difficulty of inference generation across the two text types in this condition.

c) Differences in the inferential processes of children and adults: Much of the research into the effect of text type on inference generation has employed adult participants. It is not clear whether similar results would be obtained if the question were pursued from a developmental perspective. Research has identified differences in the inferential processes of children and adults. Paris, Lindauer and Cox (1977) for example found a developmental trend in the ability to use implied consequences as a cue for recall. They found that 7- to 8-year olds were significantly poorer at using such cues than 11- to 12-year olds, who in turn were poorer than undergraduate university students. Furthermore, Inman and Dickerson (1995) replicated the findings common to adult research, that causal bridges were made more frequently in narrative text than predictive inferences. However, unlike much of the adult research into the time-course of inferences (e.g. Potts et al., 1988), Inman and Dickerson found that both types of inference were made by children during recall rather than during encoding of the text.

Aim of the present study: As a result of the bias towards the use of narrative text materials in investigations into inference generation, it is impossible to assess how well the main theories of inference generation can be generalised across text formats. Unless the differential effect of text type on inferential processing has been established, studies into inference generation should include materials created in both narrative and expository text. Based on the paradigm used by Singer et al. (1992), the current experiment was designed to investigate whether children draw coherence and elaborative inferences in narrative and expository text when materials are matched for content.

4.5 Experiment 2

4.5.1 Method

a) Participants: 22 Year 5 children from a York primary school participated in this experiment. They were recruited on the basis of their performance on the Suffolk Reading Scale (SRS; Level 2, Form A; Hagley, 1986) administered as a screening device prior to the commencement of the experimental procedures. The SRS is a test of single sentence reading. To complete this test, the participant is presented with a booklet consisting of 75 sentences (4 practice and 71 test sentences), each with one word missing. Following each sentence is a list of five possible words that could fill that blank space. The child's task is to choose the word he or she thinks most suitably completes the sentence. This screening device was used to ensure that all the children were of an adequate reading age to complete the experimental tasks. The entire class of Year 5 children was tested but only those children who obtained a score within two standard deviations of the mean were selected to participate in the study. The final sample included 22 children with a mean age of 114.90 months (standard deviation 3.68 mths) and a mean SRS standard score of 100.77 (standard deviation 6.77). The children were randomly split into two groups in order that each group could receive a different experimental list. There was no difference between the two groups in terms of reading ability or chronological age (Table 4.1).

	Group 1	Group 2
Chronological age (mths)	114.72 (4.08)	116.40 (3.96)
SRS standard score ^a	101.73 (7.5)	99.82 (6.16)

Table 4.1 Mean chronological age and Suffolk Reading Scale (SRS) standard score for each of the two participant groups (standard deviations in parentheses; a = average 100, standard deviation 15).

b) Design: This was a 2 x 2 x 2 x 2 design employing two within and two between subjects factors each with two levels. The two between subjects factors were Form and Text Format. The factor Form had two levels; Form 1 and Form 2, which referred to the two lists of experimental stimuli. The factor Text Format had two levels; Exposition and Narrative referring to the text format in which the experimental passages were written. The two within subjects factors were Test Statement and Content. Test Statement consisted of

two levels; TS1 and TS2 referring to the type of test statements following each passage e.g. a causal inference test statement (TS1) and an elaborative inference test statement (TS2). Finally, the factor Content consisted of the levels True and False referring to the empirical truth of the test statement. There were eight experimental conditions; Exposition TS1 True; Exposition TS1 False; Exposition TS2 True; Exposition TS2 False; Narrative TS1 True; Narrative TS1 False; Narrative TS2 True; Narrative TS2 False. Each child completed eight trials in each experimental condition and the trials were presented in random order.

In addition, a battery of standardised tests was used in order that performance on the experimental tasks could be related to individual differences in reading accuracy and reading comprehension.

c) Materials: 32 short passages were written, each containing two sentences that had to be joined by a causal bridging inference in order to maintain coherence. The content of these bridging inferences was based in simple real world knowledge in order to control for knowledge accessibility. An expository version and a narrative version of each passage were produced. Thus, 64 passages were written in all. The causal bridging inference required was the same in both versions of the passage. The passages were divided into two lists, A and B, such that each list contained one version of each passage. Neither list contained both the narrative and expository version of any one passage and each list contained an equal number of narrative and expository texts.

Thirty-two causal inference target sentences and 32 elaborative inference target sentences were produced. The causal inference sentences stated the mediating idea that needed to be formulated in order to draw the causal bridge whereas the elaborative inference target sentences contained information that could have been elaborated from the passage but was not necessary to maintain coherence. Half of the sentences of each inference type contained information that was consistent with real world knowledge while in half, the information was inconsistent with real world knowledge. Presentation of each passage was followed by the presentation of both inference target sentences in succession, which the child was required to rate as true or false.

Each passage contained six sentences. Within each pair of passages (i.e. each narrative passage and its expository counterpart), sentences occupying the same position contained the same number of words i.e. sentence 2 in the exposition contained the same number of words as sentence 2 in the narrative. The same target inference sentences were used for each version of a passage and the causal target sentence was matched with the elaborative target sentence in terms of number of words. Finally, the two sentences within the passage which required the causal bridging inference were always found in the 4th and 5th sentence positions and were followed by one final 'rounding off' sentence. An example of the type of stimulus used in the experiment can be seen in Table 4.2.

Thirty-two filler passages of the same length as the target passages were produced. These did not require either coherence or elaborative inferences to be made. The target statements following these passages were simple general knowledge statements related to the passage. An example of a filler passage and its corresponding test statements is given below:

“A banana is a popular fruit. It is shaped like a crescent moon. When it is ripe it is yellow in colour. Bananas have soft, creamy flesh. They are sweet and tasty. Bananas are very good for you.”

TS1: Bananas are orange.

TS2: Bananas are a favourite food of monkeys.

Both experimental lists contained the same 32 filler passages. Thus, each list contained 64 passages, 32 target and 32 filler passages. Within the lists, the order of presentation of each passage was randomised. Children assigned to group 1 received Form 1 while children assigned to group 2 received Form 2.

Passage Format		
Sentence	Exposition	Narrative
Sentence 1	Builders build houses for people that are made to last a long time.	Once there was woman who lived with her family in the forest.
Sentence 2	They use bricks to make the walls solid.	She decided to bake her children some cakes.
Sentence 3	Some people used to build houses from wood.	She lived in a pretty little wooden house.
<i>Sentence 4</i>	<i>They used to do all their cooking on big open fires.</i>	<i>To bake the cakes she lit a fire in the kitchen.</i>
<i>Sentence 5</i>	<i>Many of the wooden houses were totally destroyed.</i>	<i>The woman's little wooden house was totally destroyed.</i>
Sentence 6	Bricks are a very useful invention.	The lady was very cross with herself.
Question	Why did people start to build houses with bricks?	How did the woman's house burn down?
Target 1		fire will burn wood.
Target 2		fire keeps you warm.

Table 4.2 Example of expository and narrative test passages, target sentences and comprehension questions (Sentences in italics require a causal bridging inference).

In addition to the test passages, a list of comprehension questions was generated in order to assess the effects of general knowledge on the ability to complete the sentence judgement task. One comprehension question was devised for each test and filler passage. For the filler passages, the question was simply based on the explicit information in the passage. However, the questions derived from the test passages were designed to assess whether the child possessed the general knowledge needed to make the coherence inference. An example of a comprehension question accompanying a test passage is included in Table 4.2. A full list of the passages and corresponding comprehension questions is available in appendix I.

Standardised Tests of Reading and Language Ability.

The *Neale Analysis of Reading Ability* (NARA II; Form 1; Neale, 1989) was administered to assess reading accuracy and reading comprehension. The *Wechsler Intelligence Scale for Children* (WISC III, UK; Wechsler, 1992) vocabulary sub-test was administered as a test of vocabulary knowledge. In this sub-test, the experimenter presents a list of words verbally, one at a time, and asks the child to provide a definition of each word.

d) Procedure: Each child was seen individually over a series of three sessions. Each session lasted approximately 30 minutes. Due to the length of the test materials, the inference processing task was given over the course of the first two sessions. The task was administered using a Psycscope programme run on a Mac Powerbook 180. The first session was constructed as follows. The child saw the word 'READY' in the centre of the computer screen. On seeing this 'READY' prompt the child was required to press the space bar. They were then presented with the first sentence of the passage. To see each subsequent sentence of the passage the child pressed the space bar when they had finished reading the sentence on the screen. They were therefore in control of the pace of their reading. The end of the passage was signalled by the appearance of a small asterix in the centre of the screen after the child pressed the space bar at the end of the final sentence of the passage. Magliano, Baggett, Johnson and Graesser (1993) demonstrated that a delay of 400msec between the offset of a stimulus and the onset of a target (Stimulus Onset Asynchrony; SOA) was sufficient to allow for the generation of a causal coherence inference. Thus, the asterix remained on the screen for 500ms in order to facilitate inference generation. The asterix was immediately followed by the presentation of target sentence 1. The child had to press the button labelled 'T' if they thought the statement was true or 'F' if they thought it was false. On making the judgement, a second asterix would appear in the centre of the screen, again for 500ms, immediately followed by target statement 2. Again the child had to choose either 'T' or 'F' in response to this statement. Following this second sentence judgement, the child would once again see the READY prompt and would then be asked a comprehension question. In this way the child worked through all 64 test passages, 32 in each session, plus 4 practice passages included at the start of session 1 and 2 practice passages at the start of session 2. All text was presented in the centre of the screen in size 18 Comic Sans font.

In session one, the children were seated in front of the computer screen and given the following instructions:

" This task is going to be completed on the computer. You will be asked to read some short passages on the computer screen. Following each passage you will be given two short sentences on the screen. Your task is simply to judge whether these sentences are true or false by pressing a button on the keyboard. On the screen you will see the word 'READY'. This word indicates the start of a new passage. On the keyboard you can see a button with the word 'READY' written on it. When you see the word 'READY' on the screen I want you to press the 'READY' button on the keyboard. The passages will be presented one sentence at a time. When you press the 'READY' button you will see the first sentence of the passage appear on the screen. I want you to read that sentence to yourself. To get the next sentence you need to press the 'READY' button again. Then I want you to read that sentence to yourself. Each time you have finished a sentence and are ready to move on to the next one, I want you to press the 'READY' button. When you have read all the sentences in the passage the computer will put a little star up on the screen. Then you will see another sentence. You have to say whether this sentence is true or false. To do this you have to press the 'T' button if you think it is true, or the 'F' button if you think it is false. You have to do this as quickly as you can. When you have pressed one of those buttons you will see another star on the screen and then another sentence. Again you have to say whether this sentence is true or false by pressing the 'T' or the 'F' button as quickly as you can. I will then ask you a question about the passage you have just read. Then you will see the word 'READY' again and have to start again with the next passage."

The experimenter worked through one example with the child. The child was then given three practice trials with support and feedback from the experimenter. Following the completion of these trials, the first 32 experimental trials began. No further help was given by the experimenter.

In session two, the children were first reminded of the experimental procedure. They were then given three practice trials before completing the remaining 32 experimental trials in the exactly the same way as in session 1. In session three, children were presented

with the NARA II and the WISC III Vocabulary sub-scale according to the instructions laid out in the manuals.

4.5.2 Results.

a) Response times and error rates to test sentences following experimental passages: Initial comparisons of response time data using the between subjects factor of Form indicated that there was no significant difference in response times between groups who received different forms of the test. Thus, for the remaining analyses of test sentence response time, the data were collapsed across groups. Subject's responses were divided by each of the eight experimental conditions (TS1 True Exposition, TS1 False Exposition etc) and outliers were identified as any response time 2.5 standard deviations from the subject mean in each condition. Outliers were replaced with the subject mean for that condition. 3.8% of response times were adjusted in this way. Only response times to items answered correctly were included in the analysis.

The mean response times and number of correct responses for each test sentence was calculated and the resulting data were analysed using a 2x2x2 mixed design ANOVA with Text Format (Narrative/Exposition) as the between subjects factor and Test Sentence (Causal/Elaborative) and Content (True/False) as within subjects factors. The means of the combined data can be seen in Table 4.3. The mean number of correct responses to sentences in each condition is also shown in Table 4.3. Significant main effects of Test Sentence ($F_1(1,42)=26.35, p<.001$; $F_2(1,14)=8.35, p<.05$) and Content ($F_1(1,42)=55.06, p<.001$; $F_2(1,14)=15.77, p<.001$) were revealed. Tukeys HSD revealed that response times to TS1 (Causal) were significantly faster than to TS2 (Elaborative; $F_1 p<.001$; $F_2 <.01$) and that response times to test sentences containing true content were faster than those containing false content ($F_1 p<.001$; $F_2 p<.001$). No other main effects or interactions were significant.

Table 4.3 shows that error rates were low in all conditions. These data were analysed using a 2x2x2 ANOVA (factors as above) on the number of correct responses to sentences in each condition. A significant main effect of Format was revealed ($F_1(1,42)=7.97, p<.01$; $F_2(1,14) = 7.71, p<.05$) such that more correct responses were given to narrative test sentences than to exposition test sentences. This effect was qualified by a

Content by Format interaction by subjects but not by items ($F_1(1,42) = 5.20, p < .05$, $F_2(1,14) = 3.43, ns$). Post hoc Tukeys HSD revealed that responses to False test sentences in the Exposition condition were significantly less accurate than to any other test sentences. However, these results need to be interpreted with caution since participants were performing at ceiling on the task. No other main effects or interactions were significant.

<i>Format</i>		TS1 True	TS1 False	TS2 True	TS2 False
Exposition	Mean RT (SD)	3772.37 (1033.16)	4354.05 (1343.61)	3889.18 (1239.49)	4693.50 (1444.88)
	Mean no correct responses (max = 8)	7.82 (0.39)	7.59 (0.67)	7.91 (0.29)	7.64 (0.95)
Narrative	Mean RT (SD)	3673.99 (906.69)	4287.13 (1386.30)	4064.33 (1178.60)	4747.61 (1346.34)
	Mean no correct responses (max = 8)	8.00 (0)	7.91 (0.29)	7.86 (0.47)	8.00 (0)

Table 4.3 Mean Response Time and number of correct responses for each Test Sentence by Content condition at each Format level. (Standard deviations are given in parentheses. TS1= test sentence 1, TS2= test sentence 2.)

b) Performance on Experimental Comprehension Questions

Following each experimental trial, children were asked a comprehension question to ensure that they had understood the passage and that they possessed the general knowledge needed to make the embedded bridging inference. A correct response was given 1 mark and an incorrect response scored as zero. Performance on the experimental comprehension questions was high confirming that children were in possession of the general knowledge needed to make the bridging inference embedded in the passage. The mean number of correct responses to comprehension questions from a maximum of 16 was 13.68 (SD 1.70) for questions following narrative texts and 14.36 (SD 1.22) for questions following expository text. A t-test showed no significant difference in comprehension scores in terms of the format of the preceding passage ($t(21) = -1.54, ns$). Thus, text format did not appear to affect comprehension of the passage.

c) Analysis of coherence break: While the results of the test sentence analysis suggest that causal coherence inferences are drawn from narrative and expository text, it is not clear whether they are drawn during encoding of the passage or following completion of the passage. Magliano et al. (1993) has shown that a 400msec SOA is sufficient for the generation of a coherence inference. In this experiment, TS1 was always presented 500msec following the completion of the passage, thus allowing enough time for the inference to be drawn between passage completion and presentation of the test sentence.

In order to investigate the time course of coherence inference generation in this study, reading times were analysed for the sentence that represented the coherence break in each passage. It will be recalled that, each experimental passage was created so that, in order to fully understand the text, a causal coherence inference needed to be drawn to bridge a gap between sentences 4 and 5. In practice, this meant that unless linked with the information presented in sentence 4, sentence 5 would seem inconsistent with the rest of the text (a coherence break). If the inference required to resolve this coherence break was drawn on-line, then reading times for sentence 5 in the passages with a coherence break should be longer than for passages with no such break, allowing the reader to refer back to sentence 4 in order to form a full mental representation of the text.

Thus, the reading times for sentence 5 in the experimental passages were compared to the reading times for sentence 5 in the filler passages, constructed such that no specific inference was required to comprehend the text. Sentence 5 contained on average 9.13 words in the experimental passages and 8.16 words in the filler passages ($t(61) = 1.56, ns$). Table 4.4 summarises the mean reading time for sentence 5 in narrative and expository texts for both the experimental and filler passages.

Condition	Format	
	Exposition	Narrative
Experimental	4145.52 (1733.54)	3929.36 (1428.11)
Filler	3676.49 (3676.49)	3771.01 (1235.29)

Table 4.4 Mean coherence break reading times for passages of each format in each condition (standard deviations in parentheses.)

Initial analyses revealed no main effect of Form so the data was collapsed across Lists for the remaining analyses. A 2 x 2 mixed analysis of variance was conducted with the between subject factor Format (exposition and narrative). The within subject factor was Passage Type with two levels (experimental and filler passages). A by-items analysis was not conducted on these data as the items presented to individual participants were not comparable. For example, the experimental items presented to each participant varied depending on the list received while all participants received the same filler items in the same format. There was a significant main effect of Passage Type ($F(1,21) = 13.26, p < .01$) indicating that sentence 5 took longer to read in experimental passages than in filler passages. No other main effects were significant. However, the interaction between Format and Passage Type approached significance ($F(1,21) = 4.00, p = .059$). Post hoc Tukeys revealed that sentence 5 reading times for expository test passages were significantly slower than for either expository or narrative filler passages. Sentence 5 reading times for narrative test passages were also slower than for either filler passage type but this difference did not reach significance. There was no significant difference in sentence 5 reading times between expository test and narrative test passages, nor between expository filler and narrative filler passages.

d) Individual differences in children's on-line text processing: To assess individual differences in reading accuracy, reading comprehension and vocabulary skill, the children completed a number of standardised measures in addition to the experimental task. Mean standard scores for each group are shown in Table 4.5. T-tests showed there was no significant difference between groups on any of the measures (NARA II accuracy $t(20) = -0.34, ns$; NARA II comprehension $t(20) = -0.75, ns$; WISC Vocabulary $t(20) = 0.62, ns$). Therefore, when examining individual differences, the data were collapsed across groups.

Measure		Group 1	Group 2
NARA II Accuracy ^a	Mean	101.09 (9.77)	102.55 (10.58)
	Range	89-119	89-121
NARA II Comprehension ^a	Mean	92.27 (7.32)	94.55 (6.80)
	Range	87-100	86-105
WISC Vocabulary ^b	Mean	6.64 (2.91)	5.91 (2.55)
	Range	3-13	2-9

Table 4.5 Mean scores on standardised measures of reading, vocabulary and comprehension (a = average 100, SD 15; b = average 10, SD 3).

Table 4.6 shows the zero order correlations between response times to TS1 (Causal) and TS2 (Elaborative) and performance on each of the assessment measures. The correlations suggest a relationship between target sentence response times and NARA II reading comprehension. However, reading accuracy is also significantly correlated with target sentence response times. As discussed in chapter 3, reading comprehension is confounded by reading accuracy, because in order to comprehend a text you must first be able to decode it. Therefore, it is not clear how much of the relationship between reading comprehension and the target sentence response times is mediated by reading accuracy. A partial correlation was therefore run to control for the influence of reading accuracy. Importantly, the correlation between response times to target sentences and reading comprehension remained significant when reading accuracy was controlled.

	Chron Age	NARA II Accuracy	NARA II Comp	WISC Vocab	TS1 RT	TS2 RT
Chronological Age		0.08	0.44*	0.41	-0.33	-0.40
NARA II Accuracy			0.59**	0.48*	-0.43*	-0.48*
NARA II Comprehension	0.49*			0.43*	-0.59**	-0.60**
WISC Vocab	0.43		0.20		0.59*	-0.66***
TS1 RT	-0.33		-0.46*	-0.48*		0.96***
TS2 RT	-0.40		-0.45*	-0.56**	0.95***	

Table 4.6 Zero Order correlations between target sentence response times and performance on standardised measures (Chron Age = chronological age) and partial correlations (below diagonal) between these variables when reading accuracy is controlled for . * $p < .05$, ** $p < .01$, *** $p < .001$

Multiple regression analyses were conducted to investigate whether response times to the target sentences could predict performance on the NARA II comprehension measure. As previously discussed, it has long been accepted the ability to make inferences is a fundamental component of comprehension skills (e.g. Kintsch & van Dijk, 1978; Paris & Lindauer, 1976; Zabrocky, 1986). It was anticipated therefore that response times to both target sentences would reliably predict NARA II comprehension raw scores. NARA II accuracy was not included in the regression since the results of the partial correlation indicated that the relationship between comprehension and response times was not mediated by decoding skills.

The results of the regression analyses shown in Table 4.7 indicate that age is a significant predictor of comprehension ability when entered on the first step (43%). WISC Vocabulary accounts for 7% of the variance in comprehension scores which is marginally

significant. Response times to TS1 account for 12% of the variance when entered on step 3, and response times to TS2 account for 12% of the variance when entered on step 3, both of which are significant. These results indicate that even after age and vocabulary knowledge are taken into account, response times to inference target statements are reliable predictors of comprehension ability.

Step	Measure	R ² change	p
1	Chron Age	0.43	0.00
2	WISC Voc	0.08	0.08
3	RT TS1	0.12	0.02
3	RT TS2	0.12	0.02

Table 4.7 Multiple regression with NARA II comprehension as the dependent variable and response times to target sentences, chronological age and WISC vocabulary as the predictor variables.

4.5.3 Discussion.

Experiment 2 examined on-line inferences during children's reading of narrative and expository text using a sentence judgement task. Test sentence 1 stated a mediating idea which would have had to have been activated if the coherence inference embedded in the text were to be drawn, while test sentence 2 simply stated a possible elaboration of the text. Analysis of the response time data indicated that responses to test sentence 1 were faster than those test sentence 2. It appears therefore that the causal coherence inference that could be generated from the preceding text was drawn at a faster rate than the elaborative inference suggesting that the mediating idea referred to by the causal coherence test sentence was more readily available than that referred to by the elaborative inference test sentence. As outlined in Chapter 1 of this thesis, and in the introduction to this chapter, previous research has shown that causal coherence inferences are drawn on-line during reading while elaborative inferences are drawn off-line and only when they are specifically required by the reader. Based on this previous research, one plausible explanation for the results of the current study is that the causal coherence inferences were drawn during the reading of the passage and were therefore available in working memory when the test sentence was presented, priming the participant's response to that sentence. Conversely, the

elaborative inferences were not generated during reading and therefore the subsequent test sentences were judged without priming from the preceding text explaining the longer response times for sentences of this type. In this way, the current findings can be interpreted as adding further support to previous studies showing that causal coherence inferences are drawn from text more reliably than elaborative inferences.

The finding that TS1 response times were faster than TS2 response times is further supported by the effect of content on sentence judgement times. Response times to false test sentences were longer overall than response times to true test sentences indicating an increase in the processing time taken to judge a false test statement. This is not unexpected since a "false" sentence is incongruent with real world knowledge causing confusion and therefore a delay in responding. This finding is related to the "belief bias effect" commonly found in research on deductive reasoning whereby people find it difficult to set aside their real world knowledge in order to draw a valid deductive conclusion based on false premises. Thus the ability to deal with false information is consistently found to be problematic (e.g. Evans & Perry, 1995; Franks, 1996; Moshman & Franks, 1986). However, the fact that responses to test sentence 2 were particularly inhibited by false content indicates that some extra processing was involved in responding to these statements.

The results indicated no effect of text format. It appears that the pattern of response times to test sentences is consistent across narrative and expository text. Thus, in line with Singer et al. (1997), these results show that coherence inferences are reliably drawn in both narrative and expository text whereas elaborative inferences are not.

While these results suggest that the causal coherence inferences embedded in the passages were generated prior to the elaborative inferences, a limitation of the present design was that test sentence 1 always corresponded to the coherence inference, while test sentence 2 always corresponded to the elaborative inference. The smaller response times to TS1 over TS2 could therefore be due to an order effect. Bahri and Hussain (1997) investigated order effects in inference generation and found that the order in which inferences were presented did not effect inference processing. They presented participants with stories embedded with target statements representing either a goal inference (a form of

coherence inference) or a state inference (a form of elaborative inference). The order in which the target statements were presented was manipulated such that the goal inferences appeared first in some stories and second in others. The results demonstrated that irrespective of the order in which they were presented, reading times for the state inference target statements were consistently longer than for the goal inference statements suggesting that goal inferences were generated faster than state inferences. Thus, while an order effect may have influenced response times in the present study, there is some evidence from previous research to suggest that this is not necessarily the case. Moreover, data was collected, although not reported in the present study, which demonstrate that there was no difference in difficulty between test sentence 1 and test sentence 2. Participants were presented with the test sentences in a random order with no preceding passage context, and were asked to respond "true" or "false" by pressing a button on the keyboard. When presented with the test sentences with no preceding context, the response time difference between the two test sentences disappears. These data can be found in appendix 2. Taken together, these results suggest that, while the increased response time to TS2 might be due to an order effect, the interpretation that it is a reflection of the differences in the time-course of inference generation between coherence and elaborative inferences cannot be ruled out.

A second limitation of the present design is that the sentences constructed in the present study to test elaborative inferences were not constrained in any way. They simply stated one of any number of alternative inferences that could have been drawn from the passage. In addition to this, during the time lapse between the offset of the passage and presentation of TS2, the children were actively responding to TS1, which may have interfered with drawing the subsequent elaborative inference.

However, the results of the coherence break analysis support the assumption that coherence inferences were drawn on-line in the test passages. Narrative and Expository experimental passages invoked longer sentence 5 reading times than either of the filler passages despite the fact the passages were matched for the number of words in each sentence, although this was only significant in the case of Expository texts. Thus it can be argued that reading sentence 5 in the experimental test passages involved more processing than reading the same sentence in the filler passages. As such it seems plausible to suggest

that in the present study, coherence inferences were drawn on-line in both formats. Finally, there was no difference in sentence 5 reading times between narrative and expository text in the experimental condition. Nor was there a difference in reading time between these two text types in the filler condition. Thus, text format did not appear to influence inference generation when passages were compared within each condition.

One further criticism that may be levelled at this study is that the passages did not actually differ clearly in structure because both passage formats required a causal link. Thus, all the passages could be considered to have a narrative structure. If this were the case then the finding that there was no effect of passage format on inference generation is unsurprising. However, a lack of causal constraints is only one of the ways in which expository texts differ from narrative texts. Furthermore, it is not the case that expository texts never contain causal links. If that were true then an investigation of this nature would not be necessary. However, while the segments in narrative texts *must* be causally connected, this causal constraint is *unnecessary* in expositions, although it is entirely possible. In this case, in order to investigate the effect of text format on inference generation it was necessary for the expository texts to contain a causal link so that a comparison could be made with narrative text passages. However, the format of the passages used in experiment 3 differed in many of the aspects discussed by Olson et al. (1981). For example, the narrative texts related a story, contained characters, and were dynamic and engaging. Conversely, the expositions contained no specific characters, were passive, and were written as informative commentaries.

The multiple regression analysis indicated that inferential skill was a reliable predictor of comprehension ability. This is in line with previous research indicating that comprehension is at least partially reliant on inference generation (e.g. Kintsch & Van Dijk, 1978; Paris & Lindauer, 1976; Zabrocky, 1986). However, it is important to note that no clear causal relationships could be deduced from the statistical analyses used here. It is possible that both processes could be mediated by a third variable, although vocabulary and age were both discounted in the present study.

Finally, although text format did not affect sentence judgement response times, the content of the target sentence did prove to be an influential factor. Target sentences

containing information that was false in terms of passage content and real world knowledge were responded to slower than target sentences containing a statement of fact. Since many narratives read by children are based in fantasy or fiction, many of the inferences needed to form a representation of the text would also be false in terms of real world knowledge although true in relation to the passage. If children have difficulty dealing with empirically false information, as suggested here, then it is plausible that passage content may influence inference generation. An investigation into the effect of content on inference generation would further add to current knowledge of inference processing.

Chapter 5

On-Line Inference Generation During Children's Reading Of Fairy Tales.

5.1 Introduction

As discussed in chapter 4, Singer, Harkness and Stewart (1997) investigated whether the pattern of inference generation differed during the reading of narrative and expository texts. Contrary to previous studies (Millis & Graesser, 1994; Millis, Morgan, & Graesser, 1990; Noordman, Vonk, & Kempff, 1992) they found that the format in which the text was presented did not influence the generation of inferences. These findings were replicated in experiment 2 of this thesis. Given the inconsistency of these findings, there is clearly a need to identify factors that reliably affect the generation of inferences. One area that requires further investigation is textual content. In the majority of studies, the materials used are short stories, the content of which is generally based in real world knowledge. The inferences to be drawn from the passages are therefore also in agreement with real world knowledge and should pose no problem to the reader provided they are in possession of the general knowledge required. However, some narratives contain fictional information and are based on facts that are true only in the context of the story. In cases such as these, the inferences required for a comprehensive understanding of the passage will also be based on information that may contradict the reader's own knowledge base. In other words, readers are required on occasion to set aside their own knowledge and accept empirically false premises as true in order to form a coherent representation of a text. The content of a passage may therefore inhibit a reader's ability to generate inferences if they are unable to ignore their own understanding of the world in order to fully comprehend the text at hand.

5.2 Reasoning From False Premises: Evidence From Deductive Reasoning

Drawing inferences from false premises has largely been investigated in the context of deductive reasoning. Fundamental to this research is the distinction between **validity** and **soundness**. Validity refers to the form or structure of an argument while soundness refers to its content. In deductive reasoning, an argument is said to be **valid** if the conclusion of the argument necessarily follows from its premises. For example, argument A in Table 5.1 is a valid argument because the only possible logical conclusion based on the structure of the argument is that a sparrow has wings.

Furthermore, argument A is also **sound** because the information contained in the premises is empirically true, that is true according to our experience of the world. However, an argument can contain empirically false information, rendering it **unsound**, while still retaining its validity. For example, argument B is clearly **unsound** as the premises are empirically false. However, the conclusion necessarily follows from the premises and thus, argument B is **valid**. Therefore, when drawing deductive inferences, it is the structure of the argument, and not the content that is the key to a successful conclusion.

	Valid Inferences.	
	A	B
<i>Soundness</i>	<i>Sound</i>	<i>Unsound</i>
Premise 1	All birds have wings.	All fish have feathers.
Premise 2	A sparrow is a bird.	A trout is a fish.
Conclusion	A sparrow has wings.	A trout has feathers.

Table 5.1 Examples of valid arguments with sound and unsound premises.

Research into reasoning from false premises typically involves the presentation of a number of premises that have been manipulated in terms of their validity and soundness. The participant's task is to provide a conclusion to the argument. In the case of valid but **unsound** arguments, the question of interest is whether participants will be able to set aside their real world knowledge, ignoring the soundness of the argument, and focus instead on its structure, or validity. Research has shown that children experience difficulty with this kind of reasoning (Franks, 1996; Moshman & Franks, 1986). For example, Franks (1996) composed a number of narratives, the content of which were either empirically false, empirically true or neutral. The narratives contained embedded premise information and a deductive inference question was presented immediately following the presentation of the premises. For example, the following passage is part of a story used by Franks (1996). The story contains false information and the sentences in italics represent the premises of the argument:

At the end of our tour, we were taken to meet the beings of planet Glinko. The robot took us to a tall tower that was made of something like glass, so we could see the inside of it. When we got closer, we saw what looked like some people and some dogs inside. From the inside, we heard the sound of many voices talking at once. *I said, "Well, either the people here talk or their dogs do." The robot said, "All the people of Glinko can read each other's minds, so they never speak aloud."*

Having read the passage, participants were presented with the following forced choice question:

- Q. Who was talking, the people or the dogs?
- a) The people.
 - b) The dogs.

In order to select the correct response from the two alternatives, the participant would have had to have read and understood the premise information. Furthermore, they would have had to discard their own belief that dogs can't talk and simply followed the logic of the argument. The character in the story heard talking coming from a room in which there were people and dogs. Since the people on Glinko did not speak aloud, the only logical conclusion is that the dogs were doing the talking. While this conclusion does not fit in with our empirical knowledge of the world, it is consistent with the argument structure.

The passages were given to children of 9-10 years and 12-13 years, and to adults from 19-30 years of age. It was predicted that the younger children would find it more difficult to reason from false premises than either the older children or the adults in spite of the additional contextual information given by the surrounding narratives. Since the inference forms used in the narratives were well within even the youngest children's capabilities, this difference in performance could be attributed solely to the false content of the arguments. Consistent with this, a developmental trend was observed in the ability to draw deductive inferences from all three content conditions and this trend was strongest between the ages of 9 and 12 years. Moreover, while false premises proved more problematic than true or neutral premises for all age groups, the

effect was most notable for children of the youngest age group. Similar findings were reported in a later study (Franks, 1997).

Franks (1996, 1997) interprets her findings in terms of Moshman's developmental theory of logic (Moshman, 1990), which distinguishes between basic logical reasoning and metalogic. Metalogic can be described as the ability to recognise the importance of the relationship between the premises and conclusion of an argument over and above the empirical truth of the premises. Moshman proposes that logical reasoning develops in four stages; Explicit Content-Implicit Inference, Explicit Inference-Implicit Logic, Explicit Logic-Implicit Metalogic, and Explicit Metalogic.

1. *Explicit Content- Implicit Inference:* At this stage the child can accurately draw inferences from premise information but is not explicitly aware that they are doing so. The child focuses on the content of an argument and not its form. Stage 1 children are typically of pre-school age.

2. *Explicit Inference-Implicit Logic:* Between the ages of 6 and 10, children reach the second stage of development. At this stage children have some understanding of logical necessity and can distinguish between logically necessary and logically probable arguments. Thus, metalogic has begun to develop. However, children at this stage still have no explicit awareness of logical form and are unable to distinguish between validity and soundness.

3. *Explicit Logic-Implicit Metalogic:* An explicit awareness of the importance of logical form begins to develop in children of approximately 11 years of age. At this stage the child will be able to form a conclusion that necessarily follows from the premises, rather than relying on the empirical truth of the premise as a guide.

4. *Explicit Metalogic:* At this stage the adult thinker is able to move away from solving individual problems and begin to think about logic explicitly as a formal system, relating it to language and other formal systems. Logicians and students of logic are likely to reach this level. However, this final stage in the development of logical reasoning is one that Moshman suggests is never reached by most people.

Moshman's theory (Moshman, 1990) would predict that children are unable to reason from false premises until they reach the third stage of development as it is not until this stage that they would be able to distinguish between the form of the argument

and its content. Thus, children in a preceding stage would be unable to ignore the content of the argument and concentrate on its validity. Franks (1996, 1997) findings are clearly in line with this prediction. Furthermore, while not directly based on Moshman's developmental theory, investigations into the belief bias effect in deductive reasoning yield similar findings. The belief bias effect is a common phenomenon in studies of deductive reasoning and refers to the inability of participants' to inhibit their own beliefs in order to draw a conclusion based on logical validity. Studies have shown that children are vulnerable to the belief bias effect. Evans and Perry (1995) conducted a preliminary rating study in which children judged a list of statements as believable (i.e. cars carry people), neutral (i.e. donkeys have 14 teeth) and unbelievable (i.e. cows live in zoos)¹. In a subsequent reasoning experiment, the previously rated statements were presented as conclusions to a series of problems. Thus, children were presented with 12 problems, 4 of which had a believable conclusion, 4 with a neutral conclusion and 4 with an unbelievable conclusion. Each problem was followed by the question "Does it follow that (conclusion)?" with one of the previously judged statements inserted as a conclusion. Furthermore, 2 of each problem type (believable, neutral and unbelievable) were logically valid while 2 were logically invalid. The problems were read aloud by the experimenter but the children were also provided with a written version of each problem that they could re-read if necessary. Having heard the problem and the question the children were asked to tick either 'yes' or 'no' in a response space on a test form.

The results of the study provide evidence of premise-based reasoning suggesting that children could separate the structure of the problem from its content. However, there was also a strong belief bias effect such that more believable than unbelievable conclusions were accepted. Furthermore, there were developmental trends indicating a suppression of the belief bias effect and an increase in premise-based reasoning with increasing age. Evans and Perry suggest that while the children were clearly capable of reasoning on the basis of validity they could not completely suppress the tendency to form judgements based on prior knowledge.

In contrast to the findings of Franks (1996, 1997) and Evans and Perry (1995), other investigations have found that children can overcome the belief bias and reason

¹ Examples taken from Evans & Perry (1995) p106.

from false premises when given some supportive context. For example, Dias and Harris (1988) presented syllogistic reasoning problems to children of 4-, 5-, and 6-years of age. The content of the problems was either known (All cats miaow; Rex is a cat; Does Rex miaow?) unknown (Pogs wear blue boots; Tom is a pog; Does Tom wear blue boots?) or contrary to empirical knowledge (All cats bark, Rex is a cat; Does Rex bark?). The problems were presented either verbally, or within the context of play; that is with the use of toys and props.

The results of the study showed that children's performance on syllogistic reasoning problems was more accurate when the problems were presented in the context of play or when presented as a story, and that this effect was strongest when the content of the problems was unknown or contrary to the children's empirical knowledge. The children were asked to justify their responses and their justifications were classified into 3 categories; those that were empirical or based on prior beliefs, those that were theoretical or based on information contained in the premises, and those that were arbitrary. The results suggested that the children in the play group gave many more theoretical justifications than the children in the verbal group, particularly for the unknown and contrary problems. That is, the children in the play group focused more on the information contained in the premises than on the truth value of those premises in terms of their own beliefs.

In a later study, Dias and Harris (1990) found that encouraging children between 4 and 6 years of age to use imagery when faced with contrary to fact reasoning problems also facilitated reasoning from false premises. Richards and Sanderson (1999) found that children as young as 2-,3-and 4-years old were able to solve counterfactual syllogistic reasoning problems when encouraged to use their imagination and were more likely to justify logical responses in terms of the content of the premises rather than on the basis of their empirical knowledge.

Leevers and Harris (2000) performed a similar study with normally developing 4-year olds, children with MLD and autistic children. Like Richards and Sanderson (1999), Leevers and Harris found that normally developing children could produce logical conclusions to counterfactual syllogisms and justify their responses by referring to the premises. While performance of children in the MLD group was not comparable

to that of the normally developing children it was above chance levels indicating that children with moderate learning difficulties do have the capacity for logical reasoning from false premises. However, the performance of the children in the autistic group was at chance.

Further evidence that children can set aside prior knowledge to deal with information that is inconsistent with that knowledge comes from Power, Taylor and Nippold (2001). They presented children with proverbs that were either literally true (i.e. the baby has no teeth) or literally false (i.e. every dog is a lion at home). Each proverb was presented in a short story, which was followed by 4 possible proverb meanings, one of which was the intended meaning of the proverb and three of which were closely related foils. The literal meaning of the proverb was never given as an alternative. The children had to pick the most appropriate intended meaning for the proverb. In order to do this they had to look beyond the literal meaning of the proverb and concentrate on the context in which it was used to find the figurative meaning. Power et al. (2001) predicted that this task would be easier in the case of literally true proverbs since they were 'easier to visualize and required less imagination on the part of the reader' (p5). However, the results suggested that there was no difference in performance on the proverb comprehension task that could be attributed to the type of proverb presented. The only significant finding was a developmental trend such that 14- and 11-year olds performed better than 9-year old children, although the two older groups did not differ from each other. These findings suggest that children can ignore their empirical knowledge and find meaning beyond a literal interpretation, even when the literal interpretation is itself contrary to their own beliefs.

5.3 Generating Text-Based Inferences From False Premises.

While investigations into children's deductive reasoning skills have given some consideration to the difficulties posed by false premise information, research into children's generation of text-based inferences has not directly addressed this issue. However, children are rarely presented with logical deductive reasoning problems in isolation from any surrounding context. A more likely scenario is that children are regularly faced with reasoning tasks during reading such that they are required to make inferences in order to, for example, keep track of a sequence of events or explain the actions of a character. These inferences may not be deductive in nature. That is to say,

they may not be logically necessary and the conclusions may not be unequivocal. However, a logically possible conclusion will be drawn which is the most likely conclusion in the context of the story. Furthermore, many of the stories children are presented with contain information that is not based in reality, as is the case with fairy tales. In cases such as this, it is necessary for children to draw inferences based on information that runs counter to their empirical knowledge of the world. The conclusions themselves may not be empirically false but the premises on which they are based, and the overall context in which the inference must be drawn, may not fit with the child's knowledge of the world. Thus, in contrast to the generation of inferences from stories based solely in real world knowledge, in fairy tales the child must overcome a large degree of counterintuitive information in order to generate inferences. This is not a "belief bias" effect as investigated in the studies cited earlier in this chapter, such that the conclusions drawn may not conflict with real world knowledge. However, the generation of such inferences does require the child to suspend disbelief in order to form a coherent representation of a text. It is this interpretation of the belief bias effect which will be investigated in the present study.

Galotti, Komatsu and Voelz (1997) compared the performance of children from kindergarten and grades 2, 4, and 6 on deductive and inductive syllogisms. While deductive inferences are those for which a conclusion is logically necessary, inductive inferences are those for which a conclusion is logically possible. For example, the syllogism 'All fish like to swim. Tommy likes to swim. Is Tommy a fish?' requires an inductive inference because, while it is possible that Tommy is a fish, it is equally likely that Tommy is a person. The conclusion that Tommy is a fish does not necessarily follow from the premises. All the syllogisms presented by Galotti et al. (1997) contained fantasy content and children were asked to answer yes or no to the inference question and also to rate how confident they were in their answer. An example of a deductive syllogism and its inductive counterpart used in the study are presented below.

Deductive: All poggops wear blue boots. Tombor is a poggop. Does Tombor wear blue boots?

Inductive: Tombor wears blue boots. Tombor is a poggop. Do all poggops wear blue boots?

Galotti et al. (1997) found that children from grades 2,4 and 6 were more consistent in their responses to deductive than to inductive inferences, and responded faster to and were more confident in their responses to deductive items. The experiment was run a second time with children from grades 2,3,4 and 5 and the children were also asked to justify their responses. All children showed a tendency to refer to the premises for justification of a response to a deductive inference and to a sense of uncertainty about the premise information for justification of a response to an inductive inference. The results indicated that children were able to set aside prior knowledge in order to answer deductive inference problems based in fantasy content, but were less capable of doing so in the case of inductive inferences. As stated above, in narrative texts, many inferences are inductive in nature such that they do not involve the generation of logically necessary conclusions. While the inferences presented by Galotti et al. (1997) were not presented as narratives, the results suggest that children may have difficulty generating inductive inferences from fantasy content in a narrative text.

Aims of present study : The present study was designed to compare the ability of children to generate coherence inferences and elaborative inferences from stories containing fantasy content with those from stories based in real world knowledge. In contrast to previous studies investigating the "belief bias" effect, the inferences included in this study are coherence links, not logical deductions, in order to more closely mirror the kinds of inferences children might be faced with in their everyday reading experience. In line with the findings suggested by Galotti et al. (1997) it was predicted that children would find it easier to draw both coherence and elaborative inferences from stories containing information that was consistent with real world knowledge. However, it was also expected that the pattern of inference generation would be consistent across story types such that coherence inferences would be drawn on-line during encoding while elaborative inferences would be drawn at a later point during the reading process.

5.4 Experiment 3

5.4.1 Method

a) Participants: 39 children were selected from the Year 5 classes of two York primary schools. All the children in the Year 5 classes were given the Suffolk Reading Scale (SRS; Level 2, Form A; Hagley, 1986) to measure basic reading ability and the

children selected to participate in this study achieved a standard score on the SRS of between 95 and 115. The overall mean standard score of the sample was 104.18 (SD 7.08) and the mean age of the sample was 121.41 months (SD 4.58). The children were randomly split into two groups. Group 1 ($n = 19$) had a mean standard SRS score of 103.58 (SD 6.62) and a mean age of 121.36 months (SD 4.58). Group 2 ($n = 20$) achieved a mean standard SRS score of 104.75 (SD 7.61) and a mean age of 121.47 months (SD 3.86). The groups did not differ significantly at ANOVA on standard SRS score ($F(1,37) < 1$) or chronological age ($F(1,37) < 1$).

b) Design: This study used a $2 \times 2 \times 2$ mixed design with one between- and two within-subjects factors. Each of the three factors contained two levels. The between-subjects factor was Form and had the levels list 1 and list 2. This factor indicated which of the two versions of the experimental list the children received; children in Group 1 received list 1 and children in Group 2 received list 2. The within subjects factors were Passage Type, which had the levels Real and Fairy referring to the type of content contained in the passage, and Inference Type, which had the levels Causal Link and Static Property referring to the type of inference being tapped by each test sentence. Each child read every passage in his or her list and responded to two sentence judgement trials within each passage. Thus each child completed eight trials in each of four conditions; Real Causal Link, Real Static Property, Fairy Causal Link and Fairy Static Property. In addition to this, each child was administered a number of standardised tests to assess individual differences in reading and comprehension skills.

c) Materials: 16 test passages were created, 8 of which were based on information that was consistent with real world knowledge (Real), and 8 containing information that was inconsistent with real world knowledge (Fairy). An example of each of these passages is given in Table 5.2 and the complete set is available in Appendix III. Passages were divided into two sections and a test sentence was presented after each section. 32 test sentences were created to represent either a causal inference required to link the final two sentences (critical sentences) of each passage or a static property of a character or object that appeared in these critical sentences. Each passage contained both a causal and a static test sentence and the content of the test sentences reflected the content of the passage in which they were presented. Within each passage, the critical sentences in section 1 were matched for syllable length with

the critical sentences in section 2. The test sentence following section 1 was matched for number of syllables with the test sentence following section 2. Matching for number of syllables contained in critical and test sentences across Real and Fairy passages would have constrained the flow of the text making the passages appear stilted. Thus, the number of syllables in the critical and test sentences differed across passage type. However, t-tests showed these differences were not significant (critical sentences $t(14) = -0.37$, ns; test sentences $t(14) = -1.56$, ns).

Two versions of each passage were written. In version one, a causal link test sentence would follow section one and a static property test sentence would follow section two. In version 2 the order would be reversed. In order to create two alternate versions but maintain the flow of the passage, the wording of the different versions varied slightly and different test sentences were used. However, the passages were matched so that the critical sentences and the test sentences contained the same number of syllables in both versions. Table 5.3 shows the mean number of syllables in the critical sentences and test sentences in both versions of each passage type. The experimental passages were also assessed for readability using the book grading formula proposed by Hatcher (2000). The mean reading grades and corresponding WORD reading age equivalents (Wechsler Objective Reading Dimensions; Wechsler, 1990) are shown in Table 5.3. A 2x2 ANOVA with the between groups factor of Passage Type and the within groups factor of Form found no significant difference between the book grading of each passage type. The passages were organised into two lists each containing 16 passages.

	Real	Fairy
Section A	John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.	There once was a king who had eleven sons and one daughter. The king married a wicked queen who was jealous of the children. She sent the little girl away and made the king banish his sons from the kingdom. On the day the boys left the palace, the queen cast a spell over them. Instead of eleven boys, eleven wild swans flew from the palace gates.
Test Sentence 1	The bear had huge paws. (static)	The queen turned the boys into wild swans. (causal)
Section B	Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear fell to the ground and did not move.	The swan princes searched for their sister. When they found her she promised to break the spell. She made eleven magic shirts for her brothers. When she threw the shirts over the swans, they were transformed into eleven handsome princes. But the girl did not have enough magic cloth to finish all the shirts. The youngest prince would forever have a swan's wing in the place of an arm.
Test Sentence 2	John shot the bear dead. (causal)	His arm was covered in white feathers. (static)
Comprehension Q	Where did the bullet hit the bear?	How many sons did the king have?

Table 5.2 Examples of Real and Fairy story passage types, test sentences and comprehension questions.

Each list contained only one version of each passage and the lists were arranged so that eight passages presented causal link test sentences first and eight presented static property sentences first.

8 filler passages were written in the same format as the experimental passages. Four of these passages were based on real world knowledge (Real) and four were entirely fictional (Fairy). However, the test sentences presented in these passages required memory for literal information rather than inferential processing. Each experimental list contained four filler passages; two real world knowledge passages and two fictional passages. Due to the nature of the study, the correct response to an experimental test sentence was always TRUE. This meant that the pattern of responding to each experimental item was always TRUE/TRUE. Therefore, the filler passages were designed such that four of the passages contained a TRUE/FALSE pattern of correct responses and four contained a FALSE/TRUE pattern of responses. Within each experimental list, one knowledge-based and one fictional filler passage would contain the TRUE/FALSE pattern and the other two passages would contain the FALSE/TRUE pattern.

Finally, in order to assess whether the passages had been understood, 24 comprehension questions were written, one for each of the experimental and filler passages. The same question was used for each of the versions of the experimental passages. Examples of the types of comprehension questions asked are given in Table 5.2. A full list of the passages used in the experiment and the corresponding comprehension questions is given in Appendix III.

Mean No of Syllables	List 1		List 2	
	Real	Fairy	Real	Fairy
Section A				
Mean no syllables in critical sentences (SD)	24.5 (3.30)	25.25 (4.71)	24.5 (3.30)	25.25 (4.71)
Mean no syllables in test sentence (SD)	7.25 (1.50)	8.25 (1.04)	7.25 (1.50)	8.25 (1.04)
Section B				
Mean no syllables in critical sentences. (SD)	24.5 (3.30)	25.25 (4.71)	24.5 (3.30)	25.25 (4.71)
Mean no syllables in test sentence (SD)	7.25 (1.50)	8.25 (1.04)	7.25 (1.50)	8.25 (1.04)
Book Grade				
Mean Book Grade (SD)	17.25 (1.41)	17.94 (0.67)	16.85 (1.56)	18.04 (0.89)
Book grade range	15.08 - 19.63	16.91 - 18.87	14.15 - 18.64	16.5 - 18.93
Reading age range (WORD equivalent)	6:7 - 7:4	7:0 - 7:3	6:7 - 7:3	7:0 - 7:3

Table 5.3 Mean no of syllables and book grades for experimental passages (standard deviations in parentheses; WORD = Wechsler Objective Reading Dimensions).

Standardised Tests of Reading and Language Ability.

The *Neale Analysis of Reading Ability* (NARA II; Form 1; Neale, 1989) was administered as a test of reading accuracy and comprehension. Two sub-tests from the *Wechsler Intelligence Scale for Children* (Third Edition UK; WISC III; Wechsler, 1992) were administered; the Vocabulary sub-test as described in chapter 4 and the Similarities sub test. The Similarities sub-test requires the experimenter to say two words i.e. banana and apple, and the participant to respond by explaining how those words are related i.e. both are types of fruit. The *Test of Word Reading Efficiency* (TOWRE; Torgeson, Wagner, & Rashotte, 1999) was given to assess reading speed and accuracy. The TOWRE is a timed test in which the child is asked to read as many items as they can from a list of words within 45 seconds. The task is then repeated using a list of non-words. Three scores are obtained; a sight word reading efficiency score, a phonemic decoding score and an overall reading efficiency score.

d) Procedure: Children were tested individually over a course of two 30-minute sessions. In the first session, children completed 14 passages (2 practice, 8 test and 4 filler passages). They also completed the WISC Vocabulary and Similarities measures and the TOWRE. In the second session, children completed the remaining 12 passages (8 test and 4 filler passages) and the NARA II. The WISC sub-tests, TOWRE and NARA II were all administered in accordance with the instructions laid down in the test manuals.

The experimental paradigm ran as follows. The word READY appeared in the centre of the computer screen. The child pressed the spacebar and the 'READY' prompt would be replaced by the first half of a passage that the child read aloud to the experimenter. The child pressed the spacebar again when they had read and understood the passage. Then an asterix appeared in the centre of the screen for 500 msecs after which the first test sentence appeared. The child made a true/false judgement about the sentence by pressing the key marked True or False on the keyboard, following which, the second half of the passage automatically appeared on the screen. The child read the second half of the passage, pressing the space bar again when they had finished. Again, they saw an asterix for 500msec followed by the second test sentence which they had to judge as true or false by pressing the appropriate keyboard button. Following this the READY prompt appeared

once more. The comprehension question was then given orally signalling the end of the trial. The second trial would begin when the child pressed the spacebar. All text was presented in the centre of the screen in size 18 Comic Sans font.

At the beginning of session one, the children were seated in front of the computer and told to make themselves comfortable. They were told that they were going to read some passages and answer some questions on the computer. They were then given the following instructions:

“In a moment you will see the word READY appear on the screen. When you see that word, you must press the space bar. After you press the space bar, the first half of a story will appear in the centre of the screen. I want you to read that story out loud to me so that I can help you with any words you might get stuck on. When you have finished reading the passage on the screen, you should press the space bar again. Then a sentence will appear in the middle of the screen, and you have to decide whether the sentence is true or false. You do this by pressing the true or false buttons on the keyboard as quickly as you can. Once you have done this, the second half of the passage will appear and again I want you to read it out loud for me. When you are finished you press the space bar and look at the sentence on the screen. You then have to decide whether the sentence is true or false by pressing the true or false keys as before. Is that clear?”

Having established that the child had understood the instructions and was clear about the buttons on the keyboard the experimenter then prompted the child to complete the two practice passages. When the child had successfully completed the practice passages the experimenter then informed the child that after each passage they would be asked a question to make sure they had understood the passage. Having completed the practice and instruction phases, the child completed the 12 passages in the experimental phase.

In session two, the experimenter first ensured that the child remembered the instructions from session 1 and if not the practice trials were repeated. The child then completed the 12 remaining experimental trials.

5.4.3 Results

a) Reading times for critical sentences: A comparison of the reading times for the critical sentences in Fairy and Real stories was conducted in order to examine in more detail whether the difference in the number of syllables between Fairy and Real stories would affect reading time for those sentences. During session 1 of the experimental task, 28 of the children were recorded when reading the passages. These recordings were then digitised and the critical sentences in each of the eight experimental passages read during session 1 were identified in each separate recording. Of the 28 recordings made, 8 were eliminated at this stage due to recording errors. The reading times of the remaining 20 subjects for the critical sentences in each experimental passage were then analysed. Initial analyses indicated that reading times did not differ between groups. Therefore, the data were collapsed across groups for further analysis. The mean reading times and standard deviations for each condition can be seen in Table 5.4.

Passage Type	Place in Text	Mean Reading Time (SD)
Real	Early	9.31 (2.45)
Fairy	Early	9.34 (2.05)
Real	Late	8.69 (1.81)
Fairy	Late	9.74 (2.13)

Table 5.4 Mean reading times (secs) for critical sentences in Early and Late sections of Real and Fairy passages (standard deviations in parentheses)

A 2x2 repeated measures ANOVA was conducted on the reading time data with the factors Passage Type (Real and Fairy) and Place in Text (Early and Late). As expected there was no effect of Passage Type or Place in Text. However, a significant interaction was revealed ($F_1(1,19) = 7.90, p < .05$; $F_2(1, 6) = 0.18, ns$). Post hoc analysis using Tukeys HSD revealed that the critical sentences presented later in the text took longer to read in the Fairy passages than in the Real passages (HSD $F_1 p < .05$). No other differences were significant.

The difference in the number of syllables in the critical sentences in the Real and Fairy passages did appear to affect reading times, although this finding was only true for sentences presented later in the text. On closer analysis of individual reading times, 4 children were identified as having particularly long reading times, which probably accounted for the differences noted in the ANOVA. The performance of these 4 children on a number of other measures was examined and no specific problems were identified. It was therefore decided not to exclude these children from the sample.

b) Response times and error rates to test sentences following experimental passages: Initial comparisons of response time data using the between subjects factor of List indicated that there was no significant difference in response times between groups who received different forms of the test. Thus, for the remaining analyses of test sentence response time, the data were collapsed across groups. Subjects responses were divided into the four experimental conditions and the data was examined for outliers. Outliers were defined as any response time 2.5 standard deviations away from the mean response time of each subject in each condition and would be replaced with the subjects mean response time for that condition. No outliers were found. Analysis of the response time data was conducted on correct responses only.

		Story Type	
		Real Stories	Fairy Stories
Causal	Mean RT (SD)	3615.59 (732.71)	4149.36 (1557.62)
	Mean no correct responses (max = 8)	7.23 (0.90)	6.97 (1.18)
Static	Mean RT (SD)	4357.06 (1251.23)	4302.88 (1520.73)
	Mean no correct responses (max = 8)	4.59 (1.74)	5.23 (1.66)

Table 5.5 Mean response times and number of correct responses for target statements in each experimental condition.

The mean response time and number of correct responses for each target sentence are shown in Table 5.5. These data were analysed using a 2x2 ANOVA with the factors Passage Type and Inference Type. Analyses of the response time data showed that the main effect of Inference Type was significant ($F_1(1,38) = 10.90, p < .01$; $F_2(1,14) = 8.03, p < .05$). Thus, judgements about causal target statements were made faster than judgements about static target statements. There was no main effect of Passage Type ($F_1(1,38) = 1.81, ns, F_2(1,14) = 0.77, ns$) suggesting that the inconsistent information present in the fairy stories did not affect response time. However, there was a significant interaction between Passage Type and Inference Type by subjects ($F_1(1,38) = 4.95, p < .05$) but not by items ($F_2(1,14) = 2.56, p > .05$). Post hoc Tukeys indicated that responses to Real Causal test sentences were significantly faster than responses to Fairy Causal test sentences (HSD $F_1, p < .05$) and significantly faster than responses to both Real and Fairy Static test sentences (HSD $F_1, p < .01$). This finding indicates that there was a belief bias effect when drawing causal inferences. Analysis of the error rates revealed a main effect of Inference Type ($F_1(1,38) = 92.14, p < .001$; $F_2(1,14) = 30.57, p < .001$) showing that more correct responses were made to causal statements than to static statements. The interaction between Passage Type and Inference Type was significant by subjects but not by items ($F_1(1,38) = 6.65, p < .05$; $F_2(1,14) = 1.28, ns$). Post hoc tests revealed more correct responses in both the Real Causal and Fairy Causal conditions than either the Real Static or Fairy Static conditions (HSD $F_1, p < .01$). Furthermore, there were more correct responses to static target statements in the Fairy story condition than in the Real story condition and this difference approached significance (HSD $F_1 p < .06$). No other effects were significant.

To summarise, children responded faster and more accurately to causal statements than to static statements in both Real and Fairy stories. These results suggest that causal inferences are drawn prior to static inference when reading both Real and Fairy stories. However, responses to Fairy Causal test sentences were significantly slower than responses to Real Causal test sentences indicating that children have more difficulty making inferences from information inconsistent with real-world knowledge even when those inferences were necessary for understanding the text. It was also found that children were more accurate in their responses to static target statements in the Fairy story condition than in the real story condition. This finding was unexpected and seems counterintuitive. Given

the empirical falsity of the static target statements in the Fairy stories, one would expect more erroneous false responses than to the static target statements in the Real stories. However, it appears that children are more willing to accept imaginary static properties than they are real ones.

c) Performance on experimental comprehension questions: Following the completion of each passage, the children were asked one question designed to assess their understanding of the passage. The questions following the experimental passages were scored by giving one point for a correct answer and nothing for an incorrect answer. Each subject was awarded a comprehension score for real stories (max 8) and one for fairy stories (max 8). The mean number of correct responses to questions following real stories was 6.15 (SD 1.35) and the mean number of correct responses to questions following fairy stories was 5.87 (SD 1.20). A t-test showed no significant difference between the number of correct responses given to questions following real stories and questions following fairy stories ($t(38) = 1.02, ns$).

d) Individual differences in children's on-line text processing: In order to assess whether on-line sentence processing was related to reading and language skill, the children completed a number of standardised measures in addition to the experimental task. The mean standard scores obtained on each of these measures are shown in Table 5.6.

Measure	Mean	Standard Deviation	Range
NARA II Accuracy ^a	102.87	10.75	88 – 129
NARA II Comprehension ^a	98.62	10.09	74 – 125
WISC Similarities ^b	8.52	8.52	2 – 15
WISC Vocabulary ^b	7.54	7.54	2 – 18
TOWRE Sight Word ^a	107.33	10.45	91 – 129
TOWRE Phonemic Decoding ^a	104.13	12.38	79 - 132

Table 5.6 Mean standard scores on the standardised assessment measures (a = average standard score 100, SD 15;b = average standard score 10, SD 3).

Initial correlations between the raw scores obtained on the standardised measures of assessment, response times to test sentences, error rates and performance on the experimental comprehension questions are shown in Table 5.7. The raw scores on each of the TOWRE measures were added together to give a composite measure of reading efficiency which was entered in the correlation. A significant relationship was revealed between NARA II comprehension scores and response times to static target statements which was not reflected in the relationship between NARA II comprehension and response times to causal target statements. This suggests that general comprehension ability is more important to making static inferences than causal inferences. However, since response times to both causal and static target statements were also strongly related to scores on both TOWRE measures, it could be that the relationship between NARA II comprehension and Static target statement response times was mediated by reading fluency rather than the ability to generate an inference. In other words, children with good comprehension skills may also be able to process information faster than children with poor comprehension and were therefore able to respond more quickly to the target statements. In order to examine this possibility, a partial correlation was run controlling for age and the TOWRE composite score. (See the lower half of the correlation matrix in Table 5.7). After age and reading speed were controlled, the relationship between NARA II comprehension scores and response times to static target statements maintained significance, indicating that the relationship is not reliant on individual differences in reading fluency.

	TOWRE	NARA II	NARA II	WISC Voc	WISC Sim	RT Caus	RT Stat	Corr Caus	Corr Stat
Chron Age	0.08	Acc	-0.26	-0.07	-0.03	0.16	-0.16	0.21	-0.12
TOWRE		Comp	0.23	0.22	0.02	0.22	-0.25	-0.04	-0.21
NARA II Acc			0.69***	0.42**	0.35*	0.35*	-0.04	-0.17	-0.03
NARA II			0.68***	0.57***	0.46**	0.46**	-0.21	-0.17	-0.02
Comp									
WISC Vocab			0.44**	0.58***	0.50***	-0.03	-0.05	0.12	0.01
WISC Sim			0.37*	0.45**	0.52***	-0.22	-0.21	0.13	-0.21
RT Caus			-0.03	-0.18	-0.03	-0.16	0.73***	-0.18	0.24
RT Stat			-0.13	-0.33*	-0.06	-0.11	0.70***	0.09	0.20
Corr Caus			-0.01	-0.16	-0.13	0.11	0.26		0.31
Corr Stat			0.06	-0.03	0.06	0.24	-0.03	0.29	

Table 5.7 Zero order correlations between standardised measures, response times, error rates and experimental comprehension scores and partial correlations (below diagonal) between these variables when age and TOWRE controlled for * $p < .05$, ** $p < .01$, *** $p < .001$

A series of multiple regression analyses were conducted to look at concurrent predictors of performance on the NARA II comprehension measure. The first question of interest was whether performance on the NARA II comprehension measure could be reliably predicted from target sentence response times. The results of these analyses are shown in Table 5.8. Age did not correlate significantly with NARA II comprehension and was therefore not entered into the regression analysis. Reading accuracy as measured by the NARA II was not controlled for as it is a consistent predictor of reading comprehension and would therefore account for a large proportion of the variance, masking the effects of the other variables. Scores on the TOWRE accounted for 5% of the variance when entered on the step 1 to control for reading fluency, which was not significant. WISC vocabulary raw scores significantly accounted for 33% of the variance in NARA II comprehension performance when entered on step 2. Response times to causal target statements accounted for only 2% of the variance when entered on step 3, which was not significant. However, response times to static target statements accounted for 7% of the variance when entered on step 3, which was significant ($p < .05$).

While the generation of elaborative inferences was predictive of comprehension skill, a finding that is in disagreement with the predictions of this study, the results suggest that the generation of causal inferences in the experimental passages did not predict reading comprehension ability. However, while the response time measure used in the regression analysis indicates how quickly a child responds to an inference statement, it gives no information regarding the accuracy of the child's response. Arguably, the accuracy with which a child responds to an inference statement may be a better predictor of reading comprehension as it is a clearer indicator of how well the child has understood the passage. In order to investigate this further, accuracy rates for causal test sentences was entered on step 3 of the regression analysis. These accounted for 6% of the variance in NARA II comprehension scores which approached significance ($p = .07$). Furthermore, none of the variance was accounted for by accuracy rates to static test sentences when entered on step 3 (see Table 5.8).

<i>NARA II Comprehension Scores</i>			
<i>Step</i>	<i>Variable</i>	<i>R² change</i>	<i>P</i>
1	TOWRE	0.05	0.17
2	WISC Vocab	0.33	0.00
3	RT causal	0.02	0.27
3	RT static	0.07	0.04
3	Acc causal	0.06	0.07
3	Acc static	0.00	0.75

Table 5.8 Multiple regression with NARA II comprehension as the dependent variable and TOWRE, WISC Vocabulary, target sentence response times and target sentence accuracy as the predictor variables.

5.4.3 Discussion.

The present study asked whether children could set aside real-world knowledge in order to draw causal and static inferences based on information that was inconsistent with their experience of the world. Analysis of target sentence response times indicated that causal inference target sentences were responded to faster overall than static sentences. Moreover, this pattern of inference generation was consistent across both types of story indicating that the nature of the story did not affect the children's ability to draw inferences from the text. Furthermore, children were more accurate in responding to causal than static test sentences. The results of this study indicate that children are faster at drawing causal coherence inferences than elaborative inferences when reading. One interpretation of these results is that the children drew causal inferences on-line during the reading of the passage, while static inferences were not drawn until the test sentence was presented. This pattern is consistent with other research on inference generation and the findings are further supported by the fact that responses to causal target statements, were more accurate than to static target statements irrespective of the empirical veracity of the inference. It should be noted however, that this is not the only possible explanation for these results. For example, it may be the case that the children found the information given in the static inference test sentences less plausible than the information in the causal inference test sentences and where therefore slower to respond to the static test sentences. Furthermore, while it is clear

that causal inferences were responded to at a faster rate than static inferences, it cannot be definitely concluded that these inferences were drawn on-line. It may simply be the case that they too were drawn off-line but at a faster rate. However, as previously stated, the results of this study are consistent with previous research positing an on-line/off-line distinction between these two classes of inference. With this in mind, the tentative conclusion that the results presented here mirror this on-line/off-line distinction between the generation of causal and elaborative inferences does not seem unreasonable.

While responses to static target statements were less accurate overall than responses to causal target statements, the children were more accurate in their responses to static target statements when the content was empirically false than when the content was empirically true. Because the only correct response to a target statement following an experimental passage was TRUE, this finding suggests that children made more frequent TRUE responses to empirically false static properties than to empirically true static properties. One possible explanation for this finding could be that children were willing to accept an empirically false static property as true because it referred to an object or character in a passage that was also empirically false and therefore was consistent with the overall fictional theme of the passage. In a similar vein, when presented with a static property that they recognised as belonging to an empirically true object, character or event, they had no problem responding correctly. However, in cases where the child was unaware that the object, character or event in question possessed the static property expressed, they responded to it as if it was an empirically false property. The finding that children are willing to accept empirically false static properties as true when they are consistent with the character or object to which they refer is in line with previous research investigating the ability of children to engage in imaginary role play (see Harris 2000 for a discussion.) In short, when children engage in role-play, they actively enter into the imaginary world of the character they have chosen to portray and thus, view the world from that character's point of view. In terms of the findings reported here, it would appear that the children in this study were able to accept as true the empirically false static properties that were consistent with the theme of the story because they were able to make the judgement from the protagonist's point of view.

Finally, the number of comprehension questions answered correctly following fairy stories did not differ from the number of questions answered correctly following real world stories. This finding indicates that the children in this study had no difficulty understanding each story type. It could be argued that this was not a good measure of comprehension since only one question followed each passage. However, the overall results suggest that the children in this study were able to set aside their real world knowledge and draw inferences based in a fairy tale context. The results are in contrast with the findings of Franks (1996; 1997) and Evans and Perry (1995) outlined above.

The results of the initial multiple regression analysis indicated that response times to static target sentences but not causal target sentences predicted reading comprehension skill. This was an unexpected result and suggests that the ability to generate causal inferences does not make a significant contribution to overall reading comprehension ability. However, the ability to generate static inferences does seem to contribute to overall reading comprehension skill. Further analyses were conducted using accuracy rates as predictors of reading comprehension and the results suggested that reading comprehension could not be predicted by inference generation when inference generation is measured in terms of the accuracy with which a child responds to an inference statement. These results appear to support the notion that the generation of causal inferences is not necessary for constructing a representation of a text during reading. However, the results of the response time analysis indicated that causal inferences were automatically drawn on-line during reading while static inferences were made when prompted by the presentation of a test sentence. Given the apparent automaticity of responding to causal test sentences, a possible explanation for the findings of the multiple regression analyses might be that there was not enough variance in the inference generation measure used here to predict variance in reading comprehension performance.

A further possible limitation of this study is that the critical sentences preceding a static target statement may in some cases have also contained a potential causal inference. Therefore, the slower response times to static target statements may reflect the fact that a causal inference had already been drawn on-line and the generation of the static inference was therefore delayed by the holding of information relevant to the causal inference in

working memory. For example, in one story, the critical sentences "The men struggled to keep control of the wheel. They used all their strength to fight the wind and rain" were followed by the static test sentence "The men were soaking wet", a plausible elaboration regarding the physical state of the men given the conditions described by the sentences. However, it is also possible to draw the causal inference from these sentences that the wind and rain were causing the men to struggle with control of the wheel. In the next chapter, study 4, the materials were more constrained to ensure that a causal inference could not be drawn in the critical sentences preceding a static target statement.

A further criticism that could be directed at the study lies in the nature of the materials. In this study, stories that were used to represent counterfactual information were written as fairy tales. However, the information may not have been counterfactual in the sense that it was true in the context of the story. Moreover, while the premise information presented in the fairy tale passages was clearly incongruent with our knowledge of the world, the inferences themselves may not have been in conflict with real world knowledge. Similarly, while the content of the fairy tales was counterintuitive, it may not have been unexpected since most children have experience of fairy tales throughout their formative years. Thus, when presented with a fairy tale the children may have recognised the story as such and changed their perspective accordingly when reading to incorporate the shift in genre, thus accepting the counterintuitive information presented in the fairy tales as true. A more rigorous approach may have been to use stories that were based in a real world context but with contrary to fact information embedded in the critical sentences. Nevertheless, there is some evidence to suggest that the content of the fairy stories did have an effect such that the response times to Real Causal test sentences were significantly faster than to Fairy Causal test sentences.

While the pattern of inference generation seen in this study is consistent with that found in many previous studies the present experiment only looked at two inference types. The effect of counterfactual information may be different for inferences about the emotional, motivational or instrumental aspects of a passage for example. Furthermore, the inferences were designed to maintain local coherence. However, research has shown that global coherence is also important in building a coherent representation of a text (e.g. Long,

Oppy, & Seely, 1997). If a story is full of contradictions to real world knowledge, it may be more difficult to keep track of the global perspective than if the story is consistent with our experience and expectations.

To summarise, the results of this study suggest that the children in this study drew causal inferences on-line even when the information required to make those inferences was contrary to empirical knowledge, although inconsistent information did impede this process. Furthermore, there was no difference in the time-course of the generation of static inferences from inconsistent premises and that from consistent premises. Thus, it appears that the children in this study were able to overcome the belief bias and draw conclusions based on false premises in a narrative context.

Chapter 6

The Development of Children's On-Line Inference Generation during Reading.

6.1 The Development of Inferential Skills

Thus far this thesis has aimed to establish a pattern of on-line inference generation during reading in children with normal reading skills. Experiments 2 and 3 demonstrated that children made causal inferences on-line and elaborative inferences at a later point in reading. Moreover, experiment 3 indicated that children were able to suspend disbelief when reading in order to generate inferences from information that is inconsistent with real world knowledge. However, while this pattern of inference generation was consistent across studies, the conclusions are limited to children of approximately 9 to 10 years of age. In order to fully address the question of whether children with comprehension difficulties have difficulties with inference generation, this chapter investigates the developmental pattern of inference generation in children and how this is related to reading skill.

Research investigating developmental differences in inference generation has shown that the ability to generate inferences from text increases with age. Johnson and Smith (1981) looked at inference generation in third-grade and fifth-grade children. Using a between-subjects design, they tested children under 3 conditions; an implicit story condition, an explicit story condition and a sentence condition. In the implicit story condition, the children were presented with sets of premises embedded in a simple story. Having read the passage, the children were asked questions about information that was implied by the premises. They were also asked a number of questions based on information that was explicitly stated in the premises. The explicit story condition was similar to the implicit condition except that the inference only implied in this condition was made explicit by including it in the text. Finally, in the sentence condition, the children were presented with each set of premises and the corresponding inference question outside of the context of the passage.

In the implicit story condition, third-grade children answered 46% of inference questions correctly, and fifth-grade children 60%, suggesting that children in the younger reading group made fewer inferences than the older children. In the explicit story condition third-grade children answered 53% of inference questions correctly whereas 5th graders

were 76% correct. Performance was best for both groups of children in the sentence condition; third-grade 75% correct, fifth-grade 88% correct, indicating that the younger children were able to integrate the implied information successfully in this condition. Taken together these findings suggest that even when the information was explicit in the text, the younger children had difficulty integrating it into their story representation. The poor performance of the younger children in the implicit story condition cannot, therefore be attributed solely to an inability to generate inferences. Given that in both story conditions the children had to read the entire passage before being presented with a list of questions it may be that the younger children performed poorly because of the memory requirements of the task. In the sentence condition, this memory load was alleviated as the inference question was asked directly after each set of premises.

Casteel and Simpson (1991) conducted a study with second-, fifth- and eighth-grade children, and a group of undergraduate students. Each participant was given 8 stories to read, sentence by sentence, using a self-paced reading paradigm. Each story contained 2 forward (elaborative) inference sentences and 2 backward (coherence) inference sentences with corresponding inducing sentences. The forward inference sentences were designed to represent an action or event that could be encoded into a mental representation without knowing its consequence. For example, in the following passage taken from Casteel and Simpson (1991), the sentence in italics represents a forward inference:

"One weekend Pam went roller-skating with her friends. *She tried to roller skate for the first time.* After skating, the girls had a slumber party at Pam's."

There are many potential consequences that could be generated from the forward inference sentence given above. However, the generation of a consequence is not necessary for the sentence to fit into a mental representation of the passage.

The backward inference and corresponding inducing sentence pairs represented a sequence of events that required a bridging inference to be made in order to maintain coherence. For example:

"Pam's father told the girls a ghost story. All of her friends screamed."

This sentence pair would require the inference that the ghost story scared the girls in order to maintain coherence.

After each story was presented, a series of questions were presented on the screen. These were designed to tap the different inferences embedded in the passage. For example, the question "Did Pam fall down a lot when roller-skating?" would assess the generation of the forward inference, and "Did the ghost story scare Pam's friends?" would assess the generation of the backward inference required by the sentence pair given above. Responses and reaction times to each question were recorded.

Second-grade children performed significantly better than chance but significantly less accurately than children in all other grade levels. The results suggest that they were less likely to draw inferences than any of the other age groups although they were able to generate inferences to some extent. No other age differences were significant. Analysis of the response time data shows that 8th grade children responded faster than 5th grade children who in turn responded faster than 3rd grade children. However, there was no significant difference in response time between the 8th grade children and the undergraduate students. Finally, responses were faster and more accurate to backward than to forward inferences indicating that backward, or coherence inferences were drawn prior to forward, or elaborative inferences. These results were replicated in a later study (Casteel, 1993) and suggest that children as young as 7-years of age generate inferences that are necessary for textual coherence. Furthermore, the pattern of inference generation found in these studies mirrors that of the pattern consistently found in the present thesis.

The studies described above looked at children aged 7-years and over. However, Thompson and Myers (1985) carried out a study looking at inference generation and recall in children aged 4- and 7-years, thus comparing pre-school children with children who have some school experience. The children were given three stories to read. The stories varied in terms of the manner in which the target events in the story were connected. In one story, the target events occurred as a result of a *physical* cause, in a second story, the events were *psychologically* connected i.e. the emotional state of a character caused a target event, and in the third story, termed the *enablement* version, the events which preceded the target

event were necessary but not sufficient to have caused that event. Each story had eight corresponding questions. Two questions required the children to make a causal coherence inference and were termed 'logical inference' questions, and two questions required the application of real world knowledge to the text in order to interpret propositions and events in the story. These questions were termed 'constrained informational inference' questions. It is important at this point to distinguish between these questions and the logical inference questions described above. To illustrate, the example below is a 'physical cause' story used by Thompson and Myers (1985) followed by the corresponding 'logical cause' and 'constrained informational inference' questions:

One day Jennifer's parents gave her a dollar because she wanted to buy a turtle. But as she was walking to the pet store, *she turned a somersault (physical cause)* and lost it. Jennifer was worried that her parents would be angry with her, so she decided to search every bit of the sidewalk. For 10 long minutes she looked in all the cracks and grass. Finally, she found the dollar. When she got back to the store she was told that the last turtle had been sold just 1 minute ago.

What made Jennifer lose her dollar? (logical cause)

Why would Jennifer's parents be angry with her? (logical cause)

How old do you think Jennifer is? (constrained)

Did Jennifer have a dollar bill or a dollar in change? (constrained)

The example above shows that while the 'constrained informational inference' questions were based on information that could be easily inferred from the text, the number of possible alternative answers to these questions was greater than to the 'logical cause' questions. Thus, while both inference types could be said to aid textual coherence, it could also be argued that the restraints placed on responses to logical inference questions would create a more succinct interpretation of the story.

Two further questions were asked regarding information that could have been elaborated from the story but were not necessary for comprehension. Finally, two questions were asked which required recall of information that was explicit in the text. The children were asked to retell each story when they had finished reading and the questions

were either asked at appropriate points during the story, or following the child's retelling of it.

Seven-year old children made more correct responses to inference questions than 4-year olds in the psychological cause and enablement stories, but not in the physical cause stories. These results indicate that children as young as 4-years old are able to generate inferences when they are related to a physical cause. However, inferences based on abstract or incomplete information are not easily drawn by this age group, although the overall performance of the younger children was still better than chance on logical questions, suggesting some ability to generate causal inferences. Moreover, the 4-year old children responded most often to the unconstrained inference questions indicating a high degree of unnecessary elaboration. Conversely, the 7-year olds responded least often to these questions. Thus the 7-year old children appeared to have developed a more efficient inference generation strategy, concentrating on those inferences that were necessary for coherence. While the point at which this strategy shift takes place is not clear, it is possible that it is related to the formal reading instruction the older group has received through their school experience. Unsurprisingly, the older children's retelling of the stories suggested a greater degree of comprehension than the younger children, their accounts being more complete, accurate and well-structured. However, recall performance of both age groups was greatest for stories that contained *physically* connected events, indicating the 7-year old children were still not fully competent in generating inferences from emotional or insufficient information. These results suggest that, while overall inference generation skills begin to develop at a relatively young age, the rate at which the ability to generate of specific inference types varies with age.

Taken together these studies suggest that while young children are not as able as their older peers to generate inferences from text, they still engage in some degree of inference generation during reading. Furthermore, the limited degree of inference generation observed in younger children is not necessarily due to an absence of inferential skill, rather other factors such as memory capacity, processing speed and general knowledge could play a role. Moreover, the results of Thompson and Myers (1985) suggest that strategies for inference generation become more efficient with age, with older children becoming less likely to draw unnecessary inferences than younger children.

Clearly, the studies discussed here focus on the changes in inferential processing as a function of chronological age. Similarly, the present study was initially intended to focus on age-related differences in performance on the experimental task used in study 3 and as such the experimental task was given to three groups of children in each of years 3, 4 and 5. Nevertheless, analysis of the results revealed a pattern of performance that was unreliable. As might be expected based on the literature outlined above, children in Year 4 performed better on the experimental task than children in Year 3, thus reflecting an improvement in performance with an increase in chronological age. However, the performance of children in Year 5 was below that of children in Year 4, suggesting a developmental pattern that was not in line with previous research. Moreover, the children in Year 5 performed poorly on the individual measures of reading and comprehension relative to the children in Years 3 and 4. On reflection it was felt that this discrepancy in performance might be due to a difference in teaching strategy at the school from which the children were selected rather than differences in ability between the children in Years 3 and 4, and the children in Year 5. For example, the introduction of the literacy hour and the initiatives of a new Headteacher may have changed the teaching environment and this change was of primary benefit to the younger children coming up through the school. Thus, the reading and language skills of children in Year 5 may have confounded their performance on the experimental task, resulting in an overall developmental pattern that may not have reflected reliable age-related differences but rather differences in educational experience. With this in mind, it was felt that analysis of performance on the experimental task in terms of chronological age resulted in an unreliable picture of the development of inference generation. The aim of the study was therefore revised to investigate the influence of reading ability on inference generation. Given that children of the same chronological age may show large variations in their reading ability, the influence of reading skill on inference generation is clearly of interest. However, the literature tends to focus on age-related differences and the impact of reading ability on inference generation has not been investigated to the same degree. Moreover, since comprehension processes in reading are often constrained by decoding skill, it was decided in the present study to investigate the performance of skilled and less-skilled readers on the inference generation task used in experiment 3.

Aim of the present study

The aim of the current study was to assess whether the pattern of inference generation found in experiments 2 and 3 of this thesis varied as a function of reading age. Experiment 3 demonstrated that children aged 9-10 years of age were able to overcome their belief bias in order to generate inferences based on empirically false information. However, investigations into the belief bias effect in deductive reasoning have shown that this effect is subject to developmental change (e.g. Evans and Perry, 1995) Therefore, this study also aimed to investigate developmental differences in the belief bias effect on inference generation during reading.

6.2 Experiment 4

6.2.1 Method.

a) Participants: All the children in years 3, 4 and 5 of two York primary schools were given the *Suffolk Reading Scale* (SRS; Level 2, Form A; Hagley, 1986). Children who achieved a standard score of above 95 (a raw score over 40) were selected to take part in the study. Following selection the children were given the *Neale Analysis of Reading Ability II* (Neale, 1989). They were then assigned to two groups on the basis of their NARA II accuracy reading age. Children with a reading age between 7 years and 9 years 11 months were assigned to the Less-Skilled group. Children with a reading age between 10 years and 12 years 11 months were assigned to the Skilled group. In addition to this, children with a NARA II Comprehension standard score below 95 were excluded from the sample.

The children in each Reading Group were randomly assigned to one of two experimental groups: Group 1 received List A and group 2 received List B. However, the factor List did not reach significance in any of the initial analyses conducted and the data reported here is therefore collapsed across List. Table 6.1 shows the mean age, SRS standard score, NARA II accuracy age and NARA II comprehension standard score for each experimental group. As previously stated, the aim of this experiment was to investigate the effect of reading age on the ability to draw inferences during reading and as such the children were assigned to the Skilled and Less-skilled reading groups on the basis of their NARA II accuracy age. A series of t-tests revealed that the two groups differed significantly in NARA II accuracy age ($t(36) = -14.22, p < .001$), NARA II comprehension

standard score ($t(36) = -2.72, p < .01$), Suffolk standard score ($t(36) = -2.16, p < .05$) and chronological age ($t(36) = -2.69, p < .05$).

Measure	Less-Skilled Readers (N = 19)	Skilled Readers (N = 19)
Chronological Age (mths)	97.21 (10.93)	106.24 (9.75)
NARA II Accuracy age (mths)	100.83 (7.16)	143.77 (11.04)
NARA II Comprehension age (mths)	104.42 (6.03)	110.53 (7.72)
Suffolk Standard Score	107.32 (5.20)	111.79 (7.36)

Table 6.1 Mean standard Suffolk Reading Scale (SRS) score, chronological age (months), NARA II reading accuracy age (months) and NARA II comprehension standard score for each Reading Group (standard deviations in parentheses).

b) Design: This study was a replication of the study 3 but with the introduction of the factor Reading Group. A mixed factorial design with the between subjects factor Reading Group (Less-Skilled and Skilled) and the within-subjects factors Passage Type (Real and Fairy) and Test Sentence (Causal Link and Static Property) was used. As in study 3, each child read each passage and responded to two sentence judgement trials for each. Thus each child completed eight trials in each of four conditions; Real Causal Link, Real Static Property, Fairy Causal Link and Fairy Static Property. The trials were presented in a random order. Each child was also administered a number of standardised tests to assess individual differences in reading, comprehension and reasoning skills.

c) Materials: The materials used in experiment 3 were used in the present study with some modifications. First, the experimental passages were adjusted to overcome the confound noted in the discussion of chapter 5 that, in some cases, the critical sentences preceding the static property test sentences required a causal link. To this end, the materials were altered so that, in all cases where a static test sentence was presented, the critical sentences preceding it did not require a causal link. Examples of these modified passages and the corresponding questions are given in Table 6.2. In making these changes, the number of syllables in the critical sentences had to be altered, although the same matching criteria was applied as in the original study. Table 6.3 shows the mean number of

syllables in the critical sentences and test sentences of all the passages in both lists together with the book grading levels and equivalent WORD reading ages for the modified passages. There was no difference between Real and Fairy passages in the number of syllables in critical sentences ($t(10) = -0.2, ns$) or in the test sentences ($t(10) = -1.74, ns$). An ANOVA was run on the modified book grades using the within subjects factor List (list 1 or list 2) and the between subjects factor Story Type (Real or Fairy). There was no main effect of either List or Story Type and no interaction.

Second, an additional comprehension question was written for each experimental and filler passage. A full list of the passages, test sentences and comprehension questions can be found in Appendix IV.

Section A	Real	John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. Francis fired his pistols one after the other. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.	Fairy	The mermen lived far out to sea, where the water is blue as the bluest cornflower. The merman king had six beautiful daughters and the youngest was the prettiest of them all. Her skin was soft and delicate as a rose leaf and her eyes were as blue as the deepest sea. But the young daughter could only swim in the sea. She longed to walk on the land and feel the sand between her toes.
Test sentence 1		<i>the bear had huge paws</i>	<i>the girl had a fish tail</i>	
Section B		Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. Snarling, the bear swiped at the girls. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear fell to the ground and did not move.	One night, the mermaid rescued a prince from the stormy sea. They fell in love, but the mermaid could not live out of the water. So the mermaid sang sadly to the prince every night. The merman king could not bear to see his daughter so sad. He cast a spell over the mermaid and her tail disappeared. She ran happily to meet the prince on the shore	
Test sentence 2		<i>John shot the bear dead.</i>	<i>the spell gave the girl legs</i>	
Question 1		What was John holding as he waited in the wagon?	Who was the prettiest mermaid?	
Question 2		Who was attacking John and Henry?	Who did the mermaid rescue?	

Table 6.2 Examples of Real and Fairy story types, test sentences and comprehension questions.

	List 1		List 2	
Mean No of Syllables	Real	Fairy	Real	Fairy
Section A				
Mean no syllables in critical sentences. (SD)	25.16 (3.60)	25.67 (5.05)	25.16 (3.60)	25.67 (5.05)
Mean no syllables in test sentence (SD)	6.83 (1.47)	8.16 (1.17)	6.83 (1.47)	8.16 (1.17)
Section B				
Mean no syllables in critical sentences. (SD)	25.16 (3.60)	25.67 (5.05)	25.16 (3.60)	25.67 (5.05)
Mean no syllables in test sentence (SD)	6.83 (1.47)	8.16 (1.17)	6.83 (1.47)	8.16 (1.17)
Book Grade				
Mean Book Grade (SD)	17.27 (1.14)	18.12 (1.19)	17.40 (1.47)	18.30 (0.90)
Book grade range	15.88 - 18.64	15.9 - 19.2	15.42 - 19.83	17.11 - 18.80
Reading age range (WORD equivalent)	6.58 - 7.25	7.00 - 7.25	6.58 - 7.33	7.00 - 7.25

Table 6.3 Mean number of syllables and book grades for experimental passages (standard deviations in parentheses; WORD = Wechsler Objective Reading Dimensions.)

32 adult raters were presented with a multiple choice off-line version of the task in order to validate the inferences generated from the stories. Raters were presented with a booklet containing each of the stories in one list. The stories were presented in two sections as in the experimental task. However, after each section, raters were asked a question which tapped the inference concept generated by the critical sentences. Raters were given three possible answers to choose from; the inference used in the experimental task, a second plausible inference and a third implausible inference, and they had to indicate on the booklet which answer they had selected. The order in which the sentences were presented was randomised across stories. Participants were given 12 stories, each containing 2 questions, thus answering a total of 24 questions. It was predicted that raters would select the answer corresponding to the test sentence presented in the on-line version of the task. An example of one of the stories used in this task is shown in Table 6.4 and a full list of materials can be found in Appendix V.

Section A	John looked behind the wagon and saw the Red Indians approaching. He called for his father. Henry grabbed his rifle and pulled himself up to the front of the wagon. As the Red Indians rode closer, John could see they were carrying weapons. He raised his own weapon and aimed it at the enemy. But Henry held up his hand. John took his finger off the trigger and carefully lowered his gun.
Question	What did Henry want John to do?
a	Henry wanted John to fire his gun.
b	Henry wanted John to wait.
c	Henry wanted John to read a book.
Section B	The Indians were so close that John could see their brightly coloured faces. John's hand was firmly on his rifle as he waited in the wagon. Then Henry gave the signal and John and Henry opened fire. The battle lasted only a few minutes but seemed a lot longer. Finally, the Indians had weakened. They turned and disappeared into the mountain range.
Question	What did the Indians ride?
a	The Indians rode camels.
b	The Indians rode horses.
c	The Indians rode bicycles.

Table 6.4 Example of story and questions given to raters in the off-line validation task.

16 raters were given the stories presented in List 1 and 16 were given List 2. The modal response across raters was recorded for each story. For participants given List 1, the modal answer selected for 22 out of the 24 items corresponded to the test sentence presented in the on-line version of the task. For list 2, the answer corresponding to the on-line inference was selected as the modal response for 21 out of the 24 items. The 5 items for which the modal response did not correspond to the on-line inference sentence were all examples of static inferences and one of these items was the same in both list 1 and list 2. Due to the fact that only 5 of the 48 items were identified as problematic, the materials were not adjusted. However, in the final analysis close attention was paid to these items.

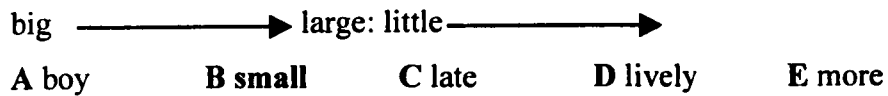
Reading Skills

To assess reading skills, the children were given the *Neale Analysis of Reading Ability* (NARA II; Form 1; Neale, 1989), and the *Test of Word Reading Efficiency* (TOWRE; Torgeson, Wagner & Rashotte, 1992).

General Cognitive Ability

The *WISC Vocabulary* sub-test (Third Edition UK; WISC III; Wechsler, 1992) and the *Verbal Reasoning* sub-test of the *Cognitive Abilities Test* (Thorndike, Hagen, & France, 1986) were administered to assess general cognitive ability.

The CAT verbal reasoning test is administered as a group test. The children are asked to solve a number of verbal analogies within an 8-minute time period. The children are given a booklet of problems and an answer sheet on which to record their responses. Each problem is presented in the following way. A pair of words is presented which are related in some way. Next to this pair of words is a third word. The child's task is to choose one word from a list of 5 possible alternatives which relates to the third word in the same way as the second word relates to the first. For example, in the problem shown below, the words big and large are related because they mean the same thing. Therefore, in order to solve the problem the child should choose the word 'small' from the list of alternatives because it relates to 'little' in the same way as big relates to large i.e. it means the same thing.



d) Procedure: The children were seen over a course of 2 x 30 minute individual sessions and 1 group session lasting 15 minutes. In sessions 1 and 2 they were given the experimental task. The procedure for this task was the same as in study 3. The NARA II, TOWRE and WISC vocabulary sub-test were also given in these sessions. In the group session, the children were administered the CAT verbal reasoning subtest.

6.2.2 Results

Response times and error rates to test sentences¹: Initial analysis of response time data indicated no significant effect of List on the response times of participants. Further analyses of response times therefore used data that was collapsed across lists. Subject's responses were divided into the four experimental conditions and the data was examined for outliers. Outliers were defined as any response time 2.5 standard deviations away from the mean response time of each subject in each condition and were replaced with the subject mean response time for that condition. 1.21% of response times were adjusted in this way. Only items to which responses were correct were included in the response time analysis.

The mean response times for correctly judged sentences and the number of correct responses overall are shown in Table 6.5. A 2x2x2 repeated measures ANOVA was conducted on the response times to items answered correctly with the between subjects factor Reading Group (Less-Skilled and Skilled), and the within-subject factors Passage Type (Fairy, Real) and Inference Type (Causal, Static). There was a significant main effect of Reading Group ($F_1(1,30) = 4.60, p < .05$; $F_2(1,10) = 5.24, p < .05$) such that the Less-Skilled group response times were significantly slower than the Skilled group. There was a main effect of Passage Type by subjects ($F_1(1,30) = 8.97, p < .01$) such that response times to Real stories were faster than response times to Fairy stories. This effect was not significant by items ($F_2(1,10) = 2.71, ns$). There was a main effect of Inference Type ($F_1(1,30) = 16.97, p < .001$; $F_2(1,10) = 12.49, p < .01$) such that response times to Causal inferences were faster than response times to Static inferences. No

¹ Response times and error rates analysed excluding the 5 items previously identified as unreliable revealed the same pattern of results. These items were therefore included in the final analysis.

other main effects or interactions were significant. These results are shown in figure 6.1.

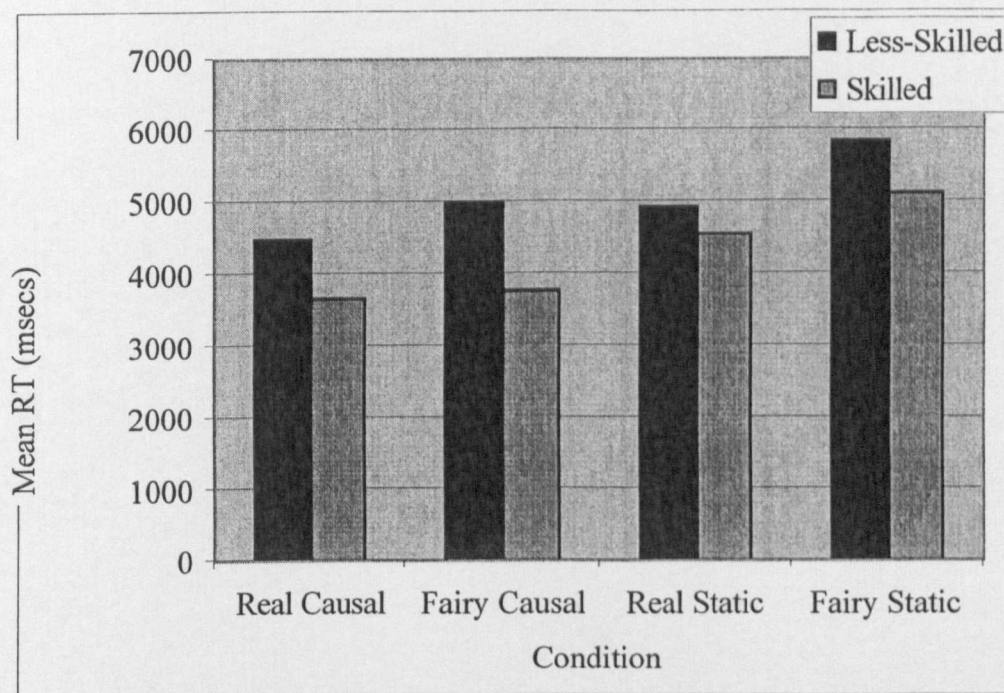


Figure 6.1: Mean response time of each experimental group in each test sentence condition

A repeated measures ANOVA was conducted on the number of correct responses given to target sentences. Results revealed a main effect of Passage Type in the by-subjects analysis ($F_1(1,36) = 9.77, p < .01$) such that more correct responses were given to Fairy than to Real stories. No such effect was found in the by-items analysis ($F_2(1,10) = 0.92, ns$). The main effect of Inference Type was also significant by subjects ($F_1(1,36) = 72.00, p < .001$) such that more correct responses were given to Causal test sentences than to Static test sentences. This effect was not significant by items ($F_2(1,10) = 1.71, ns$). There was a main effect of Reading Group by items ($F_2(1,10) = 46.28, p < .001$) but not by subjects ($F_1(1,36) = 0.53, ns$) such that overall the Skilled group were more accurate across items than the Less-Skilled group. No other main effects or interactions were significant. These results are shown in figure 6.2.

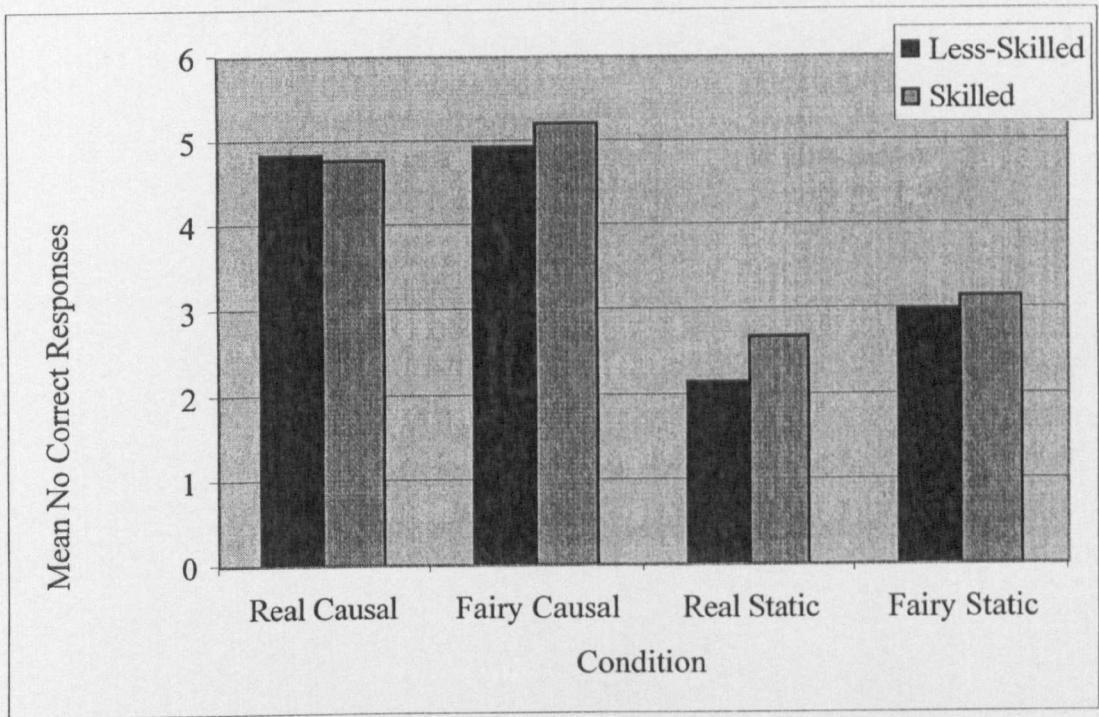


Figure 6.2: Mean number of correct responses of each experimental group in each condition

	Mean Response Times				Mean Number Correct Responses			
	Real Stories		Fairy Stories		Real Stories		Fairy Stories	
	Causal	Static	Causal	Static	Causal	Static	Causal	Static
Less-Skilled	4493.59 (960.90)	4915.25 (1563.57)	5013.07 (1149.02)	5856.88 (2022.65)	4.84 (0.90)	2.13 (1.61)	4.95 (1.03)	3.00 (1.53)
Skilled	3670.86 (1026.71)	4568.28 (1839.72)	3779.24 (731.53)	5129.29 (1274.48)	4.79 (1.18)	2.68 (2.00)	5.21 (1.08)	3.16 (1.38)

Table 6.5 Mean test sentence response times and number of correct responses for each reading age group (standard deviations in parentheses).

The results suggest that Causal sentences were responded to faster and more accurately than Static sentences; that Skilled children responded faster and more accurately than Less-Skilled children; and that Real story test sentences tended to be responded to faster but less accurately by the children than Fairy story test sentences. Given that not all effects were significant by subjects and by items it was decided to look at the relationship between speed and accuracy in order to rule out trade-off effects. Accordingly, correlations between the speed and accuracy of responding to the Causal and Static test sentences for each Reading Group were calculated. The results show a significant positive correlation between speed and accuracy of responding to static test sentences in the Less-Skilled group ($r = 0.49, p < .05$). This indicates that, for children with a lower reading age, taking time to respond to static test sentences resulted in a higher number of correct responses, suggesting that these responses are not made automatically, but require strategic processing. No other correlations were significant. However, although marginal, for Skilled readers there is a trend for faster response times to causal test sentences to be associated with higher accuracy scores indicating that there is no speed/accuracy trade-off ($r = -0.44, p = 0.63$). Figures 6.3 and 6.4 illustrate these results.

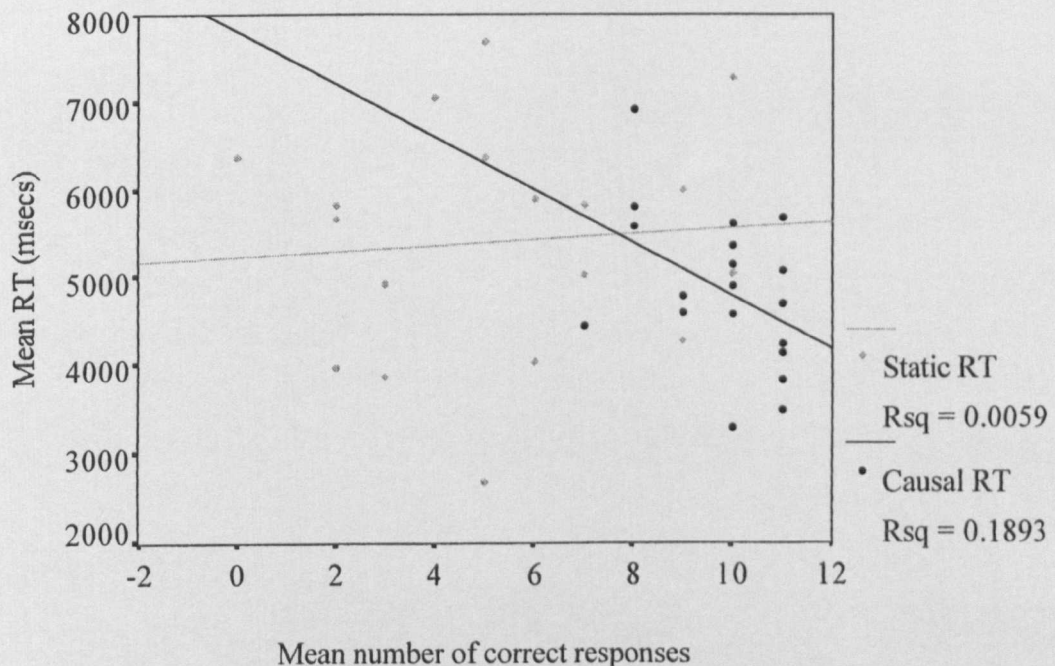


Figure 6.3 The relationship between speed and accuracy of responding to Causal and Static test sentences for the Skilled group.

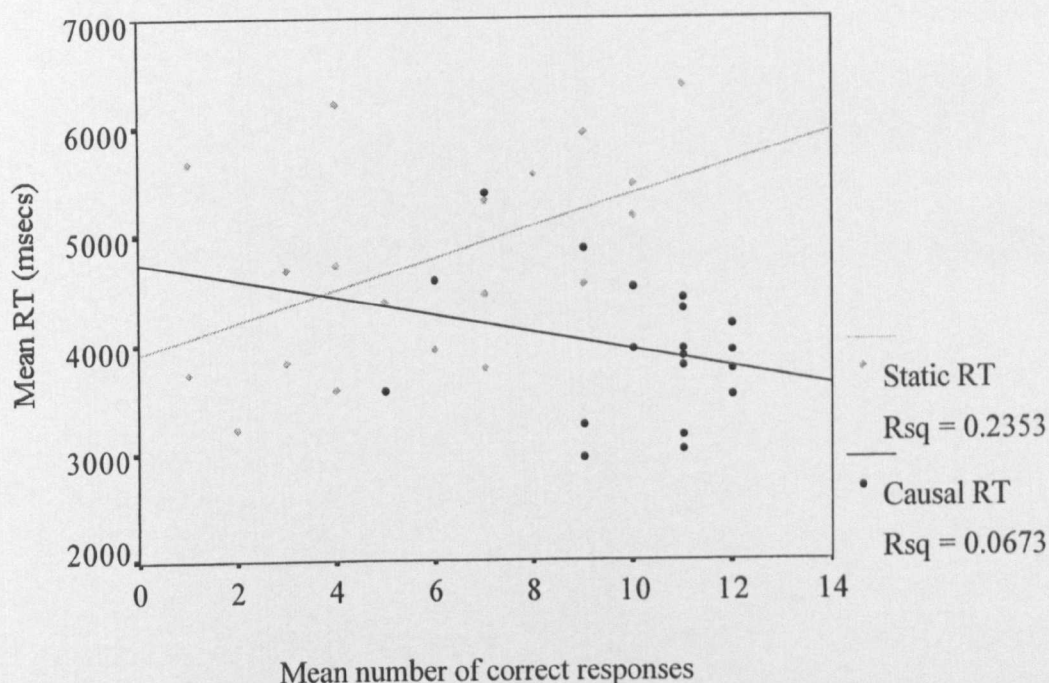


Figure 6.4 The relationship between speed and accuracy of responding to Causal and Static test sentences for the Less-Skilled group.

b) Performance on experimental comprehension questions: Children's scores on the comprehension questions asked after each passage for the three Passage Types (Real and Fairy) are shown in Table 6.6. A repeated measures ANOVA was used with the within-subjects factor of Passage Type and the between-subjects factor of Reading Age Group (Less-Skilled and Skilled). There was no effect of Reading Group indicating that both Less-Skilled and Skilled children performed equally well on the comprehension questions ($F(1,36) = 1.79, ns$). The main effect of Passage Type was not significant ($F(1,36) = 2.50, ns$) and there was no interaction between Reading Group and Passage Type ($F(1,36) = 2.50, ns$).

	Real Story (max 12)	Fairy Story (max 12)
Less-Skilled	8.53 (1.95)	7.58 (2.36)
Skilled	8.89 (2.31)	8.89 (1.94)

Table 6.6 Mean performance on experimental comprehension questions for each Reading Group (standard deviations in parentheses).

c) Individual differences in children's on-line text processing: The mean standard scores obtained on each of the standardised measures completed by participants are shown in Table 6.7. Significant differences were revealed between groups for NARA II accuracy standard score ($F(1,36) = 56.78, p < .001$), NARA II Comprehension standard score ($F(1,36) = 7.38, p < .05$); TOWRE Sight Word Reading ($F(1,36) = 6.42, p < .05$); TOWRE Phonemic Decoding ($F(1,36) = 12.91, p < .01$) and Chronological Age ($F(1,36) = 7.22, p < .05$). The two groups did not differ in performance on the CAT test of verbal reasoning ($F(1, 31) = 1.29, ns$) and WISC vocabulary task ($F(1,36) = .002, ns$).

<i>Measure</i>	<i>Group</i>			
	Less-Skilled (N =19)	Range	Skilled (N = 19)	Range
Chronological Age (mths)	97.21 (10.93)	86.04 – 125.04	106.24 (9.75)	90 – 118.92
NARA II Accuracy ^a	102.11 (4.98)	92 - 111	117.78 (7.63)	101 - 129
NARA II Comprehension ^a	104.42 (6.03)	95 - 115	110.53 (7.72)	96 - 122
WISC Vocabulary ^b	8.95 (3.03)	3 - 15	8.90 (3.23)	1 - 13
TOWRE Sight Word ^c	60.47 (6.39)	47 - 73	70.58 (5.15)	62 - 81
TOWRE Phonemic Decoding ^c	28.84 (7.16)	18 - 44	41.84 (7.14)	28 - 58
CAT Verbal Reasoning ^c	9.94 (2.54)	5 - 14	11.25 (3.85)	4 - 18

Table 6.7 Standard scores of each Reading Group on the standardised assessment measures (standard deviations is parentheses; a = standard score; b = scaled standard score; c = raw score (CAT Verbal Reasoning N = 17 for each group).

Correlations between the raw scores obtained on the standardised measures of assessment, response times to test sentences, error rates and performance on the experimental comprehension questions are shown in Table 6.8. As in section 5.6e, a TOWRE composite raw score was entered into the correlation. Consideration of the data in Table 6.8 will concentrate specifically on the relationships between test sentence response times and accuracy levels with the standardised measures of assessment.

	Towre	NARA II Acc	NARA II Comp	CAT	WISC	RT Caus	RT Stat	Corr Caus	Corr Stat
Chron Age	0.51***	0.53***	0.44***	0.28	0.51***	-0.54***	-0.35*	0.08	-0.25
TOWRE		0.71***	0.64***	0.20	-0.11	-0.47**	-0.24	-0.02	0.00
NARA II	0.65***		0.83***	0.15	0.34*	-0.55***	-0.16	0.07	0.11
Acc									
NARA II	0.54***	0.79***		0.16	0.12	-0.53***	-0.13	0.18	0.16
Comp									
CAT	0.05	-0.03	0.03		0.40*	-0.26	-0.18	0.32	0.06
WISC	0.02	0.14	0.22	0.31		-0.45**	0.00	0.51***	0.05
RT Caus	-0.37*	-0.39**	-0.38*	-0.13	-0.21		0.53***	-0.24	0.07
RT Stat	-0.14	-0.04	-0.01	-0.09	0.18	0.48**		0.07	0.15
Corr Caus	0.13	0.13	0.22	0.32	0.49**	-0.19	0.09		0.20
Corr Stat	0.20	0.28	0.33	0.13	0.33	-0.05	0.10	0.32	

Table 6.8 Zero-order correlations between response times, response accuracy, experimental comprehension scores and performance on the standardised measures and partial correlations (below diagonal) between these variables when age is controlled for. * p<.05, ** p<.01, ***p<.001

The zero-order correlations revealed that response times to causal test sentences were significantly related to both NARA II reading accuracy and comprehension, chronological age, WISC vocabulary and performance on the TOWRE. In contrast, response times to static test sentences were only significantly related to chronological age. The number of correct responses made to static test sentences did not correlate with any of the measures of interest, while the number of correct responses to causal test sentences correlated with performance on the WISC vocabulary task, indicating that responding accurately to these sentences required a good level of vocabulary.

Given the strong relationship between chronological age and all of the standardised measures, a partial correlation was performed to ascertain whether the relationships between causal test sentence RT and NARA II accuracy, NARA II comprehension and TOWRE performance were mediated by age. Due to the fact that Less-Skilled children have already been shown to be slower in responding than Skilled children, it was felt that controlling for age also controlled for differences in processing speed. Therefore, performance on the TOWRE was not partialled out on this occasion. When age was controlled, the correlations between response times to causal test sentences and NARA II accuracy, NARA II comprehension and TOWRE remained significant. Response times to static test sentences failed to correlate significantly with any of the measures of interest. Response accuracy for causal test sentences continued to correlate with WISC vocabulary.

Multiple regression analyses were conducted to examine the relationship between response times to causal and static test sentences, with NARA II reading comprehension as the dependent variable. Chronological age was entered on the first step to control for age-related differences in reading comprehension. WISC vocabulary was entered on step 2 to control for the influence of vocabulary knowledge on reading comprehension. The predictor variables of interest in these analyses were response time to causal and static test sentences, and accuracy of responding to causal and static test sentences. These four variables were entered individually on step 3 in four separate analyses. The results can be seen in Table 6.9

<i>NARA II Reading Comprehension</i>			
Step	Variable	R ² Change	P
1	Age	0.19	0.01
2	WISC Vocabulary	0.04	0.19
3	RT Causal	0.10	0.03
3	RT Static	0.00	0.91
3	Causal Accuracy	0.00	0.74
3	Static Accuracy	0.06	0.11

Table 6.9 Multiple regression analyses with NARA II comprehension as the dependent variable and Chronological Age, WISC vocabulary, RT Causal, RT Static, Causal Accuracy and Static Accuracy as the predictor variables.

When entered on step 1, Chronological Age accounted for 19% of the variance in NARA II comprehension performance ($p < .01$). WISC vocabulary accounted for 4 % of the variance when entered on step 2, which was not significant. When entered on step 3, response time to causal test sentences accounted for a significant 10% of the variance in performance on the NARA II comprehension measure. Response times to static test sentences, accuracy in responding to causal test sentences and accuracy in responding to static test sentences all failed to account for a significant proportion of the variance in comprehension performance when entered on step 3.

6.2.3 Discussion

The overall results of the response time and accuracy analysis replicate the findings of experiment 3 in that Causal test sentences were responded to faster and more accurately than Static test sentences. Moreover, this pattern was observed for both passage types. Fairy Causal test sentences were responded to faster and more accurately than Fairy Static test sentences. Real Causal test sentences were responded to faster and more accurately than Real Static test sentences. Furthermore, the finding that for the Skilled group's responses to causal test sentences, a greater degree of accuracy was associated with faster response times adds further support to the assumption that these inferences are made on-line automatically during reading in children with well-developed reading skills. As in

studies 2 and 3, this interpretation of the results is not unequivocal. However, as in study 2, given the congruence of these results with previous research into the time-course of inference generation, the conclusion that in the present study, causal coherence inferences were drawn on-line is a plausible one. Finally, in line with the findings of experiment 3, responses were faster overall to Real test sentences than to Fairy test sentences indicating that the belief bias effect found in deductive reasoning is present when generating inferences from text, even when those inferences are necessary for comprehension.

A developmental pattern was observed such that children in the Skilled group responded faster and more accurately overall than children in the Less-Skilled group. This suggests that children with less well developed reading skills are less efficient at drawing both causal and static inferences during reading than children with good reading ability. However, children in the Less-Skilled group demonstrated the same pattern of inference generation as children in the Skilled group such that responses were faster to Real and Fairy Causal test sentences than to their Static counterparts. Thus, children as young as 7-years old find it easier to draw causal inferences than elaborative inferences when reading. However, the difference in response time to Static test sentences in the Real and Fairy story conditions was larger for the Less-Skilled than the Skilled reading age group (Less-Skilled 941.63 msec, Skilled 561.01 msec), indicating a greater degree of difficulty in responding to Static test sentences for the Less-Skilled group. Furthermore, there was a trend in the data such that the difference in response time to Real Causal test sentences and Fairy Causal test sentences was greater for the Less-Skilled group than the Skilled group (Less-Skilled 519.48 msec, Skilled 108.38msec). Although not significant, these results suggest that the presentation of counterintuitive information had a greater effect on the Less-Skilled group, suggesting that when generating causal links, the Skilled group were more efficient at suspending disbelief and were able to respond with equal speed to Real and Fairy Causal test sentences. In contrast, responses of the Less-Skilled children were somewhat inhibited by the Fairy Causal test sentences. Response times of this group for Fairy Causal test sentences were slower than for Real Static test sentences, indicating that the Less-Skilled children did not draw causal inferences on-line when reading Fairy stories, but instead were generating the inferences when presented with the sentence judgement task.

Taken together the results so far suggest that a form of belief bias effect does extend to inference generation from text, although there is some evidence to suggest that children with well-developed reading skills can overcome this effect if causal links are required. There is some evidence of a developmental progression such that the children with reading ages of 10 years and over become more efficient at dealing with inconsistent information and accept it more readily as part of their mental representation. In contrast, children with less well developed reading skills were unable to process information that was inconsistent with real world knowledge and as a result, they failed to draw causal links on-line when reading fairy stories. However, this suggestion is not fully supported by the present findings and requires further investigation.

The results of the zero-order and partial correlations suggested that causal inference generation shared a relationship with reading comprehension that was not mediated by age and that this relationship was not reflected in that between static inference generation and reading comprehension. Multiple regression analyses confirmed this assumption. Causal test sentence response time was a reliable predictor of reading comprehension after age and vocabulary were controlled. Conversely static inference response time did not predict reading comprehension. Neither causal test sentence accuracy nor static test sentence accuracy significantly predicted reading comprehension. Nonetheless, these results suggest that the ability to generate causal inferences is an underlying factor in the development of reading comprehension.

The finding that causal test sentence response time was a significant predictor of reading comprehension while static test sentence response time was not is clearly in contrast to the results of study 3 in which the opposite pattern was found. One possible explanation of these findings may lie in the changes made to the materials for use in the current study. As outlined in the discussion of Chapter 5, the materials used in study 3 were confounded by the possibility that the critical sentences preceding a static target statement may in some cases have also contained a potential causal inference. The slower response times to static target statements may therefore have reflected the fact that a causal inference had already been drawn during reading. Thus, responses to the static inference test sentences might have been delayed because participants had to suppress this causal inference information before they could make a judgement about the information presented

in the static test sentence. Arguably, therefore, responses to static test sentences in study 3 might have involved a higher degree of comprehension skill than responses to causal test sentences. In the present study the materials were constrained such that a causal inference could not be drawn in the critical sentences preceding a static target statement. Thus, the hypothesised suppression of causal inference information that might have been required in responding to static test sentences in study 3 was no longer required in the present study. However, while this explanation is a plausible one, there is no systematic evidence available to support it. It may be the case that the discrepancy in results between studies 3 and 4 was simply due to noise in the data, possibly as a result of the larger age range of the participants included in study 4. A replication of the results of study 4 using the modified materials would help to clarify this issue.

To summarise, the results of the current study suggest that children with reading ages between 7 and 13 years present the same pattern of inference generation when reading stories based in real world knowledge. Both reading age groups studied in experiment 4 generated causal inferences faster than static inferences during reading. While speed and accuracy increased with reading age, the pattern of responding did not change, indicating that even the youngest children seemed to recognise the necessity of causal relationships in maintaining discourse cohesion, thus attempting to resolve the coherence breaks embedded in the stories with speed and efficiency. In-line with the findings of other investigations into children's inference generation (e.g. Casteel & Simpson, 1991; Casteel, 1993), this finding could be interpreted as indicating that children drew coherence inferences on-line during reading. Conversely, all the children seemed to generate static inferences only when required i.e. when presented with the test sentence, suggesting that none of the children found this level of information necessary for maintaining coherence when reading the text. It should be noted that, as in experiment 2, it is possible that while coherence inferences were certainly drawn faster than elaborative inferences in this study, they may still have been drawn off-line. However, while the results of the current study do not prove conclusively that coherence inferences were drawn on-line during reading, they certainly suggest that this is a reasonable conclusion.

Furthermore, while there was a slight increase in accuracy scores with increased reading age, both groups of children were responding to test sentences at a comparable

level. This suggests that both reading age groups were to some extent able to set aside their empirical knowledge to generate causal inferences based on false premise information. An alternative explanation is that the test sentences presented in the present study were too easy, enabling the Less-Skilled readers to respond with the same degree of accuracy as the Skilled readers. However, Less-Skilled children demonstrated response times to Fairy Causal test sentences that were comparable to those for Real Static test sentences suggesting that, when reading stories containing information that is inconsistent with real world knowledge, the Less-Skilled group did not draw causal inferences on-line, but instead waited until the inference was required i.e. when presented with the sentence judgement task. Therefore, the presentation of information in stories that is contrary to our empirical knowledge of the world, does inhibit inference generation during reading. This effect is strongest for children with a reading age below 9 years 11 months, indicating that the development of proficient reading skills may help to overcome the difficulties faced when generating inferences from false premises.

Finally, the results presented here suggest the ability to generate causal inferences is a fundamental skill underlying reading comprehension while generation of static inferences is not related to reading comprehension skill.

Chapter 7

Inferential Skills of Children with Specific Comprehension Difficulties.

7.1 Introduction

The work reviewed in chapter 1.1. suggests that children with specific comprehension difficulties have problems with inferential processing. For example, the work of Oakhill and colleagues demonstrated that less-skilled comprehenders do not actively integrate information across sentences (Oakhill, 1982), do not instantiate specific examples of category nouns (Oakhill, 1983) and are less efficient in their use of lexical cues than skilled comprehenders (Oakhill & Yuill, 1986; Yuill & Oakhill, 1988a). Furthermore, on standard question answering tasks in which they are asked to read a passage and answer related questions, less-skilled comprehenders perform significantly less well than skilled comprehenders when the information needed to answer the question has to be inferred from the passage. However, when the questions tap information explicitly stated in the passage, less-skilled comprehenders perform at a level comparable to age-related peers (Cain & Oakhill, 1999; Oakhill, 1984). Finally, Yuill & Oakhill (1988b) demonstrated that training in inference awareness dramatically improved the reading comprehension of children with comprehension difficulties. Taken together these results indicate that children with specific comprehension difficulties do not engage in the inferential processing activities used by RA Controls to form coherent mental representations of text.

One limitation of this research is its focus on the use of lexical cues and cohesive devices to maintain textual coherence (e.g. Oakhill, 1982; 1983; Yuill & Oakhill, 1988a). As discussed in chapter 1.2, lexical cues, such as anaphora and pronoun resolution, which are used to maintain textual cohesion are not considered examples of inference generation by all researchers (Fulcher, 1989; Graesser, Singer, & Trabasso, 1994). Second, studies that have aimed to investigate knowledge-based inferences have tended not to differentiate between different inference types. Oakhill (1984) gave children passages to read followed by two types of question; literal questions tapping information explicit in the passage, and inferential questions tapping information to be inferred from the text. The inferential questions varied in terms of the type of information required to answer the questions; while some of the questions required the generation of a causal link, other questions were not causal in nature, requiring inferences about the emotional state of a story protagonist for example. An

open question is whether the problems of children with specific reading comprehension difficulties may in fact depend upon inference type.

A further consideration is that the majority of studies tend to use narrative passages containing information that is based in real-world knowledge. As discussed in chapter 5, children can be influenced by the belief bias effect and may find it difficult to make inferences that are based on information that is empirically inconsistent. Cain, Oakhill, Barnes and Bryant (2001) addressed this issue by comparing the performance of skilled and less-skilled comprehenders on a task designed to assess their inference making ability in a situation that was not based in real life experience. The children were first taught twelve facts about a planet called Gan. These facts comprised the knowledge base and the children had to have perfect recall of all twelve facts before moving on to the next stage of the experiment. In stage 2, the children were read a short story about Gan that was divided into 6 episodes. Following the completion of each episode, the children were asked 4 questions; an elaborative inference question, a coherence inference question, a literal information question and a question which asked the meaning of a novel simile included in the story episode e.g. 'What does "Dack was like a boxer who had lost a fight" mean?' Less-skilled comprehenders performed at a lower level than skilled comprehenders on all but the simile question. Thus, it appears that overall comprehension ability rather than story content influences the generation of inferences. Skilled comprehenders were able to set aside their real world knowledge in order to generate the inferences required.

However, in this study, rather than setting aside their empirical knowledge of life on earth, the children were asked to interpret the story from the perspective of their newly acquired knowledge i.e. from the perspective of life on Gan. Cain et al. (2001) did not in this study make a comparison between the Gan story and an alternative passage based in real-world knowledge. It is possible therefore that group differences in performance on the inference questions may have varied had the information in the stories been true to real world knowledge. The experiments described in chapter 5 and 6 of this thesis included a comparison of stories based in real-world knowledge and those that were inconsistent with real-world knowledge. The results showed that children with normal reading skills were able to overcome the belief bias effect and make inferences based on inconsistent information. The results of Cain et al. (2001) suggest that children with comprehension difficulties may be more affected by the

belief bias effect and therefore be unable to suspend disbelief in order to generate inferences. However, it is not clear whether children with comprehension difficulties would be able to generate inferences consistent with real world knowledge.

Aim of present study: The present study investigated the inference processing skills of less-skilled comprehenders using the paradigm employed in experiments 3, 4 and 5 of this thesis. The main aim was to extend previous findings by using an on-line measure of inference generation including materials based on both real world knowledge and empirically false information, while concentrating on two specific inference types; causal and static inferences. It was predicted that children with comprehension difficulties would be slower and less accurate in responding to test sentences requiring a coherence inference, but that there would be no difference between groups in speed or accuracy of responding to elaborative inference test sentences. It was also predicted that children with comprehension difficulties would be less able to suspend disbelief in order to generate inferences, with responses to test sentences based on information inconsistent with real world knowledge being slower and less accurate than those based in real-world knowledge.

7.2 Experiment 5

7.2.1 Method

a) Design: A mixed factorial design was used with the between-subjects factor Comprehension Group (Control Group and Less-Skilled Comprehenders) and the within-subjects factors Passage Type (Real and Fairy) and Inference Type (Causal Link and Static Property). The within-subjects factors combined to form four experimental conditions; Real Causal; Real Static; Fairy Causal and Fairy Static. However, initial analyses revealed that Passage Type had no effect on speed or accuracy of responding and this factor was therefore dropped from the design. Each child completed twelve trials in each condition (Causal and Static) and the order of the trials was randomised. A battery of standardised tests was also administered to assess each child's reading and language skills.

a) Participants: 10 less-skilled comprehenders and 10 reading age match controls were selected to take part in this study. The less-skilled comprehenders all had age appropriate reading skills, as measured by the Neale Analysis of Reading Ability (NARA II; Form 1;(Neale, 1989) and the Test of Word Reading Efficiency (TOWRE;

(Torgeson, Wagner, & Rashotte, 1999), but their NARA II comprehension age was at least 9 months below their chronological age and at least one year below their NARA II reading accuracy age. The control group were matched with the less-skilled comprehenders on the basis of NARA II reading accuracy age and chronological age. The children were also given the WISC III vocabulary subscale (Third Edition UK; WISC III;(Wechsler, 1992) as a measure of vocabulary knowledge. Performance on these standardised tests can be seen in Table 7.1.

	Less-Skilled Comprehenders (n = 10)	RA Controls (n = 10)
Chronological Age (mths)	111.79 (12.18)	112.31 (11.54)
NARA II Accuracy Age (mths)	115.48 (13.61)	115.88 (14.81)
NARA II Comprehension Age (mths)	94.15 (7.69)	114.96 (10.49)
TOWRE Sight Word ^a	111.50 (10.70)	110.30 (4.08)
TOWRE Phonemic Decoding ^a	109.00 (11.90)	110.60 (4.55)
TOWRE Total Efficiency ^a	112.30 (11.64)	112.00 (3.94)
WISC Vocab ^b	7 (3.09)	10.80 (2.20)

Table 7.1 Mean performance of Less-Skilled Comprehenders and RA Controls on tests of reading and language ability (standard deviations in parentheses; a) standard score; b) scaled score; average 10 (SD 3).

A series of t-tests revealed significant differences between groups for NARA II Comprehension age ($t(18) = 5.06, p < .001$) and WISC III vocabulary ($t(18) = 3.17, p < .01$). No other differences were significant (Chronological Age $t(18) = 0.10, ns$; NARA II Accuracy age $t(18) = 0.06, ns$; TOWRE Sight Word Reading $t(18) = 0.33, ns$; TOWRE Phonemic Decoding $t(18) = 0.40, ns$; TOWRE Total Efficiency $t(18) = 0.08, ns$). Therefore, while the groups differed in terms of comprehension skill and vocabulary, there was no difference in terms of decoding ability. Any subsequent differences found between groups in terms of inference generation can therefore not be attributed to differences in decoding skills.

c) Materials: The computer-based inference generation task used in experiment 4 was used in the present study. Children read a series of passages presented on the

screen. At the end of each passage, children were required to judge whether a target sentence was true or false, using a key-press response on the computer keyboard. Each target sentence related to specific information embedded in the passage. Three types of test sentence were used; sentences that relayed information explicitly stated in the text, sentences that represented the mediating idea underlying a causal inference required for textual coherence, or sentences that represented a possible elaboration from the text.

d) Procedure: Following the initial screening phase, the children were seen in two individual sessions lasting approximately half an hour each. In session one, the first half of the computer-based task was completed. In session 2, the children completed part 2 of the inference-generation task and were given the WISC vocabulary test and the TOWRE.

7.2.2 Results

a) Response times and error rates to test sentences following experimental passages: Due to the lack of a significant difference in response times between groups receiving different forms of the test in the previous studies reported in this thesis, the data from the present study was collapsed across List for the purpose of analysis. Subject's response data in each of the two conditions (causal inference and static inference) was examined for outliers. Outliers were defined as any response time 2 standard deviations away from the mean response time of each subject in each condition and would be replaced with the subjects mean response time for that condition; 5.3% of responses were replaced in this way. Analysis of the response time data was performed on correct responses only.

	Response Times		No. Correct Responses	
	Causal	Static	Causal	Static
RA Controls (N = 10)	3970.44 (674.07)	4464.71 (941.30)	10.4 (0.8)	6.4 (4.27)
Less-Skilled Comprehenders (N=10)	4528.73 (1433.80)	5014.85 (1514.70)	8.2 (2.15)	6.5 (2.42)

Table 7.2 Mean response time and number of correct responses (max 12) to target sentences for Less-Skilled Comprehenders and RA controls (standard deviations in parentheses).

The mean response times and number of correct responses for each target sentence are shown in Table 7.2. These data were analysed using a 2x2 ANOVA with the between subjects factor of Group (RA Controls and Less-Skilled Comprehenders) and the within subjects factor of Inference Type (Causal and Static). Analysis of the response time data revealed a main effect of Inference Type ($F_1(1,17) = 7.32, p < .05$; $F_2(1,11) = 4.73, p = .052$) such that Causal test sentences were responded to faster than Static test sentences. A main effect of Group was revealed by items but not by subjects ($F_1(1,17) = 1.21, ns$; $F_2(1,11) = 13.51, p < .01$) such that Skilled comprehenders were faster to respond overall than Less-Skilled comprehenders. A significant Group by Inference interaction was found by items ($F_2(1,11) = 9.78, p = .01$) but not by subjects ($F_1(1,17) = 0.02, ns$). Post hoc Tukeys HSD of the by-items data revealed that RA Controls did not show the expected response time advantage for Causal over Static test sentences while Less-Skilled comprehenders responses to Causal test sentences were significantly faster than their responses to Static test sentences. RA Control's responses to Causal and Static test sentences were faster than Less-Skilled Comprehenders responses to Static sentences.

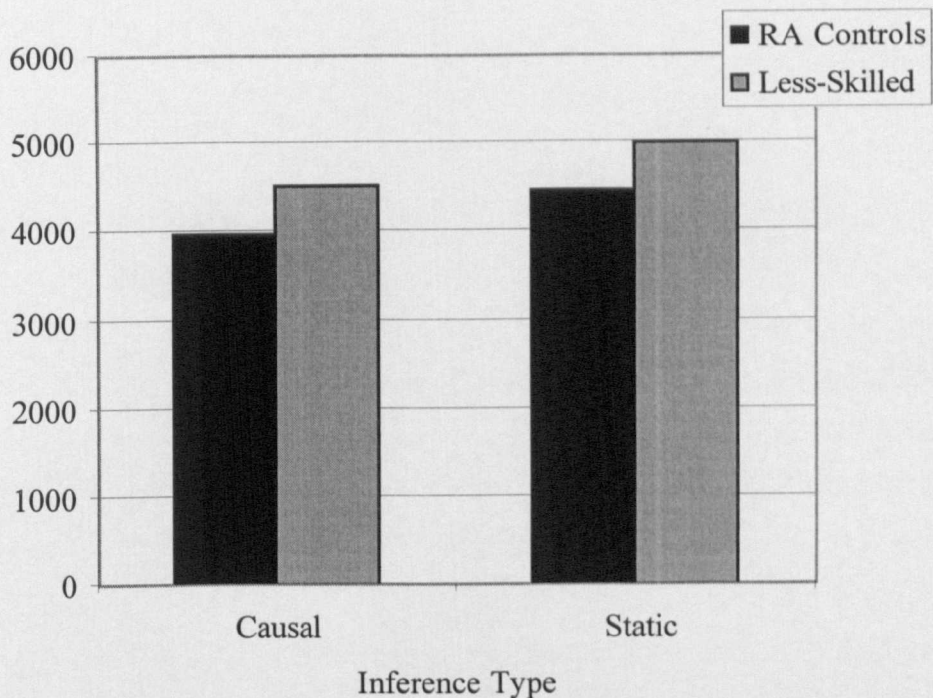


Figure 7.1 Mean response time of Less-Skilled Comprehenders and RA Controls to Causal and Static test sentences.

This pattern of results is surprising as it suggests that the time course of inference generation in the Normal Reader group is the same for both Causal and Static inferences. However, this pattern was only found in the by-items data suggesting that it rests on the nature of the materials. The by-items data was examined for any anomalous response times and it was found that for one item, RA Controls were responding 1.5 standard deviations from the mean response time in both the Causal and Static condition. The mean response time for this item in both the Causal and Static condition was replaced with the overall mean response time for items in each condition. The data were then reanalysed using a series of t-tests which revealed that both groups of children were faster when responding to Causal sentences than to Static sentences (RA Controls $t(22) = 2.06, p < .05$; Less-Skilled Comprehenders $t(22) = 2.64, p < .01$) reflecting the pattern of inference generation consistently found in this thesis. There was no difference in response times to Causal test sentences between groups ($t(22) = 1.58, ns$). However, RA Controls were faster than Less-Skilled Comprehenders at responding to Static test sentences ($t(22) = 3.10, p < .01$). The by-subjects data revealed the same pattern of performance although the differences failed to reach significance.

Analysis of the accuracy data revealed a main effect of Inference Type ($F_1(1,18) = 16.27, p < .01$; $F_2(1,11) = 80.08, p < .01$) such that more correct responses were given to Causal than to Static test sentences. A Group by Inference interaction was found in the by items analysis ($F_2(1,11) = 16.33, p < .05$) but not in the by subjects analysis ($F_1(1, 18) = 3.32, ns$). Post hoc Tukeys HSD revealed that the responses of RA Controls to Causal test sentences were significantly more accurate than to Static test sentences while the responses of Less-Skilled Comprehenders to Causal test sentences were no more accurate than to Static test sentences (HSD $p < .01$). RA Controls responses to Causal test sentences were also more accurate than Less-Skilled Comprehenders responses to Static test sentences (HSD $p < .01$). Less-Skilled Comprehenders responses to Causal test sentences were more accurate than RA Controls responses to Static test sentences (HSD $p < .05$). No other differences were significant. Thus, in line with the predictions of this study, there was a tendency for Less-Skilled Comprehenders to be less accurate in their responses to Causal test sentences than RA Controls indicating some difficulties in on-line inference generation which is not reflected in response time. However, the Less-Skilled group were more accurate in responding to Causal sentences than to Static test sentences, although this difference was not significant. Similarly, they were significantly more accurate in responding to Causal test sentences than the RA Controls were in responding to Static test sentences. Although not significant, this pattern was also found in the by-subjects data. Therefore, while they were slightly less accurate than RA Controls when responding to Causal test sentences, they showed the same pattern of performance as RA Controls.

The results of the response time and accuracy data were in contrast to our predictions in that Less-Skilled comprehenders performed at a comparable level to Skilled comprehenders on the on-line inference generation task. In order to confirm these results, a second set of analyses were conducted comparing the less-skilled comprehenders with a larger normative sample ($N = 44$) comprising children who had been tested for study 4 and the RA controls from the present study. Analyses of both the response time and accuracy data revealed the same pattern of results. Less-skilled comprehenders performed at a comparable level to the normative control group when responding to causal and static test sentences.

b) Performance on experimental comprehension questions: Children's scores on the comprehension questions asked after each passage are shown in Table 7.3. A t-test was used to compare the comprehension scores of RA Controls and Less-Skilled Comprehenders. Results revealed a significant difference such that RA Controls achieved a higher score on the comprehension questions than Less-Skilled Comprehenders ($t(18) = 3.55, p < .01$).

	Mean	Standard Deviation
RA Controls (N = 10)	19.1	3.00
Less-Skilled Comprehenders (N = 10)	13.1	4.43

Table 7.3 Mean performance of Less-Skilled Comprehenders and RA controls on experimental comprehension questions.

c) Regression analyses: The results so far suggest that Less-Skilled comprehenders perform in a similar manner to RA Controls when generating inferences on-line during reading. In order to investigate this effect more thoroughly, multiple regressions were run to assess the amount of variance the factor Group could account for in response time to Causal test sentences when verbal skills and speed of responding was controlled for. If the Less-Skilled Comprehender group are performing at a comparable level to the RA Controls, there should be no group difference in performance on the on-line task. Therefore, the factor Group should not account for significant unique variance in response time to Causal test sentences after individual differences in response time are accounted for by controlling for Static test sentence response time.

Step		<i>Dependent Variable</i>	
		R ² Change	P
1	WISC III Vocabulary	0.27	<.05
2	Static Response Time	0.37	<.001
3	Group	0.00	0.92

Table 7.4 Multiple regressions with Causal Sentence RT as the dependent variable and WISC vocabulary, Static Sentence RT and Group as the predictor variables (N=20).

The results shown in Table 7.4 indicate that with Causal test sentence response time as the dependent variable, WISC vocabulary accounts for 27% of the variance in Causal test sentence response time. Static test sentence response time accounts for 37% of the variance when entered on step 2 which is significant. When entered on step 3, Group accounts for none of the variance in Causal test sentence response time. With Static test sentence response time as the dependent variable, WISC vocabulary accounts for 23% of the variance in Static test sentence response time when entered on step 1. When entered on step 2, Causal response time accounts for a significant 39% of the variance. The factor Group does not account for any variance in Causal sentence response time when entered on step 3.

7.2.3 Discussion

The present study investigated the inference processing abilities of children with specific comprehension difficulties. At the outset of the study it was intended that Skilled and Less-Skilled comprehenders would be compared in terms of their ability to generate inferences from stories based in real world knowledge and from fairy tales. Indeed, specific predictions were made such that children with specific comprehension difficulties would be less able to suspend disbelief in order to generate inferences based on false premises. Initial analyses of the response time data failed to support this assumption such that response times were not affected by passage type. The data were therefore collapsed across passage type for inclusion in this thesis. However, while the results suggested that Less-Skilled comprehenders were no more affected by counterintuitive information than Skilled comprehenders, further investigation is needed to clarify this issue. Participants were presented with 12 experimental stories in this study, 6 of which were based in real world knowledge, and 6 of which were fairy tales. Given the small number of participants in each group, it may be the case that analysis of response times across 6 items was not a sensitive enough measure with which to assess differences in inference generation across passage type. As such, the ability of Less-Skilled comprehenders to generate inferences based on false premises is still an open question. It would be desirable to investigate this issue with a larger pool of items of each passage type.

As stated above, the results reported in this thesis were collapsed across passage type. Analysis of the response time data revealed no significant difference between

Less-Skilled comprehenders and RA Controls in response times to causal and static inference test sentences. The paradigm was sensitive however, as shown by the finding that both groups of children showed a pattern of inference generation consistent with the other studies reported in this thesis, such that responses were faster to causal than to static inferences. Furthermore, both RA Controls and less-skilled comprehenders were more accurate when responding to causal test sentences than when responding to static test sentences. Thus, it would appear that both RA Controls and less-skilled comprehenders generated causal inferences on-line while static inferences were generated at a later point in time. This pattern of results was replicated in a second set of analyses conducted on the data from a cohort of 12 less-skilled comprehenders and a larger control group selected from previous experiments reported in this thesis.

The present findings are surprising given the results of previous research into the inference processing skills of less-skilled comprehenders (e.g. Cain & Oakhill, 1999; Cain, Oakhill & Barnes, 1999; Oakhill, 1982; 1983; 1984; Yuill & Oakhill, 1988a) which suggest that children with comprehension difficulties have problems with inference generation. However, earlier studies have used off-line tasks in which children are required to read a passage and answer a series of questions following completion of the passage as a whole. Conversely, the present study uses an on-line paradigm in which the measure of inference generation is taken immediately after presentation of the inference information. The difference in performance might possibly lie in the inability of less-skilled comprehenders to retain information over a period of time. Unfortunately, the children in this study were not assessed on a working memory task and therefore, this interpretation of the results cannot be fully investigated. However, this assumption is supported by previous research on the working memory of less-skilled comprehenders that revealed deficits in verbal working memory (Nation, Adams, Bowyer-Crane, & Snowling, 1999; Yuill, Oakhill, & Parkin, 1989). In support of this hypothesis, the results of the present study also revealed that, despite the comparable performance of both groups of children on the on-line inference task, the less-skilled comprehenders performed significantly worse than both the large sample of RA Controls and the chronological age-match group when answering the off-line comprehension questions presented after completion of each on-line passage. One problem with this interpretation is that, if less-skilled comprehenders have difficulties with off-line tasks such as answering questions about the text they have just read, one might expect them to have performed poorly in the Static test sentence condition

relative to RA Controls since the inferences being tapped by static test sentences are assumed to be generated off-line. However, the performance of less-skilled comprehenders was comparable to that of RA Controls in both the Causal and Static conditions. However, while static inferences are assumed to be made off-line, the working memory demands in responding to the static inference test sentences in this task are less than would be required in a traditional question/answer paradigm. Thus, it may be that reducing the working memory load enables less-skilled comprehenders to perform as RA Controls in both on-line and off-line inference generation tasks.

An alternative explanation for these findings may be that the small number of less-skilled comprehenders included in this study meant that there was insufficient statistical power to detect group differences. Inclusion of a larger sample of less-skilled comprehenders in further investigations may reveal differences in on-line inference generation which were not apparent in this study. However, examination of the results in terms of power indicated that the factor Group accounted for little variance in either response time or accuracy of responding. Analysis of the response time data revealed that the factor Group had a small effect size ($\eta^2 = 0.067$) and low statistical power (Power = 0.179). In order to increase the statistical power of these results to a satisfactory level, a sample of 60 participants in each group would be needed. Similarly, analysis of the accuracy data revealed a small effect size ($\eta^2 = 0.025$) and low statistical power (Power = 0.104). A sample of 400 participants in each group would be needed to increase the statistical power of the accuracy data to a satisfactory level. In contrast, analysis of the factor Inference Type revealed a strong effect size in the response time data ($\eta^2 = 0.301$) and statistical power close to the level 0.80 advised by Cohen (1988) (Power = 0.72). Similarly, analysis of the accuracy data revealed a strong effect size for Inference Type ($\eta^2 = 0.475$) and sufficient statistical power (Power = 0.968). This suggests that the lack of any significant differences between groups could not be attributed to the small sample size. A significant result could only be obtained if the size of the sample used was in excess of that usually included in experimental studies of this nature. The results of such a study might be deemed unreliable due to the degree of individual variation inherent in such large samples. This finding supports the interpretation that less-skilled comprehenders perform in a similar manner to RA Controls when generating inferences on-line.

Further support for the notion that less-skilled comprehenders perform in a similar manner to RA Controls during on-line tasks comes from a recent study by Nation, Marshall and Altmann (2002). Nation et al. (2002) recorded children's eye movements as they listened to a sentence describing one of four objects presented in a visual scene. Each sentence contained the name of the target in the sentence-final position. However, in the supportive sentence condition, the verb contained in the sentence was highly predictive of the target object, while in the neutral condition, the verb contained in the sentence was equally as predictive of all four objects in the visual array. For example, the children were shown a visual array containing a picture of a hat, a dog, a lipstick and a cake. The neutral sentence accompanying that array was "Jane watched her mother choose the cake" where the verb 'choose' is equally as predictive of 'cake' as it is of 'hat', 'dog' and 'lipstick'. In the supportive sentence, "Jane watched her mother eat the cake", the verb 'eat' is highly related to the noun 'cake' but not to the nouns 'dog', 'hat' or 'lipstick'. In line with previous research (Altmann & Kamide, 1999), Nation et al. (2002) found that children made anticipatory eye movements to the target object in the supportive sentence condition but not in the neutral sentence condition. This pattern of eye movements was seen in both RA Controls and poor comprehenders. Thus, it would appear that poor comprehenders are sensitive to linguistic constraints during on-line processing.

The findings of Nation et al. (2002) add support to the findings of the current study, and suggest that the language difficulties of poor comprehenders may be further characterised by a distinction between on-line and off-line processing skills. However, it should be noted that the poor comprehenders in Nation et al.'s study showed atypical eye movement patterns following the initial anticipatory eye movements. Thus, while poor comprehenders may have normal sensitivity to linguistic constraints during initial language processing, subsequent processing may be atypical.

To summarise, the results discussed here suggest that, contrary to the predictions made at the outset of this study, less-skilled comprehenders generate causal inferences on-line during reading. These results are incongruent with previous research suggesting that less-skilled comprehenders do not actively engage in constructive processing during reading. However, the difference in research findings may be attributable to the moderate working memory demands of this task compared to more traditional question and answer paradigms.

Chapter 8

General Discussion

8.1 Summary of Findings

The aim of this thesis was to investigate inference generation during children's reading. The point of departure was an exploratory investigation into the relationship between reading comprehension and inference generation. A qualitative analysis was conducted on a set of standardised tests in order to assess the degree to which inferential ability was required to meet the task demands. Two measures of reading comprehension and a test of mathematical reasoning were included in this analysis. It was found that all three tests required some degree of inference generation although the types of inferences required by each of the three tests varied. Both tests of reading comprehension required the use of cohesive devices, the generation of knowledge-based coherence inferences and the generation of elaborative inferences. However, performance on the NARA II (Neale, 1989) reading comprehension test was more reliant on inference generation than performance on the WORD (Wechsler, 1990) reading comprehension sub-test, which in turn was heavily biased towards the retention of literal information. The WOND (Wechsler, 1996) test of mathematical reasoning also required the use of cohesive devices and the use of knowledge-based coherence inferences, but did not require the generation of elaborative inferences.

These standardised tests were administered to a group of normal readers aged between 6- and 11-years. The results of a zero-order correlation revealed that both measures of reading comprehension correlated highly with each other and with the WOND test of mathematical reasoning. When age and reading accuracy were controlled for, the correlation between the measures of reading comprehension remained significant. However, while the correlation between mathematical reasoning and NARA II comprehension remained significant, the correlation between mathematical reasoning and WORD reading comprehension did not. Nevertheless, a series of multiple regression analyses suggested that mathematical reasoning was not a unique predictor of NARA II reading comprehension and similarly, that reading comprehension was not a significant predictor of mathematical reasoning. A skilled and less-skilled comprehension group were selected from the original sample of participants and their performance on the tests of reading comprehension was compared. While the skilled group performed better overall, the less-skilled group performed better on the

WORD than on the NARA II test of comprehension, with scores on the WORD being in the average range. Finally, a detailed analysis was performed on the children's responses to the different question types identified in each test. The results showed that less-skilled comprehenders performed at a comparable level to the skilled comprehenders on literal and cohesive questions but were impaired relative to the skilled group in their performance on knowledge-based coherence inference and elaborative inference questions.

Experiment 2 was designed to investigate the effect of text format on inference generation. Participants were given a computer-based passage reading and sentence judgement task in which they had to read short passages and then respond true or false to test sentences presented at the end of the passages. The passages were written in either an expository text format or a narrative text format. The test sentences represented information that was required to generate a causal link in the preceding passage or a possible elaboration from the passage. Response time and accuracy of responding to test sentences was recorded as a measure of inference generation. The findings indicated that causal inferences were generated faster than elaborative inferences, which were generated at a later point in time. These results are consistent with previous studies of inference generation, which suggest that causal inferences are made on-line and elaborative inferences are made off-line. Moreover, the results showed that text format did not affect inference generation; the response time and accuracy data were consistent across text formats. Finally, multiple regression analysis suggested that the ability to generate inferences was a significant predictor of reading comprehension ability.

Experiment 3 investigated the effect of content on inference generation in terms of the ability to suspend disbelief in order to generate inferences from text. The children were presented with a passage reading and sentence judgement task as in Experiment 2. However, the passages and test sentences used in experiment 3 were either based on real-world knowledge or on information that was inconsistent with real-world knowledge i.e. fairy tales. Analysis of the response time and accuracy data showed that, even when information was inconsistent with real-world knowledge, children could generate inferences required to maintain causal coherence, although information inconsistent with real-world knowledge did impede inference generation to some extent.

Experiment 4 extended the results of experiment 3 by looking at the inference generation of children during reading from a developmental perspective using improved materials. The results of experiment 4 replicated those of experiment 3 such that children were faster and more accurate when responding to causal test sentences than to static test sentences. These results indicated as in experiments 2 and 3, that children generated causal inferences but not static inferences during reading. As was the case in experiment 3, this pattern of results was the same for stories based in real-world knowledge and for fairy stories. Moreover, while responses to causal test sentences were faster than to static test sentences in both real and fairy story conditions, responses were slightly slower to fairy story test sentences in the causal condition. This finding suggests that even when the inference is necessary for comprehension, readers can still be misled by the presence of counterintuitive information in stories. A developmental trend was found, such that while both groups of children showed a tendency to draw causal inferences faster than elaborative inferences, less-skilled readers were slower and less accurate than skilled readers. Finally, less-skilled readers were less able to suspend disbelief than skilled readers.

Experiment 5 investigated the inferential skills of children with specific comprehension difficulties. A group of less-skilled comprehenders was compared with both a small well-matched group of reading age-matched controls and a heterogeneous group of 44 normal readers on the inference generation task used in experiment 4. It was predicted that the less-skilled comprehenders would be slower and less accurate in responding to causal inference test sentences than the reading age-matched controls and the normal readers, given their previously reported difficulties with constructive processing. In contrast to this prediction, there was no difference between groups in the speed and accuracy of responding to causal or static inference test sentences. Like the normal readers, the less-skilled comprehenders drew causal inferences but not elaborative inferences on-line during reading and did so effectively, at least as measured by the response times to causal and static test sentences. However, their performance on the off-line comprehension questions was impaired relative to the reading-age matched control group. Taken together these findings suggest that children with specific comprehension difficulties have particular difficulties with off-line tasks, while their on-line inference processing skills may be relatively intact.

8.2 Discussion of Findings

8.2.1 The Effect of Text Genre, Content and Reading Skill on Inference Generation during Reading.

Three factors thought to influence inference generation were investigated in this thesis; text genre, content and reading skill. The results of experiment 2 suggest that text genre does not have an influence on inference generation. Causal inferences were drawn faster than elaborative inferences when reading both narrative and expository text. Thus, it would seem that when reading both narrative and expository text, readers strive for textual coherence.

These findings are in opposition to many previous investigations into the effect of text format on inference generation (e.g. Britton & Gulgoz, 1991; Millis et al., 1990, 1994; Noordman et al., 1992; Narvaez et al., 1999; Olson et al., 1981) the results of which suggest that inference generation does not take place during reading of expository text. However, the materials used in the present study differ from those used in previous investigations. For example, the passages used in the current study were written for use with children and were designed to tap inference concepts which were familiar to them. Thus, the passages were simple and accessible. The study attempted to directly compare narrative and expository texts, and moreover, unlike previous studies (e.g. Narvaez et al., 1999) the narrative and expository texts used in this study were matched for content such that for each item, a narrative and expository version of the same text was created. Other research has attempted to control the materials in this way and has generated similar results to those found in the present study. For example, Singer et al. (1997) and Perrig and Kintsch (1985) showed that when narrative and expository texts are equated in terms of difficulty, the same pattern of inference generation is witnessed in both text formats. Thus, by overcoming some of the methodological confounds of previous research, the findings of the present study provide a clear account of on-line inference generation during reading of narrative and expository text.

The finding that text format has no influence on inference generation finds further support in recent work by Duke (2002) investigating the impact of using diverse genres in the classroom on subsequent reading comprehension. In short, Duke (2002) found that providing children with a large range of text genres at the beginning of Grade

One had no effect on either their overall reading comprehension ability or genre-specific reading comprehension skills at the end of the school year. The results of the present study suggest that inference generation is a reliable predictor of reading comprehension. Therefore, if reading comprehension was affected by experience with different text genres, one might expect that inference generation would also be affected. Thus, the finding of Duke (2002) that experience with different text genres does not affect reading comprehension implies in turn that text genre does not affect inference generation.

There are a number of limitations of this study however. For example, as previously stated the materials used in this study were written to be accessible to children. As such, the inference concepts being tapped may have been accessible without the need to engage in inference generation. Thus, the task may not have been a sensitive enough measure of inference processing ability. Using more complex inference concepts may have revealed differences in inference generation across text formats that were not apparent in the results of this study. Furthermore, as a result of matching the passages for content, the materials did not represent natural texts. The language and structure of the passages was constrained and may have seem contrived during reading. Thus, the results of the study may not represent the on-line inference processes that occur during reading of natural as opposed to experimenter-generated texts. Finally, the expository texts written for use in this study were simply adapted from the narrative texts by removing references to characters, actions, and events, rendering them passive versions of their more dynamic narrative counterparts. Many types of expository text were not included in this study. Hence, whether the degree to which inferences are drawn on-line varies as a function of type of expository text remains an open question. A thorough investigation of inference generation from a variety of expository texts is required to clarify this issue.

The findings of experiments 3 and 4 suggest that children are able to generate inferences from information that is inconsistent with real-world knowledge. A belief bias effect was present such that responses were slower to causal test sentences following stories containing counterintuitive information than to those following stories containing real-world knowledge. However, responses of children in both studies were faster and more accurate to causal test sentences than to static test sentences, irrespective of the content of the preceding passages. The difference in response time to

causal test sentences across conditions may reflect a slight delay in the decision making aspect of the sentence verification task in the fairy story condition. It may be the case that children were less confident in responding to the counterfactual sentences, despite having already generated the inference concept on-line, because the information contained in the test sentences did not agree with the child's knowledge of the world. In other words, it may be that they become vulnerable to a form of the belief bias effect for a brief period at this stage in the task.

However, the fact that responses to fairy causal test sentences were still faster than fairy static test sentences supports the assumption that, at the sentence verification stage of the task, the inference concept is already active in the fairy causal condition due to the generation of the inference concept during the reading phase of the task. This prior activation facilitates the children's responses and enables them to quickly overcome the difficulties they face when presented with counterintuitive information in a narrative context. In the fairy static condition, as in the real static condition, no such inference generation takes place during reading and therefore, the child has no active inference concept in place to aid in the decision making process inherent in the sentence verification task.

The findings of the present study are in line with the results from studies of deductive reasoning suggesting that children as young as 2-years old can solve counterfactual reasoning problems (e.g. Dias & Harris, 1988, 1990; Leever & Harris, 2000; Richards & Sanderson, 1999). However, the findings contrast with those of previous research investigating inference generation during reading where it has been reported that children aged 9-10 years of age have difficulties reasoning from false premises (e.g. Franks, 1996, 1997). A possible reason for the difference in findings is that Franks (1996, 1997) embedded deductive reasoning problems in narrative contexts. Thus, the materials may not have reflected the kinds of inferences commonly required during reading. Rather, the artificial and unfamiliar structure of the passages may have had a detrimental effect on children's ability to maintain coherence which was more pronounced in the counterfactual passages. In contrast, it was intended that the passages used in the present study should reflect the kinds of inferences commonly found in children's stories. While some manipulation was required in order to meet the experimental task demands, such that each story contained a coherence break, the problems did not take on the formal structure of logical inference forms. Furthermore,

the task used by Franks (1996, 1997) was an off-line question and answer task and therefore, the findings may reflect difficulties in completion of the off-line task rather than problems with the on-line inference generation process.

Experiment 4 also identified a developmental trend in the generation of inferences which was reliant not on chronological age but on the development of proficient reading skills. Whether the relationship between reading skill and reasoning ability follows from the increased exposure to print experienced by skilled readers or a higher-level cognitive process that is used in both reading and reasoning is an issue for further investigation. An alternative explanation for the findings of this study and their incongruence with those of Franks (1996, 1997) might be that the inductive inferences used in the present study may have been easier for the children to solve than the deductive inferences used by Franks (1996, 1997). However, this explanation seems unlikely given the findings of Galotti et al. (1997) that children find inductive inference problems more difficult to solve than deductive inference problems.

The ability of children to overcome the belief bias effect in order to generate inferences is plausibly related to their willingness to engage in make-believe play. Harris (2000) reports a series of studies which suggest that when children engage in role-play, they do not maintain an objective viewpoint but rather, they actively enter into an imaginary world and adopt the viewpoint of the character they have chosen to portray. The process of reading can be seen as analogous to pretend play to the extent that the reader is required to enter into the world created by the author. Evidence suggests that adult readers place themselves in the same situation as the story protagonist during reading. For example, Glenberg, Myer and Lindem (1987, experiment 1) presented participants with short paragraphs in which a critical object was either spatially associated or spatially dissociated from the story protagonist. For example, in one story the protagonist either 'put the last flower in his buttonhole' or 'put the last flower in the vase.' Participants were then presented with a recognition task in which they had to state whether the target word (i.e. flower) had occurred in the story. Responses to this recognition task were faster when the protagonist and the object were spatially associated than when they were spatially dissociated. Similar results were found using a lexical decision task such that responses were faster when the protagonist was spatially associated with the object (Glenberg, Myer & Lindem, 1987; experiment 2).

Black, Turner and Bower (1979) also found that adult readers adopt the same point of view as the story protagonist. They conducted a series of experiments in which participants were presented with sentences designed such that the first half of the sentence established the location of the protagonist (e.g. Terry finished working in the yard) and the second half of the sentence described a movement (e.g. and then he went inside)¹. The motion verb used in the second half of the sentence was either consistent with the protagonist's point of view (e.g. went) or inconsistent (e.g. came). Participants read sentences consistent with the protagonist's point of view faster than sentences than inconsistent sentences. Furthermore, they rated consistent sentences as more comprehensible than inconsistent sentences. When asked to recall the sentences, participants recalled sentences with a consistent point of view more accurately than inconsistent sentences. Moreover, sentences that were originally contained verbs that were inconsistent with the protagonist's point of view were recalled with consistent verbs. Finally, when asked to rewrite the sentences to make them more comprehensible, participants were more likely to rewrite sentences containing verbs that were inconsistent with the protagonist's point view than consistent sentences.

Black et al. (1979) replicated the results from their recall and rewriting tasks using a second set of materials in which participants were presented with groups of three sentences describing the protagonist interacting with another character. The final sentence of the group described a movement by one of the characters using the same deictic motion verbs as used in the earlier tasks (come/go; bring/take). The results showed that participants recalled sentences with a consistent point of view more accurately than inconsistent sentences. Moreover, verbs that were inconsistent with the protagonist's point of view were replaced by consistent verbs at recall. Finally, sentences containing verbs inconsistent with the protagonist's point of view were more likely to be rewritten than sentences containing consistent verbs. Taken together, the findings of Black et al. (1979) suggest that adult readers adopt the narrative point of view during reading.

Rall and Harris (2000) conducted a similar experiment with children aged 3 and 4 years of age. The children were read two stories each containing four test sentences. Each test sentence contained a deictic verb of motion which was either consistent or

inconsistent with the point of view of the protagonist e.g. 'Little Red Riding Hood was sitting in her bedroom when her mother *came/went* in and asked her to go to Grandmother's house'. The stories were read to the children and after each test sentence they were given a cued recall task in which they had to recall what they had just heard. The cue for recall was the spatial location of the protagonist (e.g. Now remember, Little Red Riding Hood was sitting in her bedroom) followed by a general prompt (e.g. What happened next?)². The results indicated that in cases where the verb was consistent with the protagonist's point of view, children were more likely to recall the test sentence verbatim. However, if the verb was inconsistent with the protagonist point of view, children were more likely to make 'perspective-shift errors' such that they would recall the form of the verb consistent with the protagonist's point of view (e.g. recalling *came* when the verb used was *went*). These results suggest that children as young as 3- and 4-years of age adopt the protagonist's point of view when listening to a story.

Thus, young children seem to engage with a story in the same way as they engage in make-believe play. It seems plausible to suggest therefore, that young readers would engage with a text in a similar manner. Interestingly, Harris (2000) suggests that children are aware of the difference between fantasy and reality. They do not believe that the fantasy worlds are real in the empirical sense but they are able to imagine they exist. In contrast to Moshman's developmental theory of logic, this apparent understanding suggests that children as young as 3 and 4 years old possess some degree of metacognitive ability. In terms of the findings of this thesis, this would imply that children are aware of the lack of empirical truth in the inferences generated from fairy stories. However they are able to make the sentence verification decision on the basis of how true the inferences are in the context of the story. It may be the case that metacognitive ability develops with reading experience and not with chronological age as proposed by Moshman (1990). This hypothesis is consistent with the findings of the present study that the generation of inferences during reading is more dependent on reading age than chronological age.

Harris (2000) cites evidence to suggest that engaging in pretend play contributes to the development of mental state understanding. In short, he suggests that children who engage in role play become more proficient at viewing situations from a different

¹ Example taken from Black et al. (1979).

² Example taken from Rall and Harris (2000).

point of view. Interestingly, Harris (2000) cites evidence to suggest that pretend play is rarely seen in children with autism. Moreover, children with autism have been consistently shown to have difficulties with tasks requiring mental state understanding (see Happé, 1994 for a review). If, as posited above, the ability to engage in pretend play is related to children's ability to overcome the belief bias effect during reading, one might expect that children with autism would perform poorly on such a task. Interestingly in relation to this, it has been demonstrated that children with autism fail to use context to guide their reading. Frith and Snowling (1983) presented autistic children, dyslexic children and normal readers with a series of sentences to read aloud, the sentence-final word taking the form of a homograph; a pair of words or set of words which are spelt the same but which have different pronunciations and different meanings (e.g. She had a *tear* in her eye; she had a *tear* in her dress). The correct pronunciation of a homograph depends on the understanding of its use in the context of the sentence. Frith and Snowling (1983) found that the autistic children tended to produce the most frequent pronunciation of the homographs rather than using context to guide their reading. They performed at a lower level on this task than both the dyslexic children and normal readers indicating that autistic children do not engage in active sentence integration during reading. It could therefore be argued that autistic children would be unlikely to engage in inference generation during reading. With this in mind, the performance of autistic children on an on-line inference generation task such as the one used in this thesis would help to expand on current knowledge regarding the cognitive abilities of children with autism.

8.2.2 The Time Course of Inference Generation

As outlined in section 1.2. of this thesis, the constructionist theory predicts that causal inferences are drawn on-line during reading and elaborative inferences are drawn at a later point in time. The results presented in this are interpreted as supporting this prediction. This assumption is based on the consistent finding that causal test sentences were responded to faster and more accurately than static test sentences throughout the experiments of this thesis. The interpretation of these results as evidence that causal inferences are drawn on-line and elaborative inferences off-line agrees with the findings of many previous studies comparing the time-course of coherence and elaborative inferences (e.g. Fincher-Kiefer, 1996; Magliano, Baggett, Johnson & Graesser, 1993; Millis & Graesser, 1994; Valencia-Laver & Light, 2000). Moreover, this research extends the findings of previous studies by investigating elaborative inference

generation using static property inferences. Most previous work comparing coherence and elaborative inferences has used causal consequence, or predictive, inference as the elaborative inference example (e.g. Fincher-Kiefer, 1996; Magliano, Baggett, Johnson & Graesser, 1993; Millis & Graesser, 1994; Valencia-Laver & Light, 2000). However, the constructionist theory predicts that there a number of elaborative inferences that will not be drawn on-line including instantiation of a noun category, instrument inferences and static properties (Graesser et al., 1994). By incorporating one of these alternative forms of elaborative inference into the work presented here, the present thesis adds further support to the constructionist theory of inference generation. It should be noted that these findings are not inconsistent with the basic predictions of the minimalist hypothesis which state that causal inferences will be drawn on-line when necessary for local coherence, and elaborative inferences will be drawn off-line if required to meet task demands. However, as discussed in Chapter 1, the minimalist hypothesis makes no specific predictions regarding the types of inferences considered to be necessary for coherence and those that are elaborative. As such, the present thesis provides no direct evidence in support of the minimalist hypothesis.

It should be noted however, that the conclusions of this study are based on the results of a sentence verification task. Measuring the time-course of inference generation using this task does have its limitations. For example, as mentioned in chapter 2, the response time to any given target statement in the sentence verification task is comprised of the time it takes to produce the correct answer or make a judgement and the time it takes to make the response. Separating the decision making process involved in answering a question or making a true or false judgement from the actual time taken to make the response is difficult. It may be that the required inference concept is accessed off-line during the decision making process as opposed to being generated during comprehension. However, if the response time to target sentences was mediated solely by the decision making process, one might expect there to be no difference between response times to causal inference test sentences and responses to static inference test sentences. In contrast, the results presented throughout this thesis show a consistent pattern whereby responses to causal test sentences are faster than to static test sentences suggesting that the response times reported in this thesis are not confounded by the decision making process. Nevertheless, the use of reaction time as a measure of inference generation has its own limitations. Reaction time data tends to incorporate a great deal of noise, particularly when recording the reaction times of

children. For example, responses can be spoiled by the participant becoming distracted or temporarily disengaged with the task. In the present thesis, spoiled responses were recorded as outliers and removed from the final analysis in an effort to overcome this difficulty. However, the use of reaction time data in this thesis is further limited by the finding that the accuracy of responses to static test sentences was at chance. Thus, it may be the case that the longer reaction times to static test sentences was not a result of an on-line/off-line distinction between causal and static inferences. Instead it may be the case that the static test sentences were simply more difficult to judge as they were not as clearly related to the preceding passage as the causal test sentences, resulting in children making slower, more random responses. One way of addressing this issue would be to look at the response time data for only those passages where the additional comprehension questions were answered correctly. In this way, one could ensure that responses were not impeded by a lack of understanding of the passage.

The results presented in this thesis can be interpreted as demonstrating that causal inferences are drawn on-line during reading, and elaborative inferences are drawn off-line. However, given the limitations of the sentence verification task in determining the time-course of inference generation outlined above, further investigation is needed in order to clarify these results. Replication of these results using alternative on-line measures of inference generation such as lexical decision or naming speed would provide further evidence of an on-line/off-line distinction between the generation of causal and elaborative inferences. Moreover, a more sophisticated version of the paradigm used in this thesis could be created which incorporated the measurement of eye-movements during reading. The use of such a paradigm would allow one to track the eye-movements of the reader when reading a passage. In this way, one could record what happened to those eye-movements when the reader encountered a coherence break and, in principle, this might allow one to more accurately assess whether inferences are generated on-line or off-line.

Finally, there are a number of issues relating to the materials used throughout this thesis which warrant some consideration. Firstly, norming studies were not conducted on the materials from the outset of this thesis. The materials used in studies 2 and 3 were not validated in a systematic way prior to their administration in the experiments reported in this thesis. Thus, it may be the case that the materials used in these studies were not representative of the types of inferences they were designed to

measure. For example, as mentioned in Chapter 5, while the fairy stories presented in study 3 contained premise information that was counterintuitive, the inferences themselves were not necessarily in conflict with real world knowledge. Of particular concern are the inferences represented by the static inference test sentences in the fairy story condition. In some cases, these inferences reflected static properties which were stereotypical of the characters described in the corresponding stories e.g. a king wearing a gold crown. As such, while children were required to generate inferences from false premises, the classification of these inferences as counterfactual is problematic. Thus, the distinction between these inferences and those presented in the 'real world' stories is not unequivocal. An additional concern is the finding that the response accuracy for static inference test sentences was at chance. As stated above, it is possible that the inferences represented by the static test sentences were not as clearly related to the preceding passage as those inferences represented by causal test sentences. Certainly, the very nature of elaborative inferences meant that the static inferences were less constrained by the passage context than the causal inferences. Thus, the comparison between response times to causal and static test sentences may not have reflected differences in the generation of coherence and elaborative inferences. Rather, the difference in response times may have been a reflection of the fact that the static inferences presented in this study were simply more difficult to judge than the causal inferences.

It is possible that these difficulties would have been overcome had a norming study been conducted on the materials from the outset. The materials used in study 4 were subject to such a study. However, the design of this study may not have provided a rigorous enough test of the materials since the participants were asked to validate inferences that had already been generated by the experimenter. It may be the case that the inferences generated by the experimenter would not have been generated by other readers. In order to improve the materials used in this thesis, a more rigorous norming study could be conducted using a think-aloud protocol. In principle, participants could be presented with the passages used in this thesis and would be asked to 'think-aloud' while reading each passage, expressing aloud any inferences generated from the passage. The inferences most commonly generated could then be identified, and it is these inferences which would be incorporated into the materials as test sentences. Thus, one could be confident that the inference concepts used to measure inference generation were concepts which could be reliably generated from the experimental passages.

Improving the materials in this way would allow for a more comprehensive assessment of the time-course of inference generation.

8.2.3 The Relationship between Inference Generation and Reading Comprehension.

The findings from experiment 1 suggest that children with comprehension difficulties have problems generating knowledge-based coherence and elaborative inferences during reading. Less-skilled comprehenders were impaired relative to skilled comprehenders on the knowledge-based coherence and elaborative inference questions but performed at a comparable level on the literal information and cohesive inference questions. These findings support the results of Oakhill and colleagues (e.g. Cain & Oakhill, 1999; Oakhill, 1982, 1983, 1984; Oakhill & Yuill, 1986; Yuill & Oakhill, 1988) which suggest that children with comprehension difficulties do not use inference generation to actively construct a mental representation of a text during reading. Furthermore, the results presented in chapters 4 and 6 of this thesis identify causal inference generation as a significant predictor of reading comprehension. Taken together these findings suggest that the generation of coherence inferences during reading is a fundamental component of reading comprehension and moreover that the relationship between the two skills may be causal in nature such that inference generation underlies successful reading comprehension. Given these findings, it was predicted that children with specific comprehension difficulties would perform poorly given the on-line inference generation task used in this thesis. However, the findings presented in chapter 7 suggest that less-skilled comprehenders perform at a comparable level to skilled comprehenders on this task, drawing causal inference on-line during reading. This finding is clearly in contrast to previous research and undermines the assumption that inference generation is deficient in children with poor reading comprehension skills.

One way of explaining these findings is to consider the nature of the task used here in comparison to previous research. The research conducted by Oakhill and colleagues has mainly employed off-line measures of inference generation such as question and answer paradigms. These kinds of measures have a significant working memory load as the children have to retain the information they read (or infer) until they have finished reading the entire passage and then use that information to answer a series of questions about the text. Since less-skilled comprehenders have deficits in the area

of verbal working memory (e.g. Nation, Adams, Bowyer-Crane & Snowling, 1999; Palladino et al., 2001; De Beni & Palladino, 2000; Swanson & Berninger, 1995), their impaired performance on off-line measures of inference generation could be attributed to difficulties with the working memory demands of the task. In contrast, the paradigm used in the work presented here is an on-line task, which makes no such working memory demands. The test sentences are presented immediately after presentation of the critical, inference-inducing sentences. With this in mind, the results presented in chapter 7 suggest that less-skilled comprehenders do not have difficulties with inference generation per se and suggest that previous research may have confounded their difficulties on measures of inference generation with their verbal working memory difficulties.

An alternative explanation may be that verbal working memory underlies both inference generation and reading comprehension. Recent work by Calvo (2000) supports this assumption. Calvo (2000) investigated the relationship between working memory capacity and the time-course of predictive inferences. The results of a naming speed task showed that participants with a high working memory span, as measured by a reading span task, made predictive inferences faster than participants with a low working memory span. However, it should be noted that this study did not look at the influence of working memory span on inferences thought to be generated automatically during reading i.e. coherence inferences, and focused instead on those inferences that are thought to be processed strategically. Thus, the results presented by Calvo (2000) may simply suggest that a high level of working memory capacity facilitates strategic processing.

Radvansky and Copeland (2001) investigated the influence of working memory capacity on situation model updating. As discussed in chapter 1.2, the situation model is part of the mental representation of a text constructed during reading. The situation model represents the actions, events, characters and overall meaning of the text. Furthermore, because the situation model is constructed during reading it requires updating whenever new information is encountered. Updating the situation model requires the integration of explicit information in the text with information left implicit or inferred. Therefore, the construction and updating of the situation model requires the generation of inferences. Radvansky and Copeland (2001) looked at the relationship between performance on a test of situation model updating and a test of working

memory capacity. While the participants in the study clearly showed evidence of situation model updating, there was no relationship between performance on this task and working memory capacity suggesting that working memory capacity is not related to inference generation.

It should be noted that the situation model updating task used here was not a direct measure of inference generation and as such the results do not directly undermine the claim that working memory has a causal influence on inference generation. However, a situation identification test was also given in this study in which readers were given a series of sentences e.g. 'the man lost a hand of poker at the card shark's' followed by a recognition test in which they were given six alternative versions of each of the original sentences and they were asked to select the version that most closely resembled the situation described by the sentence read earlier. For example, of the six versions of the sentence 'the man lost a hand of poker at the card shark's' given below, sentence 'b' is the correct response.

- a) The man lost a hand of poker like the card shark.
- b) The man lost some money at the card shark's.
- c) The man won a hand of poker at the card shark's.
- d) The man lost some money like the card shark.
- e) The man won a hand of poker like the card shark.
- f) The man won some money at the card shark's.

Arguably, in order to choose sentence 'b' as the sentence most closely resembling the original item, the reader would have to infer that losing a hand of poker at a card shark's would in all likelihood involve losing some money and that therefore this situation is the most similar to that described in the original sentence. Thus, the situation identification task can be seen as a measure of inference generation. Interestingly, while performance on the situation identification task was related to performance on the situation updating task, it was not related to working memory capacity.

Taken together, the results presented by Radvansky and Copeland (2001) suggest that inference generation is not reliant on working memory capacity. Clearly, the relationship between working memory and inference generation requires further

investigation. However, the findings of Radvansky and Copeland (2001) supports the assumption made earlier that previous research has confounded the performance of less-skilled comprehenders on inference generation tasks with their verbal working memory difficulties. When these difficulties are controlled for by using an on-line measure of inference generation, children with comprehension difficulties perform at a comparable level to reading-age matched peers.

An alternative explanation for these findings is that normal readers have more efficient strategic processing skills than less-skilled comprehenders but that the two groups are comparable in terms of automatic processing. It could be argued that the on-line task used in this thesis requires less strategic processing than the off-line tasks used by Oakhill and colleagues. For example, the question and answer task used by Oakhill (1984) and Cain and Oakhill (1999) required the reader to consciously locate the information contained in the text, or in their mental representation of the text, that was relevant to answering the comprehension questions presented after reading the text. Similarly, Cataldo and Oakhill (2000) gave children a passage to read followed by a series of comprehension questions. However, in this study they explicitly asked children to locate the point in a text where the answer to each question could be found. Their results showed that less-skilled comprehenders were slower to locate the relevant information in the text and moreover that less-skilled comprehenders engaged in less efficient search strategies than skilled comprehenders. Finally, Yuill & Oakhill (1988a) explicitly asked children to specify the referent of each anaphor they encountered during the reading of a passage.

Tasks such as those described above clearly require a high level of strategic processing. Therefore, the difference in performance between the less-skilled comprehenders included in the studies by Oakhill and colleagues and the less-skilled sample used in the experiment described in chapter 7 might be due to the fact that the task used in the present study did not require the same degree of strategic processing as required in previous research. This assumption is supported by the finding that the less-skilled comprehender's performance was impaired relative to the normal readers on the off-line comprehension questions presented in experiment 5, despite comparable performance on the on-line task. However, one might also expect there to be a difference in the performance of normal readers and less-skilled comprehenders in the generation of static inferences during completion of the on-line task, since static

inferences are thought to require strategic off-line processing. In contrast, the performance of normal readers and less-skilled comprehenders was comparable for both causal and static inference generation. Arguably, although it seems unlikely, it could be due to a lack of sensitivity in the materials such that the static inferences were too simple to require a great deal of strategic processing. Further research using more complex inference concepts is needed to clarify this issue.

Finally, it should be noted that the less-skilled comprehenders chosen for participation in this study were selected using criteria that differed from those used in previous research. As in the current thesis, the less-skilled comprehenders included in the work of Oakhill and colleagues (Cain & Oakhill, 1996; Cain & Oakhill, 1999; Cain, Oakhill, Barnes & Bryant, 2001; Oakhill, 1982, 1983, 1984; Oakhill & Yuill, 1986; Oakhill, Yuill & Donaldson, 1990; Yuill & Oakhill, 1988a, 1988b; Yuill, Oakhill & Parkin, 1990) were matched to a chronological-age match group on the basis of NARA II accuracy age. However, the control group was also matched to the less-skilled comprehender group on the basis of vocabulary knowledge. In contrast, the children who participated in experiment 5 of this thesis showed significant differences in vocabulary such that the less-skilled comprehenders had lower vocabulary scores than the control group. Moreover, the inclusion of a comprehension-age match control group has been advocated as a rigorous method for investigating comprehension difficulties in children (Cain, Oakhill, & Bryant, 1998). However, such a group was not possible for use in this study since children with a comprehension-age comparable to the less-skilled comprehender group would have been too young to read the stories presented.

It may be the case, therefore, that the disagreement between the findings presented in this thesis and those presented previously may in part be due to the different selection criteria used. However, this explanation seems unlikely since one might expect that the weaker vocabulary knowledge possessed by the less-skilled comprehenders included in the present study would have adversely affected their performance on the inference generation task. Instead, the children included in this sample performed at a comparable level to the control group on the inference generation task, irrespective of vocabulary level. Moreover, Nation et al. (Nation, Clarke, & Snowling, in press) emphasise that there is a great deal of heterogeneity in the population of children commonly identified as poor or less-skilled comprehenders.

Thus, there may be individual differences within the samples used by Oakhill and colleagues and the sample included in the current study which have not been identified by the selection procedures. The use of a wider range of cognitive ability tests when identifying poor comprehenders would be an advantage in future research.

8.3 Conclusions

In conclusion, the findings presented in this thesis suggest that children generate inferences on-line during reading when necessary for coherence but unnecessary inferences are generated off-line. These findings agree with previous research into inference generation and provide support for the predictions of the constructionist theory. Moreover, text genre and content have been demonstrated to have little influence on inference generation in children aged between 7 and 10 years, and it is reading skill rather than chronological age that aids in the development of good inferential skills. Finally, data from skilled and less-skilled comprehenders suggests that, while inference generation is an important component of reading comprehension, the relationship between these two variables has yet to be clarified. The plausibility of a causal relationship between inference generation and reading comprehension is undermined by the finding that less-skilled comprehenders performed at a comparable level to reading-age controls on the inference generation task used in this study. If inference generation is indeed a causal factor in the development of reading comprehension skills, the less-skilled comprehenders should have shown impaired performance relative to the control group. An alternative explanation is that a third variable underlies both of these skills and a possible candidate for this role is working memory. This explanation seems highly likely in light of the reported verbal working memory difficulties of less-skilled comprehenders (Nation, Adams, Bowyer-Crane & Snowling, 1999; Palladino et al., 2001; De Beni & Palladino, 2000; Swanson & Berninger, 1995). However, the findings of Radvansky and Copeland (2001) that working memory capacity is not related to inference generation serve to cast some doubt on this hypothesis. Further investigation into the relationship between working memory, inference generation and reading comprehension is warranted in order to clarify this issue.

8.4 Educational Implications

From an educational perspective, the results presented in this thesis have implications for reading comprehension assessment, instruction and intervention. For

example, the results from experiment 1 clearly showed that different measures of reading comprehension assessment may tap different aspects of reading comprehension. Children who performed poorly on the NARA II test of reading comprehension, performed in the normal range on the WORD reading comprehension test. Similarly, although the data is not reported here, the children included in experiment 5 were given the Tests of Reading Comprehension (TORCH; Mossenson, Hill, & Masters, 1987). The TORCH is a story-based test in which children are presented with one story that they read silently. They are then presented with an alternative version of the same story, in which the structure of the text has been altered and key information has been replaced by blank spaces. The task of the child is to locate the key information in the original text and use it to fill in the blank spaces in the alternate version of the story. A key feature of the TORCH is that the child is required to use a number of different strategies when completing one story. For example, in order to fill in one of the blanks the child might need to locate information in the original text and copy it verbatim into the alternative version. However, later in the passage, the child may need to infer information from the original text in order to fill in the blank in the alternative version. Thus, the TORCH assesses a wide range of reading comprehension skills, which are not directly assessed by more traditional comprehension assessments. While the skilled and less-skilled comprehenders in experiment 5 could be clearly differentiated in terms of their performance on the NARA II, there was a great deal of overlap between the two groups in performance on the TORCH. Thus, children have individual profiles of strengths and weaknesses in reading comprehension, which emphasises its complex nature. With this in mind, the assessment of reading comprehension is clearly problematic and the findings presented in this thesis advocate the use of two or more different measures of reading comprehension to give a full and comprehensive assessment.

The finding that reading skill and not chronological age is directly implicated in the development of inferential skills suggests the rather obvious conclusion that it is clearly important to encourage reading in children to support the development of proficient reading skills. However, the lack of a significant influence of text genre and content on inference generation suggests that, while it is clearly important that children read regularly, presenting children with a diverse range of text types will not have a significant impact on their later reading comprehension.

The results presented in this thesis suggest that the generation of causal inferences is more important for comprehension than inferring static properties. Thus, intervention strategies that seek to incorporate inference skill training can be designed to enhance those aspects of inference generation that might effectively improve reading comprehension rather than training in generic inference skills. Furthermore, the results presented in chapter 7 emphasise the multi-faceted nature of reading comprehension. While inference generation is clearly important to reading comprehension, its importance relative to other skills is unclear. Thus, construction of reading comprehension instruction and intervention programmes should aim to include training in other skills such as strategic processing and working memory in order to encompass the wide range of skills of which reading comprehension is comprised.

References

- Albrecht, J.E., & O'Brien, E.J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology- Learning, Memory and Cognition.*, 19(5), 1061 – 1070.
- Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation of verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247-264.
- Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles - a problem of polysemy. *Cognitive Psychology.*, 7, 167-180.
- Anderson, R.C., Pichert, J.W., Goetz, E.T., Schaller, D.L., Stevens, K.V., & Trollip, S.R. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behaviour*, 15, 667-679.
- Bahri, T., & Hussain, A. A. A. (1997). Question type and order of inference in inferential processes during reading comprehension. *Perceptual and Motor Skills*, 85, 655-664.
- Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506-521.
- Black, J.B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behaviour*, 20, 267-275.
- Black, J.B., Turner, T.J., & Bower, G.H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behaviour*. 18, 187-198.
- Bloom, C.P., Fletcher, C.R., Van den Broek, P., Reitz, L., & Shapiro, B.P. (1990). An on-line assessment of causal reasoning during comprehension. *Memory and Cognition*, 18(1), 65-71.
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretative approach. *Cognitive Psychology.*, 3, 193-209.
- Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83(3), 329-345.
- Bower, G.H., Black, J.B. & Turner, T.J. (1979) Scripts in memory for text. *Cognitive Psychology*, 11, 177-220.
- Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology*, 14, 187-201.

- Cain, K., Oakhill, J., & Bryant, P. (1998). Investigating the causes of reading comprehension failure: The comprehension age-match design. *Reading and Writing: An Interdisciplinary Journal*, 00, 1-11
- Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, 11, 489-503.
- Cain, K., Oakhill, J.V., Barnes, M.A., & Bryant, P.E. (2001). Comprehension skill, inference-making ability and their relation to knowledge. *Memory and Cognition*. 29(6), 850-859.
- Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing*, 13(1-2), 31-56.
- Calvo, M.G. (2000). Working memory capacity and time course of predictive inferences. *Memory*, 8(1), 51-61.
- Calvo, M. G. (2000). The time course of predictive inferences depends on contextual constraints. *Language and Cognitive Processes*, 15(3), 293-319.
- Calvo, M.G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, 9(4/5/6), 365-381.
- Calvo, M. G., & Castillo, M. D. (1996). Predictive inferences occur on-line but with delay: convergence of naming and reading times. *Discourse Processes*, 22, 57-78.
- Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-line predictive inferences in reading: processing time during versus after the priming context. *Memory and Cognition*, 27(5), 834-843.
- Calvo, M..G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research*, 65(3), 158-169.
- Casteel, M. A. (1993). Effects of inference necessity and reading goal on children's inferential generation. *Developmental Psychology*, 29(2), 346-357.
- Casteel, M. A., & Simpson, G. B. (1991). Textual coherence and the development of inferential generation skills. *Journal of Research in Reading*, 12(2), 116-129.
- Cataldo, M. G., & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology*, 92(4), 791-799.

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. 2nd Edition. Hillsdale, N.J: Lawrence Erlbaum Associates.
- Curtis, M. E. (1980). Development of components of reading skill. *Journal of Educational Psychology*, 5, 656-669.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- De Beni, R., & Palladino, P. (2000). Intrusion errors in working memory tasks - are they related to reading comprehension ability? *Learning and Individual Differences*, 12, 131-143.
- DeJong, G. (1979). Prediction and substantiation: A new approach to natural language processing. *Cognitive Science*, 3, 251-273.
- De Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, 6(1), 51-77.
- Dias, M. G., & Harris, P. L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology*, 6, 207-221.
- Dias, M. G., & Harris, P. L. (1990). The influence of imagination on reasoning by young children. *British Journal of Developmental Psychology*, 8, 305-318.
- Donaldson, M. L. (1986). *Children's Explanations: A Psycholinguistic Study*. Cambridge: Cambridge University Press.
- Duffy, S.A., & Rayner, K. (1990). Eye movements and anaphor resolution: Effects of antecedent and typicality and distance. *Language and Speech*, 33, 103-119.
- Duke, N. (2002). *How does diversifying genres used in first grade literacy instruction impact on comprehension development*. Paper presented at the Society for the Scientific Studies of Reading, Chicago, IL.
- Dyer, M.G. (1983). *In-Depth understanding: A Computer Model of Integrated Processing for Narrative Comprehension*. Cambridge, MA: MIT Press.
- Ehrlich, S.F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behaviour*, 20, 641-655.
- Einstein, G. O., McDaniel, M. A., Owen, P. D., & Cote, N. C. (1990). Encoding and recall of texts: the importance of material appropriate processing. *Journal of Memory and Language*, 29, 566-581.
- Elliot, C. D., Smith, P., & McUlloch, K. (1996). *British Ability Scales Second Edition (BAS II)*, Windsor: NFER Nelson.

- Evans, J. S. B. T., & Perry, T. S. (1995). Belief bias in children's reasoning. *Current Psychology of Cognition*, 14(1), 103-115.
- Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, 22, 225-246.
- Fodor, J.D. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press.
- Franks, B. A. (1996). Deductive reasoning in narrative contexts: developmental trends and reading skill effects. *Genetic, Social and General Psychology Monographs*, 122(1), 75-105.
- Franks, B. A. (1997). Deductive reasoning with prose passages: effects of age, inference form, prior knowledge, and reading skill. *International Journal of Behavioural Development*, 21(3), 501-535.
- Frith, U. & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *Journal of Developmental Psychology*, 1, 329-342.
- Fulcher, G. (1989). Cohesion and coherence in theory and reading research. *Journal of Research in Reading*, 12(2), 146-163.
- Galotti, K. M., L.K., K., & Voelz, S. (1997). Children's differential performance on deductive and inductive syllogisms. *Developmental Psychology*, 33(1), 70-78.
- Garnham, A., Oakhill, J., & Cain, K. (1997). The interpretation of anaphoric noun phrases: time course, and effects of overspecificity. *The Quarterly Journal of Experimental Psychology*, 50A(1), 149-162.
- Garnham, A., Oakhill, J., & Johnson-Laird, P. N. (1982). Referential Continuity and the Coherence of Discourse. *Cognition*, 11(1), 29-46.
- Gernsbacher, M.A. (1997). Two decades of structure building. *Discourse Processes*, 23, 265-304.
- Gernsbacher, M., & Faust, M. E. (1991). The mechanism of suppression: a component of general comprehension skill. *Journal of Experimental Psychology- Learning, Memory and Cognition*, 17(2), 245-262.
- Gernsbacher, M. A., Hallada, B. M., & Robertson, R. W. (1998). How automatically do readers infer fictional characters' emotional states? *Scientific Studies of Reading*, 2(3), 271-300.
- Glenberg, A., & Mathew, S. (1992). When minimalism is not enough: mental models in reading comprehension. *Psychology*, 3.
- Glenberg, A.M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26, 69-83.

- Graesser, A., & Kreuz, R. J. (1993). A theory of inference generation during text comprehension. *Discourse Processes, 16*, 145-160.
- Graesser, A., Millis, K. K., & Zwann, R. A. (1997). Discourse Comprehension. *Annual Review of Psychology, 48*, 163-189.
- Graesser, A., Singer, M., & Trabasso, T. (1994). Constructing inferences during text comprehension. *Psychological Review, 101*(3), 371-395.
- Hagley, F. J. (1986). *Suffolk Reading Scale, Berkshire: NFER - Nelson.*
- Halpern, D. F., & Le May, M. L. (2000). The smarter sex: a critical review of sex differences in intelligence. *Educational Psychology Review, 12*(2), 229-246.
- Happé, F. (1994). *Autism – an introduction to psychological theory.* London; UCL Press.
- Harris, P. L. (2000). *The Work of the Imagination.* Oxford, UK; Blackwell Publishers Ltd.
- Hatcher, P. (2000). Predictors of Reading Recovery book levels. *Journal of Research in Reading, 23*(1), 67-77.
- Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing: An interdisciplinary journal, 2*, 127-160.
- Inman, T. M., & Dickerson, D. J. (1995). Children's inferences of causal events during story understanding. *The Journal of Genetic Psychology, 156*(3), 265-277.
- Johnson, H., & Smith, L. B. (1981). Children's inferential abilities in the context of reading to understand. *Child Development, 52*, 1216-1223.
- Kameenui, E. J., Carnine, D. W., & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly, 17*(3), 367-388.
- Keenan, J.M., Golding, J.M., Potts, G.R., Jennings, T.M., & Aman, C.T. (1990). Methodological issues in evaluating the occurrence of inferences. In A.C Graesser, & G.H. Bower (Eds.), *Psychology of Learning and Motivation: Inferences and Text Comprehension*, (Vol 25, pp. 295-312).
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-393.
- Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58*, 88-111.

- Leevers, H. J., & Harris, P. L. (2000). Counterfactual syllogistic reasoning in normal 4-year olds, children with learning disabilities and children with autism. *Journal of Experimental Child Psychology, 76*, 64-87.
- Long, D. L., & Golding, J. M. (1993). Superordinate goal inferences: are they automatically generated during comprehension. *Discourse Processes, 16*, 55-73.
- Long, D. L., Oppy, B. J., & Seely, M. S. (1997). A 'global-coherence' view of event comprehension: inferential processing as question answering. In T. Bourg (Ed.), *Developmental Spans in Event Comprehension: Bridging fictional and actual events* (pp. 361-384). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes, 16*, 35-53.
- Magliano, J.P., & Graesser, A.C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics, 20*, 193-232.
- McDaniel, M. A., & Einstein, G. O. (1989). Material appropriate processing: a contextualist approach to reading and studying strategies. *Educational Psychology Review, 1*(2), 113-145.
- McDaniel, M. A., Einstein, G. O., Dunay, P., & Cobb, R. E. (1986). Encoding difficulty and memory: toward a unifying theory. *Journal of Memory and Language, 25*, 645-656.
- McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1982). The effects of long-term vocabulary instruction on reading comprehension: a replication. *Journal of Reading Behaviour., 15*(1), 3-18.
- McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behaviour, 19*, 668-682.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*(1), 82-91.
- McKoon, G., & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory and Cognition. 14*(2), 331-343.
- McKoon, G., & Ratcliff, R. (1989). Assessing the occurrence of elaborative inference: Computability checking vs. compound cue theory. *Journal of Memory and Language, 28*, 547-563.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440-466.

- Merrill, E. C., Sperber, R. D., & McCauley, C. (1981). Differences in semantic encoding as a function of reading comprehension skill. *Memory & Cognition*, 9(6), 618-624.
- Millis, K.K., Golding, J.M., & Barker, G. (1995). Causal connectives increase inference generation. *Discourse Processes*, 20, 20 – 49.
- Millis, K. K., & Graesser, A. C. (1994). The time-course of constructing knowledge-based inferences for scientific texts. *Journal of Memory and Language*, 33, 583-599.
- Millis, K. K., Morgan, D., & Graesser, A. C. (1990). The influence of knowledge-based inferences on the reading time of expository text. In A.C. Graesser & G. H. Bower (Eds.), *Psychology of Learning and Motivation: Inferences and Text Comprehension* (Vol. 25, pp. 197-212). San Diego: Academic Press.
- Moshman, D. (1990). The Development of Metalogical Understanding. In W. F. Overton (Ed.), *Reasoning, Necessity and Logic: Developmental Perspectives*. Hillsdale, NJ, US: Lawrence Erlbaum Associates Inc.
- Moshman, D., & Franks, B. A. (1986). Development of the concept of inferential validity. *Child Development*, 57, 153-165.
- Mossenson, L., Hill, P., & Masters, G. (1987). *Tests of Reading Comprehension*. Victoria, Australia; Australian Council for Educational Research.
- Murray, J. D., Klin, C. M., & Myers, J. M. (1993). Forward inferences in narrative text. *Journal of Memory and Language*, 32, 464-473.
- Myers, J.M., Cook, A.E., Kambe, G., Mason, R.A., & O'Brien, E.J. (2000). Semantic and episodic effects on bridging inferences. *Discourse Processes*, 29(3), 179-199.
- Myers, J.M., Shinjo, M. & Duffy, S.A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, 26, 453 – 465.
- Narvaez, D., van den Broek, P., & Barron Ruiz, A. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, 91(3), 488-496.
- Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. S. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, 73, 139-158.
- Nation, K., Clarke, P., & Snowling, M. J. (in press). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology*.

- Nation, K., Marshall, C.M., & Altmann, G.T.M. (2002). *On-line language processing in children with poor language comprehension: Evidence from a language-mediated ey-tracking experiment*. Poster session presented at the Joint Meeting of the International Association for the Study of Child Language and the Symposium for Research on Child Language Disorders. Madison, Wisconsin.
- Nation, K., Marshall, C. M., & Snowling, M. J. (2001). Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders. *Language and Cognitive Processes*, 16(2-3), 241-259.
- Nation, K., & Snowling, M. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67, 359-370.
- Nation, K., & Snowling, M. J. (1998a). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, 39(1), 85-101.
- Nation, K., & Snowling, M. J. (1998b). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development*, 69(4), 996-1011.
- Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: evidence from semantic priming. *Cognition*, 70(1), B1-B13.
- Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics*, 21(2), 229-241.
- Neale, M. D. (1989). *Neale Analysis of Reading Ability - Revised*, Windsor: NFER-Nelson.
- Noordman, L. G. M., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, 31, 573-590.
- Oakhill, J. (1982). Constructive Processes in Skilled and Less Skilled Comprehenders Memory for Sentences. *British Journal of Psychology*, 73(FEB), 13-20.
- Oakhill, J. (1983). Instantiation in skilled and less-skilled comprehenders. *Quarterly Journal of Experimental Psychology*, 35A, 441-450.
- Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, 54, 31-39.

- Oakhill, J., & Yuill, N. (1986). Pronoun Resolution in Skilled and Less-Skilled Comprehenders - Effects of Memory Load and Inferential Complexity. *Language and Speech, 29*, 25-37.
- Oakhill, J., Yuill, N., & Donaldson, M. L. (1990). Understanding of Causal Expressions in Skilled and Less Skilled Text Comprehenders. *British Journal of Developmental Psychology, 8*, 401-410.
- Olson, G. M., Mack, R. L., & Duffy, S. A. (1981). Cognitive aspects of genre. *Poetics, 10*, 283-315.
- Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition, 29*(2), 344-354.
- Paris, S. G., & Lindauer, B. K. (1976). The role of inferences in children's comprehension and memory for sentences. *Cognitive Psychology, 8*, 217-227.
- Paris, S. G., Lindauer, B. K., & Cox, G. L. (1977). The development of inferential comprehension. *Child Development, 48*, 1728-1733.
- Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language, 24*, 503-518.
- Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: lexical decision versus naming. *Journal of Memory and Language, 27*, 399-415.
- Power, R., Taylor, C. L., & Nippold, M. A. (2001). Comprehending literally-true versus literally-false proverbs. *Child Language, Teaching and Therapy, 17*(1), 1-18.
- Radvansky, G.A., & Copeland, D.E. (2001). Working memory and situation model updating. *Memory and Cognition, 29*(8), 7073-1080.
- Rayner, K., Raney, G.E., & Pollatsek, A. (1995). Eye movements and discourse processing. *Sources of Coherence in Reading*. (pp 9-36). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rall, J., & Harris, P.L. (2000). In Cinderella's slippers? Story comprehension from the protagonist's point of view. *Developmental Psychology, 36*, 202-208.
- Revlín, R., & Hegarty, M. (1999). Resolving signals to cohesion: two models of bridging inference. *Discourse Processes, 27*(1), 77-102.
- Richards, C. A., & Sanderson, J. A. (1999). The role of imagination in facilitating deductive reasoning in 2-, 3- and 4-year olds. *Cognition, 72*, B1-B9.

- Schank, R.C. & Abelson, R. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sears, S., & Keogh, B. (1993). Predicting reading performance using the Slingerland procedures. *Annals of Dyslexia*, 43, 78-89.
- Seigneuric, A., Ehrlich, M. F., Oakhill, J., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing; an interdisciplinary journal*, 13, 81-103.
- Singer, M. (1983). Global inferences of text situations. *Discourse Processes*, 16, 161-168.
- Singer, M. (1993). Causal bridging inferences: validating consistent and inconsistent sequences. *Canadian Journal of Experimental Psychology*, 47(2), 340-359.
- Singer, M. H., & Crouse, J. (1981). The relationship of context-use skills to reading: a case for alternative experimental logic. *Child Development*, 52, 1326-1329.
- Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behaviour*, 22, 437-448.
- Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language*, 33, 421-441.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*, 31, 507-524.
- Singer, M., Harkness, D., & Stewart, S. T. (1997). Constructing inferences in expository text comprehension. *Discourse Processes*, 24(199-228).
- Snowling, M. J., Stothard, S. E., & McLean, J. (1996). *Graded Nonword Reading Test*, Bury St Edmunds: Thames Valley Test Company.
- Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly*, 19, 278-303.
- Stanovich, K. E., Nathan, R. G., & Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading Research Quarterly*, 21(3), 267-283.
- Stothard, S. (1992). *Reading Difficulties in Children: Problems of Decoding and Comprehension*. University of York, York.
- Suh, S. & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language*, 32, 279-300.

- Swanson, H.L., & Berninger, V. (1995). The role of working memory in skilled and less-skilled readers' comprehension. *Intelligence*, *21*, 83 – 108.
- Swinney, D., & Osterhout, L. (1990). Inference generation during auditory language comprehension. In A.C. Graesser & G. H. Bower (Eds.), *Psychology of Learning and Motivation: Inferences and text comprehension* (Vol. 25, pp. 17-33). San Diego: Academic Press.
- Thompson, J. G., & Myers, N. A. (1985). Inferences and recall at ages four and seven. *Child Development*, *56*, 1134-1144.
- Thorndike, R. L., Hagen, E., & France, N. (1986). *Cognitive Abilities Test, Berkshire: NFER-Nelson*.
- Thurlow, R., & van den Broek, P. (1997). Automaticity and inference generation during reading comprehension. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*, 165-181.
- Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency, Austin, Texas: Pro-Ed*.
- Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes*, *21*(3), 255-287.
- Trabasso, T., Suh, S., Payton, P., & Jain, R. (1995). Explanatory inferences and other strategies during comprehension and their effect on recall. In R.F. Lorch and E.J. O'Brien (Eds.), *Sources of Coherence in Reading*. (pp 219-239). Hillsdale, NJ: Lawrence Erlbaum Associates..
- Valencia-Laver, D. L., & Light, L. L. (2000). The occurrence of causal bridging and predictive inferences in young and older adults. *Discourse Processes*, *30*(1), 27-56.
- Van den Broek, P.W. (1990) Causal inferences in the comprehension of narrative texts. In A.C. Graesser & G.H. Bower (Eds.), *Psychology of Learning and Motivation: Inferences and Text Comprehension* (Vol 25, pp. 175-194). San Diego: Academic Press.
- Van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 539-588). New York: Academic Press.
- Van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081-1087.

- Van den Broek, P., Young, M., Tzeng, Y. & Linderholm, T. (1999). The landscape model of reading inferences and the on-line construction of memory representations. In H. van Oostendorp & S.R. Goldman, (Eds.), *The Construction of Mental Representations During Reading* (pp. 71-98). Mahwah, NJ: Lawrence Erlbaum Associates.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- Wechsler, D. (1990). *Wechsler Objective Reading Dimensions, London: The Psychological Corporation*.
- Wechsler, D. (1992). *Wechsler Intelligence Scale for Children, London: The Psychological Corporation*.
- Wechsler, D. (1996). *Wechsler Objective Numerical Dimensions, London: The Psychological Corporation*.
- Wiig, E. H., & Secord, W. (1992). *Test of Word Knowledge., The Psychological Corporation*.
- Willson, V. L., & Rupley, W. H. (1997). A structural equation model of reading comprehension based on background, phonemic and strategy knowledge. *Scientific Studies of Reading, 1*(1), 45-63.
- Woodcock, R. W., & Johnson, M. B. (1989, 1990). *Woodcock-Johnson Psycho-Educational Battery - Revised., Allen, TX; DLM Teaching Resources*.
- Yuill, N., & Oakhill, J. (1988a). Effects of Inference Awareness Training on Poor Reading- Comprehension. *Applied Cognitive Psychology, 2*(1), 33-45.
- Yuill, N., & Oakhill, J. (1988b). Understanding of Anaphoric Relations in Skilled and Less Skilled Comprehenders. *British Journal of Psychology, 79*, 173-186.
- Yuill, N., & Oakhill, J. (1991). *Children's Problems in Text Comprehension: An experimental investigation*. New York, NY: Cambridge University Press.
- Yuill, N., Oakhill, J., & Parkin, A. (1989). Working Memory, Comprehension Ability and the Resolution of Text Anomaly. *British Journal of Psychology, 80*, 351-361.
- Zabracky, K. (1986). The role of factual coherence in discourse comprehension. *Discourse Processes, 9* (197-220).
- Zinar, S. (2000). The relative contributions of word identification skill and comprehension-monitoring behaviour to reading comprehension ability. *Contemporary Educational Psychology, 25*(363-377).

Zwann, R.A. (1999). Five dimensions of narrative comprehension: The event-indexing model. In S.R. Goldman, A.C. Graesser, P. Van den Broek (Eds.), *Narrative Comprehension, Causality and Coherence: Essays in honour of Tom Trabasso*. Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix I
Experiment 2 Test Materials

List One

1. Builders build houses for people that are made to last a long time. They use bricks to make the walls solid. Some people used to build houses from wood. They used to do all their cooking on big open fires. Many of the wooden houses were totally destroyed.

Bricks are a very useful invention.

Test Sentence 1: Fire will burn wood.

Test Sentence 2: Fire keeps you warm.

Question: Why did people start to build houses with bricks?

2. Peter didn't like going to bed very much. Often he would read until very late. But Peter did not want anyone to know about it. So Peter tried to read in the dark but this hurt his eyes. Soon he found that he needed to wear glasses. He did not really mind wearing glasses but decided to read with the light on in future.

Test Sentence 1: If you have bad eyes you need glasses.

Test Sentence 2: Reading can help you to learn new things.

Question: Why did Peter decide to read with the light on?

3. You should never play with matches.

Playing with matches can be very dangerous. You can start a fire by accident. You must never go near a fire. You should

find a phone and call the fire brigade. They usually arrive quickly and know what to do.

Test Sentence 1: Firemen put out fires

Test Sentence 2: Firemen wear yellow helmets

Question: Why should you call the fire brigade when there is a fire?

4. Sam was a little dog who lived in a box. He liked living there when the sun was shining brightly. But he didn't like it in bad weather. He wished for an umbrella when it rained. It was not very nice to get wet. One day, Sam found an umbrella by the dustbins and was very happy.

Test Sentence 1: Umbrellas protect you from the rain.

Test Sentence 2: when it rains rivers can flood.

Question: Why do people carry umbrellas?

5. Windows allow the light to come in. You can also look through a window. You can watch what is happening outside. Windows can break if they are hit hard. Glass smashes all over the ground. You should always take care not to cut yourself.

Test Sentence 1: Windows are made from glass

Test Sentence 2: Glass is sharp when broken

Question: What should you be careful of if you hit a window?

6. Mr Jones liked to drive his car to work every day. He did not enjoy walking to work at all. One morning Mr Jones looked out of the window. It had been snowing all night and the snow had made the roads icy. He decided to leave his car at home and walk to work that day. On the way he saw an accident and was glad his car was safe at home.

Test Sentence 1: Icy roads are dangerous for cars.

Test Sentence 2: You can slip on icy roads

Question: Why did Mr Jones leave his car at home?

7. Making your own cards is more fun than buying them. You can cut shapes out of coloured paper to decorate them. When you cut out the shapes you should use a pair of scissors. Scissors can be very sharp. You should always get an adult to help you cut with the scissors. When you have finished your card you can write a message inside.

Test Sentence 1: Sharp scissors can be very dangerous.

Test Sentence 2: You can use glue for sticking

Question: Why should you get an adult to help you cut with scissors?

8. One summer day, Jack and Sarah were walking in the park. Jack saw an ice-cream van. He wanted to buy an ice-cream. He

did not want it to melt. It was very hot and he ate his ice-cream very quickly. But he shared some of it with Sarah.

Test Sentence 1: Ice-cream melts when it is hot.

Test Sentence 2: Ice-cream comes in many different flavours.

Question: Why did Jack eat his ice-cream quickly?

9. Some people live in the country where it is very quiet. They often live on farms where there is a lot of space. But other people live in busy towns and cities where there is a lot of traffic. The traffic on the roads can be noisy all through the night.

People often find it difficult to sleep. Some people use earplugs to help them sleep.

Test Sentence 1: Noise can stop you from sleeping.

Test Sentence 2: Driving can be dangerous at night.

Question: Why do people who live on busy roads find it difficult to sleep?

10. Lucy woke up one day to find she had a sore tooth. It made her whole mouth very sore. She was very unhappy because her mouth hurt so much. So Lucy told her mum that she had a toothache. Lucy had to go to the Dentist. She was very good and afterwards her tooth did not hurt anymore.

Test Sentence 1: Dentists look after your teeth.

Test Sentence 2: You should brush your teeth.

Question: Why did Lucy have to go to the Dentist?

11. When you are sick you can feel hot. You might have a sore throat or a runny nose. You might even come out in red spots. When you are ill you cannot go to school. There are lots of other children at your school. You might have to go to the Doctor to get some medicine.

Test Sentence 1: sometimes other people can catch your illness.

Test Sentence 2: you might sneeze with a runny nose.

Question: Why can you not go to school when you are ill?

12. John was exploring the huge jungle. It was very hot and he wanted to take his shoes off. But John thought that there might be snakes in the jungle. Sometimes snakes can bite you. Some snakes can make you very ill. So John drank water from his water bottle instead.

Test Sentence 1: Some snakes are poisonous

Test Sentence 2: Snakes shed their skin.

Question: How can some snakes make you ill?

13. People like to go on holiday each year to many different places. Some people stay in England and others go abroad.

Sometimes you need to catch a plane or a boat. But it can be more fun to catch the train. Railway stations are busy, exciting places. You can go anywhere in the country by train.

Test Sentence 1: Trains leave from stations.

Test Sentence 2: Aeroplanes leave from airports.

Question: Why might you need to go to the railway station?

14. Timmy wanted some fish to keep in his bedroom. He caught a big fish and lifted it out of the water. It was a big silver fish and Timmy was very pleased. But all fish need water to breathe. The fish could not live out of the water. Timmy put the fish straight back in the river.

Test Sentence 1: Breathing keeps you alive.

Test Sentence 2: People often keep goldfish.

Question: Why could the fish not live out of the water?

15. Some people like to spend a lot of time in their garden. They work very hard to keep them tidy. They do a lot of digging and planting. They like to grow different flowers and plants in their gardens. They have to water their plants every day. In the

summer the flowers bloom and the gardens look very colourful.

Test Sentence 1: Plants need water to grow

Test Sentence 2: Gardens usually have some grass.

Question: Why do plants need water every day?

16. Mrs Jones was making some tea for her friends. She put some biscuits on a plate. She filled the kettle with some water and started it boiling. The boiled water was very hot. Mrs Jones carefully poured tea into some cups. Then she took her friends the biscuits and sat down with them to enjoy her tea.

Test Sentence 1: Boiling water will burn you.

Test Sentence 2: People also drink coffee.

Question: Why was Mrs Jones careful when pouring the tea?

17. Cooking is a very popular hobby. People often like to make cakes or cook meals. They like to share them with friends. Bakers are people who cook for a living. Bakers have to spend a lot of time in the kitchen. Cooking can be fun but it can also be messy.

Test Sentence 1: A kitchen is for sleeping in.

Test Sentence 2: Bakers can make trees and flowers.

Question: Why do Bakers spend a lot of time in the kitchen?

18. Peter was walking home with his mum. He was not looking where he was going. He fell over a big stone. He hurt his arm and started to cry. His mum took him straight to the hospital. The doctor had to put a plaster on his arm and Peter was very brave.

Test Sentence 1: Hospitals are where you go when you are hungry.

Test Sentence 2: Sometimes doctors are men and sometimes they are cats.

Question: Why did Peter go to hospital?

19. When it is good weather, it is common for families to go on day trips. There are many places to go. Sometimes they go to the Zoo or maybe the seaside. Often they decide to pack up a picnic lunch. They sit in the sunshine and eat sandwiches, cakes and fruit. It is nice for families to spend this time together having fun.

Test Sentence 1: Picnics are meant to be eaten indoors.

Test Sentence 2: Picnics are often packed in milk bottles.

Question: What can people have in their picnics?

20. Mr Allan had taken a day off work. He did not know what to do. He decided to

spend the day relaxing. Mr Allan liked to take photographs. He put his camera in his bag and went for walk. Mr Allan had a lovely day taking pictures of rare birds.

Test Sentence 1: You take photographs with a radio.

Test Sentence 2: People hang pictures on the floor.

Question: Why did Mr Allan take a camera out walking?

21. In the Winter it is important to dress very warmly. It can very cold outside. You should wear both a shirt and a jumper. You need to keep your feet nice and warm. A pair of warm socks are a good idea in Winter. You should also wear gloves to keep your hands warm.

Test Sentence 1: You wear socks on your head.

Test Sentence 2: Shorts are worn in the winter.

Question: Why are socks a good idea in winter?

22. John was feeling very happy. He was excited because he only had one week to go until his birthday. John made a list of the names of the week. Each day of the week has a different name. John had seven different names written very carefully on his list. He put the list on his wall.

Test Sentence 1: a week has nine days.

Test Sentence 2: Saturday is a school day.

Question: Why did John write seven different names on his list?

23. Time helps to organise the day. You might have to start school at a certain time, or have lunch at a certain time. It is always better not to be late. Many people buy watches to wear on their arm. They usually arrive at their appointments on time. Some people tell the time by looking at the sun.

Test Sentence 1: You can tell the time by wearing a hat.

Test Sentence 2: People can use clocks to eat their dinner with.

Question: Why are people who were watches usually on time?

24. Mrs Jackson was having a party for her friends. She had made a big cake covered in fresh strawberries. She put lots of cream on top. But the cake needed to be kept cold. Mrs Jackson put it in the refrigerator. When her friends arrived they were all pleased with the cake.

Test Sentence 1: Refrigerators keep food warm.

Test Sentence 2: Food is cooked in a car park.

Question: Why did Mrs Jackson put the cake in the refrigerator?

25. Healthy eating is very important. It helps to stop you feeling ill. You shouldn't eat too many cakes and biscuits. Apples make a very healthy pudding. Fresh fruit is very good for you. But you can still eat cakes and biscuits as a special treat.

Test Sentence 1: Apples are a type of nut.

Test Sentence 2: Fruit is used for washing cars.

Question: Why are apples a healthy pudding?

26. Tom was writing a thank you letter to his Aunt Betty. He wanted to thank her for his birthday present. He took out his pad and pencil. When he had finished he looked in his address book. On the back of the envelope he had to write down where his Aunt lived. Then he put a stamp on the letter and put it in the post box.

Test Sentence 1: An address tells you how old someone is.

Test Sentence 2: Letters and parcels are delivered by the milkman.

Question: Why did Tom look in his address book?

27. People like to grow different things in their gardens. It can be very hard work. Flowers make a garden look very pretty. Red roses are a popular choice. But there are many flowers to choose from.

Test Sentence 1: Roses are a type of vegetable.

Test Sentence 2: Flowers grow well on the ceiling.

Question: Why do some people plant roses?

28. Beth was on holiday at the seaside. She was bored with sitting on the beach. So she decided to go for a walk. Beth saw an island that she had never seen before. She decided to hire a boat. She spent the day exploring and resting on the warm sand.

Test Sentence 1: Islands are surrounded by jelly.

Test Sentence 2: Boats float on red custard.

Question: Why did Beth decide to hire a boat?

29. People often like to go on summer holidays. But hotels can be very expensive. Some people look for different places to stay. It is often more fun to camp as you can sleep outdoors. Some people even pack tents in their luggage. They sleep in special beds called sleeping bags.

Test Sentence 1: When you go camping you sleep in a cupboard.

Test Sentence 2: Summer holidays are usually taken during the winter months.

Question: Why might you pack a tent to go on holiday?

30. Peter was having a party for his birthday. He was sitting at the tea table. His friends had come for tea. When he had blown out all the candles he made a wish. Then he shared his cake with all his friends. After tea they played some games.

Test Sentence 1: You usually have candles on a sandwich.

Test Sentence 2: Every year you become one year younger.

Question: Why did Peter not share his cake before he blew out his candles?

31. The police try to make sure that people don't break the law. But some people do still commit crimes. The police have to work very hard. They catch people who commit the crimes. They help to send them to prison. They work very hard to keep us safe.

Test Sentence 1: people who commit crimes are sent to school

Test Sentence 2: the police wear red spotty trousers and shirts.

Question: Why do the police send you to prison?

32. Billy Bat did not like the daytime. He liked to stay in bed. His friends always asked him to play. But Billy was nocturnal and only went out at night. He could find his way in the dark quite well. Billy liked to fly by the light of the moon.

Test Sentence 1: At night it is very light.

Test Sentence 2: At night most animals play football.

Question: Why did Billy Bat need to be able to find his way in the dark?

List Two

1. Once there was a woman who lived with her family in the forest. She decided to bake her children some cakes. She lived in a pretty little wooden house. To bake the cakes she lit a fire in her kitchen. The woman's little wooden house was totally destroyed. The lady was very cross with herself.

Test Sentence 1: Fire will burn wood.

Test Sentence 2: Fire keeps you warm.

Question: Why was the woman's house destroyed?

2. Some people find it very difficult to sleep. They need to find ways to relax. Some people like to read before they go to sleep. But some people try reading in the dark and this hurts their eyes. They often find that they need to wear glasses. It is always a good idea to keep a light on when you are reading at bedtime.

Test Sentence 1: If you have bad eyes you need glasses.

Test Sentence 2: Reading can help you to learn new things.

Question: Why should you read with a light on?

3. Tom was riding his bike. He thought he could smell smoke nearby. He carefully went to take a look. He could see a house

on fire. Tom ran to a phone and called the fire brigade. They arrived very quickly and knew what to do.

Test Sentence 1: Firemen put out fires

Test Sentence 2: Firemen wear yellow helmets.

Question: Why did Tom call the fire brigade when he saw the fire?

4. In Britain we do not always have nice warm weather. Often it is cold and sometimes it rains quite heavily. When it rains people often wear big raincoats. Many people carry large umbrellas in the rain. It is not very nice to get wet. If you wear wellington boots, you can jump in puddles without getting wet feet.

Test Sentence 1: Umbrellas protect you from the rain.

Test Sentence 2: When it rains rivers can flood.

Question: Why did Sam want an umbrella?

5. Jack and Jill were in the garden. They were playing catch with a ball. Jack threw the ball but Jill missed. The ball hit a window and it broke. The glass smashed all over the ground. They had to take care not to cut themselves.

Test Sentence 1: Glass is used to make windows

Test Sentence 2: Glass is made from sand

Question: What should you be careful of if you hit a window?

6. Britain is often very cold and sometimes it snows in Winter. Police warn drivers not to drive in bad conditions. They should only use the roads in an emergency. Often it is so cold that the snow freezes and the roads become icy. People often leave their cars at home and walk to work in the snow. Other people share a car so that the traffic on the roads is a lot lighter.

Test Sentence 1: Icy roads are dangerous for cars.

Test Sentence 2: you can slip on icy roads

Question: Why do people leave their cars at home?

7. John had started to make his mum a birthday card. He wanted to decorate it with lots of different coloured shapes. But the shapes had to be cut out of coloured paper with scissors. Scissors can be very sharp. John knew he should get an adult to help him with the scissors. With the help of his dad John finished the card on time.

Test Sentence 1: Sharp scissors can be dangerous

Test Sentence 2: You can use glue for sticking.

Question: Why did John ask his dad to help him cut with scissors?

8. In the summer it is common for people to eat ice-cream. Ice-cream vans are very busy. Ice-cream is very tasty and refreshing. When it melts it makes a mess. When it is very hot you have to eat ice-cream quickly. Ice-cream is very nice with chocolate sauce.

Test Sentence 1: Ice-cream melts when it is hot.

Test Sentence 2: Ice-cream comes in many different flavours.

Question: Why should you eat ice-cream quickly?

9. Mrs Brown lived in a house on a very busy road. One day Mrs Brown found a big woolly hat in her drawer. At night she pulled the hat right down over her ears before she went to bed. The traffic on the road would be noisy all through the night. Mrs Brown found it difficult to sleep. But with her hat she slept till morning.

Test Sentence 1: Noise keeps you awake.

Test Sentence 2: Driving can be dangerous at night.

Question: Why did Mrs Brown find it difficult to sleep?

10. Your teeth are important because they help you to chew your food. They also help you to speak properly. It is very important to look after your teeth carefully. If you don't look after them you get a toothache.

You have to go to the Dentist. The Dentist tries to make sure that everyone has a lovely smile.

Test Sentence 1: Dentists look after your teeth.

Test Sentence 2: You should brush your teeth.

Question: Why do you have to go to the dentist if you have a toothache?

11. Jack had to go to see the Doctor. He was covered in red spots and felt very hot. The doctor told Jack he had the measles. Jack couldn't go to school because he was ill. There were lots of other children in his school. He had to stay in bed all day and take some medicine.

Test Sentence 1: Sometimes other people can catch your illness.

Test Sentence 2: You might sneeze with a runny nose.

Question: Why could Jack not go to school when he was ill?

12. Many people are frightened of snakes. But snakes are really very shy animals who don't mean any harm. They don't like humans and run away whenever they see one. Sometimes snakes might bite you. Some snakes can make you very ill. But snakes will only bite if they cannot escape.

Test Sentence 1: Some snakes are poisonous

Test Sentence 2: Snakes shed their skin.

Question: How can some snakes make you ill?

13. Mr and Mrs Smith were going to visit their friends in London.

They had not seen them for a long time.

They packed their bags and left very early one morning. They were going to travel to London on the train. The railway station was very busy. When the train arrived they both hopped aboard excitedly.

Test Sentence 1: Trains leave from stations.

Test Sentence 2: Aeroplanes leave from airports.

Question: Why did Mr and Mrs Smith go to the railway station?

14. Fish can live in rivers, streams, lakes and oceans. Some people like to keep fish in big tanks in their houses. Fish can be lots of different colours and look very pretty. But all fish need water to breathe. Fish cannot live for long out of the water. Fish breathe through slits in their sides called gills. Test Sentence 1: Breathing keeps you alive.

Test Sentence 2: People often keep goldfish.

Question: Why can fish not live long out of the water?

15. Mr Green enjoyed spending a lot of time working in his garden. He worked very

hard to keep it tidy. When the summer came his garden looked beautiful. He liked to grow different flowers and plants in his garden. He watered his plants very carefully every day. Mr Green was very pleased with himself and his garden looked beautiful.

Test Sentence 1: Plants need water to grow

Test Sentence 2: Gardens usually have some grass.

Question: Why did Mr Green water his plants every day?

16. Lots of people in England like to drink tea. They drink it in the afternoons. To make tea you need to boil water in a kettle. Boiling water is always very hot. You should always be careful when pouring tea. You can add milk and sugar to tea but some people drink it with lemon slices.

Test Sentence 1: Boiling water will burn you.

Test Sentence 2: People also drink coffee.

Question: Why should you be careful when pouring tea?

17. It was very late one evening. Mr Smith had just had a very large dinner. But he was still feeling very hungry. So he decided to cook himself a snack. He went into the kitchen but the cupboards were all empty. So poor Mr Smith went to bed feeling very miserable.

Test Sentence 1: A kitchen is for sleeping in.

Test Sentence 2: Bakers can make trees and flowers.

Question: Why was Mr Smith in the kitchen?

18. Doctors and nurses are very skilled people. They work with lots of people every day. They help people who are sick. They also help people who are hurt. You can find doctors and nurses in hospitals. Doctors and nurses have to train very hard to learn how to do their jobs.

Test Sentence 1: Hospitals are where you go when you are hungry.

Test Sentence 2: Sometimes doctors are men and sometimes they are cats.

Question: Why do Doctors and Nurses work in hospitals?

19. It was a nice day and Jack and Kate were bored with being at home. They didn't want to be inside. So they rode their bikes down to the duck pond. They had decided to pack up a picnic lunch. They sat in the sunshine and ate their sandwiches and cakes. They were glad they had decided to go out for the day.

Test Sentence 1: Picnics are meant to be eaten indoors.

Test Sentence 2: Picnics are often packed in milk bottles.

Question: What did Jack and Kate have in their picnic?

20. People often do things in their spare time. Some people play sports or watch films. Other people have more relaxing hobbies. Many people like to take photographs. They pack up their camera and go out walking for the day. It can be a very rewarding way to spend your time.

Test Sentence 1: You take photographs with a radio.

Test Sentence 2: People hang pictures on the floor.

Question: Why might people take a camera out walking?

21. It was winter and the ground was covered in snow. Tom wanted to build a snowman. He wrapped up warm and went outside to play. The snow was cold but Tom's feet were warm. He was wearing some warm socks that he bought for Winter. He built a tall snowman with a long carrot nose.

Test Sentence 1: You wear socks on your head.

Test Sentence 2: Shorts are worn in the winter.

Question: Why did Tom wear warm socks?

22. A month has four weeks. Most people think that Monday is the first day of the week. But the first day of the week is

actually Sunday. Each day of the week has a different name. Seven different names are used for the days of the week. The days of the week are named after Greek and Roman Gods and Goddesses.

Test Sentence 1: A week has nine days.

Test Sentence 2: Saturday is a school day.

Question: Why are seven different names used for the days of the week?

23. Sam was usually late for school. He seemed to take a long time to get ready. But Sam did not like being late. One morning Sam was given a watch to wear. That morning, Sam was on time for school. Sam would not be late for school ever again.

Test Sentence 1: You can tell the time by wearing a hat.

Test Sentence 2: People can use clocks to eat their dinner with.

Question: Why was Sam on time for school when he wore a watch?

24. You can buy all your food at the supermarket. Some food comes in cans and will last a long time. It can be kept in a cupboard. But other food needs to be kept cold. Many people have refrigerators in their kitchens. Fruit and vegetables need to be eaten quickly or they become rotten.

Test Sentence 1: Refrigerators are used to keep food warm.

Test Sentence 2: Food is cooked in a car park.

Question: Why do people have refrigerators in their kitchens?

25. Beth had finished her dinner. But she still felt a bit hungry. She didn't know what to have for pudding. She decided to have an apple. So she went to the fruit bowl. She found a lovely red apple and sat down to eat it.

Test Sentence 1: Apples are a type of nut.

Test Sentence 2: Fruit is used for washing cars.

Question: Why did Beth go to the fruit bowl for her apple?

26. You can write a letter for a lot of different reasons. But you will always need to post it properly. A finished letter goes in an envelope. Then you might need to look in your address book. On the back of the envelope you need to write down where the person lives. Then you put a stamp on the letter and put it in the post box.

Test Sentence 1: An address tells you how old someone is.

Test Sentence 2: Letters and parcels are delivered by the milkman.

Question: Why do you need your address book when writing a letter?

27. Mrs Jones was feeling very tired and fed up. She had been working very hard. So Mr Jones decided to cheer her up. Mrs Jones liked colourful flowers Mr Jones bought some red roses. Mrs Jones was thrilled and felt much happier.

Test Sentence 1: Roses are a type of vegetable.

Test Sentence 2: Flowers grow well on the ceiling.

Question: Why did Mr Jones buy red roses to cheer up Mrs Jones?

28. England is a very small, cold country. But it attracts many different visitors every year. England is part of the British Isles. The British Isles is made up of two small islands. You can reach England by boat. But you can also fly or travel through the Channel tunnel.

Test Sentence 1: Islands are surrounded by jelly.

Test Sentence 2: Boats float on red custard.

Question: Why might you need a boat to reach England?

29. Kim was getting ready to go on holiday. She was busy packing her things. She wanted to make sure she remembered everything. She was going camping with her friends and she was very excited. She packed her tent carefully in her luggage.

She loved camping and was looking forward to it.

Test Sentence 1: When you go camping you sleep in a cupboard.

Test Sentence 2: Summer holidays are usually taken during the winter months.

Question: Why did Kim pack her tent?

30. Your birthday is the day you were born. You can enjoy your birthday every year. You might have a birthday party. When you blow out the candles you can make a wish. Then you can share your cake with your friends. You might even get some presents.

Test Sentence 1: You usually have candles on a sandwich.

Test Sentence 2: Every year you become one year younger.

Question: Why must you wait until after you blow out the candles before sharing your cake?

31. The thief had tried to hide the stolen diamonds in his pocket. Then he climbed out of the window. The police saw the thief running away. They caught the thief committing the crime. They helped to send him to prison. The thief had to give back the diamonds.

Test Sentence 1: People who commit crimes are sent to school

Test Sentence 2: The police wear red spotty trousers and shirts.

Who did the police send to prison?

32. Many animals come out during the day. That is when they go hunting. But not all animals like the daytime. Some animals are nocturnal and only go out at night. They are able to find their way in the dark. They use the light of the moon to guide them.

Test Sentence 1: At night it is very light.

Test Sentence 2: At night most animals play football.

Question: Why do some animals need to be able to find their way in the dark?

Filler Passages

1. Sam wanted a bike for Christmas. On Christmas morning he looked everywhere in the house. But he could not see a bike anywhere. Then his mum opened the front door. In the front garden was a shiny new mountain bike. Sam was so pleased that he gave his mum a big kiss.

Test Sentence 1: Christmas is in March.

Test Sentence 2: Bikes have a saddle and pedals.

Question: Why was Sam so pleased?

2. Polly the cat was very thirsty. But her owners were in bed. So Polly slipped into the kitchen. She opened the fridge with her paw. Then she pushed the milk bottle over. The milk went all over the floor and Polly licked it up happily.

Test Sentence 1: Cats have four legs.

Test Sentence 2: Milk comes from pigs.

Question: Where did Polly find the milk?

3. In the Autumn Jack liked to go blackberry picking. He would go out with an empty basket and try to fill it with juicy blackberries. Sometimes he would eat some on his way home. But he saved most of them for his mum. She would make his favourite blackberry pie. He liked to eat it with fresh cream.

Test Sentence 1: Blackberries are yellow.

Test Sentence 2: Blackberries grow on bushes.

Question: Why did Jack save blackberries for his mum?

4. Susan and Peter went to the circus. They sat at the front and watched the elephants in the ring. Suddenly the clowns came on and threw water at the crowd. Susan got very wet. Peter laughed out loud. Then the clowns threw water over him too.

Test Sentence 1: You can see clowns at the supermarket.

Test Sentence 2: There are lots of animals in the circus.

Question: How did Susan get wet?

5. The cowboy was riding his horse very fast. The Indians were chasing him through the mountains. The cowboy rode into a cave. He stayed very still and quiet. He watched the Indians ride straight past him. Then he rode out to safety.

Test Sentence 1: Cowboys ride bicycles.

Test Sentence 2: Horses wear saddles when you ride them.

Question: Why did the Indians ride past the cowboy?

6. The fisherman was lost in a storm. His boat was being thrown from side to side. Suddenly, he saw a light in the distance. It was the lighthouse telling him which way to

go. The fisherman sailed with all his might towards the light. Soon he was safely on dry land with the lighthouse keeper.

Test Sentence 1: Fishermen catch wild cats.

Test Sentence 2: A lighthouse tells boats which way to go.

Question: How did the fisherman know which way to go?

7. John and Pam were playing football.

John ran down the pitch with the ball. He passed the ball to Pam. Pam kicked it towards the goal. The goalkeeper jumped for the ball but he missed it. The ball hit the back of the net.

Test Sentence 1: A ball is shaped like a square.

Test Sentence 2: The goalkeeper tries to save the ball.

Question: Were John and Pam on the same team?

8. The spy crept into the room through the window. He walked over to the wall. He found the safe behind a picture. He cracked the code and opened the safe. He had found the secret plans. Now the spy could rescue the prisoners.

Test Sentence 1: A spy is someone who plays the guitar.

Test Sentence 2: A safe is where you keep valuable things.

Question: How did the spy open the safe?

9. Computer games have become very popular. There are many different types of games. Some games let you play sports. But in other games you have to solve puzzles. It is not just boys who play computer games. Girls like computer games too.

Test Sentence 1: You can play computer games with other people.

Test Sentence 2: You can play computer games with your nose.

Question: Name two different types of computer games.

10. Lots of people like to go to dancing lessons. Some people like to learn ballet or tap dancing. Other people like more modern sorts of dancing. You can even learn how people dance in other countries. You can start dancing at a very young age. Many people like to watch dancing shows.

Test Sentence 1: Boys and girls can learn to dance.

Test Sentence 2: To learn to dance you need huge feet.

Question: Name two different sorts of dancing.

11. A carnival is a special sort of party. It is usually held in the street so anyone can go. People dress up and decorate trucks. Then the trucks parade through the streets. There

is usually some music and dancing.

Families can enjoy carnivals together.

Test Sentence 1: Carnivals are a lot of fun.

Test Sentence 2: You have a party when you are naughty.

Question: Where are carnivals usually held?

12. A flute is a musical instrument. It is usually made from metal or wood. It is shaped like a long tube with holes in it. You play it by blowing through a hole in one end. If you cover different holes with your fingers you can play different notes. A flute has a very soft sound.

Test Sentence 1: A guitar is a musical instrument.

Test Sentence 2: You use musical instruments to make a cake.

Question: What shape is a flute?

13. A hippopotamus is a very large animal. It has very short legs and thick wrinkled skin. But it can move very fast. A hippopotamus is often called a hippo for short. Hippos live in Africa on the banks of the rivers. They are usually grey in colour.

Test Sentence 1: Hippos are very large.

Test Sentence 2: Hippos are red in colour.

Question: Is a hippopotamus fast?

14. A helmet is a kind of hat. You wear a helmet to protect your head. Soldiers wear helmets. But you also need helmet if you

ride a motorbike. Knights used to wear big heavy metal helmets. But they are much lighter to wear now.

Test Sentence 1: You wear a helmet on your head.

Test Sentence 2: Helmets are very soft.

Question: Name two people who might wear a helmet.

15. Breakfast is the most important meal of the day. You need to eat breakfast so that you have lots of energy. Some people have big cooked breakfasts. Others have cereal or toast. Orange juice is nice with breakfast. You should have breakfast every day.

Test Sentence 1: Breakfast is eaten before lunch.

Test Sentence 2: You usually have fish and chips for breakfast.

Question: Name two things you might have for breakfast?

16. The capital city of England is London. There are lots of important buildings in London. London is where the Queen and the Prime Minister live. London is a big city and is always very busy. The Millennium Dome is in London. It is an exciting place to live.

Test Sentence 1: London is in England.

Test Sentence 2: The Queen is called Henry.

Question: What city does the Queen live in?

17. Harry had got his kite stuck in a tree. He was very upset because it was too high for him to reach. He started to cry. Then his dad came along with a ladder. He climbed the ladder and set the kite free. Then Harry and his dad flew the kite together for the rest of the day.

Test Sentence 1: Kites fly in the sky.

Test Sentence 2: You can skip up ladders.

Question: Why did Harry's dad need a ladder?

18. Mr Walker's dog liked to chase cats. Mr Walker couldn't keep up with his dog. So Mr Walker used to wear roller skates when they went for a walk. Then he could hold his dog's lead and be pulled along on his skates. All the people used to stop and stare. But Mr Walker just waved happily as he rolled along.

Test Sentence 1: Roller skates are worn on the feet.

Test Sentence 2: Dogs like to chase elephants.

Question: Why did Mr Walker need roller skates?

19. Mr White the Milkman used to get up very early. He would climb into his float while it was still dark. Then he would deliver bottles of milk to all the people in

his town. He climbed back into bed just as everyone else was getting up. Every morning they would find fresh milk on the doorstep. And it was all thanks to Mr White.

Test Sentence 1: You can put milk in tea.

Test Sentence 2: Milkmen deliver fresh books.

Question: Why was it still dark when Mr White delivered the milk?

20. The sailor stood on the deck of his boat. He could see the pirate ship sailing towards him. He loaded his cannons. Then he fired his cannons at the pirates. The pirate ship sank and the pirates were floating in the sea. The sailor sailed over and rescued the pirates.

Test Sentence 1: Boats sail on water.

Test Sentence 2: Sailors wear ball dresses.

Question: Why did the pirate ship sink?

21. Mr Smith wanted a new car. He saw a lovely red sports car in the garage. So he went to the bank and drew out some money. Then he went back to the garage. Soon he was driving his red sports car out into the country. Mr Smith was very pleased with himself.

Test Sentence 1: You can keep your money in a bank.

Test Sentence 2: You can buy cars at a zoo.

Question: Why did Mr Smith go to the bank for some money?

22. The little girl was playing in the sand. She was building a castle with her bucket and spade. She did not notice the big wave coming towards her. The wave washed the castle away. The little girl was very upset. She started to cry.

Test Sentence 1: You can build sandcastles on a beach.

Test Sentence 2: Buckets are used for making milkshakes.

Question: What sort of castle did the little girl build?

23. Tom and Sue paddled their boat in the warm water. The sea was warm and the beach was in sight. Suddenly, a big wave came and swept the boat further out to sea. The paddles were lost overboard. Tom told Sue to use her hands to steer the dinghy. Together they guided it to safety.

Test Sentence 1: Paddles help steer a boat.

Test Sentence 2: A beach is usually made of sand.

Question: Why did Sue use her hands to steer the boat?

24. The bride looked beautiful as she walked into the church. Her bridesmaids followed her carrying flowers. It was the best day of her life. She smiled as she

walked towards the groom. They knew they would be very happy. All the people in the church were crying with joy.

Test Sentence 1: You can get married in a church.

Test Sentence 2: A bride wears a dress on her wedding day.

Question: Were the people in the church sad?

25. There are many different kinds of aeroplanes. Some aeroplanes are used to take people on holiday. But other aeroplanes are used to carry animals and other goods. Some people have their own aeroplanes. These are usually quite small. The biggest aeroplane is called a Jumbo Jet.

Test Sentence 1: Aeroplanes fly in the sky.

Test Sentence 2: Aeroplanes have wings.

Question: Name two things aeroplanes are used for.

26. When you are young you have to go to school. Many children do not like school. But going to school is very important. You learn lots of new things every day. Going to school will help you to get a good job. You could even go to college.

Test Sentence 1: Teachers work in schools.

Test Sentence 2: You can play sports at school.

Question: Name two reasons why going to school is important.

27. Elephants live in Africa and India. Elephants are very large animals. They have long noses called trunks. They are usually grey in colour. Male elephants are called bulls. They have horns on the sides of their face called tusks.

Test Sentence 1: Elephants are very good dancers.

Test Sentence 2: Elephants have big ears.

Question: Do all elephants have tusks?

28. Dolphins are a kind of whale. They usually grow to about two or three metres long. Dolphins are very friendly animals. They live in groups in the sea. They are also very clever. Many people like to swim with dolphins.

Test Sentence 1: Whales live in the forest.

Test Sentence 2: Dolphins are very clever.

Question: Do dolphins live alone?

29. A banana is a popular fruit. It is shaped like a crescent moon. When ripe it is yellow in colour. Bananas have soft creamy flesh. They are sweet and tasty. Bananas are very good for you.

Test Sentence 1: Bananas are orange.

Test Sentence 2: Bananas are a favourite food of monkeys.

Question: Tell me three things about bananas?

30. A library is a building full of books. There are all different sorts of books there. If you join a library you can borrow the books. You can take them home to read. But you have to take them back on time. Or you might have to pay a fine.

Test Sentence 1: Books are made of bread.

Test Sentence 2: You can join a library.

Question: Why might you have to pay a fine?

31. Every day people buy newspapers. Newspapers are a good way of finding out all the world news. Good newspapers usually have news from this country and lots of other countries. You can find out what the weather is going to be like from a newspaper. You can also read about films and television. Sometimes there is a crossword to do as well.

Test Sentence 1: Newspapers are written in felt-tip pen.

Test Sentence 2: Newspapers are sold at the newsagents.

Question: What can you read about in the newspapers?

32. Mount Everest is the highest mountain in the world. It is in a mountain range called the Himalayas. The Himalayas are in south east Asia. Mount Everest is nearly nine thousand metres above sea level. Many people have tried to climb the

mountain. But not very many of them have made it right to the top.

Test Sentence 1: Mount Everest is a river.

Test Sentence 2: Mount Everest is the highest mountain in the world.

Question: Where in the world is Mount Everest?

Appendix II

Experiment 2b

Experiment 2b was conducted in order to assess the relative difficulty of the test sentences included in the materials used in experiment 2. Response times to test sentences 1 and 2 were compared when the sentences were presented without the preceding passage content.

A2.1 Method

a) Participants: Eighteen of the children who participated in Experiment 2 completed this follow-up task (Mean age 115.44 months, standard deviation 4.08 mths; Mean SRS standard score 101.39, standard deviation 7.01).

b) Design: This was a one-way repeated measures design with the within-subjects factor of Test Sentence (TS1 and TS2). The factor Content was not included in this analysis since it was not directly related to the objective of the study. All participants completed 32 trials in each condition.

c) Materials: The test sentences used in the original experiment were compiled into one list for use in this study.

d) Procedure. Each child was seen individually for one session lasting approximately 20 minutes. The child was told that they would be presented with sentences on a computer screen which they had to judge as being true or false. The sentences were presented in a random order. They were presented in the middle of the screen and each was preceded by a centrally located asterix which appeared for 500msecs. The children were told to read each sentence to themselves and then to press the button on the keyboard labelled T if the sentence was true or the button labelled F if they thought it was false. As soon as they had made a response to an item the asterix would reappear followed by the next item. They were told to work their way through all the sentences and to answer as quickly and as accurately as possible.

A2.2 Results: The mean response times and error rates for each test sentence are shown in Table A2.1. A t-test was conducted on the response time and accuracy data.

No significant difference was found between response times to TS1 and TS2 ($t(17) = 0.06$, ns), nor between the number of correct responses to TS1 and TS2 ($t(17) = 1.43$, ns).

	TS1	TS2
Mean response time	4081.41 (1274.20)	4075.03 (1065.71)
Mean No Correct Responses (max 32)	31.05 (0.99)	30.78 (0.65)

Table A2.1. Mean response time and mean number of correct responses to each test sentence (standard deviations shown in parentheses).

A2.3 Summary: The data of 18 children who took part in both the original and the follow-up experiment were compared. As predicted the response times to test sentence 1 and 2 were comparable when the preceding passage context was not available. Thus, the data suggest that the test sentences were matched in terms of difficulty level.

Appendix III

Experiment 3 Test Materials

List One

1. John looked behind the wagon and saw the Red Indians approaching. He called for his father. Henry grabbed his rifle and pulled himself up to the front of the wagon. As the Red Indians rode closer, John could see they were carrying weapons. He raised his own weapon and aimed at the enemy. But Henry held up his hand. John took his finger off the trigger and carefully lowered his gun.

Test Sentence 1: Henry wanted John to wait (causal).

The Indians were so close that John could see their brightly coloured faces. John's hand was firmly on his rifle as he waited in the wagon. Then Henry gave the signal and John and Henry opened fire. The battle lasted only a few minutes but seemed a lot longer. Finally, the Indians had weakened. They turned and disappeared into the mountain range.

Test Sentence 2: The Indians rode horses. (static)

Question: What was John holding as he waited in the wagon?

2. The fishermen were lost in a storm. Their boat was being thrown from side to side by

the wind. Many of their belongings had been lost overboard and a huge wave had crashed across the deck, smashing the mast. The men struggled to keep control of the wheel. They used all their strength to fight the wind and rain.

Test Sentence 1: The men were soaking wet. (static)

The men knew they were running out of time. The boat was being forced towards the rocks. The Captain knew they could not save the ship and he ordered the men into the life boat. The ship crashed into the rocks. A huge hole appeared in the side of the hull. The men watched as the boat sank beneath the waves.

Test Sentence 2: The ship filled with water. (causal)

Question: What did the Captain order the men to do?

3. The spy dropped from the wall surrounding the large house. He pulled out his torch and made his way over to the open window. As quietly as possible, he pulled himself over the window ledge and crept into the office. He found the safe hidden under the desk. The spy entered the code into the panel on the safe door. He found

the secret plans in the safe.

Test Sentence 1: The code opened the safe

Carefully, the spy photographed the plans. Then he slipped them back into the safe where he had found them. With the camera film safely in the bottom of his backpack, the spy slipped out of the office the way he came in. In the darkness, he scaled the wall once more. His mission complete, he disappeared into the dark night.

Test Sentence 2: The spy was dressed in black. (static)

Question: What did the spy find in the safe?

4. Emily sat happily in the sand at the edge of the sea. She was building a castle with her bucket and spade. Carefully she filled her bucket to the top with sand. Then she turned the bucket over in the sand. Emily tapped it gently with her spade. Then she lifted the bucket to reveal a perfect sand castle.

Test Sentence 1: The sand in the bucket was wet. (static)

Emily was very pleased with herself. She loved building castles on the beach. Emily set about filling her bucket again. She wanted to build more castles. But Emily had

not noticed that a wind had begun to blow in from the sea. She did not see the huge wave coming towards her. The wave crashed heavily onto the sand. When the wave finally rolled away, Emily began to cry.

Test Sentence 2: The wave washed the castle away. (causal)

Question: What was Emily building?

5. The cowboy was riding his horse at great speed through the mountains. He could hear the bandits drawing closer. The cowboy could feel his horse was tiring. He knew he had to stop but he had to evade capture. Taking a risk, the cowboy rode into a cave hidden behind a rock and waited. Soon he heard the bandits ride past the cave.

Test Sentence 1: The bandits could not see the cowboy. (causal)

From the safety of the cave, the cowboy watched as the bandits searched for him in the valley below. They remained there, circling on their horses for some time. They were clearly disappointed to have lost their prey. When darkness fell, the leader ordered the men back to the hideout. They would have to think of another way to capture the cowboy.

Test Sentence 2: The bandits were carrying rifles. (static)

Question: What was wrong with the cowboy's horse?

6. The thief looked cautiously around as he lifted the money from the safe. He wanted to be sure nobody had seen him break into the bank. He put the money into his hold all and left by the back door. He had almost reached his car when he saw the two policemen. They were waiting for him in the shadows.

Test Sentence 1: The policemen wore blue uniforms. (static)

The thief threw the hold all into the bushes and fumbled with the key for his car. But it was too late. The policemen caught him before he could make his escape. At the police station the villain denied the robbery. But he was convicted because of his own foolishness. He had forgotten to wear gloves. The police were able to place the thief at the scene of the crime.

Test Sentence 2: The thief left fingerprints at the bank. (causal)

Question: What did the thief do with his hold-all?

7. The football match was coming to an

end. Neither team had scored a goal and the crowd was tense with excitement. Tom was racing towards the goal with the ball. He took a shot but it was too wide. The crowd groaned as the ball went back into play. It looked as though no one was going to win. Suddenly, Jan scored from the edge of the pitch. Tom cheered loudly and ran over to hug Jan.

Test Sentence 1: Tom and Jan were on the same team. (causal)

The whistle blew signalling the end of the match. The crowd was cheering loudly for Jan. Her team-mates gathered round her, lifted her on to their shoulders and carried her off the pitch. Jan was treated like a hero. She was declared man of the match for scoring the winning goal.

Test Sentence 2: Jan was a good football player. (static)

Question: How did everyone know that the match was over?

8. John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. Francis fired his pistols one after the other. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.

Test Sentence 1: The bear had huge paws.
(static)

Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. Snarling, the bear swiped at the girls. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear fell to the ground and did not move.

Test Sentence 2: John shot the bear dead.
(causal)

Question: Where did the bullet hit the bear?

9. The Duchess flung the baby at Alice and hurried out of the room. Alice caught the baby with some difficulty for it was a queer shaped creature and held out its arms and legs in all directions. The baby had been crying but now it started grunting. When Alice looked down the baby's nose had turned up and his eyes had become very small. Alice was holding a pig.

Test Sentence 1: The baby had become a pig. (causal)

Alice felt that she could not carry the baby any further. She thought it was quite an ugly child, but was quite handsome as a pig. She was sure the tiny child would be happier in

the wood with the other animals. So Alice set the little baby down. She felt quite relieved to see it trot away quietly into the wood.

Test Sentence 2: The baby walked on pigs trotters. (static)

Question: What did Alice think of the baby?

10. The Witch was very cruel to Dorothy and Toto. She made Dorothy work very hard all day and threatened to hit her. Once the Witch struck Toto so hard with her umbrella that he flew at her and bit her hard on the leg. But the Witch could not feel pain because she was too wicked. Instead the old woman just laughed.

Test Sentence 1: The old woman flew on a broomstick. (static)

Dorothy was very angry. She shouted at the Witch to stop being so cruel. But the Witch just sneered and ordered her to get back to work. The little girl was so mad, she picked up a bucket of water and threw it at the Witch. The Witch was soaked with water from head to foot. Dorothy watched as the Witch melted away.

Test Sentence 2: The water made the wicked witch melt. (causal)

Question: What did the Queen use to hit

Toto?

11. When the battle was over, Lucy found Edmund a little way back from the fighting line. He was covered with blood, and his face was a nasty green colour. Lucy remembered the precious red liquid she had been given for a Christmas present. She poured a few drops of red liquid into Edmund's mouth. Edmund leapt to his feet in perfect health.

Test Sentence 1: The red liquid healed Edmund's wounds. (causal)

Edmund fought well in the battle. He had beaten the giants and broken the Witch's wand. Without him the battle may have been lost. Aslan the lion walked slowly towards the boy. He was proud of Edmund. The lion told Edmund he was to become a knight. Then the lion let Edmund ride on his back

Test Sentence 2: The lion was able to talk. (static)

Question: What had happened to the Witch's wand?

12. Lucy, Edmund and Eustace were staring at a picture on the wall. It was a picture of a ship sailing on an ocean. But it was no ordinary picture. The children could hear

the wind blowing and the water splashing the side of the boat. The boat was bobbing up and down in the waves. The sea was crashing against the rocks. The picture seemed to be moving violently.

Test Sentence 1: The sea was moving in the picture. (static)

Eustace rushed to pull the picture off the wall. Edmund and Lucy tried to stop him. But the children found themselves standing on the picture frame. There was no glass in front of them, just real sea. A gust of wind blew them off the frame. They struggled to stay afloat in the stormy sea.

Test Sentence 2: The children fell into the picture. (causal)

Question: What did Eustace try to do?

13. One night Sam woke up with a start. He heard a strange sound coming from outside his window. Sam crept out of bed, hid behind the curtain and looked outside. He let out a small scream. An enormous spaceship appeared to have landed in his garden. The ship filled his garden completely. The roses in the flower-bed were completely crushed.

Test Sentence: The spaceship landed on the roses. (causal)

A door opened in the side of the ship and a strange figure appeared. It looked almost like a man but had long arms, short legs and a huge head with haunting black eyes. Sam cried out and the creature looked up in surprise. Then the door closed and the ship rose slowly off the ground. Quick as a flash it was gone.

Test Sentence 2: The ship was covered in flashing lights. (static)

Question: What did Sam see in his garden?

14. The mermen lived far out to sea, where the water is blue as the bluest cornflower. The merman king had six beautiful daughters and the youngest was the prettiest of them all. Her skin was soft and delicate as a rose leaf and her eyes were as blue as the deepest sea. But the young daughter could only swim in the sea. She longed to walk on the land and feel the sand between her toes.

Test Sentence 1: The girl had a fish tail. (static)

One night, the mermaid rescued a prince from the stormy sea. They fell in love, but the mermaid could not live out of the water. So the mermaid sang sadly to the prince every night. The merman king could not bear to see his daughter so sad. He cast a

spell over the mermaid and her tail disappeared. She ran happily to meet the prince on the shore.

Test Sentence 2: The spell gave the girl legs. (causal)

Question: Who was the prettiest mermaid?

15. There once was a king who had eleven sons and one daughter. The king married a wicked queen who was jealous of the children. She sent the little girl away and made the king banish his sons from the kingdom. On the day the boys left the palace, the queen cast a spell over them. Instead of eleven boys, eleven wild swans flew from the palace gates.

Test Sentence 1: The queen turned the boys into wild swans. (causal)

The swan princes searched for their sister. When they found her she promised to break the spell. She made eleven magic shirts for her brothers. When she threw the shirts over the swans, they were transformed into eleven handsome princes. But the girl did not have enough magic cloth to finish all the shirts. The youngest prince would forever have a swan's wing in the place of an arm.

Test Sentence 2: His arm was covered in

white feathers. (static)

Question: How many sons did the king have?

16. Long ago there lived a young girl named Brave Heart who had learned to hunt. One winter, when the young men were out hunting, the women of the village saw some warriors heading towards them. They were filled with fear. With the men out hunting they did not know who would protect them. The young girl vowed to defend the women. She took up her bow and arrow and circled the villagers.

Test Sentence 1: The girl was an indian

List 2

1. John looked behind the wagon and saw the Red Indians approaching. He called for his father. Henry grabbed his rifle and pulled himself up to the front of the wagon. As the Red Indians rode closer, John could see they were carrying weapons. He raised his own weapon and aimed at the enemy. But Henry held up his hand. John took his finger off the trigger and carefully lowered his gun.

Test Sentence 1: John wore a cowboy hat. (static)

brave. (static)

The women watched in wonder as Brave Heart transformed herself into a rainbow. The beautiful rainbow arched over the village. The warriors too saw the mighty rainbow arch above them. Brave Heart stood in the midst of the rainbow. The young girl raised her bow and started to shoot. Lightening fell as the warriors fled from the village.

Test Sentence 2: The arrows were made of lightening. (causal)

Question: Why were the women frightened.

The Indians were so close that John could see their brightly coloured faces. John's hand was firmly on his rifle as he waited in the wagon. Then Henry gave the signal and John and Henry opened fire. The battle lasted only a few minutes but seemed a lot longer. Finally, the Indians had weakened. They turned and disappeared into the mountain range.

Test Sentence 2: The Indians were beaten. (causal)

Question: What was John holding as he waited in the wagon?

2. The fishermen were lost in a storm. Their

boat was being thrown from side to side by the wind. Many of their belongings had been lost overboard and a huge wave had crashed across the deck, smashing the mast. The men struggled to keep control of the wheel. They used all their strength to fight the wind and rain.

Test Sentence 1: The boat was hard to steer (causal)

The men knew they were running out of time. The boat was being forced towards the rocks. The Captain knew they could not save the ship and he ordered the men into the life boat. The ship crashed into the rocks. A huge hole appeared in the side of the hull. The men watched as the boat sank beneath the waves.

Test Sentence 2: The men wore lifejackets. (static)

Question: What did the Captain order the men to do?

3. The spy dropped from the wall surrounding the large house. He pulled out his torch and made his way over to the open window. As quietly as possible, he pulled himself over the window ledge and crept into the office. He found the safe hidden under the desk. The spy entered the code into the panel on the safe door. He found

the secret plans in the safe.

Test Sentence 1: The spy carried a gun. (static)

Carefully, the spy photographed the plans. Then he slipped them back into the safe where he had found them. With the camera film safely in the bottom of his backpack, the spy slipped out of the office the way he came in. In the darkness, he scaled the wall once more. Without a sound, the spy disappeared into the dark night.

Test Sentence 2: The spy escaped safely. (causal)

Question: What did the spy find in the safe?

4. Emily sat happily in the sand at the edge of the sea. She was building a castle with her bucket and spade. Carefully she filled her bucket to the top with sand. Then she turned the bucket over in the sand. Emily tapped the bucket gently with her spade. Then she lifted the bucket and looked underneath.

Test Sentence 1: Emily built a sandcastle. (causal)

Emily was very pleased with herself. She loved building castles on the beach. Emily set about filling her bucket again. She

wanted to build more castles. But Emily had not noticed that a wind had begun to blow in from the sea. She did not see the huge wave coming towards her. The wave crashed heavily onto the sand. When the wave finally rolled away, Emily began to cry.

Test Sentence 2: Emily wore a swimming suit. (static)

Question: What was Emily building?

5. The cowboy was riding his horse at great speed through the mountains. He could hear the bandits drawing closer. The cowboy could feel his horse was tiring. He knew he had to stop but he had to evade capture.

Taking a risk, the cowboy rode into a cave hidden behind a rock and waited. Soon he heard the bandits ride past the cave.

Test Sentence 1: The cowboy had a gun and holster. (static)

From the safety of the cave, the cowboy watched as the bandits searched for him in the valley below. They remained there, circling on their horses for some time. They were clearly disappointed to have lost their prey. When darkness fell, the leader ordered the men back to the hideout. They would have to think of another way to capture the cowboy.

Test Sentence 2: The bandits could not hunt in the dark. (causal)

Question: What was wrong with the cowboys horse?

6. The thief looked cautiously around as he lifted the money from the safe. He wanted to be sure nobody had seen him break into the bank. He put the money into his hold all and left by the back door. He had almost reached his car when he saw the two policemen. They were waiting for him in the shadows.

Test Sentence 1: The police wanted to catch the thief. (causal)

The thief threw the hold all into the bushes and fumbled with the key for his car. But it was too late. The policemen caught him before he could make his escape. At the police station the villain denied the robbery. But he was convicted because of his own foolishness. He had forgotten to wear gloves. The police were able to place the thief at the scene of the crime.

Test Sentence 2: The policemen wore blue uniforms. (static)

Question: What did the thief do with his hold-all?

7. The football match was coming to an

end. Neither team had scored a goal and the crowd was tense with excitement. Tom was racing towards the goal with the ball. He took a shot but it was too wide. The crowd groaned as the ball went back into play. It looked as though no one was going to win. Suddenly, Jan scored from the edge of the pitch. Tom cheered loudly and ran over to hug Jan.

Test Sentence 1: Tom and Jan played in football boots. (static)

The whistle blew signalling the end of the match. The crowd was cheering loudly for Jan. Her team-mates gathered round her, lifted her on to their shoulders and carried her off the pitch. Jan was declared man of the match for scoring the final goal. She was treated like a hero.

Test Sentence 2: Jan's goal had won the football match. (causal)

Question: How did everyone know that the match was over?

8. John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. Francis fired his pistols one after the other. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.

Test Sentence 1: The bear was in pain. (causal)

Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. Snarling, the bear swiped at the girls. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear fell to the ground and did not move.

Test Sentence 2: The bear had huge paws. (static)

Question: Where did the bullet hit the bear?

9. The Duchess flung the baby at Alice and hurried out of the room. Alice caught the baby with some difficulty for it was a queer shaped creature and held out its arms and legs in all directions. Alice was holding on tightly to the crying baby. When Alice looked down the baby's nose had turned up and his eyes had become very small. The baby started to grunt.

Test Sentence 1: The baby sounded like a pig. (static)

Alice felt that she could not carry the baby any further. She thought it was quite an ugly child, but was quite handsome as a pig. She

was sure the tiny child would be happier in the wood with the other animals. So Alice set the baby down. She felt quite relieved to see the pig trot away into the wood.

Test Sentence 2: The baby had become a pig. (causal)

Question: What did Alice think of the baby?

10. The Witch was very cruel to Dorothy and Toto. She made Dorothy work very hard all day and threatened to hit her. Once the Witch struck Toto so hard with her umbrella that he flew at her and bit her hard on the leg. But the Witch was too wicked to feel anything at all. Instead the old woman just laughed.

Test Sentence 1: The old woman did not feel the pain. (causal)

Dorothy was very angry. She shouted at the Witch to stop being so cruel. But the Witch just sneered and ordered her to get back to work. The little girl was so mad, she picked up a bucket of water and threw it at the Witch. The Witch was soaked with water from head to foot. Dorothy watched as the Witch melted away.

Test Sentence 2: The wicked witch flew on a broom stick. (static)

Question: What did the Witch use to hit Toto?

11. When the battle was over, Lucy found Edmund a little way back from the fighting line. He was covered with blood, and his face was a nasty green colour. Lucy remembered the precious red liquid she had been given for a Christmas present. Aslan the lion told her to give some liquid to Edmund. Edmund leapt to his feet in perfect health.

Test Sentence 1: The lion was able to talk. (static)

Edmund fought well in the battle. He had beaten the giants and broken the Witch's wand. Without him the battle may have been lost. Aslan the lion walked slowly towards the boy. He was proud of Edmund. The lion touched Edmund on the shoulder with his sword. Then he gave the boy a fine suit of armour.

Test Sentence 2: The boy had just become a knight. (causal)

Question: What had happened to the Witch's wand?

12. Lucy, Edmund and Eustace were staring at a picture on the wall. It was a picture of a ship sailing on an ocean. But it was no

ordinary picture. The children could hear the wind blowing and the water splashing the side of the boat. The boat was bobbing up and down in the waves. The picture looked like it was moving. The sea was crashing violently against the rocks.

Test Sentence 1: The picture become real. (causal)

Eustace rushed to pull the picture off the wall. Edmund and Lucy tried to stop him. But the children found themselves standing on the picture frame. There was no glass in front of them, just real sea. A gust of wind blew them off the frame. They struggled to stay afloat in the stormy sea.

Test Sentence 2: The children were soaking wet. (static)

Question: What did Eustace try to do?

13. One night Sam woke up with a start. He heard a strange sound coming from outside his window. Sam crept out of bed, hid behind the curtain and looked outside. He let out a small scream. An enormous spaceship appeared to have landed in his garden. The ship filled his garden completely. The roses in the flower-bed were completely crushed.

Test Sentence 1: The spaceship was covered

in flashing lights. (static)

A door opened in the side of the ship and a strange figure appeared. It looked almost like a man but had long arms, short legs and a huge head with haunting black eyes. Sam saw the funny looking man and cried out. The creature looked up in surprise. Then the door closed and the ship rose slowly off the ground.

Test Sentence 2: Sam had given the creature quite a fright. (causal)

Question: What did Sam see in his garden?

14. The mermen lived far out to sea, where the water is blue as the bluest cornflower. The merman king had six beautiful daughters and the youngest was the prettiest of them all. Her skin was soft and delicate as a rose leaf and her eyes were as blue as the deepest sea. The young girl had a beautiful shiny tail. She could swim as fast as any other creature in the ocean

Test Sentence 1: The girl was a mermaid. (causal)

One night, the mermaid rescued a prince from the stormy sea. They fell in love, but the mermaid could not live out of the water. So the mermaid sang sadly to the prince every night. The merman king could not

bear to see his daughter so sad. He cast a spell over the mermaid and her tail disappeared. She ran happily to meet the prince on the shore.

Test Sentence 2: The Prince wore a gold crown. (static)

Question: Who was the prettiest mermaid?

15. There once was a king who had eleven sons and one daughter. The king married a wicked queen who was jealous of the children. She sent the little girl away and made the king banish his sons from the kingdom. On the day the boys left the palace, the queen cast a spell over them. Instead of eleven boys, eleven wild swans flew from the palace gates.

Test Sentence 1: The boys were covered in white feathers. (static)

The swan princes searched for their sister. When they found her she promised to break the spell. She made eleven magic shirts for her brothers. When she threw the shirts over the swans, they were transformed into eleven handsome princes. But the girl did not have enough magic cloth to finish all the shirts. The youngest prince would forever have a swan's wing in the place of an arm.

Test Sentence 2: The magic shirt was

missing a sleeve. (causal)

Question: How many sons did the king have?

16. Long ago there lived a young girl named Brave Heart who had learned to hunt. One winter, when the young men were out hunting, the women of the village saw some warriors heading towards them. They were filled with fear. With the men out hunting they did not know who would protect them.. Brave Heart took up her bow and arrow and started to sing. A lovely rainbow appeared in the sky

Test Sentence 1: Brave Heart had conjured a rainbow. (causal)

The women watched in wonder as Brave Heart transformed herself into a rainbow. The beautiful rainbow arched over the village. The warriors too saw the mighty rainbow arch above them. Brave Heart stood in the midst of the rainbow. The young girl raised her bow high and started to shoot. The lightening arrows fell as the warriors fled.

Test Sentence 2: The bow and arrow were magic.(static)

Question: Why were the women frightened?

Filler Passages

17. In ancient times Noah was collecting together two of every animal to put in his ark. But the she-cat refused to go in before she had caught a mouse. After all, she thought, there might be no mice where she was going, and she loved meat. So while all the other animals were lining up two by two, she was nowhere to be seen.

Test Sentence 1: All the animals lined up.

When all the other beasts were aboard, Noah could not wait any longer. He began to close the door, when up ran the she-cat. She made a great leap and squeezed through, but the closing door sliced her tail clean off. No-one bothered to mend the tail, so to this day the descendants of the she-cat go tail-less.

Test Sentence 2: Noah fixed the she-cat's tail.

Question: What did the she-cat do before she went into the arc?

18. Polly had discovered that if you opened a little door in the box-room attic of her house you would find a dark place behind it which you could get to by a little careful climbing. The dark place was a long tunnel

with brick wall on one side and sloping roof on the other. There was no floor in this tunnel: you had to step from rafter to rafter. Otherwise you would find yourself falling through the ceiling of the room below.

Test Sentence 1: The tunnel had no floor.

Polly had used a bit of the tunnel as a smugglers cave. She had made a floor out of broken boxes. In her cave she kept a box containing various treasures, and a story she was writing and usually a few apples. Polly had placed empty bottles in the cave to make it look more like smugglers cavern.

Test Sentence 2: Polly kept full bottles.

Question: What did Polly use to make the floor?

19. A beautiful princess was walking along a country road when she saw a large pig sitting on the grass verge. Being of a kindly nature she stopped to speak to it. She greeted the pig with a smile. But the pig did not smile. He told the princess that he was really a handsome prince. He had upset the local witch who had changed him into a pig.

Test Sentence 1: The pig upset the witch.

The pig explained that only a kiss from a beautiful princess could break the witch's spell. The princess was a little doubtful. She did not know whether to trust the pig. But being a kind-hearted princess, she bent down, shut her eyes and planted a kiss on the end of the pig's rubbery snout. Immediately the princess also turned into a pig.

Test Sentence 2: The princess kicked the pig.

Question: Why had the prince been turned into a pig?

20. Eric longed to be a policeman. He loved snooping around and was always getting into trouble for listening at doors. But when Eric's grandma was burgled, his curiosity paid off. Eric raced over to his gran's house to search for clues. In the flower-bed below the kitchen window Eric found a large footprint.

Test Sentence 1: Eric searched for clues in the house

Eric traced the footprint onto a piece of paper. Then he ran down to the local shoe shop and showed it to the shopkeeper. The shopkeeper explained that the footprint was unusual as the markings on the sole

belonged to a rare make of shoe. In fact only one person had ever bought a shoe like that from the shop. Excitedly, Eric took this information to the local police who used it to catch the villain. Eric had single-handedly solved the crime.

Test Sentence 2: Eric went to the fire station.

Question: What did Eric want to be?

21. Once a dog stole a juicy bone from a butcher's shop and ran off with it in his mouth. He stopped by a calm river to catch his breath. When the dog looked down, he was surprised to see another dog staring up at him. The other dog also had a bone in his mouth. But his bone looked far more meaty and juicy than the stolen bone.

Test Sentence 1: The other dog had a meaty bone.

The first dog growled at the dog in the river. He bared his teeth and demanded the other dog's bone. Then he lunged for the meaty juicy bone in the other dog's mouth. But as he did so his own bone fell into the river. The dog realised he had been growling at his own reflection. But it was too late and the dog watched his bone being swept away.

Test Sentence 2: The dog was growling at a rabbit.

Question: Where did the dog get the bone?

22. Harry had been given a new kite for his birthday. It was a magnificent kite, the biggest Harry had ever seen. It was shaped like a diamond, painted in lots of different colours and had two long silver streamers that would trail behind it when it flew.

Harry and his dad took the kite out for its first flight.

Test Sentence 1: The kite had two long silver streamers.

Harry held on to the string as the wind picked up the kite and it soared into the air. The streamers looked like lightening as they twisted in the wind. Suddenly, a gust of wind blew the kite into a tree. Harry could not reach it but Harry's dad had an idea. He raced home and fetched his painting ladder. It was just tall enough to reach the kite. Harry soon had the kite soaring up into the clouds once more.

Test Sentence 2: The ladder was not quite tall enough.

Question: What shape was Harry's kite?

23. Once a farmer decided to build a scarecrow to stop the birds eating his crops.

He made the scarecrow's head first and as soon as he painted on the ears, the scarecrow could hear the farmer talking. Then the farmer painted on an eye and the scarecrow began looking around. It was his first look at the world and he was very curious.

Test Sentence 1: The farmer painted on an eye.

The farmer painted the second eye a little bigger than the first, so the scarecrow could see much better than before. The scarecrow watched as the farmer painted his nose and mouth. Then he saw the farmer make his body and attach his arms and legs. Finally the farmer fixed the scarecrow's head onto the body. The scarecrow was complete, and was quite disappointed when he was put on a pole in the farmer's field.

Test Sentence 2: The scarecrow was not completed.

Question: What part of the scarecrow did the farmer make first?

24. John and Sam paddled their dinghy in the warm rock pool. The water was calm and the beach was just the other side of the rocks. The children lazily basked in the warm sun, their bare feet trailing in the sea. Suddenly, without warning, a huge wave crashed into the rock pool and pushed the

boat over the rocks and into the sea.

Test Sentence 1: A wave pushed the boat.

John and Sam clung tightly to the sides of the boat. Sam looked around for the paddles but the wave had washed them away. The little boat bobbed uncontrollably on the waves. Finally, the wind dropped and the sea was calm again. The boys used their hands to steer the dinghy. Together they managed to guide it to the safety of the beach.

Test Sentence 2: The boys used their feet.

Question: What happened to the paddles?

25. The North Wind was very conceited and believed himself to be much stronger than the Sun. One day the Sun challenged him to prove his strength by blowing the coat off a man who was walking below. The North Wind began to blow with all his might around the man, trying to tear off his coat. But the man just pulled the coat more firmly around him and shivered in the cold.

Test Sentence 1: The North Wind blew around the trees.

When the North Wind was too tired to blow

anymore, the Sun took up the challenge. And he began to shine with warm, soft, golden rays. Soon the man began loosening his coat. The sun continued to shine, sending bright light everywhere. The man gasped with the heat, took off his coat and slung it over his shoulder. The Sun had shown that the strongest do not always win.

Test Sentence 2: The man gasped and took off his coat.

Question: How did the Sun want the North Wind to prove his strength?

26. William had a favourite teacher at school. Certain members of his class always put pretty little posies of flowers on her desk every morning. William was determined to outshine his classmates. The next morning before he set out for school, he went out into the garden carrying a large basket and a pair scissors. He headed straight for the greenhouse.

Test Sentence 1: William carried a shopping bag.

The greenhouse was a riot of colour. William worked very hard and finally staggered out of the greenhouse beneath a piled-up basket of blooms. The greenhouse looked barren and deserted. Then William

set off for school. On entering the classroom, his teacher was amazed to see her desk and chair covered in withering blooms.

Test Sentence 2: Then William set off for his school.

Question: Why did William want the flowers in the greenhouse?

27. Mr Twit was a very hairy-faced man. The whole of his face except for his eyes, his nose and his forehead, was covered in thick hair. The stuff even sprouted in revolting tufts out of his ear-holes. Mr Twit felt that his hairiness made him look terrifically clever and grand. But in truth he was neither of these things.

Test Sentence 1: Mr Twit thought he looked clever.

Mr Twit was a twit. He was born a twit. And now at the age of sixty he was a bigger twit than ever. The hair on Mr Twit's face grew in spikes that stuck out straight like the bristles of a nailbrush. And how often did Mr Twit wash this bristly nailbrushy face of his? The answer is never, he hadn't washed it for years.

Test Sentence 2: Mr Twit never washed his face.

Question: What was sticking out of Mr Twit's ears?

28. The sailor stood on the deck of his ship and held up his telescope. He could see the pirate ship sailing towards him. The Jolly Roger was blowing in the wind. The sailor wanted to protect his cargo from the thieves. The maps he was carrying would enable his navy to win the battle against the pirates.

Test Sentence 1: The sailor did not want his cargo.

The sailor loaded his canon. He aimed the barrel directly at the enemy vessel and fired. The first shot missed the pirates and they carried on advancing. The sailor fired again. This time the missile found its target. The sailor watched as the pirates jumped into the sea to escape the blast. He and his crew picked up the pirates and threw them below decks.

Test Sentence 2: The pirates jumped into the sea.

Question: What did the sailor fire at the pirates?

29. Owl was at home warming himself by the fire. He was eating hot buttered toast and pea soup for supper. Owl heard a loud

sound at the front door. He wondered who could possibly be banging and pounding on his door on such a cold night. But when Owl opened the door wide, only the snow and the cold of Winter were outside.

Test Sentence 1: Owl did not pull the door open wide.

Owl invited Winter inside to warm himself by the fire. But Winter came in very fast and a cold wind pushed Owl against the wall. Winter ran around the room and blew out the fire. The snow whooshed up the staircase and into all the rooms in Owl's house. Owl was most annoyed. He opened his door and angrily told Winter to leave. Winter rushed wildly out of the house and Owl gladly slammed the door behind him.

Test Sentence 2: Winter rushed wildly out of the house.

Question: What was owl having for supper?

30. The fox crept up the dark tunnel to the mouth of his hole. He poked his long handsome face out into the night and sniffed the air once. He moved an inch or two forward and stopped. He sniffed again. He was always very careful coming out of his hole. The fox inched forward a little more. The front half of his body was now out in

the open.

Test Sentence 1: The fox inched out a lot.

His black nose twitched from side to side, sniffing for the scent of danger. He found none and was just about to go trotting off into the wood when he heard a rustling in the bushes. Then a loud bang rang out and the smoke from a gun floated into the air. The fox darted back into his hole and lay very still and quiet.

Test Sentence 2: The smoke floated up.

Question: What was the fox sniffing for?

There once was a little girl who had a magic finger. Whenever she got cross she would get very hot all over. Then the tip of the forefinger of her right hand would tingle terribly. Suddenly a flash would come out of the tip of her finger and touch the person who had made her cross. Afterwards, all sorts of strange things would start to happen.

Test Sentence 1: A flash would come right out of her toes.

One day, at school, the little girl's teacher made her very cross. The little girl put the magic finger on her teacher. All of a sudden cat whiskers began growing out of the

teacher's face. They grew so fast that soon they were out of her ears. When she turned around she had grown a huge bushy tail as well. The teacher was never quite the same again.

Test Sentence 2: The teacher had grown a bushy tail.

Question: What happened to the little girl's teacher?

32. Robert and Miss Cannon were enjoying a quiet picnic on the riverbank. Robert was just offering Miss Cannon a plate of sandwiches when his eyes spotted something on the road leading from the village to the riverside. He froze with horror. A curious figure was cycling towards them with a blackened face, a few drooping feathers on its head and a door-mat flying like a cape in the wind.

Test Sentence 1: A figure was skating towards them.

Robert could see it was William. William came to the end of the road. He did not stop or turn on to the other roads by the riverside. Instead, with a set face he rode onto the riverbank and straight towards them. William rode over the sandwiches, rolls and cakes and straight into the river.

Test Sentence 2: William rode straight towards them.

Question: What was Robert offering to Miss Canon when he first saw William?

Appendix IV

Experiment 4 and Experiment 5 Test Materials

List One

1. John looked behind the wagon and saw the Red Indians approaching. He called for his father. Henry grabbed his rifle and pulled himself up to the front of the wagon. As the Red Indians rode closer, John could see they were carrying weapons. He raised his own weapon and aimed at the enemy. But Henry held up his hand. John took his finger off the trigger and carefully lowered his gun.

Test Sentence 1: Henry wanted John to wait. (causal)

The Indians were so close that John could see their brightly coloured faces. John's hand was firmly on his rifle as he waited in the wagon. Then Henry gave the signal and John and Henry opened fire. The battle lasted only a few minutes but seemed a lot longer. Finally, the Indians had weakened. They turned and disappeared into the mountain range.

Test Sentence 2: The Indians rode horses.(static)

Question 1: What was John holding as he waited in the wagon?

Question 2: Who was attacking John and Henry?

2. The fishermen were lost in a storm. Their boat was being thrown from side to side by the wind. Many of their belongings had been lost overboard and a huge wave had crashed across the deck, smashing the mast. The men struggled to keep control of the wheel. They used all their strength to fight the wind and rain.

Test Sentence 1: The men were soaking wet. (static)

The men knew they were running out of time. The boat was being forced towards the rocks. The Captain knew they could not save the ship and he ordered the men into the life boat. The ship crashed into the rocks. A huge hole appeared in the side of the hull. The men watched as the boat sank beneath the waves.

Test Sentence 2: The ship filled with water. (causal)

Question 1: What happened to the mast?

Question 2: What did the Captain order the men to do?

3. The spy dropped from the wall surrounding the large house. He pulled out his torch and made his way over to the open window. As quietly as possible, he pulled himself over the window ledge and crept

into the office. He found the safe hidden under the desk. The spy entered the code into the panel on the safe door. He found the secret plans in the safe.

Test Sentence 1: The code opened the safe door. (causal)

Carefully, the spy photographed the plans. Then he slipped them back into the safe where he had found them. With the camera film safely in the bottom of his backpack, the spy slipped out of the office the way he came in. In the darkness, he scaled the wall once more. His mission complete, he disappeared into the dark night.

Test Sentence 2: The spy was dressed all in black. (static)

Question 1: What did the spy find in the safe?

Question 2: What did the spy do when he found the plans?

4. Emily sat happily in the sand at the edge of the sea. She was building a castle with her bucket and spade. Carefully she filled her bucket to the top with sand. Then she turned the bucket over in the sand. Emily tapped it gently with her spade. Then she lifted the bucket to reveal a perfect sand castle.

Test Sentence 1: The sand in the bucket was wet. (static)

Emily was very pleased with herself. She loved building castles on the beach. Emily set about filling her bucket again. She wanted to build more castles. But Emily had not noticed that a wind had begun to blow in from the sea. She did not see the huge wave coming towards her. The wave crashed heavily onto the sand. When the wave finally rolled away, Emily began to cry.

Test Sentence 2: The wave washed the castle away. (causal)

Question 1: What was Emily building?

Question 2: How was Emily feeling?

5. The cowboy was riding his horse at great speed through the mountains. He could hear the bandits drawing closer. The cowboy could feel his horse was tiring. He knew he had to stop but he had to evade capture. Taking a risk, the cowboy rode into a cave hidden behind a rock and waited. Soon he heard the bandits ride past the cave.

Test Sentence 1: The bandits could not see the cowboy. (causal)

From the safety of the cave, the cowboy watched as the bandits searched for him in

the valley below. They remained there, circling on their horses for some time. They were clearly disappointed to have lost their prey. When darkness fell, the leader ordered the men back to the hideout. They would have to think of another way to capture the cowboy.

Test Sentence 2: The bandits were carrying rifles. (static)

Question 1: What was wrong with the cowboy's horse?

Question 2: Where were the bandits looking for the cowboy?

6. John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. Francis fired his pistols one after the other. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.

Test Sentence 1: The bear had huge paws. (static)

Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. Snarling, the bear swiped at the girls. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear fell to the ground and

did not move.

Test Sentence 2. John shot the bear dead. (causal)

Question 1: Where did the bullet hit the bear?

Question 2: What did John do?

7. The Witch was very cruel to Dorothy and Toto. She made Dorothy work very hard all day and threatened to hit her. Once the Witch struck Toto so hard with her umbrella that he flew at her and bit her hard on the leg. But the Witch was too wicked to feel any pain. So Toto had not hurt the cruel old woman.

Test Sentence 1: The old woman flew on a broomstick. (static)

Dorothy was very angry. She shouted at the Witch to stop being so cruel. But the Witch just sneered and ordered her to get back to work. The little girl was so mad, she picked up a bucket of water and threw it at the Witch. The Witch was soaked with water from head to foot. Dorothy watched as the Witch melted away.

Test Sentence 2: The water made the wicked witch melt. (causal)

Question 1: What did the Witch use to hit Toto?

Question 2: Why was Dorothy angry with the Witch?

8. When the battle was over, Lucy found Edmund a little way back from the fighting line. He was covered with blood, and his face was a nasty green colour. Lucy remembered the precious red liquid she had been given for a Christmas present. She poured a few drops of red liquid into Edmund's mouth. Edmund leapt to his feet in perfect health.

Test Sentence 1: The red liquid healed Edmund's wounds. (causal)

Edmund fought well in the battle. He had beaten the giants and broken the Witch's wand. Without him the battle may have been lost. Aslan the lion walked slowly towards the boy. He was proud of Edmund. The lion pinned a gold medal to Edmund's jacket. Then the lion let Edmund ride on his back.

Test Sentence 2: The lion was able to talk. (static)

Question 1: What had happened to the Witch's wand?

Question 2: Who was proud of Edmund?

9. One night Sam woke up with a start. He heard a strange sound coming from outside

his window. Sam crept out of bed, hid behind the curtain and looked outside. He let out a small scream. An enormous spaceship appeared to have landed in his garden. The ship filled his garden completely. The roses in the flower-bed were completely crushed.

Test Sentence 1: The spaceship landed on the roses. (causal)

A door opened in the side of the ship and a strange figure appeared. It looked almost like a man but had long arms, short legs and a huge head with haunting black eyes. Sam cried out and the creature looked up in surprise. Then the door closed and the ship rose slowly off the ground. Quick as a flash it was gone.

Test Sentence 2: The ship was covered in flashing lights. (static)

Question 1: What did Sam see in his garden?

Question 2: What colour were the alien's eyes?

10. The mermen lived far out to sea, where the water is blue as the bluest cornflower. The merman king had six beautiful daughters and the youngest was the prettiest of them all. Her skin was soft and delicate as a rose leaf and her eyes were as blue as

the deepest sea. But the young daughter could only swim in the sea. She longed to walk on the land and feel the sand between her toes.

Test Sentence 1: The girl had a fish tail. (static)

One night, the mermaid rescued a prince from the stormy sea. They fell in love, but the mermaid could not live out of the water. So the mermaid sang sadly to the prince every night. The merman king could not bear to see his daughter so sad. He cast a spell over the mermaid and her tail disappeared. She ran happily to meet the prince on the shore.

Test Sentence 2: The spell gave the girl legs. (causal)

Question 1: Who was the prettiest mermaid?

Question 2: Who did the mermaid rescue?

11. There once was a king who had eleven sons and one daughter. The king married a wicked queen who was jealous of the children. She sent the little girl away and made the king banish his sons from the kingdom. On the day the boys left the palace, the queen cast a spell over them. Instead of eleven boys, eleven wild swans flew from the palace gates.

Test Sentence 1: The queen turned the boys into wild swans. (causal)

The swan princes searched for their sister. When they found her she promised to break the spell. She made eleven magic shirts for her brothers. When she threw the shirts over the swans, they were transformed into eleven handsome princes. But the girl did not have enough magic cloth to finish all the shirts. The youngest prince would forever have a swan's wing in the place of an arm.

Test Sentence 2: His arm was covered in white feathers. (static)

Question 1: How many sons did the king have?

Question 2: What did the girl make for her brothers?

12. Long ago there lived a young girl named Brave Heart who had learned to hunt. One winter, when the young men were out hunting, the women of the village saw some warriors heading towards them. They were filled with fear. With the men out hunting they did not know who would protect them. The young girl vowed to defend the women. She took up her bow and arrow and circled the villagers.

Test Sentence 1: The young girl was an indian brave. (static)

The women watched in wonder as Brave Heart transformed herself into a rainbow. The beautiful rainbow arched over the village. The warriors too saw the mighty rainbow arch above them. Brave Heart stood in the midst of the rainbow. The young girl raised her bow and started to shoot. Lightning fell as the warriors fled from the village.

Test Sentence 2: The arrows were made of lightening. (causal)

Question 1: Why were the women frightened?

Question 2: What did Brave Heart use to protect the village?

List Two

1. John looked behind the wagon and saw the Red Indians approaching. He called for his father. Henry grabbed his rifle and pulled himself up to the front of the wagon. As the Red Indians rode closer, John could see they were carrying weapons. He raised his own weapon and aimed at the enemy. But Henry told John to wait so John took his finger off the trigger and slowly lowered his gun.

Test Sentence 1: John wore a cowboy hat.
(static)

The Indians were so close that John could see their brightly coloured faces. John's hand was firmly on his rifle as he waited in the wagon. Then Henry gave the signal and John and Henry opened fire. The battle lasted only a few minutes but seemed a lot longer. Finally, the Indians had weakened. They turned and disappeared into the mountain range.

Test Sentence 2: The Indians were beaten.
(causal)

Question 1: What was John holding as he waited in the wagon?

Question 2: Who was attacking John and Henry?

2. The fishermen were lost in a storm.

Their boat was being thrown from side to side by the wind. Many of their belongings had been lost overboard and a huge wave had crashed across the deck, smashing the mast. The men struggled to keep control of the wheel. They used all their strength to fight the wind and rain.

Test Sentence 1: The boat was hard to steer. (causal)

The men knew they were running out of time. The boat was being forced towards the rocks. The Captain knew they could not save the ship and he ordered the men into the life boat. The ship crashed into the rocks. A huge hole appeared in the side of the hull. The ship filled with water and began to sink.

Test Sentence 2: The men wore lifejackets.
(static)

Question 1: What happened to the mast?

Question 2: What did the Captain order the men to do?

3. The spy dropped from the wall surrounding the large house. He pulled out his torch and made his way over to the open window. As quietly as possible, he pulled himself over the window ledge and crept into the office. He found the safe hidden under the desk. The spy entered the code

into the panel and the door opened. He found the plans in the safe.

Test Sentence 1: The spy carried a gun.
(static)

Carefully, the spy photographed the plans. Then he slipped them back into the safe where he had found them. With the camera film safely in the bottom of his backpack, the spy slipped out of the office the way he came in. In the darkness, he found the wall once more. He jumped safely to the ground and ran into the dark night.

Test Sentence 2: He climbed over the wall.
(causal)

Question 1: What did the spy find in the safe?

Question 2: What did the spy do when he found the plans?

4. Emily sat happily in the sand at the edge of the sea. She was building a castle with her bucket and spade. Carefully she filled her bucket to the top with sand. Then she turned the bucket over in the sand. Emily tapped the bucket gently with her spade. Then she lifted the bucket and looked underneath.

Test Sentence 1: Emily built a sandcastle.
(causal)

Emily was very pleased with herself. She loved building castles on the beach. Emily set about filling her bucket again. She wanted to build more castles. But Emily had not noticed that a wind had begun to blow in from the sea. She did not see the huge wave coming towards her. The wave crashed heavily onto the sand and washed away Emily's sandcastle. Emily began to cry.

Test Sentence 2: Emily wore a swimming suit. (static)

Question 1: What was Emily building?

Question 2: How was Emily feeling?

5. The cowboy was riding his horse at great speed through the mountains. He could hear the bandits drawing closer. The cowboy could feel his horse was tiring. He knew he had to stop but he had to evade capture. Taking a risk, the cowboy rode into a cave hidden behind a rock and waited. The bandits could not see him in the cave.

Test Sentence 1: The cowboy had a gun and holster. (static)

From the safety of the cave, the cowboy watched as the bandits searched for him in the valley below. They remained there,

circling on their horses for some time. They were clearly disappointed to have lost their prey. When darkness fell, the leader ordered the men back to the hideout. They would have to think of another way to capture the cowboy.

Test Sentence 2: The bandits could not hunt in the dark. (causal)

Question 1: What was wrong with the cowboy's horse?

Question 2: Where were the bandits looking for the cowboy?

6. John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. Francis fired his pistols one after the other. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.

Test Sentence 1: The bear was in pain. (causal)

Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. Snarling, the bear swiped at the girls. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear was badly hurt by John's bullet.

Test Sentence 2: The bear had huge paws. (static)

Question 1: Where did the bullet hit the bear?

Question 2: What did John do?

7. The Witch was very cruel to Dorothy and Toto. She made Dorothy work very hard all day and threatened to hit her. Once the Witch struck Toto so hard with her umbrella that he flew at her and bit her hard on the leg. But the Witch was too wicked to feel anything at all. The old woman just laughed out loud.

Test Sentence 1: The old woman did not feel the pain. (causal)

Dorothy was very angry. She shouted at the Witch to stop being so cruel. But the Witch just sneered and ordered her to get back to work. The little girl was so mad, she picked up a bucket of water and threw it at the Witch. The Witch was soaked with water from head to foot. Dorothy watched as the water melted the witch away.

Test Sentence 2: The wicked witch flew on a broomstick. (static)

Question 1: What did the Witch use to hit Toto?

Question 2: Why was Dorothy angry with

the Witch?

8. When the battle was over, Lucy found Edmund a little way back from the fighting line. He was covered with blood, and his face was a nasty green colour. Lucy remembered the precious red liquid she had been given for a Christmas present. Aslan the lion told her to give some liquid to Edmund. The liquid helped to heal Edmund's wounds.

Test Sentence 1: The lion was able to talk. (static)

Edmund fought well in the battle. He had beaten the giants and broken the Witch's wand. Without him the battle may have been lost. Aslan the lion walked slowly towards the boy. He was proud of Edmund. Aslan put something on Edmund's jacket. Edmund stared proudly at the shiny gold medal pinned there.

Test Sentence 2: Aslan gave Edmund a medal. (causal)

Question 1: What happened to the Witch's wand?

Question 2: Who was proud of Edmund?

9. One night Sam woke up with a start. He heard a strange sound coming from outside his window. Sam crept out of bed, hid

behind the curtain and looked outside. He let out a small scream. An enormous spaceship appeared to have landed in his garden. The huge spaceship filled his garden completely and the flower-beds were all completely ruined.

Test Sentence 1: The spaceship was covered in flashing lights. (static)

A door opened in the side of the ship and a strange figure appeared. It looked almost like a man but had long arms, short legs and a huge head with haunting black eyes. Sam saw the funny looking man and cried out. The creature looked up in surprise. Then the door closed and the ship rose slowly off the ground.

Test Sentence 2: Sam had given the creature quite a fright. (causal)

Question 1: What did Sam see in his garden?

Question 2: What colour were the alien's eyes?

10. The mermen lived far out to sea, where the water is blue as the bluest cornflower. The merman king had six beautiful daughters and the youngest was the prettiest of them all. Her skin was soft and delicate as a rose leaf and her eyes were as blue as the deepest sea. The young girl had a

beautiful shiny tail. She could swim as fast as any other creature in the ocean

Test Sentence 1: The girl was a mermaid. (causal)

One night, the mermaid rescued a prince from the stormy sea. They fell in love, but the mermaid could not live out of the water. So the mermaid sang sadly to the prince every night. The merman king could not bear to see his daughter so sad. The king cast a spell over the mermaid and her tail disappeared. Two perfect legs appeared in the place of her tail.

Test Sentence 2: The King wore a gold crown. (static)

Question 1: Who was the prettiest mermaid?

Question 2: Who did the mermaid rescue?

11. There once was a king who had eleven sons and one daughter. The king married a wicked queen who was jealous of the children. She sent the little girl away and made the king banish his sons from the kingdom. On the day the boys left the palace, the queen turned the boys into swans. Instead of eleven boys, eleven wild swans flew from the palace gates.

Test Sentence 1: The boys were covered in

white feathers. (static)

The swan princes searched for their sister. When they found her she promised to break the spell. She made eleven magic shirts for her brothers. When she threw the shirts over the swans, they were transformed into eleven handsome princes. But the girl did not have enough magic cloth to finish all the shirts. The youngest prince would forever have a swan's wing in the place of an arm.

Test Sentence 2: The magic shirt was missing a sleeve. (causal)

Question 1: How many sons did the king have?

Question 2: What did the girl make for her brothers?

12. Long ago there lived a young girl named Brave Heart who had learned to hunt. One winter, when the young men were out hunting, the women of the village saw some warriors heading towards them. They were filled with fear. With the men out hunting they did not know who would protect them.. Brave Heart took up her weapon and started to sing. Soon a lovely rainbow appeared in the sky.

Test Sentence 1: Brave Heart had conjured up a rainbow. (causal)

The women watched in wonder as Brave Heart transformed herself into a rainbow. The beautiful rainbow arched over the village. The warriors too saw the mighty rainbow arch above them. Brave Heart stood in the midst of the rainbow. The young girl raised her bow high and started to shoot lightning arrows. The warriors ran away in fear.

Test Sentence 2: The bow and arrow were enchanted. (static)

Question 1: Why were the women frightened?

Question 2: What did Brave Heart use to protect the village?

Filler Passages.

13. A beautiful princess was walking along a country road when she saw a large pig sitting on the grass verge. Being of a kindly nature she stopped to speak to it. She greeted the pig with a smile. But the pig did not smile. He told the princess that he was really a handsome prince. He had upset the local witch who had changed him into a pig.

Test Sentence 1: The pig upset the witch.

The pig explained that only a kiss from a beautiful princess could break the witch's spell. The princess was a little doubtful.

She did not know whether to trust the pig. But being a kind-hearted princess, she bent down, shut her eyes and planted a kiss on the end of the pig's rubbery snout.

Immediately the princess also turned into a pig.

Test Sentence 2: The princess kicked the pig.

Question 1: Why had the prince been turned into a pig?

Question 2: Where did the princess kiss the pig?

14. Harry had been given a new kite for his birthday. It was a magnificent kite, the biggest Harry had ever seen. It was shaped like a diamond, painted in lots of different colours and had two long silver streamers that would trail behind it when it flew. Harry and his dad took the kite out for its first flight.

Test Sentence 1: The kite had two long silver streamers.

Harry held on to the string as the wind picked up the kite and it soared into the air. The streamers looked like lightning as they twisted in the wind. Suddenly, a gust of wind blew the kite into a tree. Harry could not reach it but Harry's dad had an idea. He raced home and fetched his painting ladder.

It was just tall enough to reach the kite.
Harry soon had the kite soaring up into the clouds once more.

Test Sentence 2: The ladder was not quite tall enough.

Question 1: What shape was Harry's kite?

Question 2: How did Harry's dad get the kite down?

15. Mr Twit was a very hairy-faced man. The whole of his face except for his eyes, his nose and his forehead, was covered in thick hair. The stuff even sprouted in revolting tufts out of his ear-holes. Mr Twit felt that his hairiness made him look terrifically clever and grand. But in truth he was neither of these things.

Test Sentence 1: Mr Twit thought he looked clever.

Mr Twit was a twit. He was born a twit. And now at the age of sixty he was a bigger twit than ever. The hair on Mr Twit's face grew in spikes that stuck out straight like the bristles of a nailbrush. And how often did Mr Twit wash this bristly nailbrushy face of his? The answer is never, he hadn't washed it for years.

Test Sentence 2: Mr Twit never washed his face.

Question 1: What was sticking out of Mr Twit's ears?

Question 2: How old was Mr Twit?

16. The sailor stood on the deck of his ship and held up his telescope. He could see the pirate ship sailing towards him. The Jolly Roger was blowing in the wind. The sailor wanted to protect his cargo from the thieves. The maps he was carrying would enable his navy to win the battle against the pirates.

Test Sentence 1: The sailor did not want his cargo.

The sailor loaded his canon. He aimed the barrel directly at the enemy vessel and fired. The first shot missed the pirates and they carried on advancing. The sailor fired again. This time the missile found its target. The sailor watched as the pirates jumped into the sea to escape the blast. He and his crew picked up the pirates and threw them below decks.

Test Sentence 2: The pirates jumped into the sea.

Question 1: What did the sailor fire at the pirates?

Question 2: What did the pirates do when their ship was hit?

17. Owl was at home warming himself by the fire. He was eating hot buttered toast and pea soup for supper. Owl heard a loud sound at the front door. He wondered who could possibly be banging and pounding on his door on such a cold night. But when Owl opened the door wide, only the snow and the cold of Winter were outside.

Test Sentence 1: Owl did not pull the door open wide.

Owl invited Winter inside to warm himself by the fire. But Winter came in very fast and a cold wind pushed Owl against the wall. Winter ran around the room and blew out the fire. The snow whooshed up the staircase and into all the rooms in Owl's house. Owl was most annoyed. He opened his door and angrily told Winter to leave. Winter rushed wildly out of the house and Owl gladly slammed the door behind him.

Test Sentence 2: Winter rushed wildly out of the house.

Question 1: What was Owl having for supper?

Question 2: What happened to Owl's fire?

18. The fox crept up the dark tunnel to the mouth of his hole. He poked his long handsome face out into the night and sniffed the air once. He moved an inch or two forward and stopped. He sniffed again. He

was always very careful coming out of his hole. The fox inched forward a little more. The front half of his body was now out in the open.

Test Sentence 1: The fox inched out a lot.

His black nose twitched from side to side, sniffing for the scent of danger. He found none and was just about to go trotting off into the wood when he heard a rustling in the bushes. Then a loud bang rang out and the smoke from a gun floated into the air. The fox darted back into his hole and lay very still and quiet.

Test Sentence 2: The smoke floated up.

Question 1: What was the fox sniffing for?

Question 2: What did the fox hear?

Appendix V

Experiment 4 Off-Line Validation Materials

List One

Story 1

John looked behind the wagon and saw the Red Indians approaching. He called for his father. Henry grabbed his rifle and pulled himself up to the front of the wagon. As the Red Indians rode closer, John could see they were carrying weapons. He raised his own weapon and aimed at the enemy. But Henry held up his hand. John took his finger off the trigger and carefully lowered his gun.

What did Henry want John to do?

- a Henry wanted John to fire his gun
- b Henry wanted John to wait
- c Henry wanted John to read a book

The Indians were so close that John could see their brightly coloured faces. John's hand was firmly on his rifle as he waited in the wagon. Then Henry gave the signal and John and Henry opened fire. The battle lasted only a few minutes but seemed a lot longer. Finally, the Indians had weakened. They turned and disappeared into the mountain range.

What did the indians ride?

- a the indians rode camels
- b the indians rode horses
- c the indians rode bicycles

Story 2

The fishermen were lost in a storm. Their boat was being thrown from side to side by the wind. Many of their belongings had been lost overboard and a huge wave had crashed across the deck, smashing the mast. The men struggled to keep control of the wheel. They used all their strength to fight the wind and rain.

Which of the following best describes the men?

- a the men were freezing cold
- b the men were soaking wet
- c the men were boiling hot

The men knew they were running out of time. The boat was being forced towards the rocks. The Captain knew they could not save the ship and he ordered the men into the

life boat. The ship crashed into the rocks. A huge hole appeared in the side of the hull. The men watched as the boat sank beneath the waves.

- Why did the ship sink?
- a the ship filled with water
 - b the ship filled rocks
 - c the ship filled with seaweed

Story 3

The spy dropped from the wall surrounding the large house. He pulled out his torch and made his way over to the open window. As quietly as possible, he pulled himself over the window ledge and crept into the office. He found the safe hidden under the desk. The spy entered the code into the panel on the safe door. He found the secret plans in the safe.

- What did the code do?
- a the code blew up the safe
 - b the code opened the safe
 - c the code locked the safe

Carefully, the spy photographed the plans. Then he slipped them back into the safe where he had found them. With the camera film safely in the bottom of his backpack, the spy slipped out of the office the way he came in. In the darkness, he scaled the wall once more. His mission complete, he disappeared into the dark night.

- How was the spy dressed?
- a the spy was dressed in black
 - b the spy was dressed in bright green
 - c the spy was dressed in red

Story 4

Emily sat happily in the sand at the edge of the sea. She was building a castle with her bucket and spade. Carefully she filled her bucket to the top with sand. Then she turned the bucket over in the sand. Emily tapped it gently with her spade. Then she lifted the bucket to reveal a perfect sand castle.

- What was in the bucket?
- a the bucket was filled with sand

- b the bucket was filled with salt
- c the bucket was filled with shells

Emily was very pleased with herself. She loved building castles on the beach. Emily set about filling her bucket again. She wanted to build more castles. But Emily had not noticed that a wind had begun to blow in from the sea. She did not see the huge wave coming towards her. The wave crashed heavily onto the sand. When the wave finally rolled away, Emily began to cry.

Why was Emily crying?

- a the wave washed Emily away
- b the wave washed the bucket away
- c the wave washed the sandcastle away

Story 5

The cowboy was riding his horse at great speed through the mountains. He could hear the bandits drawing closer. The cowboy could feel his horse was tiring. He knew he had to stop but he had to evade capture. Taking a risk, the cowboy rode into a cave hidden behind a rock and waited. Soon he heard the bandits ride past the cave.

Why did the bandits ride past the cave?

- a the bandits could not hear the cowboy
- b the bandits could not see the cowboy
- c the bandits were hungry

From the safety of the cave, the cowboy watched as the bandits searched for him in the valley below. They remained there, circling on their horses for some time. They were clearly disappointed to have lost their prey. When darkness fell, the leader ordered the men back to the hideout. They would have to think of another way to capture the cowboy.

What were the bandits carrying?

- a the bandits were carrying rifles
- b the bandits were carrying handbags
- c the bandits were carrying water pouches

Story 6

John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. Francis fired his pistols one after the other. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.

What sort of paws did the bear have?

- a the bear had yellow paws
- b the bear had huge paws
- c the bear had small paws

Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. Snarling, the bear swiped at the girls. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear fell to the ground and did not move.

Why did the bear fall to the ground?

- a John shot the bear
- b John hit the bear
- c John kissed the bear

Story 7

The Witch was very cruel to Dorothy and Toto. She made Dorothy work very hard all day and threatened to hit her. Once the Witch struck Toto so hard with her umbrella that he flew at her and bit her hard on the leg. But the Witch was too wicked to feel any pain. So Toto had not hurt the cruel old woman.

Which of the following best describes the old woman?

- a the old woman flew on a broomstick
- b the old woman flew on an umbrella
- c the old woman had a black cat

Dorothy was very angry. She shouted at the Witch to stop being so cruel. But the Witch just sneered and ordered her to get back to work. The little girl was so mad, she picked up a bucket of water and threw it at the Witch. The Witch was soaked with water from head to foot. Dorothy watched as the Witch melted away.

What happened to the witch?

- a the water made the witch melt
- b the water made the witch explode
- c the water made the witch grow

Story 8

When the battle was over, Lucy found Edmund a little way back from the fighting line. He was covered with blood, and his face was a nasty green colour. Lucy remembered the precious red liquid she had been given for a Christmas present. She poured a few drops of red liquid into Edmund's mouth. Edmund leapt to his feet in perfect health.

How did Edmund recover so quickly?

- a the lion helped to heal Edmund's wounds
- b the red liquid helped to heal Edmund's wounds
- c the doctor helped to heal Edmund's wounds

Edmund fought well in the battle. He had beaten the giants and broken the Witch's wand. Without him the battle may have been lost. Aslan the lion walked slowly towards the boy. He was proud of Edmund. The lion pinned a gold medal to Edmund's jacket. Then the lion let Edmund ride on his back

What was special about the lion?

- a The lion was able to talk
- b the lion was very gentle
- c the lion was small as a mouse

Story 9

One night Sam woke up with a start. He heard a strange sound coming from outside his window. Sam crept out of bed, hid behind the curtain and looked outside. He let out a small scream. An enormous spaceship appeared to have landed in his garden. The ship filled his garden completely. The roses in the flower-bed were completely crushed.

What happened to the roses?

- a the spaceship landed on the roses
- b Sam landed on the roses
- c a car drove over the roses

A door opened in the side of the ship and a strange figure appeared. It looked almost like a man but had long arms, short legs and a huge head with haunting black eyes. Sam

cried out and the creature looked up in surprise. Then the door closed and the ship rose slowly off the ground. Quick as a flash it was gone.

What did the ship look like?

- a the ship was shaped like a saucer
- b the ship was covered in flashing lights
- c the ship had curtains at the window

Story 10

The mermen lived far out to sea, where the water is blue as the bluest cornflower. The merman king had six beautiful daughters and the youngest was the prettiest of them all. Her skin was soft and delicate as a rose leaf and her eyes were as blue as the deepest sea. But the young daughter could only swim in the sea. She longed to walk on the land and feel the sand between her toes.

What did the girl look like?

- a the girl had a fish tail
- b the girl had long golden hair
- c the girl had very big ears

One night, the mermaid rescued a prince from the stormy sea. They fell in love, but the mermaid could not live out of the water. So the mermaid sang sadly to the prince every night. The merman king could not bear to see his daughter so sad. He cast a spell over the mermaid and her tail disappeared. She ran happily to meet the prince on the shore.

What happened to the girl?

- a the spell gave the girl legs
- b the spell gave the girl shoes
- c the spell gave the girl measles

Story 11

There once was a king who had eleven sons and one daughter. The king married a wicked queen who was jealous of the children. She sent the little girl away and made the king banish his sons from the kingdom. On the day the boys left the palace, the queen cast a spell over them. Instead of eleven boys, eleven wild swans flew from the palace gates.

What happened to the boys?

- a the queen turned them all into frogs
- b the queen told them all to leave
- c the queen turned them into wild swans

The swan princes searched for their sister. When they found her she promised to break the spell. She made eleven magic shirts for her brothers. When she threw the shirts over the swans, they were transformed into eleven handsome princes. But the girl did not have enough magic cloth to finish all the shirts. The youngest prince would forever have a swan's wing in the place of an arm.

What happened to the prince's arm?

- a his arm was very sore
- b his arm was covered in white feathers
- c his arm was covered in bandages

Story 12

Long ago there lived a young girl named Brave Heart who had learned to hunt. One winter, when the young men were out hunting, the women of the village saw some warriors heading towards them. They were filled with fear. With the men out hunting they did not know who would protect them. The young girl vowed to defend the women. She took up her bow and arrow and circled the villagers.

Which of the following best describes Brave Heart?

- a the young girl was a good hunter
- b the young girl was a good seamstress
- c the young girl was an indian brave

The women watched in wonder as Brave Heart transformed herself into a rainbow. The beautiful rainbow arched over the village. The warriors too saw the mighty rainbow arch above them. Brave Heart stood in the midst of the rainbow. The young girl raised her bow and started to shoot. Lightening fell as the warriors fled from the village.

Why did lightening start to fall?

- a there was a storm
- b the arrows were made of lightening
- c there was a fireworks display

List Two

Story 1

John looked behind the wagon and saw the Red Indians approaching. He called for his father. Henry grabbed his rifle and pulled himself up to the front of the wagon. As the Red Indians rode closer, John could see they were carrying weapons. He raised his own weapon and aimed at the enemy. But Henry told John to wait so John took his finger off the trigger and slowly lowered his gun.

What was John wearing?

- a John wore a three piece suit
- b John wore a cowboy hat
- c John wore a gun and holster

The Indians were so close that John could see their brightly coloured faces. John's hand was firmly on his rifle as he waited in the wagon. Then Henry gave the signal and John and Henry opened fire. The battle lasted only a few minutes but seemed a lot longer. Finally, the Indians had weakened. They turned and disappeared into the mountain range.

Why did the Indians disappear?

- a the indians were beaten
- b the indians were hungry
- c the indians were frightened

Story 2

The fishermen were lost in a storm. Their boat was being thrown from side to side by the wind. Many of their belongings had been lost overboard and a huge wave had crashed across the deck, smashing the mast. The men struggled to keep control of the wheel. They used all their strength to fight the wind and rain.

Why did the men need all their strength to fight the wind and rain?

- a the boat was hard to steer
- b the wheel was hard to turn
- c the boat was easy to steer

The men knew they were running out of time. The boat was being forced towards the rocks. The Captain knew they could not save the ship and he ordered the men into the

life boat. The ship crashed into the rocks. A huge hole appeared in the side of the hull. The ship filled with water and began to sink.

- Which of the following best describes the men?
- a the men were boiling hot
 - b the men were freezing cold
 - c the men were soaking wet

Story 3

The spy dropped from the wall surrounding the large house. He pulled out his torch and made his way over to the open window. As quietly as possible, he pulled himself over the window ledge and crept into the office. He found the safe hidden under the desk. The spy entered the code into the panel and the door opened. He found the plans in the safe.

- What did the spy carry?
- a the spy carried a handbag
 - b the spy carried a gun
 - c the spy carried a torch

Carefully, the spy photographed the plans. Then he slipped them back into the safe where he had found them. With the camera film safely in the bottom of his backpack, the spy slipped out of the office the way he came in. In the darkness, he found the wall once more. He jumped safely to the ground and ran into the dark night.

- How did the spy escape?
- a he climbed over the wall
 - b he climbed over a gate
 - c he climbed up a tree

Story 4

Emily sat happily in the sand at the edge of the sea. She was building a castle with her bucket and spade. Carefully she filled her bucket to the top with sand. Then she turned the bucket over in the sand. Emily tapped the bucket gently with her spade. Then she lifted the bucket and looked underneath.

- What did Emily see under the bucket?
- a Emily had made a cake
 - b Emily had built a sandcastle
 - c Emily had captured some crabs

Emily was very pleased with herself. She loved building castles on the beach. Emily set about filling her bucket again. She wanted to build more castles. But Emily had not noticed that a wind had begun to blow in from the sea. She did not see the huge wave coming towards her. The wave crashed heavily onto the sand and washed away Emily's sandcastle. Emily began to cry.

What was Emily wearing?

- a Emily wore a swimming suit
- b Emily wore a party dress
- c Emily wore flip flops

Story 5

The cowboy was riding his horse at great speed through the mountains. He could hear the bandits drawing closer. The cowboy could feel his horse was tiring. He knew he had to stop but he had to evade capture. Taking a risk, the cowboy rode into a cave hidden behind a rock and waited. The bandits could not see him in the cave.

What was the cowboy wearing

- a the cowboy wore a gun and holster
- b the cowboy wore a tie
- c the cowboy wore blue jeans

From the safety of the cave, the cowboy watched as the bandits searched for him in the valley below. They remained there, circling on their horses for some time. They were clearly disappointed to have lost their prey. When darkness fell, the leader ordered the men back to the hideout. They would have to think of another way to capture the cowboy.

Why could the bandits not capture the cowboy?

- a the bandits were tired
- b the bandits were scared of the dark
- c the bandits could not hunt in the dark

Story 6

John and Francis watched as the bear approached the girls. The bear was growling and showing its teeth. The girls were terrified but were frozen with fear. Francis fired his pistols one after the other. A bullet struck the bear's ear and it reared up on its hind legs. It let out a terrifying howl.

- Why did the bear howl?
- a the bear was frightened
 - b the bear was singing
 - c the bear was in pain

Cathie began to scream. The bear was moving slowly, getting closer and closer. Now that it was in pain it was even more dangerous. Snarling, the bear swiped at the girls. John knew he had to act fast. He quickly raised his rifle to his shoulder and fired once. The bear was badly hurt by John's bullet.

- What sort of paws did the bear have?
- a the bear had huge paws
 - b the bear had small paws
 - c the bear had yellow paws

Story 7

The Witch was very cruel to Dorothy and Toto. She made Dorothy work very hard all day and threatened to hit her. Once the Witch struck Toto so hard with her umbrella that he flew at her and bit her hard on the leg. But the Witch was too wicked to feel anything at all. The old woman just laughed out loud.

- Why did the old woman laugh?
- a the old woman thought Toto was funny
 - b the old woman did not mind the pain
 - c the old woman did not feel the pain

Dorothy was very angry. She shouted at the Witch to stop being so cruel. But the Witch just sneered and ordered her to get back to work. The little girl was so mad, she picked up a bucket of water and threw it at the Witch. The Witch was soaked with water from head to foot. Dorothy watched as the water melted the witch away.

- Which of the following best describes the witch?
- a the witch had a black cat
 - b the witch flew on a broomstick
 - c the witch flew on an umbrella

Story 8

When the battle was over, Lucy found Edmund a little way back from the fighting line. He was covered with blood, and his face was a nasty green colour. Lucy remembered the precious red liquid she had been given for a Christmas present. Aslan the lion told

her to give some liquid to Edmund. The liquid helped to heal Edmund's wounds.

What was special about the lion?

- a the lion was very gentle
- b The lion was able to talk
- c the lion was small as a mouse

Edmund fought well in the battle. He had beaten the giants and broken the Witch's wand. Without him the battle may have been lost. Aslan the lion walked slowly towards the boy. He was proud of Edmund. Aslan put something on Edmund's jacket. Edmund stared proudly at the shiny gold medal pinned there.

Where did Edmund get his medal?

- a Aslan gave Edmund the medal
- b the PE teacher gave Edmund the medal
- c the Witch gave Edmund the medal

Story 9

One night Sam woke up with a start. He heard a strange sound coming from outside his window. Sam crept out of bed, hid behind the curtain and looked outside. He let out a small scream. An enormous spaceship appeared to have landed in his garden. The huge spaceship filled his garden completely and the flower-beds were all completely ruined.

What did the ship look like?

- a the ship had curtains at the window
- b the ship was shaped like a saucer
- c the ship was covered in flashing lights

A door opened in the side of the ship and a strange figure appeared. It looked almost like a man but had long arms, short legs and a huge head with haunting black eyes. Sam saw the funny looking man and cried out. The creature looked up in surprise. Then the door closed and the ship rose slowly off the ground.

Why did the ship start to fly away?

- a Sam had given the creature a fright
- b Sam had made the creature happy
- c Sam had made the creature angry

Story 10

The mermen lived far out to sea, where the water is blue as the bluest cornflower. The

merman king had six beautiful daughters and the youngest was the prettiest of them all. Her skin was soft and delicate as a rose leaf and her eyes were as blue as the deepest sea. The young girl had a beautiful shiny tail. She could swim as fast as any other creature in the ocean

Why was the girl able to swim as fast as the other sea creatures?

- a the girl was a diver
- b the girl was a mermaid
- c the girl was a fish

One night, the mermaid rescued a prince from the stormy sea. They fell in love, but the mermaid could not live out of the water. So the mermaid sang sadly to the prince every night. The merman king could not bear to see his daughter so sad. The king cast a spell over the mermaid and her tail disappeared. Two perfect legs appeared in the place of her tail.

Which of the following best describes the king?

- a the king wore high heels
- b the king wore an luxurious robe
- c the king wore a gold crown

Story 11

There once was a king who had eleven sons and one daughter. The king married a wicked queen who was jealous of the children. She sent the little girl away and made the king banish his sons from the kingdom. On the day the boys left the palace, the queen turned the boys into swans. Instead of eleven boys, eleven wild swans flew from the palace gates.

What did the boys look like?

- a the boys had big noses
- b the boys were covered in white feathers
- c the boys had webbed feet

The swan princes searched for their sister. When they found her she promised to break the spell. She made eleven magic shirts for her brothers. When she threw the shirts over the swans, they were transformed into eleven handsome princes. But the girl did not have enough magic cloth to finish all the shirts. The youngest prince would forever have a swan's wing in the place of an arm.

Why did the prince have a swan's wing in the place of an arm?

- a the magic shirt was too small
- b the magic shirt was missing a button
- c the magic shirt was missing a sleeve

Story 12

Long ago there lived a young girl named Brave Heart who had learned to hunt. One winter, when the young men were out hunting, the women of the village saw some warriors heading towards them. They were filled with fear. With the men out hunting they did not know who would protect them. Brave Heart took up her weapon and started to sing. Soon a lovely rainbow appeared in the sky.

How did the rainbow appear?

- a the women made the rainbow appear
- b Brave Heart made the rainbow appear
- c the hunters made the rainbow appear

The women watched in wonder as Brave Heart transformed herself into a rainbow. The beautiful rainbow arched over the village. The warriors too saw the mighty rainbow arch above them. Brave Heart stood in the midst of the rainbow. The young girl raised her bow high and started to shoot lightning arrows. The warriors ran away in fear.

What was special about the bow and arrow?

- a the bow and arrow were magical
- b the bow and arrow were ancient
- c the bow and arrow were expensive