A Sheffield Hallam University thesis

The Development of Real-World and Counterfactual-World Inference Generation Abilities during Adolescence

Rebecca Hancock

A thesis submitted in partial fulfilment of the requirements of Sheffield Hallam University for the degree of Doctor of Philosophy

September 2016

Declaration

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

Name: Rebecca Hancock

Signed:

Date: 23rd September 2016

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

ABSTRACT

Government statistics suggest that around one-third of children and young people do not make the expected progress in reading comprehension and text production during secondary school (Department for Education, 2015a; 2015b). Inference generation abilities are at the core of skilled models of comprehension (e.g. Kintsch, 1988; Zwaan, 2003). However, previous research exploring the development of inference generation abilities typically compares a child group (9years or less) with an adult group (18years or over). To assess literal, coherence inference and elaborative inference processing in real-world and counterfactual-world conditions in a wide age-range a new measure was created: the Image Selection Task (IST). The IST embeds a self-paced reading methodology and forced-choice picture-selection task into short stories to explore changes in inference generation skill (number of errors) and time-course (speed). The IST was found to be a valid measure of inference generation abilities and, due to reducing demands on translation and expressive language processes, potentially a purer measure than existing tasks which utilise verbal response methods. The IST was used to explore age-related changes in inference generation abilities in Year 5, Year 7, Year 9 and adults using a cross-sectional design. Findings suggest that inference generation skill plays a role in both reading comprehension and text production during adolescence, with skill increasing until Year 7. Error patterns suggest that coherence inferences were no more difficult than elaborative inferences for Years 5, 7 and adults. Year 9, however, found coherence inference generation more difficult than elaborative inference generation. Inference generation speed was found to improve until Year 9 for both real-world and counterfactual-world information, with time-course patterns comparable across all age groups, such that, in real-world and counterfactual-world conditions, coherence inferences were generated online and elaborative inferences offline. Real-world coherence inference generation skill was found to be underpinned by knowledge during adolescence. Both knowledge and inhibitory control appear to play a role in real-world elaborative inference generation skill during adolescence. Counterfactual-world inferences were both found to be underpinned by belief biases and inhibitory control. However, the direction of effect of inhibitory control reversed (from positive to negative) in Year 9. Further research is needed to explore whether this effect is specific to the current sample or more generalisable. Based on the findings, educational recommendations are provided, including suggestions for assessment and activities at specific points in the reading process. The recommended activities focus on promoting those skills underpinning inference generation.

LIST OF TABLES

LIST OF FIGURES

OVERVIEW

This thesis is split into three parts. Part 1 is concerned with reviewing the existing body of research exploring inference generation abilities. Therefore, a literature review is presented in Chapter 1. Part 2 of this thesis focuses on inference generation task selection, design, and evaluation. A review of current inference generation measures is presented in Chapter 2. This review revealed that no one measure is suitable for the assessment of both inference generation skill and time-course across a wide-range of age groups. Therefore, a new researcher-designed inference generation measure is introduced at the end of Chapter 2: The Image Selection Task (IST). The study reported in Chapter 3 evaluated the appropriateness of the inference-evoking texts and pictorial stimuli used in the IST. The study reported in Chapter 4 evaluated the utility of the IST by comparing patterns of results from the IST with those of an existing measure using the same methodology. The IST was found to be a valid measure of inference generation. Part 3 of this thesis presents the data and analysis of one large study conducted using the IST to assess inference generation abilities during adolescence. The methodology underpinning the large study is reported in Chapter 5. In Chapter 6 a relationship between inference generation and reading comprehension and text production is established. The reading comprehension ages of each age group were also explored, resulting in the construction of four new groups, split by reading comprehension age instead of chronological age. These reading comprehension age groups were used to conduct additional analysis in Chapters 7 and 8. The developmental trajectory of inferential skill (number of errors) and time-course (speed) was explored in Chapter 7 and the development of those cognitive skills underpinning inference generation during adolescence was explored in Chapter 8. A general discussion of findings is presented in Chapter 9.

Part 1: A Literature Review

CHAPTER 1 LITERATURE REVIEW

The ability to successfully comprehend a text is one of the most important skills a child or young person (CYP) must develop, given the importance of reading comprehension to educational success and later professional success (Cain & Oakhill, 2006a; Meneghetti, Carretti, De Beni, 2006; National Center for Educational Statistics, 2009; 2011; Salahu-Din, Persky, & Miller, 2008). However, between 2010 and 2015, around one third (31-35%) of CYP finished Secondary School without achieving the literacy levels expected (Department for Education, 2015a). Government statistics suggest that some of these CYP did achieve age-related expectations (ARE) at the end of Primary School, such that only 11-17% of CYP completed Primary School without achieving literacy ARE between 2005 and 2010. This suggests that for some CYP Secondary School presents challenges concerning the development of literacy skills. This is further evidenced by progress rates, such that between 2010 and 2015, less than one fifth (8-17%) of CYP failed to make the expected progress during Primary School, whereas around one third (27-31%) of CYP failed to make the expected progress during Secondary School (Department for Education, 2015a). This suggests that some CYP have the literacy skills needed for success during Primary School but begin to struggle during Secondary School. The primary purpose of this thesis is to build upon the existing body of literature used to inform reading comprehension teaching and intervention during adolescence[1].

Reading comprehension is multi-faceted, resulting from the interaction of several skills, the exploration of all would be beyond the scope of this thesis (Kintsch 2012; Oakhill & Cain, 2011; Perfetti & Adlof, 2012; van den Broek, 2012). Therefore, for the reasons listed below, inference generation abilities are the focus of this thesis. Zwaan & Singer (2003) suggest that almost all aspects of comprehension are at least partially inferential, with inference generation at the heart of many skilled models of reading comprehension (Kintsch, 1988; van den Broek, Rapp, & Kendeou, 2005; Zwaan, 2003). Demands for inferential, as opposed to simply literal understanding, also increase as the CYP moves from Primary to Secondary School (AQA, 2013; Department of Education, 2013; Leach, Scarborough & Rescorla, 2003). Similarly, research exploring poor comprehenders, finds that whilst poor comprehenders are a heterogeneous group, many often have weak

---

[1] In this thesis adolescence refers to 9-10 years to adulthood (18 years and over). This range was selected as, as highlighted in Sections 1.1.2. and 1.2.2. this appears to be a potential period of qualitative change.

inference generation skills (Cain & Oakhill, 1999; Cain & Oakhill, 2006a; Cornoldi, de Beni, and Pazzaglia, 1996; Fletcher & Vaughn, 2009). Whilst teachers often identify weak inference generation skills as a need, in this researcher's experience, teachers are often unable to go any further than this – i.e. state a specific class of inferences the CYP is struggling with, state which aspect of inference generation a CYP is struggling with, or provide a specific strategy to support the CYP. This suggests a lack of understanding of what inference generation is and what teachers should expect from a CYP at any given time. It is vital that a comprehensive understanding of the role of inference generation in the reading comprehension process during adolescence is held as this could support those CYP who are not making the progress expected during the Secondary School years. The typical development of inference generation abilities during adolescence is the focus of this thesis. See Figure 1.1 for an overview of this Chapter.

Real-World Inference Generation
This section defines inference generation and the various inference types. The Constructionist Theory is drawn upon to categorise inference types and make predictions about the associated time course (Graesser, Singer & Trabasso, 1994). Whilst several other theories exist (e.g., the minimalist hypothesis; McKoon & Ratcliffe, 1992), the Constructionist Theory is drawn upon in this study as it is currently the most prominent and supported theory of inference generation. The development of the different inference types is then discussed, with a specific focus on those studies exploring age-related changes during the Secondary School years.

Counterfactual-World Inference Generation
This section defines counterfactual-world inferences and their use in the classroom. Given a lack of research focusing specifically on counterfactual-world inference generation, research exploring counterfactual-world text processing and counterfactual-world reasoning is discussed. The development of different inference types is discussed, with explanations for age-related changes being drawn from all three bodies of literature.

Components of Inference Generation
This section outlines the neurological basis of inference generation in order to highlight its multifaceted nature. Developments in those brain areas implicated in inference generation are explored. Four key skills thought to underpin real-world and counterfactual-world inference generation are identified: Amount of Knowledge, Accessibility of Knowledge, Inhibitory Control, and Belief Biases. These skills and their role in inference generation are discussed.

Inference Generation and Literacy Skills
This section outlines the role of inference generation abilities in two key higher-order literacy skills: Reading Comprehension and Written Language. Inference generation is discussed within the constraints of the Simple View of Reading and the Simple View of Writing as these models are prominent within the literature and used to inform educational practice (Gough, Hoover, & Peterson, 1996; Juel, 1988). Limitations of the application of these models to the Secondary School years are discussed. Hayes model of writing development is used to hypothesise about the role of inference generation abilities in writing development.

Summary of Research
This section presents a summary of the research and research aims.

Figure 1.1. Summary of Chapter 1

1.1. Real-World Inference Generation

Comprehension is more than the simple decoding of words. It is the ability to extract the relevant (and often implicit) meaning of a text. Van Dijk and Kintsch (1983) propose three levels of mental representation of a text: surface-level, propositional-level, and situational-level – see Figure 1.2. The surface-level is a verbatim representation of the words and phrases presented in the text. The propositional-level represents the explicit semantic meaning of the text but exact words and phrases are often lost. The situational-level represents an integration of both the semantic meaning of the text and a reader's relevant background knowledge. Subsequently, the structure of the text is often lost, with information fitting into the reader's pre-existing knowledge and beliefs. Unlike the preceding levels, the situational model is thought to contain both explicitly and implicitly stated information regarding the characters, objects, settings, events and actions (Graesser, Singer, & Trabasso, 1994; Schmalhofer, McDaniel & Keefe, 2002). The situational model is associated with true comprehension. The construction of a situational model of the text is thought to require more effort than the construction of either a surface- or propositional-level model, however (e.g. Friese, Rutschmann, Raabe & Schmalhofer, 2008). The situational-level representation, though, is found to be more robust and long-lasting, being stored in long-term memory (Kintsch et al., 1990). For comprehension in the classroom, it is the information contained in the situational model that the CYP is likely to acquire and store (i.e. learn) and thus be able to retrieve at a later date (i.e. recall for use in an assessment).
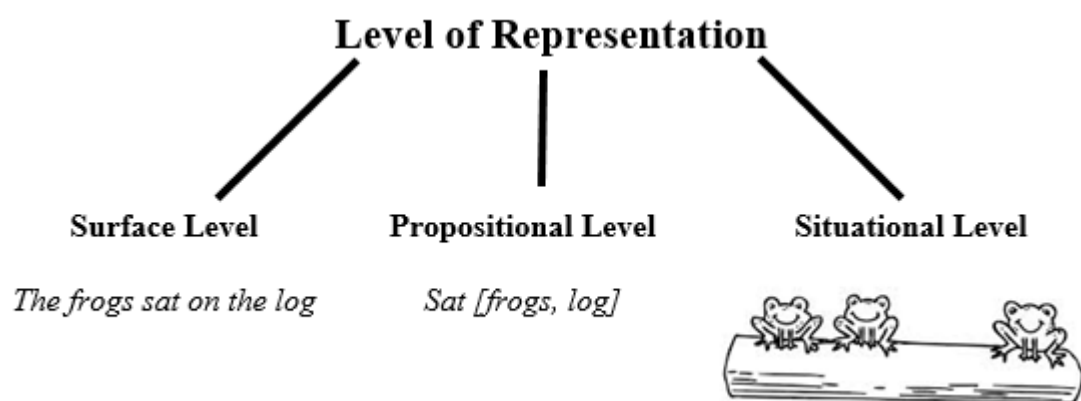
Figure 1.2. A Sample of a Schematic Representation of the Three Levels of Mental Representation

Several theories have been proposed to explain the construction of a situational model (e.g. Kintsch, 1988; 1998; Kintsch & Van Dijk, 1978; van den Broek et al., 2005; van den

Broek, Young, Tzeng, & Linderholm, 1999; Zwaan, 2003). The fundamentals of most skilled models are the same, therefore, whilst discussion and evaluation of all of these models is beyond the scope of this thesis it is possible to briefly outline the general principles of mental model construction. Knowledge is thought to be activated via an associative mechanism whilst reading. The knowledge and text activated in the reader's working memory are integrated into the reader's mental model following several cycles of iteration, whereby strong connections become stronger and weaker connections are lost. Additionally, most models converge to suggest that for full comprehension – i.e. for an enriched situational model to be constructed – inferences are needed. For example, consider the passage below:

"Lucy was in the woods. It was getting dark and she had not eaten all day. Lucy spotted a rabbit to her right. She crouched down low and then slowly began to sneak up on the rabbit, one paw at a time. When she was close enough, she pounced. Her tail wagged as she tucked into her meal."

To fully understand even this short passage several inferences must be made. The reader must infer that Lucy is a dog, it is early evening, Lucy is hungry, therefore she wanted to catch the rabbit and eat it, Lucy snuck up on the rabbit so it didn't run away, Lucy caught the rabbit and ate it. The reader must also make a connection between 'Lucy' and 'she'. The reader could also make additional inferences concerning the breed and colour of dog, the colour and size of rabbit, etc. Inferences are thus fundamental to successful mental model construction, with the number and range of possible inferences numerous.

Whilst inference generation is positioned at the heart of situational model construction, inference generation appears to be a distinct process, above and beyond literal understanding. Neuroimaging research finds that when participants read a text that requires an inference, those brain areas associated with literal text comprehension are activated along with several other areas (typically the inferior prefrontal gyrus and superior temporal gyrus; e.g. Jin, Liu, Mo, Fang, Zhang & Lin, 2009; Jung-Beeman, 2005; Kim, Yoon, Kim, Lee & Kang, 2012; Mason & Just, 2004; Prat, Mason & Just, 2011; Virtue, Parish, & Jung-Beeman, 2008). Those additional areas of the brain activated when reading texts requiring an inference are related to skills and processes such as the selection of appropriate knowledge, semantic integration and detection of inconsistencies (Tyler & Marslen-Wilson, 2008; Zhang, Feng, Fox, Gao & Tan, 2004; Zhu, Zhang, Wang, Xiao, Huang & Chen, 2009). Given its unique and distinct nature, specific exploration of the

inference generation process during adolescence may prove fruitful in furthering understanding of reading comprehension during this time. Understanding which inferences CYP generate whilst reading and the processes underpinning this is vital if specific literacy needs are to be identified for those making slow progress during the Secondary School years.

1.1.1. Categorisation of Inference Types

Although different terminology may be used, two distinct types of inference are consistently identified: 1) coherence and 2) elaborative (e.g. Graesser et al., 1994; Singer & Ferreira, 1983). Coherence inferences refer to those inferences essential for the construction of a coherent mental representation of a text. They may link a referent to its owner or link explicitly stated information by drawing upon existing background knowledge to bridge a gap – see Table 1.1. for examples and Section 1.1.1.2.1. for further discussion. Elaborative inferences are not necessary for maintaining coherence, but do create a more enriched representation of the text – see Table 1.1. for examples and Section 1.1.1.2.2. for further discussion. Several different types of inference fall into each category. Consequently, comprehensive categorisation systems have been developed. Table 1.1 provides an overview of the thirteen different classes of inference outlined by Graesser et al. (1994).

Table 1.1. Graesser, Singer, & Trabasso's (1994) Inference Taxonomy

Inference Type	Description	Example	Needed For
Referential	Links referent to its corresponding element in the text	Billy picked up the ball and walked away with it Inference: 'it' refers to 'the ball'	Local coherence
Case Structure Role Assignment	The assignment of a particular role to a noun	Ben leant his bike against the wall. Inference: against the wall is assigned a location role.	Local coherence
Causal Antecedent	Bridges two ideas or concepts together	It was Sally's friend's Birthday party in a few hours. Sally rushed to the shopping centre Inference: Sally was going to the shopping centre to buy her friend a present	Local coherence AND Explanations
Superordinate Goal	The goal that motivates an action or event	Sally dreamed of the new family car as she raced off to her second job. Inference: Sally needs money to buy a car	Explanations AND Global coherence
Thematic	The main point or moral of the text	Loose lips sink ships: Inference: Don't gossip as you may be talking to a spy who could then use said information against you.	Global coherence
Character Emotional Reaction	The emotional response of a character	Billy was being bullied. Inference: Billy was sad	Global coherence
Causal Consequence (predictive)	The prediction of a potential consequence of an event	The man threw the vase at the wall Inference: The vase smashed	Elaboration
Instantiation of Noun Category	Activation of examples or subcategories associated with a noun or verb.	The farmer tended to his animals Inference: The farmer's animals are cows, pigs and sheep.	Elaboration
Instrument	Instantiation of the tool used to complete an action.	The man dug a hole Inference: The man used a spade to dig the hole	Elaboration
Subordinate Goal-Action	The action or goal explaining how an action is achieved.	Sally bolted down her dinner Inference: Sally grasped her fork and quickly moved it back and forth between her plate and mouth.	Elaboration
Static	Elaboration of the characteristics of a character or objects, traits, knowledge, or beliefs. These elaborations are not causally related to the story.	The tomato rolled down the hill. Inference: The tomato was red	Elaboration
Emotion of Reader	The emotional response of the reader	The sink was overflowing with sewage, rotting fish heads and mouldy cheese floated on top Inference: The reader is disgusted.	Pragmatic communicative exchanges
Author's Intent	Authors reason for writing the passage	Loose lips sink ships Inference: Discourage the public from gossiping	Pragmatic communicative exchanges

As can be seen from Table 1.1., the number of inferences that could be drawn whilst reading is numerous. However, a reader's processing capacity is limited. Therefore, it is unlikely that all possible inferences are drawn whilst reading. One of the most prominent and controversial areas of investigation in inference generation is thus determining the time-course of different inference types. Specifically, which inferences are generated online and which inferences are generated offline. *Online inferences* are those inferences drawn whilst reading. The time-course of their generation is argued to be rapid, since they are drawn with no or little strategic effort (e.g. Bowyer-Crane & Snowling, 2010; Calvo, Castillo, & Schmalhofer, 2006; Ford & Milosky, 2008). *Offline inferences* are those that require strategic effort and, as a result, are not drawn whilst reading. The time-course of offline inference generation is argued to be significantly longer than the inferential time-course of online inferences (e.g. Campion, 2004; Calvo et al., 2006).

Online inferences are those inferences generated whilst reading, with minimal strategic effort – i.e. automatically. However, this definition seems to conflict with the definition of reading comprehension being the result of a complex interaction between a variety of skills and processes (e.g. Kintsch 2012; Oakhill & Cain, 2011; Perfetti & Adlof, 2012; van den Broek, 2012). The nature of an online inference must be understood within the wider context of reading comprehension. Automaticity is not absolute but relative, such that automatic processing occurs faster and as the result of less conscious effort than strategic processing (Thurlow & van den Broek, 1997). Knowledge-based inferences depend on the activation and integration of the necessary knowledge from long-term memory (Graesser et al., 1994). Therefore, if the knowledge necessary for inference generation is not automatically activated the reader may have to engage in a conscious, strategic search of long-term memory (Ericsson & Kintsch, 1995). In this case, inference generation would be offline. When reading, knowledge related to the text is thought to be automatically activated via associative mechanisms (Kintsch, 1988; van den Broek et al., 1999; 2004). The knowledge must then be integrated with text nodes and a link formed (Graesser et al., 1994; Kintsch, 1998). This process is more demanding than making no connection between knowledge and real-world information, such that texts evoking inferences result in longer reading speeds and higher levels of brain activity (e.g. Casteel, 1993; Virtue, Haberman, Clancy, Parrish, & Jung-Beeman, 2006). However, when a strategic search of long-term memory is not required, this process is routinely found to occur within 400ms (Garrod, O'Brien, Morris, & Rayner, 1990). Skilled models of reading comprehension suggest that text is processed in cycles (occurring at the end of a

sentence, clause or phrase), with each cycle taking no more than 400ms (Kintsch, 1998). Therefore, within the bounds of reading comprehension, an online, automatic inference is one that is activated and fully integrated into the situational model within 400ms. Conversely, an offline, strategic inference can be defined as an inference that, whilst not integrated into the situational model within 400ms, can, with a strategic and conscious search of long-term memory, still be generated when the reader is given more time.

1.1.1.1. Experimentally Determining the Online-Offline Nature of an Inference

Inference generation measures assessing the online-offline nature of different inference types are underpinned by two assumptions. First, as discussed above, the generation of a link between two or more text-based ideas and the reader's existing background knowledge is more demanding than literal processing (e.g. Casteel, 1993; Virtue et al., 2006). Therefore, measures that tap the inferential time-course whilst reading (e.g. reading speeds, eye movements, brain-imaging) conclude that an inference has been drawn online if more cognitive effort is observed compared to the processing of explicit information (measured by slower reading speeds, increased regressions and longer fixations, and increased neuronal activity). Second, online inferences are integrated into the situational model within 400ms at which point they should be readily accessible in the reader's working memory in the same way as explicitly stated information. Therefore, inference generation measures that tap the inferential time-course by recording responses after reading (e.g. lexical decision, naming, validation) conclude that an inference has been drawn online if the inferential item is responded to just as fast as explicitly stated information. This is known as a facilitation effect. See Chapter 2 for discussion of current inferential assessment tools.

1.1.1.2. The Constructionist Theory

The Constructionist Theory, proposed by Graesser et al. (1994) has three critical assumptions:

1. *The reader goal assumption* – the reader will attempt to construct a situational model of the text that addresses their specific goals.
2. *The coherence assumption* – the reader will attempt to construct a situational model that is both locally and globally coherent.
3. *The explanation assumption* – the reader will attempt to construct a situational model that explains why particular events and actions have occurred.

As can be seen from Table 1.1., Graesser et al. (1994) propose three general classes of inferences: coherence inferences, elaborative inferences, and pragmatic communicative exchange inferences. Based on the three assumptions outlined above the Constructionist Theory predicts that coherence inferences are routinely drawn online whereas elaborative and pragmatic communicative exchange inferences are drawn offline. These predictions were supported by Graesser et al. (1994) using a three-pronged method, combining verbal protocols (think-aloud and question-answering, question-asking tasks), timed behavioural measures (reading speeds, response latencies on naming and lexical decision tasks) and theoretical predictions. The Constructionist Theory acknowledges that there will be exceptions to these predictions. For instance, if the reader's goal is simply to proofread, as opposed to comprehend the text; they are unlikely to be seeking coherence or explanations. Moreover, whilst, for the sake of simplicity, Graesser et al. (1994) make a dichotomous distinction between online and offline inferences, they recognise that in reality the online-offline nature of inferences will run on a continuum.

1.1.1.2.1. Coherence Inferences

Coherence inferences are those that form a link between ideas both explicitly stated in the text and those that can be implied. Typically, this requires the activation and integration of background knowledge from long-term memory. As can be seen from Table 1.1., according to Graesser et al. (1994), a number of inferences fall into this category: Referential, Case Structure Role Assignment, Causal Antecedent, Superordinate Goal, Thematic, and Character Emotional Reaction. The most commonly researched coherence inference is the causal antecedent inference which bridges two ideas or concepts together.

"Jack wanted a cup of tea, but there was no milk. Jack put on his coat and headed to the shop."

To understand *why* Jack headed to the shop it is necessary to link the second sentence with the first by inferring that Jack was going to the shop to buy some milk. To do this the reader must first activate the mediating idea that shops sell milk from their background knowledge. Without this link, the two sentences would appear unrelated and incoherent.

Coherence inferences are consistently found to be drawn online (e.g. Bowyer-Crane & Snowling, 2010; Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Casteel, 1993; Ford & Milosky, 2008; Kim et al., 2012; Long, Seely, & Oppy, 1996; Prat et al., 2011; Suh & Trabasso, 1993; Virtue et al., 2006; 2008). Specifically, the knowledge needed for

coherence inferences is thought to be automatically activated via associative mechanisms and have strong and numerous connections to other nodes within the model. In addition, the Constructionist Theory suggests that constructionist processes may heighten the strength of a connection if it provides coherence or an explanation. Therefore, when iteration occurs the coherence inference is strongly activated and quickly integrated into the situational model (within 400ms).

1.1.1.2.2. Elaborative inferences

As can be seen from Table 1.1., a number of inference types are classified as elaborative: Causal Consequence, Instantiation of Noun Category, Instrument, Subordinate Goal-Action, and Static. Like coherence inferences, the generation of elaborative inferences is dependent on the activation and integration of background knowledge from long-term memory.

"The young girl was dusting when a huge wolf crept up behind her. As the girl turned around the wolf snarled and let out a terrifying howl."

The reader could generate numerous elaborative inferences from the text above. For instance, it could be inferred that the wolf is going to eat the girl (causal consequence), the girl was dusting with a cloth (instrumental), or the wolf has big sharp teeth (static). Whilst these inferences serve to enrich the text, they are not necessary for coherence. Moreover, as is often the case, that due to limited contextual constraints these inferences are not certain – e.g. the girl could be dusting with a feather duster, old rag, or even teddy bear. Due to limited working memory capacity, it is unlikely that the reader could generate and maintain all possible elaborative inferences whilst reading without overloading working memory (Baddeley, 1986; Just & Carpenter, 1992). Even if they could, this would likely be an inefficient use of resources as, due to multiple possibilities, as the text unfolds the explanations are likely to be revealed as erroneous (e.g. Albrecht, O'Brien, Mason, & Myers, 1995; Calvo et al., 2006). This is supported by the Schmalhofer et al. (2002) model of discourse processing, which - developed within Kintsch's Construction-Integration model (Kintsch, 1988; 1998) - suggests that coherence and elaborative inference generation result from the same construction processes, but differ at the point of integration. Specifically, whilst coherence inferences are confirmed by the text and thus integrated, due to multiple possibilities, elaborative inferences are not. The Constructionist Theory (Graesser et al., 1994) predicts that elaborative inferences will be drawn offline (coherence assumption). There is a large body

of research supporting this, such that, typically, a facilitation effect is only observed for elaborative inferences after a time interval of 800ms – 1250ms (e.g. Calvo & Castillo, 2001a; 2001b; Calvo et al., 2006; Campion & Rossi, 2001; Long & Golding, 1993).

The Constructionist Theory (Graesser et al., 1994) suggests that constraining the context and thus minimising the number of plausible predictive inferences allows for the online generation of predictive inferences, however, as there are significantly fewer alternatives to forecast and thus working memory is not overloaded. This is supported by a number of studies finding that elaborative inferences are generated online when the stimuli used results in possibilities that are highly constrained by the text and/or knowledge needed to generate the elaborate inference is readily accessible (Calvo, 2000; Calvo, Castillo & Estevez, 1999; Jin et al., 2009; Klin, Guzmán, & Levine, 1999; Cook et al., 2001; Till, Mross, & Kintsch, 1988; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Weingartner, Guzman, Levine, & Klin, 2003). Similarly, Virtue et al. (2008) have shown that predictability underpins the neural processing engaged in, such that highly constrained text results in a qualitatively different process to less constrained text. This led Virtue et al. (2008) to conclude that individuals only generate inferences when textual constraints are low when they are forced to do so.

Some studies have found facilitation effects for elaborative inferences after time intervals of 500ms or less without a highly-constrained context (e.g. Cook et al., 2001; Peracchi & O'Brien, 2004). These studies typically employ a self-paced reading methodology - through a series of button presses, the participant moves through the text at their own speed. It is argued that when a self-paced methodology is used the reader is able to anticipate *when* the test item will appear and predict *what* the intended elaborative inference will be (Calvo et al., 2006; Keenan, Baillet, & Brown, 1984). Consequently, participants are able to engage in the strategic processing needed to generate this inference *before* moving onto the test item. Therefore, by the time the test item is presented the reader has already activated the information needed to respond correctly. Elaborative inference items may appear to be responded to just as fast as coherence and literal items, but the elaborative inferential processing may actually be occurring offline.

Facilitation effects have been observed for elaborative inferences after time intervals of 400ms or less, even when anticipation effects have been controlled for using a Rapid Serial Visual Presentation methodology – see Chapter 2 (e.g. Fincher-Kiefer, 1995; 1996; Keefe & McDaniel, 1993; McKoon & Ratcliff, 1986, 1989). Whilst at first this appears

contradictory, the level to which an inference is encoded must be considered. Skilled models of reading comprehension suggest that the knowledge activation process is dumb (e.g. Kintsch, 1998). Subsequently, initially, knowledge needed for the generation of a range of elaborative inferences may be activated, but only minimally – i.e. low activation levels, limited connections. After several cycles of spreading activation, these nodes are likely to be pruned and thus not be integrated into the final situational model. In this way, whilst reading, an elaborative inference can be activated but only minimally encoded into the situational model. Facilitation effects for elaborative inferences after a time interval of 400ms or less may thus reflect inferences that have been minimally activated, but during the process of integration have not been encoded into the situational model, such that when tested after longer time intervals, integration has begun and a facilitation effect is no longer observed.

In sum, numerous inference types exist with most of these falling within either the coherence inference category or elaborative inference category. Coherence inferences are thought to be generated with minimal effort due to the automatic activation of the necessary knowledge and strong and multiple links to other ideas and knowledge. For the most part, elaborative inferences are thought to be generated offline, such that whilst the necessary knowledge may be activated initially, due to multiple possibilities that are not constrained by the text, this knowledge is quickly pruned from the model, meaning the target elaborative inference must be drawn with strategic effort after reading. If a highly constraining context is presented, meaning few possibilities are likely, and/or knowledge needed is readily accessible elaborative inferences may also be generated online. Similarly, it is possible that elaborative inferences are minimally activated when reading, but through the process of integration are lost due to weak activation levels. These factors must be key considerations when assessing the generation of elaborative inferences.

1.1.2. The Development of Inference Generation Abilities

When exploring age-related changes this thesis refers to two types of development, quantitative and qualitative. Quantitative development refers to changes in quantity – i.e. increases or decreases in a particular skill. For example, if adults read critical texts faster than CYP this would be described as a quantitative change. Qualitative development, on the other hand, refers to different patterns across groups. For example, if, regardless of differences in speed, CYP generated coherence and elaborative inferences offline, whereas adults generated coherence and elaborative inferences online this would be

described as a qualitative change. Research finds that inference generation skill quantitatively improves with age, with adults displaying superior inference generation abilities compared to CYP (e.g. Ackerman, 1986, 1988; Barnes, Dennis, & Haefele-Kalvaitis, 1996; Kendeou, Bohn-Gettler, White & van den Broek, 2008; Oakhill, Cain, & Bryant, 2003). This may be due to qualitative changes in inference generation abilities. Research finds that whilst very young children have the ability to generate inferences from a narrative, they do not do this spontaneously (e.g. Omanson, Warren & Trabasso, 1978; Paris, Lindauer & Cox, 1977), with the ability to spontaneously generate inferences whilst reading thought to emerge at approximately 9 years (Casteel, 1993; Casteel & Simpson, 1991; Paris & Lindauer, 1976; Paris et al., 1977). After 9-10 years, there are thought to be no qualitative developments in the inference generation time-course. Consistent with the Constructionist Theory, research utilising online methodologies to explore the inferential time-course of CYP suggests that, CYP engage in a qualitatively similar process to adults, such that coherence inferences are generated online and elaborative inferences offline (e.g. Bowyer-Crane & Snowling, 2010; Casteel, 1993; Ford & Milosky, 2003; 2008).

For instance, Casteel (1993) asked participants to read short texts containing a target sentence that evoked either a coherence inference, elaborative inference, or no inference (explicit). Reading speeds of the target sentence were recorded to obtain an online measure of inference generation. Casteel found that 9 year olds, 11 year olds, 14 year olds, and adults read sentences requiring a coherence inference slower than sentences that did not require an inference for understanding – i.e. elaborative and explicit. The slower reading speed observed for coherence-inference evoking sentences compared to other sentence types is indicative of the additional cognitive demands needed to generate an inference. This suggests that CYP as young as 9 years are able to spontaneously generate coherence inferences whilst reading. However, as will be discussed shortly, the texts used by Casteel were not reflective of texts used in the classroom. Additionally, sample sizes were small comprising 'average' readers, only such that participants were reading at the level expected for their age. It is possible that those CYP who are not making the expected progress during the Secondary School years are engaging in a different process to those CYP achieving age-related expectations. Therefore, inclusion of all typical developers (where typical development refers to those CYP with no known learning difficulty, performing within 2 standard deviations of the mean) is key to the development of a

comprehensive understanding of age-related changes in the inference generation time-course.

Whilst research suggests no qualitative developments in the inference generation time-course after 9-10 years, quantitative improvements have been observed (e.g. Barnes et al., 1996; Casteel & Simpson, 1991; Casteel, 1993). However, the age at which adult-like performance is achieved is unclear. For instance, Casteel and colleagues found that adult-like performance is achieved in early adolescence (between 10 and 14 years; Casteel, 1993; Casteel & Simpson, 1991). Conversely, Barnes et al. (1996) suggest that the development of inference generation abilities continues to develop throughout adolescence. Differences could be due to the complexity of stimuli used, such that whilst Casteel and colleagues employed short texts (the texts used by Casteel (1993) are just four sentences long) Barnes et al. (1996) employed longer texts, more indicative of texts used in the classroom. Therefore, basic inference generation skills may reach adult-like levels at an early age, but other factors such as distance between information to be integrated and syntactic complexity may determine how successfully inference generation skills can be applied (Duffy & Rayner, 1990; Ehrlich & Rayner, 1983; German & Nichols, 2003). Moreover, whilst Casteel and colleagues used yes or no questions, Barnes et al. (1996) used open-ended questions. Cain and Oakhill (2006b) suggest that yes or no questions are easier than open-ended questions since when answering a yes or no question the inference is explicitly stated. In addition, Barnes et al. (1996) primed the knowledge needed to successfully generate the target inferences to control for knowledge. The results of Barnes et al. may reflect a different process to that employed when reading naturally.

In sum, those studies that have explored inference generation at different points during adolescence, are inconsistent, with adult-like performance being achieved at different ages in different studies. This may be due to methodological differences. Research converges to suggest quantitative gains in inference generation abilities between childhood and adulthood, but no qualitative differences after 9-10 years. Current research is largely limited to 'average' readers, however. Therefore, those CYP who are making slower than expected progress during the Secondary School years are not currently represented. Further research exploring age-related changes in the inference generation abilities of an adolescent sample reflective of the current typical UK school-aged population is needed.

1.2. Learning from Fiction: Counterfactual-World Inference Generation

To be successful in the classroom, and beyond, CYP must be able to engage with not just non-fiction texts, but fictional resources too. Fictional texts are central to the enrichment and reinforcement of subject knowledge. For example, CYP may learn about historical events by reading texts such as 'Goodnight Mr Tom' (Magorian, 1981). CYP may learn about and reinforce their understanding of difficult scientific concepts that often require 'what if' thinking by engaging with text-based materials such as 'The Sarah Jane Adventures' (bbc bitesize, 2014; Chandrasekharan & Nersessian, 2007), with teachers advocating the use of science fiction texts in the classroom to encourage scientific thinking (Dubeck, Bruce, Schmuckler, Moshier, & Boss, 1990). Fictional texts, particularly fantasy and science-fiction texts are often read for pleasure by CYP (Clark & Foster, 2005; Clark & Rumbold, 2006; Zirinksy & Rau, 2001). Reading for pleasure allows for the practice and development of sophisticated comprehension skills, with positive correlations observed between reading habits and ability to comprehend complex texts (Cain & Oakhill, 2011; Moje et al., 2008 Cipielewski & Stanovich, 1997).

Fictional texts can present ideas, and thus require a level of understanding, that runs counter to actual events and/or an individual's real-world knowledge. Subsequently, a counterfactual-world inference is needed. A counterfactual-world inference is an inference drawn from information that violates real-world knowledge. It has been argued that there are two types of counterfactual-world information: plausible and implausible (e.g. Johnson-Laird & Byrne, 2002; McMullen & Markman, 2002). Plausible counterfactual-world information includes "if only" statements (e.g. "if only I hadn't missed the bus, I wouldn't have been late for work") and are also known as 'hypotheticals'. Implausible counterfactual-world information includes information that attempts to manipulate physical laws such as time – i.e. events that would be impossible in the real-world (Byrne, 2007; Zwaan, Langston, & Graesser, 1995). Implausible counterfactual-world information is thought to be more difficult to process than plausible counterfactual-world information (e.g. Kulakova, Aichhorn, Schurz, Kronbichler, & Perner, 2013; Zwaan et al., 1995). However, perhaps due to their commonality, fairy-tales and the events within them – i.e. talking animals – although impossible in the real-world are thought of as plausible counterfactual-world information (e.g. Zwaan et al., 1995). Counterfactual-world plausibility is thus best considered on a scale, with completely implausible information at one end, very plausible information at the other, and the information in most fairy-tales and fantasy texts falling somewhere in the middle.

Counterfactual-world inferences, particularly those generated from implausible information, may require the reader to integrate two ideas that would not normally be connected and may share a connection within the book that is impossible in the real-world. It is important to understand (a) how real-world inferences are generated when presented in a counterfactual-world context and (b) how counterfactual-world inferences are generated in fantasy and fiction texts. Exploration of (a) is important as information regarding cultural, political, and historical climates may be inferred from fantasy and fiction texts. Moreover, information regarding scientific processes may be inferred from science-fiction texts. Exploration of (b) is important as the counterfactual-world inference generation process may differ from the real-world inference generation process. Understanding how real-world and counterfactual-world inferential processes differ and the additional challenges counterfactual-world inferences present is essential to providing recommendations for the promotion of these inferences in the classroom. This section explores counterfactual-world inference generation by drawing upon not only research specifically exploring counterfactual-world inference generation, but also research exploring counterfactual-world text processing and counterfactual-world reasoning.

1.2.1. Counterfactual-World Text Processing

Research exploring counterfactual-world processing routinely finds that counterfactual-world processing results from a different, more demanding process than real-world processing (e.g. Ferguson, 2012; Ferguson & Sanford, 2008; Ferguson, Sanford & Leuthold, 2008; Kulakova et al., 2013; Robinson & Beck, 2000). Specifically, counterfactual-world processing is thought to result in the representation of both the actual (real-world) and the alternate (counterfactual-world) state of events (Johnson-Laird & Byrne, 2002; Kulakova et al., 2013). To prevent confusion between what is real and what is not, it is argued that the individual dissociates real-world and counterfactual-world information by constructing at least two mental models: real-world and counterfactual-world (Byrne, 2007). Some argue that the models are represented simultaneously (see the Mental Model Theory, Johnson-Laird & Byrne, 2002). Others argue that the models are created and represented sequentially (see the Suppositional Theory; Evans, Over & Handley, 2002). Regardless of when the two models are represented, Roese, Sanna, & Galinsky, (2005) argue that if the counterfactual-world is to be fully comprehended, the two models must be compared at some point. The real-world information, though, is likely to directly contrast with the counterfactual-world information, resulting in conflict, which must be resolved if the counterfactual-world is

to be processed successfully (e.g. Byrne, 2007; Ferguson, 2012). This effect will be referred to as real-world interference.

When recalling a counterfactual-world text participants are found to exhibit real-world interference, such that they recall information consistent with real-world knowledge rather than the text (Ceci, Caves, & Howe, 1981; Dorfman, 1989; Dorfman & Brewer, 1988). For instance, Ceci et al. (1981) found interference effects when participants recalled a counterfactual-world text a week after hearing. However, they did not find interference effects when texts were recalled after an immediate delay. This suggests interference may not be present *whilst reading* counterfactual-world texts, or at least not at the same level. As highlighted by Bartlett's (1932) seminal study, the texts individuals read and hear are argued to be *stored* within existing real-world schemas, such that recall reflects participants' beliefs and experiences more so than the original text. It may thus be the storage process that results in interference, not comprehension. Additionally, recall measures are subject to conscious construction processes that do not operate whilst reading (see Chapter 2 for a full discussion). The participant may thus consciously evaluate any real-world violations when recalling a text, although these violations may not have been considered when reading.

Many choose to read fantasy and fiction texts for pleasure (e.g. Clark & Foster, 2005). If the processing of counterfactual-world text was cognitively taxing due to the need to continuously resolve conflict experienced due to real-world interference it is unlikely that these texts would be so popular, it is even less likely that readers would report an immersion into the fantasy world (Green & Brock, 2000). Those studies suggesting that readers experience disruption due to real-world interference may not reflect the natural online reading process. Some suggest that when reading naturally, counterfactual-world text comprehension does not require the comparison of real-world and counterfactual-world models that counterfactual-world reasoning does (Gilbert, Krull, & Malone, 1990; Markman & McMullen, 2003; 2005; McMullen, 1997; McMullen & Markman, 2000; Rader & Sloutsky, 2002). This is supported by a growing body of research that employs online methods such as reading speeds, tracking of eye movements and brain-imaging to explore counterfactual-world text processing.

Recent research suggests that when processing counterfactual-world texts, some initial real-world interference may be experienced, as the reader sets up a counterfactual-world model, however, once the counterfactual-world model is set up, the conflict experienced

is quickly accommodated (e.g. Ferguson, 2012; Hald, Stennbeek-Planting & Hagoort, 2007; Warren, McConnell & Raynor, 2008). For instance, Ferguson (2012) employed measures of eye movements to explore the representation of the plausible counterfactual-world mental model whilst reading. Participants read critical sentences that were preceded by either a sentence providing a counterfactual-world context or a sentence providing a real-world context. Within each critical text was a critical word that was either consistent or inconsistent with the preceding context. This created four conditions: 1) real-world consistent, 2) counterfactual-world consistent, 3) counterfactual-world inconsistent and 4) real-world inconsistent. See Table 1.2. for example stimuli. Eye movements and reading speeds were recorded. Reading speeds of the critical word were found to be longer in the counterfactual-world consistent condition compared to all other conditions. This suggests some form of additional processing and possible disruption. However, regressive eye movements were significantly higher upon encountering the critical word in the counterfactual-world inconsistent condition compared to counterfactual-world consistent condition and real-world inconsistent condition. Ferguson suggests that participants did not regress back to the critical word in the counterfactual-world consistent condition as the additional reading time was used to set up a counterfactual-world model, upon which new counterfactual-world information can be directly mapped.

Table 1.2. Example Stimuli from Ferguson (2012), showing regions of analysis

Factual consistent
    Because Joanne had remembered her umbrella,$_1$| she had avoided the rain.$_2$| By the time she arrived at school Joanne's hair was$_{pre-critical}$| dry$_{critical}$| and some$_{post-critical}$| of her friends laughed.$_{wrap-up}$|
Factual inconsistent
    Because Joanne had remembered her umbrella,$_1$| she had avoided the rain.$_2$| By the time she arrived at school Joanne's hair was$_{pre-critical}$| wet$_{critical}$| and some$_{post-critical}$| of her friends laughed.$_{wrap-up}$|
Counterfactual consistent
    If Joanne had remembered her umbrella,$_1$| she would have avoided the rain.$_2$| By the time she arrived at school Joanne's hair was$_{pre-critical}$| wet$_{critical}$| and some$_{post-critical}$| of her friends laughed.$_{wrap-up}$|
Counterfactual inconsistent
    If Joanne had remembered her umbrella,$_1$| she would have avoided the rain.$_2$| By the time she arrived at school Joanne's hair was$_{pre-critical}$| dry$_{critical}$| and some$_{post-critical}$| of her friends laughed.$_{wrap-up}$|

Research has found that the strength of preceding context mediates the level of interference faced, with stronger (i.e. longer or explicit) contexts resulting in less

interference – i.e. immediate accommodation - and weaker contexts resulting in accommodation after a delay (e.g. Nieuwland, 2013; Nieuwland & Martin, 2012; Niewland & van Berkum, 2006). For instance, Nieuwland and van Berkum (2006) explored the processing of implausible counterfactual-world information in counterfactual-world contexts and real-world contexts in two ERP experiments. They used stories in which inanimate objects experienced human emotions and engaged in human actions (e.g. falling in love, dancing). They found that anomaly effects (N400 effects) were observed for implausible information in a real-world context, but not implausible information in a counterfactual-world context. The critical word in Nieuwland and van Berkum's (2006) study was presented to participants alongside the predicate that made the inanimate object animate four times before it is presented as the critical text. It is likely, then, that this facilitated the set-up of a counterfactual-world model before the critical text was encountered, resulting in minimal interference effects.

It appears that counterfactual-world models can also be stored and become part of the reader's knowledge-base, supporting later counterfactual-world text processing. Filik and colleagues explored counterfactual-world text processing using fictional, but familiar scenarios (e.g. Harry Potter, Tom & Jerry) – see Table 1.3. for example stimuli. Whilst the nature of the scenarios used by Filik and colleagues was implausible (e.g. magic, talking animals), the characters were already known to participants, with most participants having expectations about characters' behaviours and actions. Filik and colleagues found minimal real-world interference for counterfactual-world critical texts in counterfactual-world contexts, such that the fixation on the critical world, indicative of counterfactual-world mental model construction, was not observed (Filik, 2008; Filik & Leuthold, 2008; 2013). Interference was observed in counterfactual-world conditions when characters did not behave in the way expected, however. This suggests that additional set-up time was not needed initially as the participants already possessed the counterfactual-world model needed.

Filik and colleagues also found that anomalies were more easily accommodated for fictional characters in the counterfactual-world inconsistent condition than real-world characters in the real-world inconsistent condition. This suggests that the counterfactual-world models possessed by participants were more flexible than the real-world models, such that fewer constraints were placed on the characters observed (Filik, 2008; Filik & Leuthold, 2008; 2013).

Table 1.3. Example Stimuli for Filik & Leuthold (2013)

Condition	Example
Counterfactual-World Consistent	The Incredible Hulk was annoyed at the traffic in front of him. The angry man picked up the lorry and continued on his way.
Counterfactual-World Inconsistent	Shaggy was annoyed at the traffic in front of him. The angry man picked up the lorry and continued on his way.
Real-World Inconsistent	Terry was annoyed at the traffic in front of him. The angry man picked up the lorry and continued on his way.

In sum, it appears that when reading counterfactual-world texts some initial real-world interference may be experienced as the reader sets up a counterfactual-world model. However, the counterfactual-world model is then used to map incoming counterfactual-world information onto, meaning disruption is minimal. Research suggests that counterfactual-world models can be stored in the reader's knowledge-base allowing the reader to process new text, set within the counterfactual-world, without the additional set-up costs. The research discussed above largely focused on literal understanding, however. Research exploring counterfactual-world inference generation appears inconsistent. However, as discussed next, the differences observed may reflect age-related changes in counterfactual-world inference generation processing.

1.2.2. The Development of Counterfactual-World Inference Generation

Research exploring the development of implausible counterfactual-world inference generation suggests that rather than developing with age, the ability to generate implausible counterfactual-world inferences declines with age (Ceci et al., 1981; Dorfman, 1989; Dorfman & Brewer, 1988). Dorfman and colleagues found that adults displayed more difficulty when generating counterfactual-world inferences than CYP. CYP aged 7-8 years, 9-10 years, 11-12 years, and adults were asked to read known and novel fables. Participants were then asked to explain the point or moral of the fable – i.e. generate a thematic inference. Adults struggled to generate the thematic inference when it was inconsistent with their real-world knowledge more so than CYP, with a negative relationship observed between the strength of the necessary knowledge-base and ability to draw thematic inferences from information that violated said knowledge-base. However, offline measures of inference generation were employed. As discussed above, measures of counterfactual-world processing taken after reading may overestimate the real-world interference faced whilst reading.

Bowyer-Crane and Snowling (2010) used a timed sentence verification task (SVT) to compare children's ability to draw inferences from real-world stories with their ability to draw inferences from fairy stories. Thirty-nine 9-10 year olds were asked to read real-world and fairy stories. Bowyer-Crane and Snowling (2010) found that accuracy of inference generation for both coherence and elaborative inferences did not differ in the real-world and fairy story conditions, suggesting CYP do not find it more difficult to generate inferences based on counterfactual information. However, response times were significantly faster in the real-world compared with the fairy stories condition, suggesting that whilst CYP are able to generate counterfactual-world and real-world inferences with equal ease, they find the process of counterfactual inference generation more taxing than the generation of real-world inferences. However, in both the real-wold and fairy story condition, coherence inference items were responded to as fast as literal items and faster than elaborative inference items. This suggests coherence inferences are generated online and elaborative inferences offline, regardless of plausibility of the information.

Graesser, Kassler, Kreuz and McLain-Allen (1998) asked adults to read fantasy texts. After each text, participants completed an SVT with sentences reflecting four conditions: (1) real-world literal, (2) counterfactual-world literal, (3) real-world inference, (4) counterfactual-world inference. Response times were recorded. Reading speed of the whole passage was also recorded. Participants also completed a test battery assessing literacy expertise, reading skill, reading time and working memory. Whilst all adults struggled with counterfactual-world inference statements, those with higher levels of literacy expertise were able to successfully draw these counterfactual-world inferences, but only when they reduced their reading speed. All participants showed an advantage for real-world rather than counterfactual-world inference generation on the SVT. Moreover, all adults took much longer to make the true or false decision when confronted with counterfactual-world inferences. Taken together these results suggest that the counterfactual inferences may be generated offline, due to their more taxing nature.

Despite both using a timed SVT the results of Bowyer-Crane and Snowling (2010) and Graesser et al. (1998) appear to directly conflict. Bowyer-Crane and Snowling suggest that CYP are able to generate counterfactual-world inferences with ease such that coherence inferences can be generated online, whereas Graesser et al. suggest that adults find counterfactual-world inference generation so difficult all inferences are generated offline. Counterfactual-world inference generation thus appears to be harder for adults than children. Moreover, the differing patterns of counterfactual-world inference

generation suggest that children and adults engage in qualitatively different processes when generating counterfactual-world inferences. Research exploring the processing underpinning the development of counterfactual-world reasoning may provide an explanation for the possible conceptual shift observed.

Due to inconsistent findings regarding the development of counterfactual-world reasoning, Rafetseder and colleagues sought to explore the development of counterfactual-world reasoning and the processes underpinning this in CYP aged 6 years to adulthood (Rafetseder, Schwitalla, & Perner, 2013; Rafetseder, Cristi-Vargas, & Perner, 2010). Participants were presented with story worlds like the example below:

Carol came home and did not take her dirty shoes off, and she made the floor dirty with her shoes. Max then followed Carol across the clean floor in his muddy shoes.

Q: If Carol had taken her shoes off, would the floor be dirty or clean?

If the second sentence is not included, the question can be answered correctly using basic conditional reasoning only – i.e. someone takes their shoes off, floors generally stay clean. If the second sentence is included, though, basic conditional reasoning results in the incorrect answer. Subsequently, counterfactual-world reasoning must be employed for the correct answer to be given. Stories were acted out using small props and read to participants. Participants were then asked questions about the story worlds that required either basic conditional reasoning or counterfactual-world reasoning. Results show all groups performed well on the questions requiring only basic conditional reasoning. However, for counterfactual-world reasoning adult-like performance was only observed in the 13-15 year olds. Performance did significantly improve between 5-6 years, 7-10 years, 9-11 years, and 13-15 years, however. Taken together, it appears counterfactual-world reasoning abilities develop throughout childhood and into adolescence, such that they are not fully developed (i.e. to an adult-like level) until around 13-15 years, when CYP shift from basic reasoning to counterfactual-world reasoning.

Raefetseder et al. (2010; 2013) suggest that basic conditional reasoning requires the construction of just one representation – i.e. the real-state of events or imagined state of events. Van Reet, Pinkham, & Lillard, (2007) suggest that young children may create just one mental model when processing counterfactual-world information – i.e., engaging in pretend play. If children are only constructing one mental model, a counterfactual-world model, when reading they will face less real-world interference than those CYP and adults

who are constructing both a counterfactual-world model and a conflicting real-world model. The construction of only one mental model also explains the ease with which children are able to engage in other activities requiring counterfactual-world processing – e.g. pretend play (Currie, 1998; Harris & Kavanaugh, 1993; Jarrold, Carruthers, Smith, & Boucher, 1994; Riggs & Peterson, 2000). In addition to employing different processing strategies, Rafetseder et al. (2010) suggest the information contained within the counterfactual-world mental models constructed by CYP may differ to those constructed by adults. Specifically, Rafetseder et al. (2010) have shown that there is a lack of consistency between what adults and children believe should be changed. Less consistency was also observed between the children about what needs to be changed. This suggests children are more flexible and adults more rigid when manipulating real-world states. This may explain the conceptual shift observed in counterfactual-world inference generation. Whilst children create counterfactual-worlds where anything is possible, adults may change only one event at a time as the text necessitates. Therefore, adults may have to deal with conflict each time a counterfactual-world event is encountered.

Research exploring counterfactual-world text processing suggests that adults may also be able to create flexible models, however. Filik (2013) found that anomalies were more easily accommodated for fictional characters than real-world characters. This suggests that counterfactual-world contexts are more flexible than real-world contexts – i.e. readers are more accepting of atypical information when processing fantasy information. This may be because when reading counterfactual-world texts some argue that readers actively suspend disbelief – i.e. ''the impossible becomes possible'' (Warren et al., 2008). This is supported by Huang and Gordon (2011) who suggest that readers are able to determine whether a text is fictional or non-fictional and then employ different processing strategies accordingly. If so, then adults would also be expected to engage with counterfactual-world inference generation with the same ease as children. This is supported by the growing body of research utilising online methods to explore counterfactual-world text processing in adults, such that, as discussed above, adults appear to be able to quickly accommodate disruption experienced due to real-world interference when processing counterfactual-world texts (e.g. Ferguson, 2012; Hald et al., 2007; Warren et al., 2008).

At present there is no one study exploring the developmental trajectory of implausible counterfactual-world inference generation, within a story context, in CYP and adults using the same methodology. As a result, the differences observed between adults and

CYP could be attributable to methodological discrepancies, as opposed to a conceptual shift. Bowyer-Crane and Snowling used short, simple texts suitable for 9 year olds (114-144 words long, with readability ratings of seven years and four months or less) whereas Graesser et al. used longer, more complex texts (609-763 words long, with readability ratings ranging from approximately 20-40 years). It has been argued that complex texts are often processed at a surface-level, rather than a deep level (Baker, 1979; Glenberg, Wilkinson, & Epstein, 1982). Whilst this allows the individual to feel they have comprehended the text, it is argued that the representation created is ignorant to any discrepancies between the text and the reader's real-world knowledge (Markman, 1979; Otero & Campanario, 1990; Vosniadou, Pearson, & Rogers, 1988). Therefore, the difficulty experienced by adults could be reflective of the complex texts they were asked to read as opposed to the processes underpinning counterfactual-world inference generation. However, this is unlikely given that real-world texts were of similar difficulty. It is possible, though, that the complexity of the text affected counterfactual-world inference generation abilities, more so than real-world inference generation abilities.

Second, whilst Bowyer-Crane and Snowling (2010) employed an almost immediate time interval (500ms) between the presentation of the critical text and SVT, Graesser et al. (1998) did not present the SVT until the end of each chapter. Therefore, the time interval between critical text and test item was longer than that employed by Bowyer-Crane and Snowling and inconsistent across texts. Research suggests that as the time interval increases so does the probability of the inference being judged against real-world knowledge (Singer, 1994). Therefore, adults may have faced more real-world interference than CYP when validating counterfactual-world inferences.

Third, research shows that for real-world information coherence inferences are generated online whereas elaborative inferences are not (Calvo & Castillo, 1996; 1998; 2001a; 2001b; Calvo et al., 2006; Casteel, 1993; Casteel & Simpson, 1991; Long, Golding, & Graesser, 1992; Magliano, Baggett, Johnson, & Graesser, 1993; Millis & Grasser, 1994). This is the pattern of results observed by Bowyer-Crane and Snowling (2010) for counterfactual-world information. Whilst Bowyer-Crane and Snowling specify the types of inferences explored, Graesser et al. (1998) do not. It is possible that a) the inferences explored by Graesser et al. were predominantly elaborative inferences or b) the lack of discrimination between coherence and elaborative inferences resulted in skewed results. However, the timed-SVT is not a true measure of online inference generation (see Chapter

2). Therefore, further research using true online measures such as reading speeds and tracking of eye movements needed.

Finally, Bowyer-Crane and Snowling (2010) used fairy tales in the counterfactual-world condition, conversely, Graesser et al. (1998) used texts that violated the laws of time and space. The processing of counterfactual-world information within fairy tales is thought to be less demanding than less familiar implausible information (Zwaan et al.,1995). The counterfactual-world information processed by adults may have been more difficult than that processed by children. It is unclear if adults would still display as much difficulty generating counterfactual-world inferences if given shorter, simpler texts comprising more plausible counterfactual-world information, with exploration and comparison of coherence and elaborative inferences using online methods.

In sum, counterfactual-world processing appears to be qualitatively different for adults and children with a conceptual shift in processing occurring during adolescence. Specifically, whilst children may adopt a basic conditional reasoning strategy, such that they construct only one mental model, older CYP and adults may adopt a more complex counterfactual-world reasoning strategy, such that they construct and compare two models. However, at present, due to limited research utilising more than one age group, it is unclear if the differences in counterfactual-world inference generation abilities are due to methodological differences between those studies exploring adults and CYP or a conceptual shift in processing. Further research exploring the developmental trajectory of counterfactual-world inference generation in CYP and adults is, therefore, needed.

1.3. Components of Inference Generation

Inference generation is a multi-faceted skill resulting from the interaction of several processes. Exploration of the underlying process of inference generation is key to informing educational recommendations and the design of specific targeted interventions. This section outlines the neurological basis of inference generation and then discusses those skills central to real-world and counterfactual-world inference generation during adolescence.

1.3.1. Neurological Basis of Inference Generation

Inference generation appears to be a distinct skill, resulting from unique processing, above and beyond that associated with literal comprehension (e.g. Ferstl, Neumann, Bogler, & von Cramon, 2008; Ferstl, Rinck, & von Cramon, 2005; Ferstl, & von Cramon,

2001; Kim et al., 2012; Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006; Prat et al., 2011; Prat & Just, 2010; Prat, Keller, & Just, 2007; Virtue et al., 2006; 2008). It is possible, then, that the developmental trajectory of inference generation differs to that of reading comprehension. Specific and additional teaching of inference generation skills may be needed. Distinct patterns of brain activations are also observed for coherence and elaborative inferences, suggesting that different strategies and interventions may be needed to support the different inference types (Beeman, Bowden, Gernsbacher, 2000; Chow, Kaup, Raabe, & Greenlee, 2008; Jin et al., 2009; Virtue & van den Broek, 2005). For coherence inferences, the left hemisphere is strongly activated, with the superior temporal gyrus initially activated, followed by the inferior frontal gyrus at the coherence break (e.g. Jung-Beeman, 2005; Mason & Just, 2004; Virtue et al., 2006). Activation of the superior temporal gyrus is associated with skills such as cognitive mapping and semantic retrieval (Robertson, Gernsbacher, Guidotti, Robertson, Irwin, Mock, & Campana, 2000; Wagner, Parâe-Blagoev, Clark, & Poldrack, 2001). Activation of the inferior frontal gyrus is routinely associated with skills such as inhibitory control – specifically information selection from competing alternatives – and language and semantic processing (Barch, Braver, Sabb, & Noll, 2000; Seger, Desmond, Glover, & Gabrieli, 2000). Consistent with behavioural research, neuroimaging research thus suggests that for coherence inferences, knowledge is activated whilst reading and then, at the point of a coherence break, selection processes operate to isolate the necessary knowledge and integrate this with the text and/or related knowledge (e.g. Schmalhofer et al., 2002).

For elaborative inferences, the right hemisphere is strongly activated, with the superior temporal sulcus and the inferior frontal gyrus routinely activated (Chow et al., 2008; Jin et al., 2009). Chow et al. (2008) argue that knowledge activation during the elaborative inference generation process is controlled by top-down processes similar to those implicated in selected semantic retrieval, evidenced by the activation of the inferior frontal gyrus and superior temporal sulcus (Wagner et al., 2001; Barch et al., 2000; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). Similar to coherence inference generation, relevant knowledge is then integrated with the text and other knowledge to generate an inference. Therefore, whilst the processes underpinning coherence and elaborative inference generation are the same or similar, their time-course may differ. The knowledge activation process may also differ, such that it is more of a conscious, guided process when generating elaborative inferences compared to coherence inferences. This

is supported by the hemispheric differences, such that Virtue and Joss (2012) suggest that the right hemisphere, heavily implicated in elaborative, but not coherence inference generation, is utilised when the language resources of the left hemisphere have been exhausted. Whilst inference generation is a multifaceted skill (e.g. Virtue et al., 2006), exploration of all of these skills is beyond the scope of this thesis. Therefore, given their importance and possible different roles in both coherence and elaborative inference generation, this thesis focuses on the role of knowledge activation and inhibitory control.

1.3.2. Knowledge

For both adults and CYP, those with high topic knowledge generate more inferences on texts relating to said topic than those with low topic knowledge, regardless of IQ (Birch & Bloom, 2004; Tarchi, 2010, 2012). The knowledge a reader possesses thus determines the inferences they can generate, such that if the necessary knowledge is not possessed by the reader, the target inference cannot be generated. Graesser et al. (1994) argue that background knowledge and the inferences this facilitates determine the meaning a reader can derive from a text. For instance, when reading 'the girl picked up the bat' with no supporting context, the reader could derive a meaning related to a piece of sports equipment or a small animal depending on the knowledge they possess. General knowledge has been found to improve with age, this may thus explain developments in inferential skill (Schroeders, Schipolowski, & Wilhelm, 2015; Thorsen, Gustafsson & Cliffordson, 2014).

Inference generation, particularly whilst reading, is argued to be an automatic process (e.g. Graesser et al., 1994; Kintsch, 1993; McKoon & Ratcliffe, 1992). Knowledge accessibility refers to how fast knowledge can be activated (Glucksberg, Brown & McGlore, 1993). Difficulty rapidly accessing the necessary knowledge has been found to result in inference generation failure and may explain developments in inference generation skill during adolescence (Barnes et al., 1996; Cromley & Azevedo, 2007; Kintsch, 1994). Barnes et al. (1996) explored how accessibility of knowledge affects knowledge-based inference generation abilities in CYP aged 6-15years. To account for the effects of amount of knowledge Barnes et al. asked children to learn facts about a fictional planet. After controlling for age, Barnes et al. found, in all age groups, information that could be retrieved quickly was twice as likely to be used when drawing inferences as information that was more difficult to retrieve. Both inference generation skill and knowledge accessibility were found to improve with age, despite all groups

possessing the necessary knowledge. Improvements in inference generation between 6-15 years thus appear to be underpinned by improvements in accessibility of knowledge.

Barnes et al. (1996) found that the role of knowledge accessibility varied according to inference type. The importance of knowledge accessibility was found to increase for real-world elaborative inferences, but decline for real-world coherence inferences. Barnes et al. suggests that this is due to improvements in accessibility of knowledge and comprehension monitoring. As the CYP develops, they are more likely to detect coherence breaks whilst reading (Singer & Flavell, 1981). The Constructionist Theory suggests that readers aim to maintain coherence whilst reading and may engage in a laborious search of long-term memory if the necessary knowledge is not readily accessible (Graesser et al., 1994). Therefore, accessibility of knowledge becomes less important. However, Barnes et al. primed the necessary knowledge to control for amount of knowledge. When generating coherence inferences in the Barnes et al. study, the necessary knowledge, even if not readily accessible, may have been weakly activated, thus requiring minimal effort to be fully activated. Participants may have thus freely, and potentially consciously, engaged in this additional processing, knowing that they would find the knowledge they would need. However, when reading naturally, the knowledge necessary for coherence inferences is thought to be activated through associative processing (Cook, Halleran, & O'Brien, 1998; Kintsch, 1988; 1998; van den Broek et al., 2005). When not readily available, not wanting to disrupt the reading process and not sure how laborious of a search will be needed or if this search will be successful, older readers may not engage in this potentially time-consuming yet possibly fruitless task. Additionally, Barnes et al. used an offline measure of inference generation, meaning participants had time to engage in the strategic processing they may not have engaged in whilst reading. Subsequently, accessibility of knowledge would be essential to successful real-world coherence inference generation. Further research into the role of knowledge when reading naturally is warranted.

Much of the difficulty associated with counterfactual-world inference generation is attributed to the activation of related but contradictory real-world knowledge (Alter, Oppenheimer Eply, & Eyre, 2007; Byrne, 2002; 2005; 2007; Evans, 2006; Evans et al., 2005). When building a situational model of the text the reader is thought to activate related but contradictory real-world knowledge, however, this knowledge may have stronger and more numerous connections than those nodes associated with a counterfactual-world interpretation (McNamara & McDaniel, 2004; Zwaan & Truitt,

2000). This is supported by Alvermann, Smith and Readance (1985) who argue that when the information in the text is incompatible with real-world knowledge it is unlikely to be used to update the situational model being constructed. Alvermann et al. (1985) asked all participants to read a text. Prior to this, the required real-world knowledge was activated for one group – thus increasing the accessibility of that knowledge. They found when previous knowledge was easily accessible participants performed worse on those questions designed to assess understanding of text that was incompatible with real-world knowledge, with the representation being more reflective of real-world knowledge than the counterfactual-world text. Therefore, if knowledge is easily accessible more interference is likely to be experienced. The aforementioned research focuses on plausible counterfactual-world information or ambiguous real-world information. This type of information typically evokes participants to compare models (Markman & McMullen, 2003; 2005; 2007). Implausible information, however, may not always evoke a comparison (Filik, 2008; Filik & Leuthold, 2008; Nieuwland & Van Berkum, 2006). Further exploration of the role of knowledge in implausible counterfactual-world information, in which real-world knowledge may not be strongly linked to the text, particularly if there is a preceding context would be useful.

1.3.3. Inhibitory Control

Knowledge is thought to be associative, resulting in the activation of both related and contradictory knowledge when reading (van den Broek et al., 2005; van den Broek, Risden, & Husebye-Hartman, 1995; van den Broek et al., 1999). Working memory capacity is limited, however (Baddeley, 1986; Just & Carpenter, 1992). Subsequently, the maintenance and integration of large amounts of knowledge would heavily tax working memory capacity, leaving little capacity for the execution of those other cognitive processes needed to successfully generate an inference and comprehend a text (Cain, 2006; Oakhill & Cain, 2011; van den Broek, 2012). The amount of knowledge activated, maintained, and integrated into the situational model whilst reading must be constrained in some way. Inhibitory control has been implicated in the management of knowledge during the inferential process (e.g. Barnes, Faulkner, Wilkinson, & Dennis, 2004; Cain, 2006; Gernsbacher, 1993; Harnishfeger, 1995).

Research suggests that for comprehension to be successful, irrelevant and contradictory knowledge must be suppressed in the early stages of the inferential process (Cain, 2006;

Harnishfeger, 1995). This creates a more manageable workspace, such that processing costs are reduced as only the most relevant information is processed (e.g. Borella & de Ribaupierre, 2014; DeBeni & Palladino, 2000; Gernsbacher & Faust, 1991). The role of inhibitory control in the inference generation process is well-established. For instance, inhibitory control has been implicated in real-world inference generation across the lifespan (Cain, 2006; Barnes et al., 2004; DeBeni & Palladino, 2000; Gernsbacher, 1993). Similarly, poor comprehenders have been found to display difficulties inhibiting information, such that they displayed significantly more intrusion errors than good comprehenders (Caretti et al., 2004). However, to this researcher's knowledge, not one study has explored age-related changes in the role of inhibitory control in real-world inference generation during the Secondary School years.

Inhibitory control has been found to develop throughout adolescence, with development plateauing between around 14-17 years of age (see Romine and Reynolds, 2005 for a review). Developments in success in inhibiting irrelevant or contradictory information and developments in speed of inhibition have been observed (Bjorklund & Harnishfeger, 1990; Romine & Reynolds, 2005; Tamm, Menon, & Reiss, 2002). As the CYP ages, the amount of knowledge they are able to readily access whilst reading has been found to increase (e.g. Barnes et al., 1996). Stronger and more efficient inhibitory mechanisms may thus be needed to maintain a clear workspace. Developments in inhibitory control may thus underpin developments in inference generation skill and inference generation efficiency during adolescence. Exploration of the potentially changing relationship between inhibitory control and knowledge may be fruitful in explaining the inference generation process during adolescence.

The role of inhibitory control in counterfactual-world inference generation may be mediated by belief biases. A belief bias is a preference to process information in ways that match beliefs rather than logic – i.e. attend to, process and retrieve real-world information more readily than atypical information (e.g. Torrens, 1999). The process of initial suppression of knowledge when reading is thought to be guided by belief biases, as such, much of the contradictory and irrelevant knowledge, which is automatically activated due to associative knowledge activation mechanisms, is suppressed (e.g. Graesser et al., 1998). This frees up working memory capacity resulting in more efficient inference generation. However, when processing counterfactual-world information, belief biases have been found to hinder processing, resulting in errors and less efficient processing in both adults and children (Houde & Guichart, 2001; Markovits & Schroyens,

2007; Moutier, Plagner-Cayeux, Melot, & Houde, 2006; Richards & Sanderson, 1999). There is a growing body of research showing that counterfactual-world information processed in a counterfactual-world context is not treated as contradictory, however – i.e. belief biases are overcome (e.g. Nieuwland, 2012; 2013; 2015, Nieuwland & Berkum, 2006; Nieuwland & Martin, 2012). This accommodation is only observed after an initial delay in which the counterfactual-world is set up, suggesting that belief biases may be operating in the initial stages of text processing.

Whilst inhibitory control may at first be guided by belief biases, research suggests that in the later stages of processing inhibitory control may be essential to overcoming the effects of belief biases, such that, through correlations, training and brain imaging studies, Moutier and colleagues have highlighted a positive relationship between inhibitory control and ability to overcome belief biases (Houde, Zago, Crivello, Moutier, Pineau, Mazoyer, & Tzourio-Mazoyer, 2001; Houde & Moutier, 1996; 1999). Developments in inhibitory control during adolescence may thus support suppression of belief biases and ultimately successful counterfactual-world processing (Frank, 1996; 1997). The developing relationship between inhibitory control, belief biases and the processing of counterfactual-world information may not be as simple as the studies above suggest, however. Mitchell, Robinson, Isaacs and Nye (1996) suggest that there is a shift in the way belief biases are applied between 9 years and adulthood.

Mitchell et al. (1996) explored the false belief abilities of a group of 5 year olds, 9 year olds and adults. Participants heard a story in which a character was given a true or a false message that was consistent or counter to the character's existing experience. Mitchell et al. found that whereas children based character judgments on what the character had previously experienced, regardless of the truthfulness of the information in the message, adults based character judgements on the truthfulness of the information. Adults thought characters would believe the message if it was true even if this was not consistent with the character's existing experiences. Mitchell et al. suggest that the difference is due to a qualitative shift in the way information violating real-world beliefs is processed, such that children rigidly apply a 'seeing is believing rule' – i.e. the character must see an event to believe it and any other form of contradictory evidence is not believed even if the reader knows the real state of events. Conversely, adults are argued to adopt a more complex processing strategy, drawing upon their existing previous knowledge to consider those factors that may lead one to accept another's message. This leaves adults more vulnerable to belief biases. This would explain the possible conceptual shift in counterfactual-world

inference generation discussed in Section 1.2.2. Further exploration of the developing relationship between inhibitory control, belief biases and counterfactual-world inference generation is warranted.

1.4. Inference Generation and Literacy Skills

In Secondary School classrooms information is presented in various ways, many of which require the CYP to read and understand text – e.g. textbooks, worksheets, websites. Similarly, even if a CYP has acquired the necessary knowledge, if they are unable to produce a coherent text they may struggle to achieve age-related expectations as typically, CYP are assessed via a written exam or written coursework. Learning and assessment across the curriculum are thus dependent on good reading comprehension and text production skills. Between 2010 and 2015, on average, just short of one third (27-31%) of CYP failed to make the expected progress between Key Stage 2 (Year 6) and Key Stage 4 (Year 11), despite only 11-17% of CYP not achieving age-related expectations by the end of Year 6 (Department for Education, 2015a; 2015b). This suggests that a) many of those who entered Secondary School with poor literacy skills did not close the attainment gap and thus continued to struggle throughout their school careers, and b) some CYP who were developing typically during Primary School failed to progress as expected during Secondary School. Without appropriate identification and intervention, literacy difficulties have been found to persist into adulthood (Rapp, van den Broek, McMaster, Kendeou & Espin, 2007). This next section outlines the role of inference generation in reading comprehension and text production.

1.4.1. Inference Generation and Reading Comprehension

Currently the National Curriculum for reading is based on the most prominent model of reading comprehension development, the Simple View of Reading (SVR; DFES, 2006; Gough, Hoover, & Peterson, 1996; Gough & Turner, 1986, Hoover & Gough, 1990; Rose, 2006). The SVR suggests that reading comprehension is underpinned by two key components: decoding and linguistic comprehension. Decoding refers to the automatic recognition of a word and its semantic meaning. Linguistic comprehension refers to those language skills associated with the comprehension of oral sentences – i.e. inference generation, identification of text structure, meaning of propositions, and overall gist. Proficient reading comprehension occurs as decoding skills become fully automatized and fluent, meaning resources can be freed up to allow for the operation of linguistic comprehension skills (National Reading Panel, 2000; Perfetti, 1998; Samuels & Flor,

1997; Spear-Swerling & Sternberg, 1994). Consistent with this, the predictive utility of the two components is not equal and fluctuates with age (e.g. Storch & Whitehurst, 2002). The role of decoding is found to decline after 9-10years whereas the role of linguistic comprehension is found to increase (e.g. Adlof, Catts, & Lee, 2010; Catts, Hogan, & Adlof, 2005; Chen & Vellutino, 1997; Kershaw & Schatschneider, 2012).

The SVR was originally proposed to explore individual differences in reading comprehension, not reading comprehension development. Whilst the two key components of the SVR has been found to explain variance in reading comprehension abilities in populations aged up to 18 years and over (Savage & Wolforth, 2007), Floyd Meisinger, Gregg, & Keith (2012) highlight a lack of research exploring the developmental trajectory of decoding and linguistic comprehension abilities across the lifespan. Specifically, further research exploring the development of linguistic comprehension skills during the Secondary School years is needed as a) reading comprehension continues to improve even after decoding skills are automatic and b) Secondary School is the point when these higher-order skills appear to play a central role in reading comprehension. Researchers thus highlight the need for a comprehensive model of reading comprehension development that explores the role and development of those skills central to mental model construction (Cassidy, Valadez, Garrett, & Barrera, 2010; Pike, Barnes, & Barron, 2010).

Zwaan and Singer (2003) state that *"almost every facet of comprehension is at least partly inferential"* (Zwaan and Singer, 2003, pp. 100). Inference generation also lies at the heart of the most prominent skilled models of comprehension and has been implicated in reading comprehension success across the lifespan (e.g. Cain, Oakhill & Byrant, 2004; Dewitz, Carr & Patberg 1987; Kintsch, 1998; McGee & Johnson, 2003; Zwaan, 2003). Inferential abilities have also been found to play an indirect role in the text comprehension process, mediating the role of other skills. For instance, a relationship has been observed between vocabulary and inference skill (Cain, 2007; Cain, Oakhill, & Lemmon, 2004). Whilst vocabulary is initially a significant predictor of reading comprehension, with age, this relationship becomes reciprocal (Verhoeven & van Leeuwe, 2008). This is thought to be due to shifts in the mechanism of vocabulary acquisition, such that older readers are able to use the context surrounding an unknown word to *infer* meaning (e.g. Pretorius, 2000). Subsequently, limited vocabulary does not always impair comprehension. (Freebody & Anderson, 1983; Wittrock, Marks, & Doctorow, 1975). This is supported by Nash and Snowling (2006) who found that vocabulary could be improved by teaching

CYP to infer meaning from the context. This led to greater gains in comprehension, compared to a control group who were only taught the meaning of target words.

The importance of inference generation in classroom comprehension is also likely to increase as the CYP moves from Primary to Secondary School. As the CYP progresses through the school system, there is a shift from learning-to-read to reading-to-learn (Chall 1996; Snow, Scarborough, & Burns, 1999) and from adult-led learning to pupil-led learning (Baker, Gertsen, & Scanlon, 2002). This increases the importance of good reading comprehension skills, and specifically inference generation abilities, for several reasons. The text-based materials CYP are presented with become longer and more challenging and complex with regards to vocabulary and syntactic structure, meaning the reader is expected to infer the meanings of new words using the surrounding context. (Shanahan & Shanahan, 2008, 2012). In addition, texts become less literal, with meaning more implicit (AQA, 2013; Department for Education, 2013). Subsequently, as a CYP progresses from Primary to Secondary School, the level of inferential processing required increases. This is mirrored by the reading comprehension assessment tools used.

Whereas many reading comprehension assessments for Primary-School-aged CYP focus on explicit understanding and literal retrieval, many reading comprehension assessments for Secondary-School-aged CYP focus on implicit understanding and inference generation. For instance, Leach et al. (2003) found that poor comprehenders performed well on standardised measures of reading comprehension aged 9 years. However, by 11-12 years the performance of poor comprehenders on the same standardised task had significantly decreased, dropping 53 percentile points, on average. Leach et al. used the Peabody Individual Achievement Test—Revised (PIAT-R; Dunn & Dunn, 1997). The items increase with difficulty, such that the early questions designed to assess the abilities of children aged under 9-10 years focus on basic reading skills such as sentence comprehension and literal thinking. However, the later questions designed to assess the abilities of CYP over 9-10 years focus on higher-order reading skills such as knowledge integration and inference generation. The same distinction also exists in the UK school system.

To achieve a Level 4 (expected level) on the Key Stage 2 Reading Comprehension SATs paper (in 2013) 19 points must be achieved. 68% (13 points) of these points can be obtained using literal comprehension skills – i.e. identifying relevant information in the text. The remaining 6 points can come from the other four assessment focuses – see Figure

1.3. The GCSE English Language tests used at the end of Secondary School combine both exams and coursework, and reading and writing. Therefore, the exact number of comprehension marks needed for a GCSE grade C (expected level) cannot be determined. However, the assessment objectives on the 2013 AQA English Language foundation are presented in Figure 1.4. Taken together with the example marking scheme in Figure 1.4., it evident that for reading comprehension, the emphasis is on implicit understanding thus requiring inference generation. Taken together, there exists a clear shift from Primary to Secondary School in the assessment of reading comprehension proficiency, such that the focus goes from literal comprehension and explicit identification to full and implicit understanding.

	Assessment Focuses					
	AF2	AF3	AF4	AF5	AF6	AF7
Definition	Understand, describe, select or retrieve information, events or ideas from texts and use quotation and reference to text	Deduce, infer or interpret information, events or ideas from texts	Identify and comment on the structure and organisation of texts, including grammatical and presentational features at text level	Explain and comment on writer's uses of language, including grammatical and literary features at word and sentence level	Identify and comment on writers' purposes and viewpoints, and the overall effect of the text on the reader	Relate texts to their social cultural and historical contexts and literary traditions

Adapted from Key Stage 2 English 2013 Marking Scheme: Reading – Wolf Pack (Department of Education, 2013).

Figure 1.3. – Key Stage 2 Reading Assessment Focuses

	Assessment Objectives	
	AO1	AO2
Definition	Understand, describe, select or retrieve information, events or ideas from texts and use quotation and reference to text Deduce, infer or interpret information, events or ideas from texts Identify and comment on the structure and organisation of texts, including grammatical and presentational features at text level	Explain and comment on writer's uses of language, including grammatical and literary features at word and sentence level Identify and comment on writers' purposes and viewpoints, and the overall effect of the text on the reader Relate texts to their social cultural and historical contexts and literary traditions
	Example Marking Scheme	
	Band 1 –'limited' '	Band 2 – 'some' 'attempts'	Band 3 – clear' 'relevant'
Definition	• limited evidence that the text is understood • simple engagement with the text • may offer limited quotation, textual detail or copying out • simple reference to the way the hand will change Matthew's life	• some evidence that the text is understood • attempts to engage with the text and makes an inference • offers a relevant quotation to support what has been understood • makes a statement on the way(s) the hand will change Matthew's life	• clear evidence that the text is understood • clear engagement with the text and makes inferences • offers relevant and appropriate quotations • makes clear statements on the ways the hand will change Matthew's life

Figure 1.4. – GCSE (Key Stage 4) Assessment Objectives and Example Marking scheme

It is unsurprising then that inference generation abilities have been found to account for unique variance in the reading comprehension abilities of Secondary School aged CYP. Barth, Barnes, Francis, Vaughn and York (2015) found that inferential skill was a predictor of reading comprehension in both adequate and struggling adolescents aged 11-18 years, accounting for 2-5% of unique variance. Cromley, Snyder-Hogan, & Luciw-Dubas, (2010) found that inference generation accounted for a much larger 19.1% unique variance in the reading comprehension skill of 14-15year olds. However, rather than exploring age-related changes in the role of inference generation, these studies controlled for age. As a result, it is unclear how the role of inference generation in the reading comprehension process changes with age. Exploration of the potentially changing role of inference generation abilities in the reading comprehension process during adolescence is thus warranted.

1.4.2. Inference Generation and Text Production

A literate individual is characterised by both the ability to understand and produce text (e.g. Bain, 2006; Hand, Wallace, & Yang, 2004). Proficient written text production is a complex task. First an idea is generated (e.g. Hayes & Flower, 1986). This idea is then elaborated on via the activation of real-world knowledge (e.g. Baker, Gersten, & Graham, 2003; Harris, Graham, & Mason, 2006). The words to represent this idea must then be retrieved from the mental lexicon, ordered to create a syntactically correct sentence, the orthographic representation of the necessary words must be retrieved, and finally, the motor skills needed to produce these representations must be planned and executed (Torrance & Galbraith, 2006). To be coherent, the production of any one sentence must be easily associable to any previous sentence – i.e. ideas must be organised in a way that is both locally and globally coherent. Therefore, the writer must have in mind the sentences they have produced, the sentence they are producing, and their intended message and structure for the completed text – i.e. a mental model must be constructed and maintained (Torrance & Galbraith, 2006). As new ideas are added the coherence and structure of the text is continually updated and monitored as the writer pauses and revises (e.g. Chenoweth & Hayes, 2001; Hayes, 2012; Hayes & Chenoweth, 2006). Similar to reading comprehension, the construction of a mental model thus appears essential to successful text production. It is surprising, then, that the role of inference generation

abilities in the text production process is scarcely explored explicitly. Consequently, inference generation is not a prominent feature of the Simple View of Writing (SVW), the most influential and widely cited model of text production development (Juel, 1988).

The SVW suggests that proficient writing results from the automatization of low-level transcription skills and increasing working memory capacity. This allows for the execution of higher-level oral language skills (McCutchen, Covill, Hoyne, & Mildes, 1994). Although the SVW has received much support, it is primarily focused on the early stages of text production development (e.g. Connelly, Dockrell, & Barnett, 2005; Kellog, 1999; McCutchen, 1996; Ransdell & Levy, 1996). Research exploring text production in adolescent and adult writers finds that even after the automatization of these skills, text quality continues to improve, with even proficient writers often finding the task of producing text very effortful (Scardamalia & Bereiter, 1991; Zimmerman & Martinez-Pons, 1990). The predictive utility of the SVW has thus been found to decline with age (Babayigit & Stainthorp, 2010; 2011; Swanson & Berninger, 1996). Developments in text production quality during adolescence thus appear to be driven by skills and interactions not accounted for by the SVW.

Building on Bereiter and Scardamalia's (1987) model of children's writing, Hayes (2012) proposes three distinct developmental stages of writing: 1) knowledge-telling (5-9 years), 2) knowledge-structuring (10-13 years), and 3) knowledge-transforming (13 years and over). In the knowledge-telling stage, the writer produces a series of statements about a particular topic. This process relies on long-term memory and transcription skills. There is no consideration of the reader in this stage. During the knowledge-structuring stage writers not only produce a series of statements about a particular topic, they elaborate on these, such that a situational model is constructed. However, the reader is not thought to be considered. Finally, in the knowledge-transforming stage, writing is constructed to meet the readers' needs meaning coherence is central. Stage 1 shows very little organisation or adaptation of knowledge, it is a random recall of ideas, by Stage 2 these ideas have been organised, and by Stage 3, these ideas have been elaborated on and integrated with other knowledge. The transition between the three stages reflects a growing need for coherence and integration of ideas. The development of these stages thus reflects the growing role of inference generation abilities. As with text comprehension, inference generation may not take a pivotal role in text production until those lower-level skills (transcription, spelling) are proficient. This model is supported by research exploring individual differences and developmental changes in text

production, such that with age, text production has been found to become more coherent with lower quality texts having less coherence (Berman & Slobin, 1994; Cain, 2003; Cragg & Nation, 2006; Fayol, 1991; Chanquoy, Foulin, & Fayol, 1990; Stein & Trabasso, 1981) and contain more ideas (Keys, 1999; MacArthur, 2012; Wigglesworth, 1997).

The role of inference generation in the development of text production appears key. However, what is not clear is what drives this development. Specifically, is text production initially constrained by limited inference generation skills or fewer available resources as resources are allocated to lower-level skills? Exploration of the role of inference generation in the text production and reading comprehension process, particularly during the Secondary School Years, may prove fruitful in furthering understanding of literacy development during what, for some, is a critical time.

1.5. Summary of Research

Government statistics suggest that around one third of CYP are not making the literacy progress expected during Secondary School. Whilst a variety of skills are implicated in mental model construction, inference generation appears to be at the core. Research exploring reading comprehension suggests that the importance of inference generation increases between Primary and Secondary School. Similarly, Hayes' model of writing development suggests an increasing need for coherence and subsequent requirement for inference generation. Whilst inference generation has been found to be a predictor of literacy skills during the Secondary School years, a developmental approach has rarely been taken. Many studies explore only one or two age groups, typically an adult and/or child group. Moreover, the methodology used to explore CYP and adults can differ, meaning it is unclear if differences observed between the age groups are due to age-related changes or methodological discrepancies. Consequently, age-related changes in the role of inference generation during adolescence have not been fully explored. Additionally, CYP are presented with a variety of different texts from multiple sources and genres. Therefore, to fully comprehend all of the texts they read, CYP must also be able to generate counterfactual-world inferences. However, existing research is conflicting - this could be due to the lack of consistency in the stimuli used to assess adults and CYP or a conceptual shift in counterfactual-world inference generation. At present a full and comprehensive model of inference generation, and underpinning skills, during adolescence does not exist. The studies reported in this thesis thus explore age-related changes in real-world and counterfactual-world inference generation abilities

during adolescence to inform educational practice during this time. Part 2 of this thesis goes on to identify a measure suitable for the assessment of inference generation abilities throughout adolescence in order to investigate the following research questions:

1. How does the role of inference generation in higher-order literacy skills change during adolescence?
2. How do inference generation abilities change during adolescence?
3. How do the cognitive underpinnings of successful inference generation change during adolescence?

Part 2: The Selection, Development and Evaluation of a New Inference Generation Task

CHAPTER 2 THE ASSESSMENT OF INFERENCE GENERATION: SELECTING AN APPROPRIATE TASK

2.1. Introduction

The study of inference generation has driven the development of a variety of research tools. However, researchers explore inference generation from different theoretical perspectives and with different aims. Since different theoretical assumptions often result in different assessment methods, the range of reading comprehension and inference generation tasks that exist are not interchangeable, such that they vary in their conceptual underpinnings, format, and related task demands (Carver, 1992; Cutting & Scarborough, 2006; Keenan & Betjemann, 2006; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997; Sabatini, Albro, & O'Reilly, 2012). Subsequently, they do not necessarily assess the same underlying skills and processes and depending on the research aim, each task has both strengths and weaknesses. Selection of the most suitable task must be guided by consideration of these strengths and weaknesses and the specific research aims. For this thesis, a new task was developed by combining a self-paced reading methodology with a forced-choice picture-selection task. This Chapter begins by highlighting some key methodological considerations, the assumptions underpinning each task and rationale for selection are then discussed.

2.2. The Assessment of Real-World and Counterfactual-World Inference Generation Abilities during Adolescence: Key methodological issues and considerations

This research aims to explore the development of real-world and counterfactual-world inferential abilities during adolescence. It is essential that the task(s) selected is able to assess:

- a range of age groups
- both inferential skill and time-course
- the real-world interference thought to be experienced when generating counterfactual-world inferences

2.2.1. Assessment of a Range of Age Groups

Research shows that, due to differences in tasks demands, even the same group of participants can perform differently on two tasks designed to assess the same skill (Boot, Becic, & Kramer, 2009; Burkart & Rueth, 2013; Hinze, Bunting & Pellegrino, 2009). Many of these additional task demands are associated with skills thought to improve with

age (e.g. working memory; Hulme & Tordoff, 19889). Subsequently, younger age groups may display weaker performance than older age groups; this may be reflective of differences in working memory, for example, though, and not the key skill being assessed. The tasks used in developmental studies need to be carefully considered to ensure they target the intended age group. However, Miller (1998) argues that the use of different tasks for each age group may result in a problem of measurement equivalence, such that the tasks used may not be comparable. Miller (1998) argues that to compare particular skills in different age groups, one task should be designed that can adequately tap into the intended behaviour of all of the age groups being studied. This thesis thus aims to use the same methodology (stimuli and task) for all age groups.

2.2.2. Assessing Both Inferential Skill and Time-Course

Inference generation failure is thought to be one of the main causes of poor comprehension (e.g. Cain, Oakhill, Barnes, & Bryant, 2001; Cain, Oakhill, & Elbro, 2003; Perfetti, Marron, & Foltz, 1996). Therefore, the assessment of the inferential *product* (i.e. skill) is essential. Comprehension is also thought to be impaired if a reader is unable to generate the inferences needed to maintain coherence *whilst* reading (Graesser et al., 1994). Similarly, readers may be employing faulty inference generation strategies. Assessment of the unfolding inferential *process* (i.e. time-course) is also necessary. The assessment of both the inferential product and time-course not only allows for quantitative changes in inferential skill to be observed, but also any qualitative changes in the underlying process to be detected. This is particularly important when exploring counterfactual-world inference generation abilities given that research suggests that adults and CYP engage in a qualitatively different process (Bowyer-Crane & Snowling, 2010; Graesser et al., 1998).

In addition to assessing both inferential skill and time-course, it is important that the measure used taps the natural reading process (as much as possible). Research shows that if the text is too difficult and/or presented rapidly, comprehension suffers as participants process the text at a surface- or propositional-level only, with inconsistencies not being detected (Royer, 2001). Given that most inferences are thought to be generated at the situational-level, this could lead to inaccurate conclusions regarding inference generation abilities being made. Research also shows that reading goal can influence comprehension, such that if the reader is not reading with a goal of comprehension, a situational-level representation is unlikely to be constructed (Kendeou, Bohn-Gettler, & Fulton, 2011;

Linderholm & van den Broek, 2002; Narvaez, van den Broek, & Ruiz, 1999; van den Broek, Lorch, Linderholm, & Gustafson, 2001). Gordon, Hendrick and Foster (2000) have shown that task demands can affect reading goals, such that participants adopt the simplest strategy available to them to complete a task – i.e. simply remembering the words in the critical text (creating a surface-level model) instead of comprehending the critical text (creating a situational-level model). Similarly, task instructions have been found to affect the level of comprehension. When reading naturally, elaborative inferences are not drawn whilst reading (Graesser et al., 1994). However, when participants are instructed to predict what will happen next, elaborative inferences are found to be drawn online (Calvo et al., 2006). In sum, it is fundamental that the text, task and instructions used promote a natural reading process, whereby construction of a situational-level representation of the text is promoted.

2.2.3. Counterfactual-World Inference Generation: Assessing the Interference Experienced due to the Activation of Real-World Knowledge.

When generating counterfactual-world inferences the activation of related but contradictory real-world information is thought to result in interference (e.g. Ferguson, 2012). Resolving this interference is thought to be cognitively demanding, meaning counterfactual-world inference generation is often found to be more difficult than real-world inference generation (e.g. Ferguson & Sanford, 2008). During adolescence there is thought to be a qualitative shift in the processing of counterfactual-world information (Bowyer-Crane & Snowling, 2010; Graesser et al., 1998; Rafetseder et al., 2010; 2013). This shift may be due to, at least in part, changes in the real-world interference experienced when generating counterfactual-world inferences. Therefore, exploring developmental differences in the interference faced is likely to aid understanding of the counterfactual-world inferential process and potential causes of inference generation failure.

2.3. Selecting a Task

Inference generation tools can be roughly split into two categories: 1) online measures - used to assess the unfolding inference generation process and 2) memory measures – used to assess the situational model created as a product of the inference generation process. This division means that whilst one task may be suitable for the assessment of the inferential time-course, this same task is unlikely to be suitable for the assessment of inferential skill, and vice versa. Therefore, to meet the task selection criteria set out in

this thesis, at least two tasks were needed. A self-paced reading task was selected to assess the inferential time-course of inference-evoking critical sentences and a forced-choice picture-selection task was selected to assess inferential skill.

2.3.1. Using a Self-Paced Reading Tasks to Assess the Inferential Time-Course

A self-paced reading task (SPRT) is a processing load measure. Other processing load measures include moving window tasks, measurements of eye movements and neuro-imaging methods. Processing load measures provide information about the cognitive costs associated with processing a particular piece of text (Zwann & Singer, 2003). These tasks are underpinned by the assumption that information which is more cognitively taxing takes longer to process (Donders, 1969). The format of these tasks is very simple, with participants reading through a text at their own pace. Using computer software, the time taken to read and/or study the text is then recorded, along with additional measurements depending on specific methods. When exploring inference generation, research has shown that even if two sentences are equal with regards to syllable length (see Table 2.1.), if one sentence requires an inference for coherence to be maintained and the other does not, the sentence requiring an inference will result in greater processing costs – e.g. longer reading speeds, increased neuronal activation (Casteel, 1993; Poynor & Morris, 2003; Virtue et al., 2006; 2008). This effect is thought to be due to the increased cognitive effort needed to generate an inference and coherently integrate the text into a situational model.

Table 2.1. Sample Stimuli used in Processing Load Measures

Condition	Critical Text
Inference-Evoking	A kernel was planted. A century later a huge tree was stood in its place.
Control	Later, a century was planted in place. Stood in its kernel was a huge tree.
Literal Control	A kernel was planted. A century later it had grown into a huge tree.

In a SPRT, the researcher splits the passage into readable segments, using natural breaking points in the text. Segments are presented to the reader one at a time, with participants pressing a button to move through the text at their own speed. The SPRT is, therefore, able to provide a measure of processing costs associated with specific phrases and sentences. However, it is not as sensitive as the Moving Window Task (MWT) and

measurement of eye movements. The MWT works by presenting text on a computer screen, with only one word visible at any time. All other words are replaced with symbols or dashes. The reader is able to move through the passage by pressing a button, such that once the participant has read a word they press a button that then replaces the current word with symbols/dashes and reveals the next word. For example:

A ****** *** *******. * ******* ***** * **** **** *** ***** ** *** *****.

button press leads to:

* *kernel* *** *******. * ******* ***** * **** **** *** ***** ** *** *****.

button press leads to:

* ****** *was* *******. * ******* ***** * **** **** *** ***** ** *** *****.

Eye movement data show that words are rarely read in such a smooth one-by-one, left-to-right manner (Reichle, Rayner, & Pollatsek, 2003). Therefore, Witzel, Witzel, and Forster (2012) argue that MWTs do not promote a natural reading process. Moreover, Zwaan and Singer (2003) suggest that participant's button presses can become rhythmic, failing to reflect underlying processing. Compared with a MWT, a SPRT more closely mirrors the natural reading process. Moreover, since the segments of text presented to the reader can vary in length, the likelihood of participants engaging in rhythmic button pressing is minimised.

When eye-tracking is used, the text is presented on the screen, with small cameras used to monitor the participant's eye movements as they read. This means that the participant is not required to press any buttons to read through the text. Consequently, the data obtained do not include any additional processing costs associated with the motor responses needed to press a button and/or the effort associated with switching between reading and pressing a button. Eye tracking methodologies are, therefore, thought to result in a purer measure of the processing costs associated with the natural reading process than the MWT and SPRT. Eye-tracking methodologies are also able to provide much more in-depth detail, exploring fixations, saccades, and regressions. Given the time constraints of this PhD, it was thought that the use of an eye tracker would be too labour intensive in terms of the amount of data generated, however. Brain-imaging methods were not

selected for the same reason[2]. Fortunately, Zwaan and Singer (2003) suggest that in many cases the SPRT produces data precise enough for the assessment of many skills and processes of interest when exploring inference generation. As such reading speeds have been used to assess the inferential time-course in previous research (e.g. Casteel, 1993). As discussed below, compared to other online measures, the SPRT is also able to better fulfil the task selection criteria outlined at the start of this Chapter.

Other online measures include activation measures and information-content measures. Activation measures, such as the lexical decision task, naming task and probe recognition task, are used to assess the reader's unfolding mental representations, specifically, the level of activation of a particular idea or concept (Zwaan & Singer, 2003). Typically, participants read a sentence followed by a critical word that participants must make some form of decision about. Information-content measures are used to provide detailed information about the skills and processes a participant is using when reading a text (Zwaan & Singer, 2003). Perhaps the most commonly used information-content measure is the think-aloud task. Think-aloud tasks require participants to verbalise their thought process and understanding as they read a text. When assessing a range of age groups, the SPRT is superior to information-content measures, such that SPRTs simply require participants to read a text (Zwaan & Singer, 2003). Therefore, as long as the texts used are age appropriate, SPRTs are suitable for the assessment of a range of age groups. Conversely, whilst training is given, performance on think-aloud tasks is heavily dependent on expressive language abilities. Since expressive language abilities typically improve with age (e.g. Berninger & Abbott, 2010), any differences observed between different age groups could be due to either differences in inference generation abilities or differences in expressive language skills.

A SPRT is also likely to promote a more natural reading process than activation measures and information-content measures. Activation measures require participants to respond to a critical test item *after* reading the text. Therefore, it is unclear if the processing assessed reflects the processing taking place whilst reading or at the time of test. Moreover, Gordon et al., (2000) have shown that when probe recognition tasks are used participants simply remember possible probe words, processing the text at a surface-level

---

[2] Additionally, brain-imaging techniques require participants to remain still, given that the inference generation task designed for this thesis took around 30 minutes, it was felt that younger participants may struggle to remain still for this amount of time.

only. Whilst information-content measures are categorised as an online measure, Zwaan and Singer (2003) suggest that many of the inferences participants report to be generating as part of the think-aloud task, may not be generated when reading normally. Instead, some of the inferences generated may be the result of strategic processing promoted by task instructions. The SPRT, however, does not require an explicit response from the participant and measures are taken *whilst* the participant is reading, not after. Consequently, the natural reading process is not disrupted. Since a SPRT causes minimal disruption to the natural reading process, unlike activation measures and information content measures, a valid measure of the level of real-world interference a participant faces whilst reading naturally can be obtained (e.g. Ferguson 2012; Ferguson & Sanford, 2008; Ferguson et al., 2008). In sum, the SPRT is able to assess a range of age groups, with minimal disruption to the natural reading process. In addition, a measure of interference whilst reading can be obtained.

2.3.2. Using a Forced-Choice Picture-Selection Task to Assess the Inferential Product

Forced-choice picture-selection tasks (FCPST) present participants with a critical sentence followed by a grid displaying two or more pictures – one of which depicts the target (see Figure 2.1). Participants select the picture that best matches the preceding text.
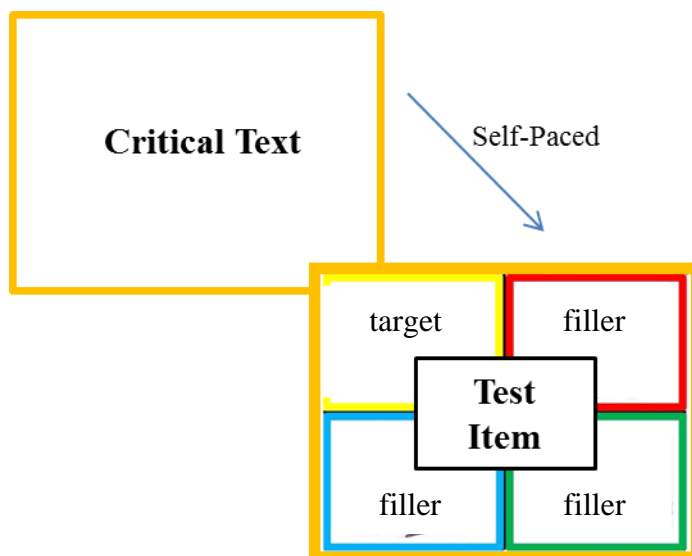
Figure 2.1. Schematic Representation of the Forced-Choice Picture-Selection Task

A FCPST is a type of multiple-choice task, which is a forced-choice memory measure. Other commonly used forced-choice memory measures include the sentence recognition task (SRT) and sentence verification task (SVT). In the SRT participants have to decide if the test sentence appeared in the preceding text or not. In the SVT participants have to decide if, based on the preceding passage, the test sentence is true or false. If the correct

inference has been drawn participants are predicted to falsely recognise the test sentence on the SRT, but correctly validate or select the test sentence on the SVT and multiple-choice task (MCT), respectively.

A FCPST is more suitable for the assessment of the inferential product than any other forced-choice memory measure for several reasons. First, research suggests that the SRT may encourage surface-level processing, such that participants can successfully complete the task without creating a situational level model (Gordon et al., 2000). Second, when using the SVT, interference is thought to be present if the participant fails to validate the counterfactual-world target and instead incorrectly validates the real-world alternative or if the participant's responses to counterfactual-world items are correct, but significantly slower than responses to real-world items. However, the SVT requires participants to not only generate an inference but validate this inference too. Research suggests that inference validation may not be a part of the inferential process when reading counterfactual-world texts naturally, as this would interfere with their engagement of the text (e.g. Gillbert, 1991; Gillbert et al., 1990; Green & Brock, 2000; Rader & Sloutsky, 2002). The need to validate inferences, then, may confound results and lead to the amount of interference experienced being over estimated.

As discussed in Section 1.2.1, theories of counterfactual-world processing suggest that, at least initially, both counterfactual-world and real-world information is mentally represented (e.g. Ferguson, 2012). The real-world knowledge and the counterfactual-world information are thus thought to be in direct competition when comprehending counterfactual-world texts. The SVT and SRT only allow for the assessment of one piece of information at a time. Therefore, if an error is made on the SVT or SRT or response times are slower in counterfactual-world conditions than real-world conditions, it can only be said, with confidence, that the correct inference has not been drawn or in the case of correct but slow response times, the inference has not been drawn online. This could be due to real-world interference, but also faulty inference generation processes. A major advantage of employing a MCT, therefore, is the ability to simultaneously present several ideas per test item – i.e. both the counterfactual-world target and real-world alternative can be placed in direct competition, just as they are in readers' internal representations. If the participant routinely selects the real-world alternative instead of the counterfactual-world target, or correctly selects the counterfactual-world target but significantly slower than they select the correct target in real-world conditions, it can be said, with much more confidence, that these patterns are due to the increased cognitive effort associated with

resolving the conflict experienced due to real-world interference. However, as highlighted by Cain and Oakhill (2006b) even though verbal responses are not required, since multiple options have to be processed and compared, the processing demands associated with a MCT are much higher than those associated with other forced-choice memory measures. As discussed in the next section, the use of pictures minimises these demands, meaning a FCPST is superior to other verbal forced-choice memory measures as processing demands are reduced. As discussed below, compared to other offline measures, the FCPST is also better able to fulfil the task selection criteria outlined at the start of this Chapter.

2.3.3. Assessing a Range of Age Groups: Reducing Output Demands Via the Use of Pictures

Spooner, Baddeley and Gathercole (2004) found that children's comprehension performance was significantly better when measured using forced-choice responses as opposed to constructive memory measures, such as self-constructed answers to open-ended questions. Spooner et al. (2004) suggest that when constructive memory measures are used, poor comprehension performance may not always be attributable to poor comprehension abilities. Poor performance may instead be due to the difficulties associated with constructing an answer. Spooner et al. concluded that since comprehension is an input process (i.e. construction of a mental model), the measure used to assess these processes should not make heavy output demands. The ability to deal with heavy output demands is likely to increase with age, due to improving working memory (e.g. Conklin, Luciana, Hooper, & Yarger, 2007) and expressive language skills (e.g. Berninger & Abbott, 2010). When completing tasks that place heavy output demands on participants, the comprehension performance of younger children is thus likely to be more confounded than the performance of older children and adults. Tasks that minimise output demands should thus be used to assess inference generation in a range of age groups. As discussed in this subsection, a pictorial response method places fewer output demands on participants than verbal response methods.

A large body of research exists exploring the processing differences of words and pictures, with several theories being proposed. Typically, these theories fall into two camps. Some suggest that pictures and words are processed differently (e.g. Glaser & Glaser, 1989). Others argue that pictures and words are processed similarly, both having access to one unified semantic system (e.g. Kiefer & Pulvermuller, 2011). Despite these theoretical differences (a full discussion of which is beyond the scope of this thesis), research

consistently finds superior performance (e.g. faster response times, fewer errors) when pictorial, as opposed to verbal stimuli, are used – i.e. a picture superiority effect (Hinojosa, Carretie, Valcarel, Mendez-Bertolo & Pozo, 2009; Schacht & Sommer, 2009).

Pictures are thought to be processed with more ease than words since additional translational processes are not needed. Text needs to be decoded – i.e. letters matched to their sounds and words deciphered (e.g. Samuels, 1988). Similarly, the words must be matched with their referent meanings (e.g. Coltheart, Curtis, Atkins, & Haller, 1993). Pictures, though, often allow for certain elements to be represented with relatively effortless maintenance, meaning working memory is freed up and can be used to complete other tasks (Glenberg & Langston, 1992). Since decoding ability, referent search ability, and working memory capacity are all found to improve with age (e.g. Conklin et al., 2007; Luciana, Conklin, Hooper, & Yarger, 2005) it follows that children show superior performance when processing pictures compared to words (e.g. Rafetseder et al., 2013; Wright & Berch, 1992). Developmental studies find that, regardless of task, the magnitude of the picture-superiority effect declines with age (Rosinksi, Pellegrino, & Siegel, 1977; McGonigle and Chalmers, 1984). This suggests that when words are used, any differences observed between different age groups could be due to developmental differences in processes associated with the additional skills needed to translate text to meaning and not real changes in the skill being assessed. It is unsurprising then that many have used pictorial stimuli, specifically FCPSTs, to assess adults and children (e.g. amount and accessibility of knowledge, Barnes et al., 1996; pronoun resolution, Sekerina, Stromswold, & Hestvik, 2004; vocabulary, Dunn & Dunn, 1997; counterfactual-world reasoning, Rafetseder et al., 2013; inference generation, Ford & Milosky, 2003).

Wright and Berch (1992), however, noted some fundamental differences between the pictures and words commonly used in research. They found that pictures are often larger than the words used. Therefore, perceptually, pictures may be easier to interpret (Theios & Amrhein, 1989). Additionally, pictures of items from the same category are typically more visually similar than items from different categories, whereas words from one category are no more or less visually similar to words from a different category (Snodgrass & McCullough, 1986). Therefore, pictures are likely to aid categorisation more so than words. Similarly, Schlochtermeier, Kuchinke, Pehrs, Urton, Kappelhoff, & Jacobs (2013) found that whilst a word and picture can contain the same information they are not necessarily equivalent with regards to the level of detail conveyed. It is thought that pictures represent more detail regarding position, size, and relationships than words.

71

As such, when Schlochtermeier et al. (2013) adapted their text so it was just as detailed as a picture, they found that the picture superiority effect disappeared.

It appears that pictures aid access to meaning since they provide more detail than words, which is more easily accessed since additional translational processes are not needed. Many of the additional skills associated with the processing of verbal information have been found to improve with age. The effect of some of these developmental confounds can be removed by using a literal control. However, unless the literal controls and test items are lexically identical, which could be potentially very repetitive, some differences may still exist. Similarly, the confounding effects of decoding could be removed by presenting test items orally, however, this could prove problematic when presenting more than one idea. The participant would need to remember all of the words or ideas presented to them, increasing the demands placed on short-term memory. In contrast, the use of pictures allows for the confounding effects of developmental differences in decoding, reading speed, associations between terms and referents, and short-term memory and working memory capacity to be minimised and in some cases completely eliminated, even when presenting multiple ideas. Subsequently, a FCPST is likely to provide a purer measure of abilities when assessing a range of age groups compared with other memory measures.

2.3.4. Using a 2*2 Forced-Choice Picture Selection Task to Promote a Natural Reading Process

When reading naturally a situational model of the text is thought to be created (e.g. Singer, 1990). Many of the tasks reviewed above promoted the creation of a surface- or propositional-level representation due to simplistic response methods. In this thesis, the term picture refers to an image that depicts not only a single object, but an event, concept or relationship. Therefore, whilst the FCPST response method is simple, the correct picture cannot be selected by simply remembering the preceding words or phrases as there will be no direct match between the picture and preceding critical text. Instead, the critical text must be processed at the situational-level for the correct picture to be selected.

The use of pictures has been found to aid CYP's construction of mental models of the text, such that they serve as cues for information likely to be necessary to maintain coherence (Glenberg & Langston, 1992; Gyselinck & Tardieu, 1999), with this information influencing younger CYP more so than older CYP (Pike et al., 2010). As discussed above, pictures more directly depict relationships between various elements

(Gyselinck & Tardieu, 1999) and thus may facilitate the transformation of text to a mental model. Moreover, the direct and clear depiction of said relationships is thought to reduce the load placed on working memory (Marcus, Cooper, & Sweller, 1996) and thus allow resources to be allocated to higher-order skills such as inference generation if the picture is unambiguous. Subsequently, some may argue that the use of a pictorial task serves to facilitate younger children's mental model construction, thus over estimating their skills for text-only comprehension and biasing the results. However, in the current research, the pictures are not presented simultaneously with the critical text. They are presented after reading the text. Thus whilst the pictures for coherence items, in particular, will depict the relationships between elements, this picture will not be able to serve as a cue for online mental model construction as it is presented *after* reading. The use of a pictorial response method may encourage participants to create more pictorial situational models than they may have done if a verbal response method was used. However, situational models are thought to contain information from a variety of sensory modalities, including vision (e.g. Zwaan, 2003).

Overall, the use of a FCPST is likely to promote the creation of a situational model. Whilst the simultaneous presentation of text and pictures has been shown to facilitate situational model construction, this is unlikely to occur on the FCPST, since pictures are presented after situational model construction has begun. Subsequently, the FCPST is likely to provide a relatively valid assessment of those inferential abilities associated with the natural reading process.

2.3.5. Using Pictures to Evaluate the Mental Representations Constructed From Verbal Stimuli

Traditionally situational models were thought to contain only linguistic information (e.g. Fodor, 1975; Newell & Simon, 1972). Subsequently, the use of pictures, as opposed to text, to assess a verbal mental representation could be problematic since additional translational demands would be placed on participants. However, Zwaan (2003) proposed the Immersed Experiencer Framework, which suggests mental models contain information from a variety of sensory modalities. This idea has much support with research finding situational models contain information from a variety of sensory modalities (Zwaan, 1996; Zwaan & Taylor, 2006), including vision (Kaup, Yaxley, Madden, Zwann, & Ludke, 2007; Zwaan & Yaxley, 2003a; 2003b). The inclusion of visual information in the situational models created by readers is also supported by studies

which find that participants often report the use of imagery when reading (e.g. Richardson, 1980).

Schnotz and Bannert (2003) have created a multimedia model of situational model construction that focuses on the comprehension of textual and pictorial stimuli (see Figure 2.2.). As can be seen from this model, there are two distinct branches, a textual (descriptive) branch and a pictorial (depictive) branch, with both branches cumulating in a depictive mental model which continually interacts with a propositional representation and is guided by the reader's conceptual organisation. Therefore, whether the initial input is depictive or descriptive both a propositional representation and depictive mental model are created, with constant interactions occurring between descriptive and depictive information.

Figure 2.2. Schnotz and Bannert's (2003) Integrated Model of Text and Picture Comprehension

The branches and pathways in Schnotz and Bannert's model are bi-directional, meaning that a picture can be used to create a situational model, and existing situational models (whether created from an initial text or pictorial input) can be evaluated using pictorial stimuli. Therefore, the use of a pictorial response method is likely to be suitable for assessing the situational models constructed whilst reading. This is supported by the numerous studies that have effectively used pictorial naming and categorisation tasks to assess the mental representations created from text (e.g. Ford & Milsoky, 2008; Kaup et al., 2007). Those using a FCPST to assess inference generation suggest that the FCPST is an effective measure of inference generation abilities (e.g. Ford & Milsoky 2003; 2008; Schmidt & Paris, 1978).

For example, Ford and Milosky (2003) used a four-choice picture response method to explore inference generation abilities in CYP aged 5 to 6 years. CYP heard short scenarios evoking emotional inferences, with the last word missing (e.g. Twinky was bouncing a ball. A bully took the ball. Twinky was _____). Children were asked to select the facial expression that completed the scenario from a choice of four pictures (e.g. happy, sad, mad, and surprised), presented using a 2*2 grid format. Ford and Milosky found that all children, even those with a language impairment, understood the task and were able to effectively use the pictorial response method. This suggests that a 2*2 format picture selection method can be used to assess inference generation abilities, even in young children with weak linguistic abilities. This suggests the FCPST is suitable for the assessment of inference generation abilities.

2.3.6. Using a 2*2 Forced-Choice Picture-Selection Task to Assess Real-World Interference

When assessing counterfactual-world processing using constructive memory measures, if incorrect information is recalled or a participant answers a question incorrectly, errors can be analysed to determine the underlying cause of inference generation failure, specifically, the presence of real-world interference (e.g. Cain et al., 2001; Cain et al., 2003; Carlson, Seipel, & McMaster, 2014; Norbury & Bishop, 2002). However, the level of real-world interference present in participants' reconstructions may not be an accurate reflection of the interference present whilst engaging with the text. When reading fantasy and fiction texts, some argue that readers actively suspend beliefs and use the counterfactual-world context as a base for all comparisons (e.g. Filik, 2008; Filik & Leuthold, 2008; Nieuwland & Van Berkum, 2006). Subsequently, real-world interference is minimal. However, when asked a question or asked to recall the text, the reader may

more consciously consider real-world violations. Subsequently, when constructing an answer, the real-world interference experienced may be stronger than when reading a fantasy or fiction text. A FCPST may be a useful tool when assessing real-world interference, such that assessment is made immediately after reading the critical text and processes associated with constructing an answer are not needed. Specifically, both the counterfactual-world target and real-world alternative can be presented. Given that processing costs are reduced by the use of pictures instead of words, several possibilities can be presented, allowing even firmer conclusions to be made regarding the cause of inference generation failure – i.e. in addition to the counterfactual-world target and real-world alternative, fillers indicative of, for example, integration failure or literal processing, are also included. Error analysis could then be conducted allowing much firmer conclusions regarding the cause of inference failure to be made.

2.4. Conclusion

A review of existing inference generation tasks revealed that no single task was able to meet the following task selection criteria:

- Assess a range of age groups
- Assess both the inferential skill and time-course associated with the natural reading process
- Assess the real-world interference thought to be experienced when generating counterfactual-world inferences

Many previous measures place high output demands on participants, despite comprehension and thus inference generation being primarily input processes. When assessing a range of age groups this is problematic as ability to overcome output demands improves with age, meaning differences in performance between older and younger age groups could be due to either genuine differences in inference generation abilities or differences related to task demands. Whilst some existing tasks use simple verbal response methods with few output demands (e.g. lexical decision task), it has been argued that successful completion of these tasks is possible by creating a surface-level model of the text only. The use of forced-choice picture-selection task (FCPST) which utilises a pictorial, as opposed to verbal, response method reduces task demands, such that translational demands (e.g. decoding) are minimised. However, a complete idea is still depicted meaning participants cannot respond correctly by simply remembering the text verbatim. A pictorial response method, unlike some of the simple verbal response

methods, thus promotes a natural reading process. Typically, existing inference generation tasks assess either inferential skill or time-course. Simultaneous assessment of the inferential process and product was desired, however, to minimise the time CYP spent out of their learning environments. Unlike offline assessment tools which measure the inferential process after reading and require an explicit response, the self-paced reading task assess the inferential process whilst reading and does not require an explicit response. The natural reading process is thus uninterrupted and by measuring reading speed of critical texts that precede the FCPST a measure of the inferential process can be obtained without increasing participant assessment time. Finally, due to minimal output demands, several ideas can be presented simultaneously in the FCPST without exhausting participants' resources. Both the target and real-world alternative can be presented in counterfactual-world conditions. Additional diagnostic errors, reflecting typical reasons for inference generation failure can also be included. Error analysis can be conducted to explore real-world interference when processing counterfactual-world information in addition to other faulty inference generation strategies. This new task, which combines a self-paced reading methodology and forced-choice picture selection task will be known as the Image Selection Task.

CHAPTER 3 THE DEVELOPMENT OF THE IMAGE SELECTION TASK - A NEW INFERENCE GENERATION PARADIGM

3.1. Introduction

This thesis is concerned with the exploration of age-related changes in real-world and counterfactual-world inference generation abilities during adolescence. To obtain a comprehensive picture of inference generation abilities during this time, it is essential that the task(s) used is able to assess a range of age groups, both inferential skill and time-course, and the real-world interference experienced when generating counterfactual-world inferences. However, as discussed in Chapter 2, many of the most commonly used inference generation tasks are unable to fulfil all of the above task selection criteria. Consequently, the creation of a new task – the Image Selection Task (IST) - was necessary. This Chapter outlines the development of the IST.

3.1.1. Outline of the Image Selection Task

The IST combines a self-paced reading methodology with a forced-choice picture-selection task (FCPST), embedding them both into a story (see Figure 3.1 for a schematic representation). Participants move through the story at their own speed until a 2*2 grid with four pictures is presented. Although not explicitly highlighted to participants, each 2*2 grid is preceded by a screen displaying a critical text. When the 2*2 grid is presented, participants select the picture they feel best matches the text they have just read. The time taken to read each critical text provides a measure of the inferential time-course whereas inferential skill was assessed via number of errors made on FCPST. An analysis of errors was also conducted to determine the potential cause of inference generation failure and levels of real-world interference in the counterfactual-world conditions.
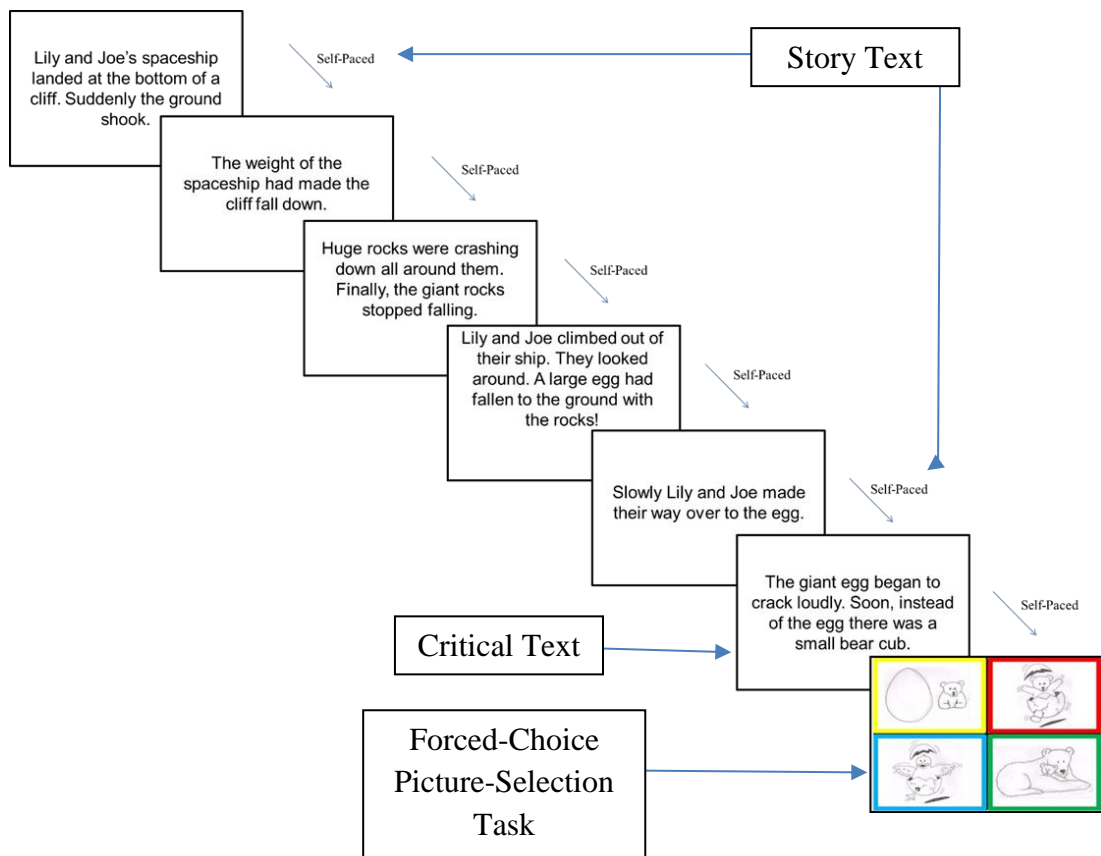
Figure 3.1. Schematic Representation of the Image Selection Task

3.1.1.1. Story Development

Many studies do not provide context before or after the critical text. Instead, they simply present the critical text followed by the test stimuli, this is then immediately followed by another item (e.g. Calvo et al., 2006; Ferguson, 2012; Gernsbacher & Faust, 1991). However, this is very different to reading in real-life, where typically a chunk of text is read to extract information. The standard coherence framework suggests that reading goals determine the level of coherence a reader seeks to gain from a text (van den Broek et al., 1995; 2001). Subsequently, reading a list of sentences with no real purpose is unlikely to lead to a reading goal of comprehension, and thus experimental stimuli are unlikely to be processed to the same level as natural texts. Embedding the critical texts and FCPST into stories provides a context and increases engagement, creating reading goals similar to those associated with the natural reading process.

This thesis is primarily concerned with those inferences drawn when engaging with the science-fiction texts commonly read for pleasure by CYP and used in the classroom. Originally, the researcher sought to use 'real texts.' However, due to their complexity and variability this was not possible. Subsequently, story plots were developed based on the ideas depicted in the BBC Bitesize Key Stages 2 (7-11years) and 3 (11-14years) Science

resources and in books in the 'young readers' fantasy' section of the researcher's local library. A predominate theme in this literature was space travel. Consequently, stories were designed to focus on the space adventures of two characters – Holly and Ben - with each story relating to the characters' adventures on a new planet.

All stories were written to be 190-210 words long to prevent fatigue and boredom, but at the same time allow a full plot to develop and engage participants. Royer (2001) suggests if a text is too difficult participants' comprehension suffers. Subsequently, stories were designed to be readable by the youngest participants assessed in this thesis – 9 year olds. Hatcher's (2000) book grading formula was used to assess the readability of stories[3]. All stories received a readability level of 9 years or less (see Appendix 3.1 for a full list of story ratings).

Six stories were written to assess real-world inference generation abilities (real-world stories), six stories were written to assess counterfactual-world inference generation abilities (counterfactual-world stories), and six stories were written to assess memory for information explicitly stated in the text (literal stories). Each story was split into three sections. The first two sections ended with a critical text. The final section provided a conclusion to the story. See Table 3.1. for an example story. Therefore, each story contained two critical texts. Each real-world and counterfactual-world story contained one critical text designed to evoke a coherence inference and one critical text designed to evoke an elaborative inference. In the literal condition, one critical text assessed memory for real-world information and the other, memory for counterfactual-world information. Two versions of each story were written, such that the second version of stories had the converse arrangement of critical texts. This was to ensure any differences between different conditions could not be attributed to the position of critical sentence. This meant that the story content had to be changed slightly. The number of words in Version 1 and 2 of the story always remained exactly the same. The general gist of the story also remained consistent, as did the readability rating (mean rating Version 1 = 8.79, SD = 0.19; mean rating Version 2 = 8.83, SD = 0.18; t (17) = 0.72, p = 0.48). There was also no difference in readability ratings across the three story conditions (mean rating real-world stories = 8.78, SD = 0.23; mean rating counterfactual-world stories = 8.74, SD =

---

[3] This formula is based on five key variables: number of words in story, number of pages in story, number of words with 6 or more letters, number of words in the longest sentence, and number of syntactic features. Syntactic features refer to contractions, use of negatives, change of verb tense, and so on.

0.14; mean rating literal stories = 8.89, SD = 0.13; F (2, 35) = 2.60, p = .09). The administration of programs 1 and 2 was also counterbalanced across participants, see table 3.2 for a summary of the counterbalancing and randomisation of critical texts and stories.

Table 3.1. Example Story Text with Critical Sentences Embedded

	Real-World Condition
Section 1	Holly and Ben had flown into an asteroid storm. Giant rocks were crashing into the spaceship. It was going to get damaged. CRASH!! WHACK!!! A rock hit the wing. The wing fell off and landed in the ocean below. The ship was out of control. Holly and Ben had to make an emergency landing. They were able to land on a beach next to the huge ocean. There was a boat tied up next to a statue.
Critical Sentence	*The gigantic clay statue shone in the sun light. It must have been at least ten feet high.*
Section 2	Holly and Ben could see the wing in the ocean. However, they could also see some very evil sharks. Holly and Ben climbed into the boat. They made it to the wing in no time. Sadly, the heavy wing had started to sink. Holly and Ben could not reach the wing from the boat. One of them would have to jump in for it. Holly dove into the sea. She grabbed the wing.
Critical Sentence	*However, a shark was approaching. It attacked. Blood began to spill out of Holly's arm.*
Section 3	Ben pulled Holly and the wing into the boat. Ben put the wing back on the spaceship. They then flew away from the awful planet.

Table 3.2. Counterbalancing and Randomisation of materials

Type of passage	Program 1 Critical sentence presented first	Program 2 Critical sentence presented first	First target picture in quadrant...	Second target picture in quadrant...
literal	CFW	RW	2	3
CFW	coherence	elaborative	2	1
RW	coherence	elaborative	4	3
CFW	elaborative	coherence	2	1
RW	coherence	elaborative	1	2
literal	RW	CFW	4	1
RW	elaborative	coherence	2	3
CFW	coherence	elaborative	3	2
CFW	elaborative	coherence	4	3
literal	CFW	RW	1	4
RW	elaborative	coherence	4	3
literal	CFW	RW	3	2
CFW	elaborative	coherence	1	4
CFW	coherence	elaborative	3	4
literal	RW	CFW	3	4
literal	RW	CFW	1	2
RW	elaborative	coherence	1	1
RW	coherence	elaborative	2	4

CFW = counterfactual-world, RW = real-world

3.1.1.2. Development of Critical Texts

Critical text refers to the sentence(s) which evokes a target inference, or in the literal conditions, explicitly states the information needed to respond correctly. By assessing responses to the different types of critical texts, the patterns of inferential skill (number of errors) and time-course (reading speed) can be obtained. To explore these patterns six types of critical text were created:

1. Real-world Coherence
2. Real-world Elaborative
3. Counterfactual-world Coherence

4. Counterfactual-world Elaborative

5. Literal Real-world

6. Literal Counterfactual-world

3.1.1.2.1. Selection of Coherence and Elaborative Inferences Types

There are several types of coherence and elaborative inferences (e.g. Graesser et al., 1994). Exploration of all of these inferences is beyond the scope of this thesis. The coherence and elaborative inferences selected for exploration were those deemed most important during adolescence.

During Key Stages 2 – 4, CYP are expected to not just read a text, but to understand it and thus extract relevant information from it (Department of Education, 2013). Consequently, it is essential that CYP are able to create coherent situational models of the texts they are presented with. Graesser et al., (1994) define global coherence as localised chunks of information which have been hierarchically organised and connected into higher-order chunks. Thus, global coherence is essential for summarising a text. However, research suggests that global coherence is unlikely to be achieved without local coherence (see Graesser et al., 1994 coherence assumption). Graesser et al. (1994) define local coherence as "structures and processes that organise elements, constituents, and references of adjacent clauses or short sequences of clauses" (Graesser et al., 1994, pg. 371). Graesser et al. (1994) describe three types of local coherence inferences: referential, case structure role assignment, and causal antecedent. Referential and case structure role assignment inferences, although thought of as essential for local coherence (e.g. Kintsch & van Dijk, 1978), are fairly basic and as such are thought to be developed to an almost adult-like level by children aged 9-10 years (Ackerman, 1986). Causal antecedent inferences, on the other hand, refer to those inferences which link or bridge together current actions/events/states with previous ones (Graesser et al., 1994). Research suggests that causal antecedent inference-making ability continues to develop throughout childhood and into adulthood (e.g. Ackerman, 1986; Casteel, 1993). Since these inferences are necessary for establishing explanations for events occurring in a text (e.g. a character's behaviour or the cause of an element changing state), they are essential for both narrative and expository texts, making them particularly important for success across the curriculum.

CYP must also be able to elaborate on a text during Key Stages 2-4, adding information which, whilst not necessary for coherence, aids understanding by creating a more detailed

picture (Department of Education, 2013). Graesser et al. (1994) describe five types of elaborative inference: causal consequence inferences (more commonly known as predictive inferences), instantiation of noun category inferences, instrumental inferences, subordinate goal-action inferences, and static inferences. Static inferences are likely to be particularly useful during Key Stages 2-4 since they refer to inferences made about a character's or object's, traits, knowledge, beliefs and/or properties (Graesser et al., 1994). When drawing opinions and conclusions about why a character may react in a certain way it is necessary to infer what the agents in the text may have known or believed. In science it is necessary to infer static properties of materials to understand and predict potential uses (e.g. which material will be the best insulator), changes in state (e.g. solid to liquid), and outcomes. Consequently, causal (coherence) and static (elaborative) inferences were explored in this thesis.

3.1.1.2.2. Criteria for Critical Texts

All critical texts were designed to meet the following criteria:

      1. All critical texts must read naturally

Typically, previous research uses the same critical text in all conditions, varying only one target word (e.g. Cook et al., 2001; Ferguson, 2012). This allows for strong experimental control. However, in opposition to previous research, the critical texts in the IST were embedded into stories to provide the critical text with a context and encourage a natural reading process. The use of identical critical texts was likely to result in the creation of contrived and repetitive stories. This issue could have been overcome by creating two additional stories for each critical text and counterbalancing these across participants. Due to time constraints this was not possible[4]. Subsequently, 36 unique critical texts were created.

With regards to interference, two identical sentences are usually employed to explore differences in responses to the real-world alternative compared to the counterfactual-world target. However, since the IST presents both the counterfactual-world target and real-world alternative, real-world alternative control sentences were not needed.

---

[4] 54 as opposed to just 18 stories would have had to be created.

2. All critical texts must have the same syllable length, syntactic structure, tone, and readability level

To ensure response patterns reflect differences in inferential processing and not linguistic processing features (e.g. text length, grammatical structures, reading difficulty) critical texts were designed to be as similar as possible. Firstly, all critical text were designed to be exactly 21 syllables in length. Critical texts were matched on syllable length, as opposed to word length, as syllables are thought to provide a more robust measure of reading speed than words (e.g. Just & Carpenter, 1980). Critical texts were written using the same syntactic structures and tone and rated using Hatcher's (2000) book grading formula[5]. All critical texts received readability ratings of 9 years or less, with no significant differences between the six conditions (real-world coherence mean = 7.60, SD = 0.48; real-world elaborative mean = 7.51, SD = 0.33; counterfactual-world coherence mean = 7.77, SD = 0.81; counterfactual-world elaborative mean = 7.75, SD = 0.68; literal real-world mean = 7.78, SD = 0.57; literal counterfactual-world mean =7.68, SD = 0.69; F (5, 35) = 0.19, p = 0.97). Subsequently, all critical texts should be readable by even the youngest children in the study.

3. Elaborative critical texts should not promote online processing.

Research finds that elaborative inferences can be drawn online under certain conditions, such as high contextual constraints resulting in few alternate possibilities, easily accessible knowledge due to preceding activation or strong semantic constraints, or when given enough processing time (e.g. Calvo, Castillo, & Schmalhofer, 2006). To ensure elaborative inferences were not primed by individual words in the critical text, the semantic associations between the words in the critical text and intended inference were checked using the Edinburgh Associative Thesaurus database (EAT, 2013). If a word was found to have a high semantic association with the target the word was changed. Research shows that when tasks are repetitive and consistently require a similar elaborative inference, participants are able to anticipate this and engage in the strategic processing needed to complete the task when reading the critical sentences (e.g. Calvo et al., 2006; Keenan et al., 1984). Consequently, results could suggest that an elaborative inference

---

[5] Words in a story was set to 101 when calculating readability grades for critical texts since critical texts were embedded into a story. All other factors (e.g. words in longest sentence and syntactic features) were calculated based solely on critical text.

has been drawn whilst reading, when in fact, if reading naturally, this may not have been the case. For this reason, critical texts were embedded into a story as opposed to appearing at the end of a story. The need to draw inferences based on several topics of knowledge should also minimise this anticipatory effect as participants cannot predict the type of static property they should be making an inference about. Some participants may still anticipate the generation of an inference related to the properties of the object being described. However, there are so many potential possibilities the reader is unlikely to try and generate them all.

Science-fiction texts are often used during Key Stages 2 and 3 to teach and reinforce scientific principles (BBC Bitesize, 2014 – see science section). Subsequently, all critical texts were designed to reflect information on one of six key scientific topics *(reproduction, transport and movement, communication, animal attacks, plant and human anatomy, physical properties)* currently taught during Key Stages 2 and 3, identified through consultation with Key Stage 2 and 3 Science National Curriculums and the BBC Bitesize Science website (see Appendix 3.2).

3.1.1.3. Development of the Forced-Choice Picture-Selection Task

Several picture databases exist. However, it was not possible to use these as implausible, and in many cases novel, ideas needed to be depicted – e.g. a bear hatching from an egg. Subsequently, it was necessary to create new pictures. The researcher read the critical texts and then drew pictures based on her mental representation. The process of picture creation is summarised in Figure 3.2.
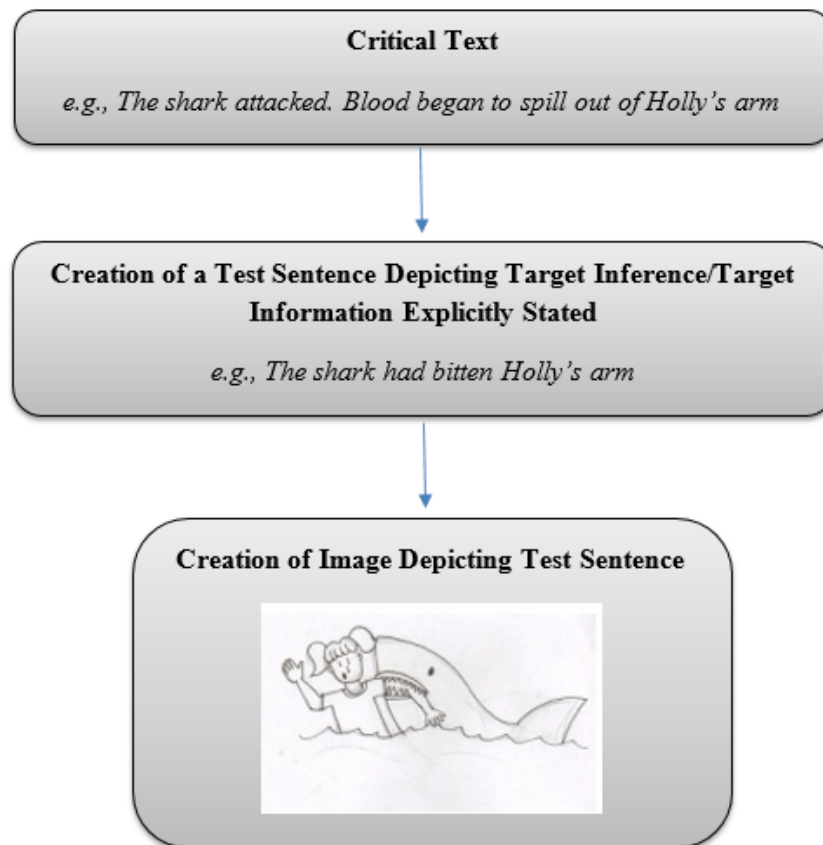
Figure 3.2. The Process of Picture Creation

As discussed in Section 2.3.5., pictorial test stimuli are likely to be particularly advantageous when assessing the real-world and counterfactual-world inference generation abilities of a range of age groups as meaning can be extracted with more ease from pictures than words. This is only the case if the picture adequately depicts the idea intended. Consequently, all pictures were designed to meet the following criteria:

1. All pictures contain two to four key components

Research suggests that busy pictures (i.e. those containing a large number of elements) can be difficult to interpret and can be more ambiguous than simpler pictures (e.g. Székely, D'amico, Devescovi, Federmeier, Herron, Iyer, Jacobsen, & Bates, 2003; Snodgrass & Vanderwart, 1980; Snodgrass & Yuditsky, 1996). Typically, the more detailed the picture, the more difficult it is to interpret. Subsequently, simple drawings were used, with the number of elements in each picture kept as consistent as possible, such that each picture comprised only 2-4 elements – in the picture above there are three elements: the shark, Holly and the sea.

2. All pictures are simple line drawings

Research shows that line drawings are easier to comprehend than real photographs (Ryan & Schwartz, 1956). Hochberg (1972) argues that this is because line drawings create a prototype – i.e. they convey only the key distinctive features of the object making object identification easier. To ensure pictures were easy to interpret simple line drawings were used.
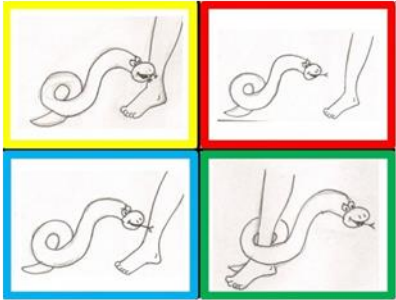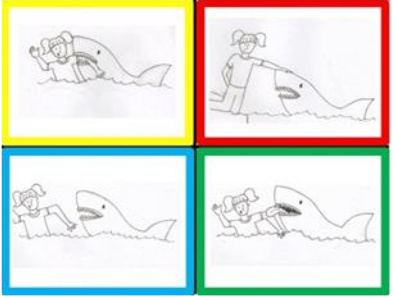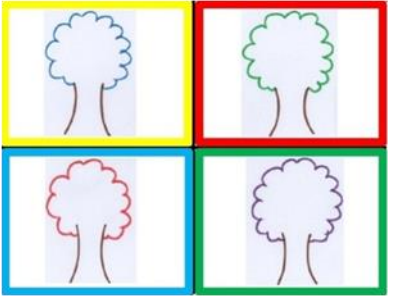
3. Whenever possible, all pictures are black and white

Research shows that visually salient features such as colour can affect picture processing, with visually distinctive elements attracting more attention than less distinctive elements (e.g. Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002). Unless colour was one of the static properties being assessed, all pictures were presented in black and white.

3.1.1.3.1. Design of Fillers

Three filler pictures were created for each item, with one filler in the counterfactual-world conditions always depicting the real-world alternative. In both real-world and counterfactual-world conditions, one filler also depicted the literal state of events in the inference conditions. Research shows that processing of a visual scene can be affected by both visually salient features (Itti & Koch, 2000) and those features likely to be interesting or meaningful to the participant (Wright, 2005). Therefore, all fillers were designed to be as similar as possible to target pictures, comprising the same composition and components – see Table 3.3. for examples. The positioning of the target item was pseudo-randomised so that the target item appeared in each quadrant nine times (see Table 3.3.). To do this a random 36 item number sequence containing the numbers 1 to 4 was created using a random number generator.

Table 3.3. Examples of Target pictures and related fillers

	Test Item
Test Sentence: *The snake had bitten Holly's ankle* **Target:** *Top Left Quadrant*	
Test Sentence: *The shark had bitten Holly's arm* **Target:** *Top Left Quadrant*	
Test Sentence: *The leaves on the tree were red* **Target:** *Bottom Left Quadrant* **Real-world Alternative:** *Top Right Quadrant*	

3.1.1.4. Additional considerations

3.1.1.4.1. Comprehension Questions

In addition to embedding the critical sentences into story text, two comprehension questions were created to follow each story. Consistent with previous research, this was to encourage participants to engage with the stories and read carefully with the goal of understanding the text (e.g. Calvo et al., 2006; Long, Oppy, & Seely, 1994). One of the comprehension questions required knowledge explicitly stated in the text. The other comprehension question required an emotional inference; emotional inference questions were used as these inferences had not been assessed in any other part of the task. Emotional inferences refer to an emotional response of the character – e.g. when reading 'Billy is being bullied' a reader may infer that Billy is feeling sad. Emotional inferences fall into the coherence inference category and thus should be drawn whilst reading. However, emotional inferences are often linked to global coherence and thus often relate

89

to comprehension of the whole story. In addition, emotional inferences were selected since this inference was applicable to all stories and unlike many of the other coherence inferences did not overlap with the causal coherence inferences being assessed. The location in the passage to which the question pertained was varied to encourage deep level processing of the whole text.

The comprehension questions were presented one at a time using a true or false format, similar to that used by Long et al. (1994) and Till et al., (1988). Participants were instructed to press the key ('k') with a ✅ sticker on it if they thought the answer was true and the key ('s') with a ❌ sticker on it if they thought the answer was false. Examples of the comprehension questions used can be found in Table 4.3. The answer to the comprehension questions was true half of the time. True or false response methods, as opposed to free recall comprehension questions, are particularly beneficial in the IST as it means all elements of the task can be presented and responded to on the laptop. The physical act of task switching is reduced, meaning the task can run more efficiently.

3.1.1.4.2. Instructions

Research suggests that reading aloud disrupts the natural reading process, leading to the text being processed at a surface-level only (e.g. Keefe & McDaniel, 199). To promote a natural reading process participants were asked to read silently. Full instructions can be found in Section 4.2.1.4.

3.1.1.4.3. Practice Trials

Participants were presented with 20 practice trials to allow them to become familiar with the coloured button response method. In these practice trials a shape appeared in one of the coloured quadrants; participants had to indicate, using the coloured buttons, where the shape had appeared as fast as possible - see section 4.1.1 for further details on response method. Ford and Milosky (2008) used a similar practice procedure when using a FCPST. If the researcher felt participants were struggling with the response method during practice trials, further instruction was given.

The task was presented on a HP ProBook Laptop using E-Prime software (Psychology Software Tools, Pittsburugh, PA) to present stimuli and record number of errors and reading speed (recorded in milliseconds).

3.1.2. Current Study

To evaluate the validity of the newly designed test stimuli, Stage 1 of this study explored the critical text's ability to elicit the type of processing expected. Stage 2 explored whether pictorial stimuli depicted the idea intended, ease of interpretation of pictures, and plausibility of ideas depicted. Since this study is concerned with the evaluation of test stimuli, as opposed to comparing different age groups, it was only necessary to use a child and an adult group.

3.2. Stage 1: Evaluation of Inferences Evoked by the Critical Text

Research shows that inferential skill and time-course vary depending on the type of inferential processing required (e.g. Bowyer-Crane & Snowling, 2010; Calvo et al., 2006; Virtue et al., 2008). To explore these patterns, the critical texts were carefully designed to evoke coherence-inference, elaborative-inference, or literal processing. Participants were asked to indicate the type of processing they engaged in when presented with the critical sentences intended for use in the IST. This was assessed using a difficulty scale. In accordance with Graesser et al.'s (1994) definitions of different inference types, a difficulty score of 1 (very easy) referred to literal processing, a difficulty score of 2 (easy) referred to coherence inferential processing, and a difficulty score of 3 (medium) referred to elaborative inferential processing. A difficulty score of 4 (hard) was also used, which reflected no or a limited connection between the critical sentences and test sentence. If critical sentences evoke the type of processing intended, then they should consistently receive the associated rating. It was predicted that all critical sentences would receive the expected rating at least 80% of the time (this cut off is in line with Calvo & colleagues' criteria for judging the validity of inference-inducing sentences; Calvo, 2000; Calvo et al., 1999; Calvo, Castillo & Schmalhofer, 2006).

3.2.1. Method

3.2.1.1. Participants

10 adults (mean age = 24.90 years, SD = 7.06, range = 18 – 40 years, 5 females) and 10 CYP (mean age = 9.60 years, SD = 0.84, range = 9-11 years, 6 females) were recruited. Adults were recruited from Sheffield Hallam University undergraduate and postgraduate

courses and the researcher's local area. All undergraduate students received research credits for their participation. CYP were recruited from sports groups at a local leisure centre. All participants were native speakers of English and did not have a known learning difficulty. All participants had normal or corrected to normal vision (e.g. contact lenses or glasses) and none were colour blind.

3.2.1.2. Design and Materials

3.2.1.2.1. Critical Text Booklet

A 36 item paper-based booklet was created to assess the type of inference evoked by each critical text. Each item consisted of the critical text followed by the expected inference (test sentence). 12 items reflected those critical sentences designed to evoke coherence inferences, 12 items reflected those critical sentences designed to evoke elaborative inferences, and 12 items reflected those critical sentences used in the literal stories. See Table 3.4. for examples and Appendix 3.3 for a full list of critical texts and test sentences. The order in which sentences were presented in the questionnaire was randomised using a random number generator. All participants received the sentences in the same order.

Table 3.4. Example Critical Text and Test Sentences for Literal, Coherence and Elaborative Conditions

	Literal	**Coherence**	**Elaborative**
Critical Text	*The fish ran across the green grass, leaving a trail of pink sticky slime as he went.*	*Holly planted a tiny seed. A century later an oak tree stood in its place.*	*Whilst shaped like a human hand, it was like a bird's wing in every other way possible.*
Test Sentence	The wet fish ran across the grass.	The seed had grown into a huge oak tree.	The man's hands were covered in feathers.

For each item participants rated how easy it was to make the connection between the critical sentence and the test sentence (a processing rating) on a 4-point likert scale, where 1 = very easy and 4 = hard. Definitions for each point on the scale were given at the start – see Table 3.5.

Table 3.5. Inference Ratings and Definitions

Rating	Definition
1 (very easy)	The decision was very easy and I knew the answer straight away. This is because the information in the sentence was exactly the same as that stated in the preceding passage
2 (easy)	The decision was easy. The idea in the sentence was not explicitly stated in the passage, but it did come into my mind whilst I was reading the passage. This is probably because the idea in the sentence connected the ideas in the passage
3 (medium)	The decision was slightly harder. I had not come up with the same idea to that in the sentence whilst I was reading the passage. However, when I read the sentence I was able to associate the characteristics being described in the sentence to one of the objects/characters in the passage
4 (hard)	The decision was hard. I couldn't decide if the sentence was linked to the passage or not, so I had to make a guess

The first three definitions were adapted from those given by Graesser et al. (1994) and correspond to literal, coherence, and elaborative inferential processing, respectively.

The definitions were read aloud and then explained in simpler terms for the CYP. Before starting the questionnaire all participants were asked to explain to the researcher what each rating meant. Participants did not complete the task until the researcher was happy that they fully understood each definition. Answers were scored as correct if the participant selected the rating thought to be elicited by the critical text. A high accuracy rating indicates a critical sentence which reliably evoked the intended inferential (or literal) processing. A low accuracy rating indicates that a critical sentence did not evoke the intended processing.

3.2.1.3. Procedure

Written consent was also gained from all adults. The three-tiered method of consent detailed in Section 5.3.3. was used to obtain consent from all CYP – this includes written organisational, parental[6], and child consent. All participants completed the booklet individually in a quiet area (e.g. a lab room or break out space in the CYP's leisure centre/sports club). All participants were given the following instructions:

> *'In this booklet there are 36 short passages about some fictitious planets. I would like you to read these passages very carefully. The passages will be in italics. A short sentence in bold font will follow every passage. I would like you to tell me how easy or difficult it was to make the connection between the passage and sentence using the following scale.'*

The rating scale and associated definitions were then explained to participants. Once the researcher was sure the participant understood the rating scale and definitions, an example was given. Participants then completed the booklet. The researcher was always close by in case the CYP needed any help reading the sentences. Adults took approximately 15-20 minutes and CYP took approximately 20-25 minutes to complete the questionnaire. Once participants had completed the questionnaire they were thanked and debriefed, CYP were also given a goody bag and letter of thanks for their parents.

---

[6] The term parental/parent is used in this thesis to refer to the wide range of individuals who care for CYP.

3.2.2. Results

If a critical text received the rating representing the inferential processing intended, it was scored as correct. A mean accuracy percentage was then calculated for each item. When looking at the ratings given by the children, 34 out of 36 items received the inference rating expected at least 80% of the time (see Appendix 3.4 for critical sentence accuracy ratings).

Items 13b (furry fish) and 15b (running fish) were designed to evoke literal processing, however, item 13b was only rated as requiring literal processing 50% of the time, and item 15b, 40% of the time. It was thought that this was due to the wording of the test sentence as opposed to the wording of the critical text. Both items 13b and 15b are counterfactual-world items, but the test sentence refers to 'this planet' (e.g. *'The fish on this planet are covered in fur'*). This phrase was meant to refer to the planet upon which the *'furry fish/running fish'* lived. However, CYP may have thought the sentence was referring to Earth, causing confusion and a more difficult decision-making process. The wording of the test sentences in these items was changed to remove the ambiguity caused by the use of *'this planet'* before the booklet was given to adults, all other items remained the same in the booklet given to adults and children. The newly modified item 13b and 15b were given a rating of 1 (the expected rating associated with literal processing) at least 80% of the time by adults. Inspection of individual item ratings given by adults found that all items received the processing rating expected at least 80% of the time (see Appendix 3.4 for a full list of percentage ratings).

After removing responses to items 13b and 15b (5.56% of the child data) from the children's data set a participant mean accuracy percentage was calculated for both adults and children for each condition. As can be seen from Table 3.6. the mean accuracy percentage was 80% or over for all conditions. This suggests that for all conditions the critical texts were consistently rated as evoking the inferential processing intended. Therefore, the critical sentences appear suitable for use in the IST.

Table 3.6. Mean Critical Text Accuracy Percentages (and Standard Deviations) across All Conditions for Both Adults and Children

	Real-World Information			Counterfactual-World Information		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Children	82.00	91.67	83.33	86.00	91.67	86.67
	(4.47)	(7.53)	(5.16)	(8.95)	(7.53)	(8.17)
Adults	88.33	91.67	90.00	88.33	88.33	88.33
	(8.37)	(7.53)	(8.94)	(8.94)	(9.83)	(7.53)

3.3. Stage 2: Evaluation of the Interpretation, Processing Difficulty, and Plausibility of Pictorial Stimuli

As discussed in Section 2.3.3., research suggests pictures, as opposed to text, allow direct access to meaning and may be more suitable for the assessment of a range of age groups. However, this is only the case if the pictures adequately depict what they are supposed to and are easy to interpret. Stage 2 of the study evaluated whether the pictures:

- depicted the idea intended,
- were easy to interpret,
- in the case of real-world target items (and real-world alternative filler items in counterfactual-world conditions), depicted plausible ideas,
- in the case of counterfactual-world target items, depicted implausible ideas.

To determine if the pictures depicted the ideas intended participants were asked to describe the picture. Participants then rated how difficult it was to make this decision and finally participants rated the plausibility of the idea depicted by each picture. If a picture reliably depicted the idea intended, then the picture should consistently be identified correctly. It was predicted that all of the pictures intended for use on the IST would be identified correctly at least 80% of the time, in accordance with cut off rates used by Calvo and colleagues (Calvo, 2000; Calvo et al., 1999; 2006). Since each picture was only rated by five adults and five children, 80% is the equivalent of just one adult or one child identifying the picture incorrectly, creating a very stringent criterion.

A difficulty rating scale ranging from 1 = extremely easy to 5 = extremely hard was used to obtain a measure of picture difficulty. It was predicted that all pictures intended for use in the IST would receive a mean difficulty rating of 2 or less. A plausibility rating scale ranging from 1 = impossible to 4 = very likely was used to obtain a measure of picture plausibility. It was predicted that all real-world target pictures would receive a mean

plausibility rating of 3 or more and all counterfactual-world target pictures would receive a mean plausibility rating of 2 or less. It was predicted that the real-world alternatives to the counterfactual-world target pictures would also receive a mean rating of 3 or more.

3.3.1. Method

3.3.1.1. Participants

20 adults (mean age = 33.55 years, SD = 13.88, range = 18-65 years; 11 females) and 20 CYP (mean age = 9.90 years, SD = 1.07, range = 9-12 years; females) were recruited. Adults were recruited from Sheffield Hallam University undergraduate and postgraduate courses and the surrounding area. All undergraduate students received research credits for their participation. CYP were recruited from sports groups at a local leisure centre. All participants were native speakers of English and did not have a known learning difficulty. All participants had normal or corrected to normal vision and none were colour blind. None of the participants recruited for Stage 2 had taken part in Stage 1 of this study.

3.3.1.2. Materials and Design

3.3.1.2.1. Picture Validation Booklet

Four 36-item booklets were created to evaluate the 144 pictures designed by the researcher for use in the IST. The 144 pictures were randomly divided across the four booklets to prevent fatigue. The distribution of the four booklets was counterbalanced across participants.

The picture validation booklet was designed to assess whether or not the pictures depicted what they were supposed to, picture difficulty, and picture plausibility, thus each item consisted of three parts. First, using an open-answer method, participants stated what they thought the picture was. Participants then rated picture difficulty on a 5 point likert scale, where 1 = very easy and 5 = very difficult[7] - see Table 3.7. for definitions of each rating.

---

[7] Participants were not told what the picture was supposed to be before making their difficulty rating as if the participant did not identify the picture correctly this data was removed from analysis.

Table 3.7. Picture Difficulty Ratings and Definitions

Rating	Definition
1 (very easy)	I knew what the picture was straight away
2 (easy)	I knew what the picture was almost immediately
3 (medium)	I wasn't sure at first but then after a few moments I realised what the picture was
4 (hard)	I had to study the picture for quite a while before I could work out what it was meant to be
5 (extremely hard)	I had to study the picture for a long time and even after looking for so long I still had little idea. I just had to guess

Finally, participants were asked to rate picture plausibility on a 4 point likert scale where 1 = impossible and 4 = very likely – see Table 3.8. for definitions for each rating.

Table 3.8. Picture Plausibility Ratings and Definitions

Rating	Definition
1 (impossible)	'It would be impossible to see the idea in the picture in real-life
2 (unlikely)	'It would be unlikely to see the idea in the picture in real-life,'
3 (likely)	'It would be likely to see the idea in the picture in real-life,'
4 (very likely)	'It would be very likely to see the idea in the picture in real-life.'

3.3.1.3. Procedure

Written organisational, parental, and child consent were gained for all CYP. Written consent was also gained from all adults. All participants completed the picture validation booklet individually in a quiet area. All participants were given the following instructions:

> *You're going to see some pictures. Next to each picture is a dotted line. I would like you to write down what you think each picture is on this line. I would then like you to tell me how easy it was to decide what each picture was using the scale below. After you've done this, I would then like you to tell me how likely you think it is that you would see the idea in the picture in real-life using the scale below.*

The rating scales and associated definitions were then explained to participants, with each being followed by an example. Once the participant had confirmed that they understood what to do, they completed the booklet. In most cases, adults and children completed the booklet themselves. However, some children who were conscious of their handwriting or spelling asked to dictate and let the researcher fill the booklet in for them. Upon completion of the booklet, the researcher quickly scanned the booklet, if any pictures had been incorrectly identified the researcher explained what the picture was meant to be and asked the participant how they thought the picture could be improved. No changes were made at this point. The booklet took adults 15-20 minutes to complete and children 20-30minutes to complete. Once participants had completed the booklet they were thanked and debriefed, children were also given a goody bag and letter of thanks and debrief to take home to their parents.

3.3.2. Results

3.3.2.1. Accuracy of Picture Depiction

All answers given by participants were scored as correct/incorrect by the researcher and a second scorer (psychology graduate) using a set of strict criteria. Table 3.9. provides examples of the criteria used and some examples of both correct and incorrect answers given by participants. Inter-rater reliability between the researcher and the second scorer was found to be very high, with 97.08% agreement between the two scorers for children and 98.75% agreement for adults.

Table 3.9. Examples of criteria used to score picture identification answers and the types of answers marked as correct/incorrect

Item	Criteria	Correct Answers	Incorrect Answers
Item 1a - target	Target: Bear hatching from egg.	Bear hatching	Bear wearing egg
	Participant must make it clear that the bear is a product of/coming from the egg	Bear coming out of egg	
Item 3b - target	Target: A hand covered in feathers.	Feathery hand,	Hand and feathers,
	Participant must make clear that the feathers and the hand were connected	Hand covered in feathers,	Glove made of feathers
		Hand made of feathers,	
		Hand growing feathers	
Item 11b - filler	Target: Penguin rowing in a boat	Penguin rowing boat	Penguin and a boat
			Penguin moving
	Participant must indicate that the penguin is rowing a boat. Any variation on rowing can be accepted - e.g., sailing. Answer can still be marked as correct if participant refers to the penguin as a bird. It is central that participant makes reference to both mode of transport and animal.	Penguin sailing	Penguin travelling
		Penguin boating in the sea	

To explore whether or not the pictures created adequately depicted the idea intended, a mean correct response percentage was calculated for each picture[8]. Inspection of these percentages shows that for children, 136 of the 144 pictures received a mean correct response percentage of 80% or above. The target picture for item 4a (Holly and Ben walking on water/across the lake), all four pictures for item 10b (Red poppies, Blue

---

[8] the researcher's own marks were used not the seconder marker's

poppies, Orange poppies, Yellow poppies) and the target and two fillers for item 14a (Hard rock, Bendy rock, Cracked rock) did not receive a mean correct response percentage of 80% or above. For adults, 138 of the 144 pictures received a mean correct response percentage of 80% of above. The fillers for item 10b (Blue poppies, Orange poppies, Yellow poppies) and the target and two fillers for item 14a (Hard rock, Bendy rock, Cracked rock) did not receive a mean correct response percentage of 80% or above (see Appendix 3.5 for accuracy percentages for every picture). Several changes, all listed below, were made to rectify these issues.

The target picture for item 4a was edited so that the waves in the lake were blue - more clearly representing water. This picture was rated again by five children who had not previously participated in the study (mean age = 9.20 years, SD = 0.45, 3 females; these new participants were also asked to provide a difficulty and plausibility rating for the new picture). This edited picture received a 100% identification ratimg from the five new participants. Water was represented in this way (i.e. blue) in all relevant pictures to improve clarity and consistency across items.

Item 10b was changed from *'The enormous field ahead of them was filled to the brim with beautiful poppies.'* to *'The Blib did not use a pencil to draw Ben. He used a small stick covered in honey.[9]'* Consequently, the intended inference went from '*poppies are red*' to '*honey is golden yellow.*' Four new pictures were created for this new item. These four pictures were then rated again by five children (mean age = 9.60 years, SD = 0.89, 3 females) and five adults (mean age = 25.80 years, SD = 11.71, 1 female), none of which had previously participated in the study (these new participants were also asked to provide a difficulty and plausibility rating for the new pictures). The new children identified the target correctly 80% of the time, filler 1 correctly 100% of the time, filler 2 correctly 100% of the time, and filler 3 correctly 80% of the time. The new adults correctly identified the target 100% of the time, filler 1 was correctly identified 80% of the time, filler 2 100% of the time, and filler 3 was correctly identified 100% of the time.

Item 14a was changed from '*However, they were trapped. There was no way out. The thick walls of rock were completely solid.*' to '*The people in this place wore tops made*

---

[9] These item was also evaluated to ensure they evoked the target inference using the methodology outlined in Stage 1.

from elastic. Elastic is quite stretchy[10].' Consequently, the intended inference went from *'rocks are hard and inflexible'* to *'elastic is flexible and stretchy.'* Four new pictures were created for this new item. These four pictures were then rated again by five children (mean age = 9.40 years, SD = 0.55, 2 females) and five adults (mean age = 22.40 years, SD = 3.58, 3 females), none of which had previously participated in the study (these new participants were also asked to provide a difficulty and plausibility rating for the new pictures). The new children identified the target correctly 100% of the time, filler 1 was identified correctly 100% of the time, filler 2 80% of the time, and filler 3 was also identified correctly 80% of the time. The new adults correctly identified the target 100% of the time, filler 1 80% of the time, filler 2 80% of the time, and correctly identified filler 3 100% of the time.

Using the new stimuli, mean correct response percentages were calculated for adults and children for each condition. These mean correct response percentages can be found in Table 3.10.

Table 3.10. Mean correct response percentages (and Standard Deviations) across All Conditions for Adults and Children

	Real-World Items			Counterfactual-world Items		
	Literal	**Coherence**	**Elaborative**	**Literal**	**Coherence**	**Elaborative**
Children	95.83	95.00	96.67	98.33	99.17	95.83
	(8.30)	(8.85)	(7.61)	(5.65)	(5.65)	(8.30)
Adults	95.83	97.50	95.83	98.33	94.17	95.83
	(8.30)	(6.76)	(8.30)	(5.65)	(9.29)	(8.30)

As can be seen from Table 3.10. in all conditions, the mean percentage accuracy rating is very high (above 90%) for both adults and children. This suggests that the pictures intended for use in the IST reliably convey the ideas intended. Overall, it appears that after some modifications, all pictures adequately depicted the intended ideas and are suitable for use in the IST.

---

[10] These items were also evaluated to ensure they evoked the target inference using the methodology outlined in Stage 1.

3.3.2.2. Picture Difficulty and Plausibility Ratings

3.3.2.2.1. Preliminary Analysis

The data were inspected prior to analysis and those pictures identified incorrectly were removed (this resulted in 2.92% of child data and 3.89% of adult data being removed from analysis), meaning only difficulty and plausibility ratings of those pictures correctly identified were used to calculate means. To ensure that pictures in all conditions were of equal difficulty, mean difficulty ratings for each condition were calculated. Mean difficulty ratings comprised all pictures a participant rated within that condition - i.e. both targets and fillers – and can be found in Table 3.11. To ensure real-world target pictures depicted plausible ideas and counterfactual-world target pictures depicted implausible ideas mean plausibility ratings were calculated for each condition. Mean plausibility ratings comprised the plausibility ratings for target items only and can be found in Table 3.12.

There were no outliers in the data and homogeneity of variance was not violated, such that the largest standard deviation was less than four times larger than the smallest standard deviation (Howell, 1987). However, for difficulty ratings, the distribution for all conditions was positively skewed - this was expected since it was hoped that most pictures would receive fairly low ratings. For plausibility ratings, the distribution was positively skewed for counterfactual-world conditions and negatively skewed for real-world conditions. Again, this was expected since it was hoped that counterfactual-world items would receive low scores, indicative of implausibility, and that real-world alternative filler pictures would receive high scores, indicative of plausibility. ANOVA is believed to be robust enough to deal with skewed distributions of data, even when small samples are used (Field, 2013). Even though Likert scales produce ordinal data, it is argued that once items are summed or averaged across the scale the data becomes interval (Carifio & Perla, 2008). Parametric ANOVAs were conducted on all difficulty and plausibility data obtained in Stage 2 of the study.

3.3.2.2.2. Picture Difficulty Ratings

First, mean difficulty ratings were calculated for each picture. Inspection of these means showed that mean ratings for all pictures were less than 2 for the children. Since a score of 1 equals very easy and 2 = easy this suggests children found all pictures easy to interpret. However, the target picture of item 3a (man flying over cliff) received a mean rating of 2.33 from the adults (see Appendix 3.6 for mean difficulty ratings for each item).

104

Consequently, lines to indicate movement were added to this picture. It was then rated again by five adults (mean age = 21.20 years, SD = 3.35, 4 females) who had not previously taken part in the study. The new adult participants gave the edited picture a mean rating of 1.6. Ratings for the amended stimuli were used when calculating a mean difficulty rating for each condition. As can be seen from Table 3.11. the mean difficulty ratings in all conditions, for both adults and children, are less than 2. Therefore, it seems all pictures, regardless of condition, are easy to interpret.

Table 3.11. Mean Difficulty Ratings (and Standard Deviations) across All Conditions for Both Children and Adults

	Real-World Items			Counterfactual-World Items		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Children	1.35	1.37	1.57	1.42	1.47	1.42
	(0.57)	(0.50)	(0.69)	(0.55)	(0.61)	(0.57)
Adults	1.43	1.52	1.58	1.50	1.46	1.50
	(0.36)	(0.25)	(0.41)	(0.36)	(0.45)	(0.40)

A 2 (Age: adults, children) * 2 (plausibility: real-world, counterfactual-world) * 3 (inference type: literal, causal, elaborative) mixed ANOVA was conducted. No significant effects of age ($F_{(1,38)} = 0.53$, $p = 0.47$, $\eta p^2 = 0.01$), plausibility ($F_{(1, 38)} = 0.003$, $p = 0.96$, $\eta p^2 < 0.001$), or inference type ($F_{(2,76)} = 2.167$, $p = 0.12$, $\eta p^2 = 0.05$) were found. There were also no significant interactions ($p > 0.05$ for all interactions). This suggests that there is no difference in picture difficulty across conditions. Overall, it appears that all pictures are easy to interpret with no differences in picture difficulty between conditions or between adults and children observed.

3.3.2.2.3. Picture Plausibility Ratings

First, mean plausibility ratings were calculated for each target picture. The definitions assigned to the ratings mean that a score of 3 or more reflects a plausible idea, whereas a score of 2 or less reflects an implausible idea. Inspection of the plausibility ratings given to individual pictures shows that all real-world target items received a mean plausibility rating of 3 or more and all counterfactual-world target items received a mean plausibility rating of 2 or less from adults (see Appendix 3. for individual plausibility ratings for each picture). However, this is not the case for children. Whilst all counterfactual-world target items received a mean plausibility rating of 2 or less, three real-world items received a mean plausibility rating of less than 3: item 8a (shark biting Holly), item 16a (Holly and

Ben walking on the road) and item 18a (shark biting Holly's foot). This suggests that children view the ideas depicted in these pictures as implausible, however, exploration of the explanations given by children who gave these pictures a score of 2 or less suggests this may not be the case.

Items 8a and 18a were rated as unlikely by some children as they felt that they personally were unlikely to see a shark or snake biting someone in England. However, the children did acknowledge that being bitten by a shark or snake was plausible, they just felt that they were unlikely to see it. Item 16a was also rated as unlikely by some children as they felt that they would be unlikely to see people walking/running in the middle of the road as 'this is dangerous'. The answers given may be due to the wording of the question - 'how likely would *you* be to see…'. When answering, some children appear to have taken an idiosyncratic approach - i.e. 'how likely is it that *I* would see this?' Since all three items were acknowledged as plausible, even if unlikely to be witnessed by the children, these items were not changed. Overall, it appears that all real-world target pictures and all real-world alternative filler pictures depict plausible ideas whereas all counterfactual-world target pictures depict implausible ideas.

Mean plausibility ratings were then calculated for each condition. As can be seen from Table 3.12. mean plausibility ratings for the real-world conditions are above 3 and plausibility ratings for counterfactual-world conditions are below 2, for both adults and children. This suggests that the real-world pictures depict plausible ideas, whereas the counterfactual-world pictures depict implausible ideas.

Table 3.12. Mean Plausibility Ratings (and Standard Deviations) across All Conditions for Children and Adults

	Real-World Items			Counterfactual-World Items		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Children	3.37	3.21	3.30	1.46	1.37	1.58
	(0.62)	(0.77)	(0.65)	(0.51)	(0.49)	(0.50)
Adults	3.64	3.55	3.32	1.18	1.27	1.55
	(0.58)	(0.80)	(0.95)	(0.40)	(0.55)	(0.91)

A 2 (Age: adults, children) * 2 (plausibility: real-world, counterfactual-world) * 3 (inference type: literal, causal, elaborative) mixed ANOVA was conducted to explore any differences in the plausibility of ideas depicted by real-world and counterfactual-world targets. A main effect of plausibility was found ($F(1,38) = 621.01, p < 0.001, \eta p^2 = 0.94$)

with real-world items (mean = 3.35, SD = 0.22) receiving significantly higher plausibility ratings than counterfactual-world items (mean = 1.41, SD = 0.20). No significant effect of inference type (F (2,76) = 0.18, p = 0.84, ηp^2 = 0.005) or age (F (1,38) = 0.08, p = 0.79, ηp^2 < 0.01) was found. There were no significant interactions observed (p > 0.05 for all interactions). It appears that real-world target items are more plausible than counterfactual-world items for both adults and children.

Finally, counterfactual-world items were designed to include the counterfactual-world target and three fillers, with one filler depicting the real-world alternative to the counterfactual-world target. To ensure these real-world alternatives depict plausible ideas subsidiary analysis was conducted to explore their plausibility. Mean plausibility ratings were calculated for each real-world alternative picture. Inspection of these means found that all pictures received mean plausibility ratings above 3 from adults and children. An overall mean was then calculated for real-world alternative pictures, real-world target items, and counterfactual-world target items. A 2 (Age: Adults, Children) * 3 (Picture Type: Real-world Target, Counterfactual-world Target, Real-world Alternative) mixed ANOVA was conducted to explore any differences in the plausibility of ideas depicted by real-world targets, counterfactual-world targets, and real-world alternatives. A main effect of picture type was observed (F (2,76) = 519.02. p < .001, ηp^2 = 0.93). Post-hoc pairwise comparisons (conducted using a Bonferroni adjustment (0.05/3)) revealed that there was no significant difference between the plausibility ratings for real-world alternative (mean = 3.50, SD = 0.31) and real-world target items (mean = 3.35, SD = 0.43; t (39) = 1.95, p = .058). Counterfactual-world target pictures (mean = 1.41, SD = 0.32) were found to receive significantly lower plausibility ratings than both real-world target (t (39) = 34.46, p < .001) and real-world alternative pictures (t (39) = 24.21, p < .001). There was no main effect of age (F (1,38) = 0.64, p = .43, ηp^2 = 0.02). There was found to be no significant interaction between picture type and age (F (2,76) = 2.27, p = .11, ηp^2 = .056). This suggests that real-world alternative pictures, like real-world target pictures, depict plausible ideas. Overall, real-world target pictures and real-world alternatives appear to depict plausible ideas whereas counterfactual-world target pictures appear to depict implausible ideas.

3.4. Discussion

All critical sentences and pictures appear to be suitable for use in the IST. Critical texts were found to elicit the type of processing intended. Pictures were found to depict the

ideas intended, be easy to interpret, and in the case of real-world targets, depict plausible ideas but in the case of counterfactual-world targets depict implausible ideas. Whilst, all critical texts were found to elicit the type of processing intended, some minor modifications had to be made to the critical text booklet. Moreover, it was after some modifications that all pictures were also found to adequately depict the idea intended. In most cases, the modifications made were minor, for example, adding blue lines to more clearly depict water. However, two items had to be replaced with new critical texts being created.

All pictures were found to be easy to interpret, with no differences found between conditions or adults and children. All real-world target pictures (and real-world alternatives, a special filler used in the counterfactual-world conditions) were found to depict plausible ideas. Conversely, all counterfactual-world pictures were found to depict implausible ideas. Some children approached the task in a very idiosyncratic manner, such that three real-world target pictures were rated as implausible by some children, since they felt that they personally would never see, for instance, a shark attack. Overall, it seems the newly designed stimuli is fit for purpose and should allow for a valid assessment of real-world and counterfactual-world inference generation abilities.

CHAPTER 4 AN EVALUATION OF THE IMAGE SELECTION TASK

4.1. Introduction

This Chapter presents initial findings from a new inference generation assessment tool, the Image Selection Task (IST), which was developed to assess real-world and counterfactual-world inference generation abilities in a range of age groups. To determine the validity of the IST, as a measure of inference generation, patterns of data from the IST are compared with patterns observed on an already widely used measure of inference generation, the Sentence Verification Task (SVT) to determine if the IST results in similar time-course and skill patterns to the SVT.

4.1.1. Designing the Image Selection Task

The Image Selection Task (IST) combines a self-paced reading (SPR) methodology and forced-choice picture-selection task (FCPST). The IST is a computer-based assessment that comprises 18 short stories (195-205 words long). Each story was split into three sections, with the first two sections of each story ending with a critical text (inference-evoking/explicit). Reading speeds of the critical text were recorded to provide a measure of the inferential time-course. Each critical text was followed by a 2*2 FCPST, in which the target and three related fillers were presented. One of the fillers depicted explicitly stated information, to tap a common reason for inferential failure – literal processing. In the counterfactual-world conditions, one of the fillers also depicted the real-world alternative, to provide a measure of the real-world interference experienced when processing counterfactual-world information. Number of errors was recorded to provide a measure of inference skill. Figure 4.1. provides a schematic representation of the IST.

The IST was designed to capitalise on the strengths of existing measures. To allow for the assessment of a range of age groups, additional demands were kept to a minimum by using a simple pictorial response method to assess the inferential product. Additionally, all text was created to be readable by CYP aged 9 years or less. Two tasks were combined to allow for the efficient assessment of both inferential skill and time-course. To promote a natural reading process, critical texts were embedded into stories and a SPR methodology was employed to assess the inferential time-course whilst reading, with minimal disruption. Finally, the use of diagnostic, as opposed to random, filler pictures allowed for an error analysis to be conducted, providing insight into the real-world

interference experienced when generating counterfactual-world inferences and possible causes of inference generation failure when generating real-world inferences.

4.1.2. Evaluating the IST

The IST is based on similar tasks used by Ford and Milosky (2003; 2008) and Schmidt and Paris (1978). The IST, with the addition of reading speeds and more complex stimuli, is, a novel paradigm, however, and has not previously been used to assess inference generation abilities. The validity of the IST as an inference generation measure needed to be ascertained. Typically, the validity of a comprehension measure is determined by:

- exploring whether the measure taping those skills underpinning comprehension correlates with and/or predicts variance in a standardised reading comprehension measure (predictive/concurrent validity; Marcotte and Hintze, 2009; Shapiro, Fritschmann, Thomas, Hughes, & McDougal, 2014), or
- comparing performance on the target measure with an already standardised measure (construct validity; e.g. Carlson et al., 2014; Reed & Vaughn, 2012).

The primary aim of this Chapter was to ascertain whether the IST is able to measure inference generation (construct validity). Therefore, the patterns obtained from the IST were compared with patterns obtained on a Sentence Verification Task (SVT). The SVT was used as it has previously been employed to assess real-world and counterfactual-world inference generation skill and time-course in adults and children (Bowyer-Crane & Snowling, 2010; Graesser et al., 1998).

To be suitable for the assessment of real-world and counterfactual-world inference generation abilities during adolescence the IST must also meet the following criteria:

1. Assess the inference generation skills of CYP aged 9-10 years to adults
   a. The IST should not result in ceiling or floor effects
   b. The IST must be simple to use and all instructions clear
2. Assess the real-world interference thought to be experienced when processing counterfactual-world information.

4.1.3. The Current Study

The primary aim of the current study was to evaluate the newly designed IST to determine if the IST is a valid measure of inference generation abilities. Patterns of time-course

(reading speed) and skill (number of errors) data obtained from the IST were compared with patterns of time-course (response time) and skill (number of errors) data obtained on the SVT. The IST was designed to be suitable for the assessment of a wide range of age groups, specifically CYP aged 9-10 years to adults. For efficiency, two age groups were assessed in the current study 9-10 year olds and adults, such that if the IST is suitable for these two groups, it should also be suitable for those intermediate age groups. Table 4.1. presents predictions made for each task.[11]

---

[11] The list of predictions in Table 4.1 is exhaustive. However, this is due to the complexity of the two tasks used, such that faster response speeds on the SVT would be indicative of an online inference whereas slower reading speeds on the IST would be indicative of an online inference.

Table 4.1. Predictions for Time-Course and Skill

	Time-Course
Real-World Coherence	Coherence items will be responded to just as fast as literal items on the SVT, but read significantly slower than literal items on the IST, indicating coherence inferences are generated whilst reading by adults and children
Real-World Elaborative	Elaborative items will be responded to significantly slower than literal items on the SVT, but read just as fast as literal items on the IST, indicating elaborative inferences are generated after reading with strategic effort by adults and children
Counterfactual-World Coherence	Children will respond to coherence items just as fast as literal items on the SVT, but read coherence items significantly slower than literal items on the IST, indicating coherence inferences are generated whilst reading Adults will respond to coherence items significantly slower than literal items on the SVT, but read coherence items just as fast as literal items on the IST, indicating coherence inferences are generated after reading, with strategic effort
Counterfactual-World Elaborative	Adults and children will respond to elaborative items significantly slower than literal items on the SVT, but read elaborative items just as fast as literal items on the IST, indicating elaborative inferences are generated after reading, with strategic effort
Counterfactual-World vs Real-World	Adults and children will respond to counterfactual-world items significantly slower than real-world items on the SVT and read counterfactual-world items significantly slower than real-world items on the IST, indicating that the processing of counterfactual-world information is more demanding than the processing of real-world information

Skill
Inferential skill will be significantly better for coherence than elaborative items on both the IST and SVT
Significantly fewer real-world inference errors will be made by adults than children on both the IST and SVT
Significantly more counterfactual-world inference errors will be made by adults than children on both the IST and SVT

4.2. Stage 1 – Comparing Time-Course and Skill Patterns on the Sentence Verification Task and Image Selection Task

4.2.1. Method

4.2.1.1. Participants

Thirty adults (mean age = 20.22 years, SD = 3.26, range = 18-33 years; 21 females) and thirty children (mean age = 9.95 years, SD = 1.03, range = 9-10 years; 17 females) were recruited. Adults were recruited from Sheffield Hallam University undergraduate and postgraduate courses, Sheffield and surrounding areas, and Nottinghamshire. All undergraduate students received research credits for their participation. Children were recruited from local schools and a sports group at a local leisure centre. All participants were native speakers of English and did not have a known learning difficulty. All participants had normal or corrected to normal vision and none were colour blind. None of the participants recruited had taken part in Study 1 (Chapter 3).

4.2.1.2. Design

An experimental mixed factorial design was employed. The between-participant independent variables were age (Children, Adults) and task (SVT, IST) – see Table 4.2. for distribution of participants. The within-participant independent variables were plausibility (Real-world, Counterfactual-world) and inference type (Literal, Coherence, Elaborative). The dependent variables were number of errors made on SVT and IST, time taken to respond to the test item (measured in milliseconds) on SVT, and time taken to read critical text (measured in milliseconds) on IST.

Table 4.2. Number of Participants Completing Each Task

	Adults	Children
Image Selection Task	15	15
Sentence Verification Task	15	15

4.2.1.3. Materials

4.2.1.3.1. Image Selection Task

A detailed description of the stimuli used in the Image Selection Task (IST) can be found in Chapter 3.

The IST comprises three parts: 1) story text, 2) critical text, and 3) 2*2 forced-choice picture-selection grid. The participant moves through the task at their own pace, selecting the picture that they feel best matches the critical text. A schematic representation of the IST can be found in Figure 4.1.

Eighteen stories (all 195-210 words long; mean = 203.44, S.D = 1.98) deemed suitable in terms of content and readability were created by the researcher (see Appendix 4.1 for readability and word length scores for all stories). All 18 stories designed for the IST received a readability rating of 9 years or less (mean = 8.81, S.D = 0.18). Six stories were created to assess real-world inferential processing abilities and six stories were created to assess counterfactual-world inferential processing abilities. Six additional stories were created to assess participant's short-term memory of information explicitly stated in the text (literal stories), examples of each story type can be found in Table 4.3. All stories were designed in three sections, such that the first two sections ended with a critical text and the final section provided a story ending. All critical texts were 21 syllables in length and either evoked an inference or, in the literal stories, explicitly stated the information needed to respond to the test items correctly.
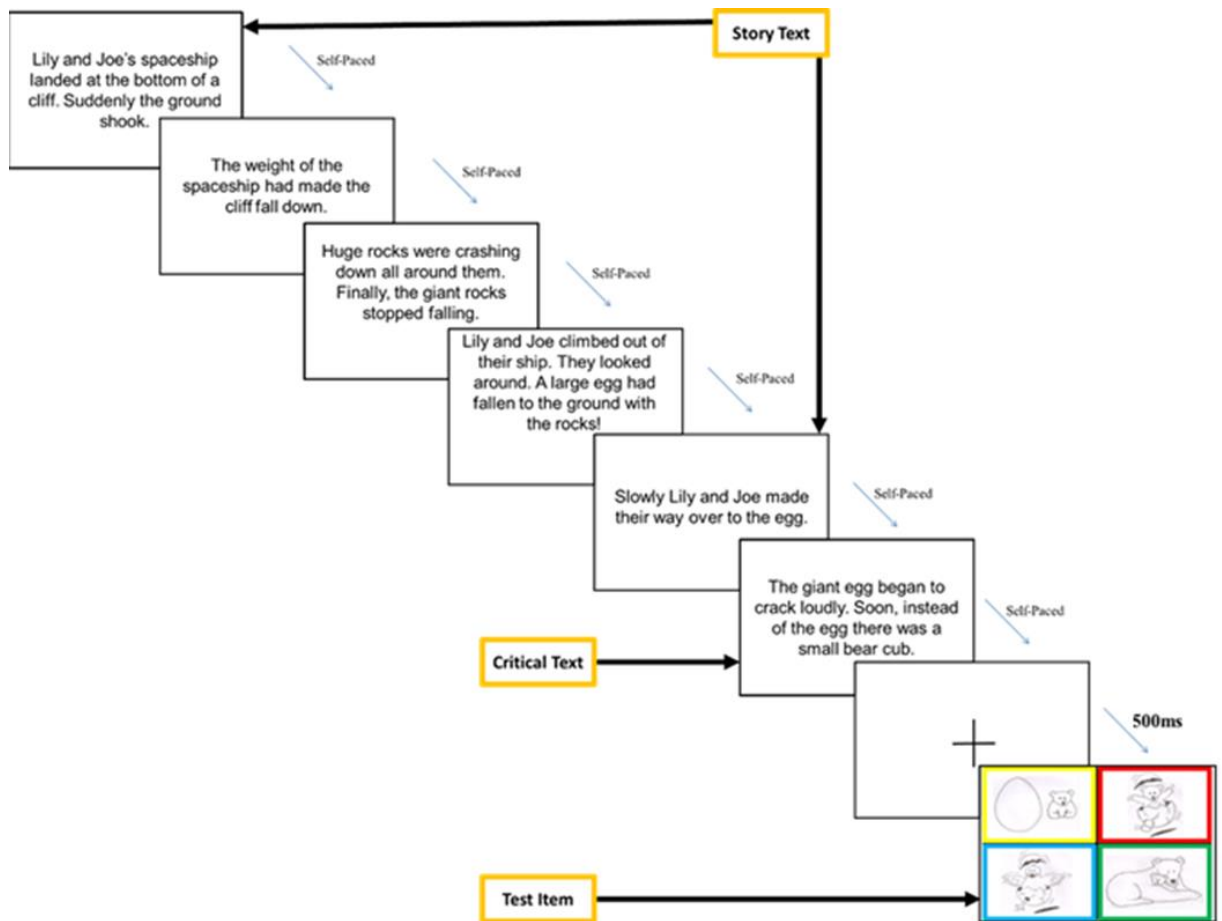
Figure 4.1. Schematic Representation of the Image Selection Task

Table 4.3. Sample Text - Real-World, Counterfactual-World and Literal Conditions

	Real-World Condition	Counterfactual-World Condition	Literal Condition
Section 1	Holly and Ben had flown into an asteroid storm. Giant rocks were crashing into the spaceship. It was going to get damaged. CRASH!! WHACK!!! A rock hit the wing. The wing fell off and landed in the ocean below. The ship was out of control. Holly and Ben had to make an emergency landing. They were able to land on a beach next to the huge ocean. There was a boat tied up next to a statue.	Lily and Joe's spaceship landed at the bottom of a cliff. Suddenly the ground shook. The weight of the spaceship had made the cliff fall down. Huge rocks were crashing down all around them. Finally, the giant rocks stopped falling. Lily and Joe climbed out of their ship. They looked around. A large egg had fallen to the ground with the rocks! Slowly Lily and Joe made their way over to the egg.	Holly and Ben landed on a planet in chaos. There had been a murder. The kind old bear that lived on a small island just outside of the city had been found dead. The island was surrounded by a fast flowing river. There was no bridge and no boats. The only way to reach the island was to swim. A slimy fish had been seen at the scene. But he had not been seen since. Holly saw the fish.
Critical Sentence	*The gigantic clay statue shone in the sun light. It must have been at least ten feet high.*	*The giant egg began to crack loudly. Soon, instead of the egg there was a small bear cub.*	*The large fish ran across the green grass. He left a trail of gunky slime as he went.*
Test Sentence	**The statue was a burnt orange colour.**	**The tiny bear cub had hatched from the egg.**	**The large fish had ran across the green grass.**
Section 2	Holly and Ben could see the wing in the ocean. However, they could also see some very evil sharks. Holly and Ben climbed into the boat. They made it to the wing in no time. Sadly, the heavy wing had started to sink. Holly and Ben could not reach the wing from the boat. One of them would have to jump in for it. Holly dove into the sea. She grabbed the wing.	Lily ran over to the tiny bear cub. He was very upset. "He must have fallen from the cliff. We have to take him back to his mum," said Joe. So Lilly and Joe set off up the cliff. The cliff was very steep. Climbing it was hard work. To make it worse it was quite a cold dark day. When the sun came out Lily and Joe saw something weird.	"I didn't do it!" cried the fish, "I can't swim." So the fish couldn't have killed the bear. But if he didn't do it who did?

Ben had an idea. Holly, Ben and the fish went to the bear's house. Ben shook the bear. It began to move. "He's not dead!" said Ben, "he was hibernating." "Well I could have told you that," said a large rhino that lived with the bear. |
Critical Sentence	*However, a shark was approaching. It attacked. Blood began to spill out of Holly's arm.*	*The tiny bear cub was quite odd. He was not covered in fur. He was covered in cheese.*	*The rhino was huge. It had huge feet. The rhino was also very very heavy.*
Test Sentence	**The evil shark had bitten Holly's arm.**	**The bear cub was a yellowish colour.**	**The rhino was very very heavy.**
Section 3	Ben pulled Holly and the wing into the boat. Ben put the wing back on the spaceship. They then flew away from the awful planet.	After a few hours Lily and Joe came to a huge nest. They put bear cub in to the nest. Lily and Joe then rushed back to their spaceship.	Holly, Ben, the fish, and the bear all returned to the city. Once there they told the police what had really happened.
Comprehension Question 1	The statue was tiny	The spaceship had caused the cliff to fall down.	The fish could swim
Comprehension Question 2	The sharks were hungry	Lily and Joe were very tired after climbing the cliff	The fish was scared of the police

Real-world and counterfactual-world stories were designed to assess coherence and elaborative inferential abilities. Subsequently, each story contained one coherence-inference-evoking critical text and one elaborative-inference-evoking critical text. Literal stories contained one critical text based on real-world information and one critical text based on counterfactual-world information. Inferential abilities were indexed by number of errors and speed at which critical text was read. The order of critical sentences within

each story was pseudo-randomised (see Table 3.2.), so that any differences in inferential performance could not be attributed to the position of the critical sentence. Additionally, two versions of the stories were written, such that the second version of stories had the converse arrangement of critical sentences. This meant that the story content had to be changed slightly. The number of words in version 1 and 2 of the story always remained the same. The general gist of the story also remained consistent, as did the readability rating (mean rating version 1 = 8.79, SD = 0.19; mean rating version 2 = 8.83, SD = 0.18; t (17) = 0.72, p = 0.48).

The stories were sorted into two programs due to two versions of each story being created to ensure any differences between different conditions could not be attributed to the position of critical sentence. Each program comprised 18 stories. Each program comprised only one version of each story, with programs organised such that each was made up of:

- 6 stories presenting the coherence-inference-evoking critical text first (3 real-world and 3 counterfactual-world),
- 6 stories presenting the elaborative-inference-evoking critical texts first (3 real-world and 3 counterfactual-world),
- 3 stories presenting the real-world literal critical texts first, and
- 3 stories presenting the counterfactual-world critical texts first

Whilst all participants received stories in the same order, the order of real-world, counterfactual-world and literal stories was randomised. The administration of programs 1 and 2 was also counterbalanced across participants, see Table 3.2. for a summary of the counterbalancing and randomisation of critical texts and stories.

A FCPST followed each critical sentence. Participants were required to select the picture that they felt best matched the preceding text as quickly and as accurately as possible. Figure 4.2. shows the format of the FCPST, as can be seen the task is comprised of four quadrants, each containing a picture. The top left quadrant has a yellow border, the top right quadrant has a red border, the bottom left quadrant has a blue border, and the bottom right quadrant has a green border. On the number pad, blue, green, yellow and red stickers were placed on keys 1, 2, 4, and 5, respectively, to mirror the quadrants on the screen. Participants used these keys to make their selection.
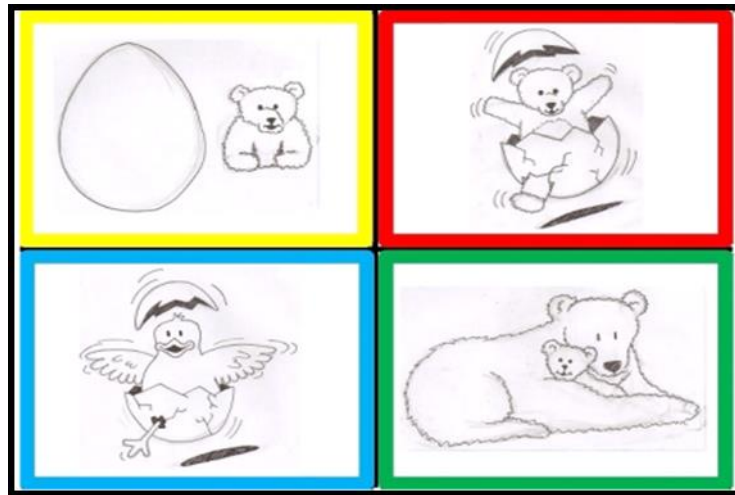
Figure 4.2. The Format of the Image Selection Task Test Items

As can be seen from Figure 4.2, the FCPST comprised four pictures: the target (the correct item which reflects the inference which could have been drawn or the information explicitly stated in the critical sentence, in the case of literal items), and three fillers.

The task was presented on a HP ProBook Laptop using E-Prime software to present stimuli and record number of errors and response time (recorded in milliseconds).

4.2.1.3.2. Sentence Verification Task

A schematic representation of the sentence verification task (SVT) can be found in Figure 4.3. The stories, critical sentences and comprehension questions used in the SVT were identical to those used in the IST. Where relevant, all of those design considerations that aimed to minimise order effects in the IST were also applied to the SVT. However, the critical text was followed by a test sentence that participants had to judge as true or false, examples of these sentences can be found in Table 4.3. The test sentences appeared in bold and two font sizes larger than the story text. If participants thought the sentence was true they were instructed to press the key ('k') with a ✅ sticker on, if they thought the sentence was false participants were instructed to press the key ('s') with a ❌ sticker on. Inference generation abilities were measured by the time taken to make this true/false judgment, response times were measured from the presentation of test stimuli to the onset of tick or cross button press by the participant. Number of errors was also recorded. The correct response in all of the inference items was always true. Thus, a TRUE/TRUE response pattern was always correct for the inference items. Therefore, in half of the literal items the correct answer was false, such that correct response pattern was

TRUE/FALSE for half of the literal trials. This is the same design used by Bowyer-Crane and Snowling (2010) and has been used to minimise the number of stories a participant has to read.
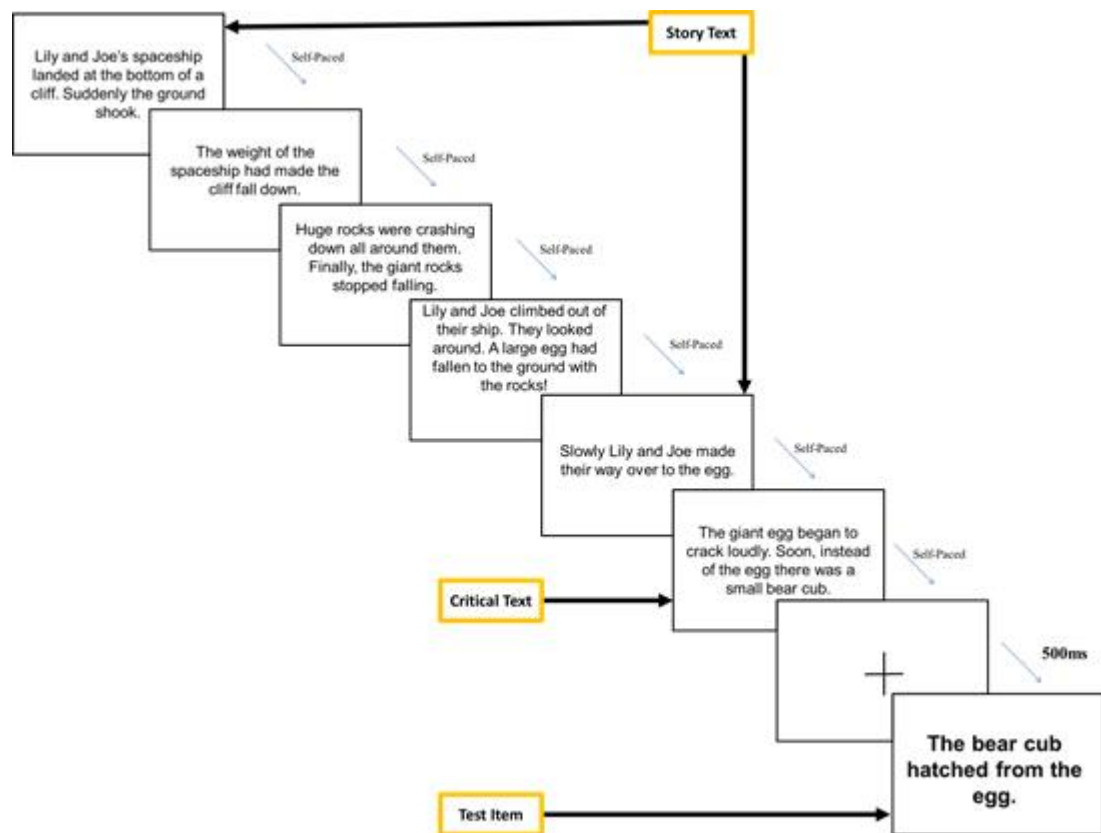
Figure 4.3. A Schematic Representation of the Sentence Verification Task

4.2.1.4. Procedure

Written consent was gained from all adults. The three-tiered method of consent, detailed in Section 5.3.3., was used to obtain consent from all children. All participants were tested individually in a quiet area. First, the researcher engaged participants in small talk to make them feel comfortable. Participants were then asked to move the laptop around on the desk so that it was in a comfortable position for them. Participants were informed that they would be helping the researcher to make sure that all of instructions were clear and made sense. Participants were encouraged to let the researcher know if anything was unclear or needed further explanation, or if they had any suggestions that would make the instructions clearer.

Once the laptop was in place all participants were given the following instructions (instructions were read aloud and appeared on the screen):

Today I would like you to read some stories about Holly and Ben. Holly and Ben are space adventurers that travel to planets all over the universe.

At the end of each story, you'll be asked two questions so you need to make sure that you're reading the stories carefully. You answer these questions using the tick and the cross keys. If you think the answer is true you press the tick and if you think the answer is false, you press the cross. Does that sound OK?

The whole story won't appear on the screen at once, instead just one or two sentences will appear on the screen. Once you have finished reading and understood the sentences on the screen just press the spacebar and you will then see the next set of sentences. Does that make sense? Please just read through these sentences at your normal reading speed. Try and read the stories in your head.

Participants completing the SVT then received the following instructions:

Every now and then you will see a sentence in bold. So this sentence will be a bit bigger than the text before it. You need to tell me if you think the sentence in bold is true or false based on the story you have just been reading. If you think the sentence is true based on the text you've just read you need to press the tick. If you think the sentence in bold is false based on the text you've just read you need to press the cross. You need to try and make this true or false decision as quickly and as accurately as possible. Does all of that make sense?

OK, please put a finger on the cross and a finger on the tick.

Conversely, participants completing the IST received the following instructions:

Every now and then, the screen will be split into four sections, with a picture appearing here, here, here, and here. You need to tell me which picture you feel best matches the story you have been reading. You need to try and make this decision as quickly

and as accurately as possible. To do this you will use these coloured buttons. If you think the picture that best matches the text you've just read is in this yellow box you press the yellow button, if you think it's in this red box you press the red button, the blue box you press the blue button, and if you think the picture that best matches the text you've just read is in the green box you press the green button. Does all of that make sense?

OK good, so put one hand over the coloured buttons and your other hand over the space bar. Press the spacebar for a little practice

Participants were then given 20 practice trials to allow them to become familiar with the coloured button response method – see Section 3.1.1.4.3. The researcher monitored participants' performance during these practice trials and gave further instruction if necessary. The researcher was confident that all participants had mastered the response method by the end of the practice trials.

The researcher ensured that all participants had understood the instructions and had a clear idea of what they had to do before starting the experiment proper. Participants worked through the story text and critical text at their own pace. Once participants had finished reading the critical text a screen with a fixation point in the centre was presented for 500ms, the test item then immediately followed this. Participants controlled the presentation of the text using the spacebar. Each press of the spacebar caused the current text to be erased and replaced with the next set of sentences.

A screen with the text 'PLEASE GET READY FOR YOUR NEXT STORY' appeared for 1000ms before each story began. After every third story, a screen would appear advising participants to take a break and have a stretch. Adults completed the IST/SVT in around 20minutes, whereas children took on average 30minutes to complete the IST/SVT. Response accuracy was not signalled to ensure the tasks provided a measure of current as opposed to learning ability. Positive encouragement was given to all participants, regardless of performance (e.g. you're doing well). Participants were thanked and debriefed. Each child was also given a goody bag, letter of thanks, and debrief letter to take home to their parents.

4.2.2. Results

4.2.2.1. Preliminary Analysis

Two versions of each story were created to ensure any differences between different conditions could not be attributed to the position of critical sentence. Therefore, story effects were investigated[12]. On the SVT, no difference in response times was found between Set 1 and Set 2 for adults (t (17) = 0.87, p = 0.40) or children (t (17) = 0.64, p = 0.53). On the IST, there was found to be no significant difference between reading speeds on Set 1 and Set 2 for adults (t (17) = .14, p =.89) or children (t (17) = .41, p = .69). For both the SVT and IST, there was found to be no difference between the number of errors made on Set 1 and Set 2 for adults (t (17) = 0.53, p = 0.59) or children (t (17) = 0.44, p = 0.67). The data from both sets of stories was, therefore, collapsed.

4.2.2.2. Assessing the Inferential Time-Course: IST Reading Speed and SVT Response Time Measures

First, the data were inspected prior to analysis and all incorrect responses were removed. Second, a data trim was conducted, with all scores 2 standard deviations above or below the participant's grand reading speed (IST) or response time (SVT) mean removed and all scores 2 standard deviations above or below each items grand reading speed (IST) or response time (SVT) mean removed. Whilst most outliers were evenly distributed across all items and participants, some items resulted in more than 50% of errors/outliers. For children, item 5b was removed from the SVT, and items 5b, 6a, and 11a were removed from the IST. For adults, items 4b, 6a, and 10b were removed from the SVT, and item 6a was removed from the IST. Overall, for children, 26.48% of data were removed from the SVT, and 29.06% of data were removed from the IST. For adults, 19.20% of data were removed from the SVT, and 18.51% of data were removed from the IST – a breakdown of the percentage of data removed at each stage can be found in Table 4.4. The error rates in this study are higher than those found in similar inference generation studies with adults (e.g. Calvo & Castillo, 2001a; 2001b; Casteel, 1993) and children (e.g. Bowyer-Crane & Snowling, 2010; Casteel, 1993). However, since this was the first time these tasks were administered, it was expected that some modifications would be necessary.

---

[12] analysis was conducted using the trimmed data set (see Section 4.2.2.2.)

Table 4.4. Percentage of Data Removed

	Children		Adults		Children and adults combined	
	IST	SVT	IST	SVT	IST	SVT
Percentage of data removed due to incorrect responses	21.48	19.44	11.11	15.43	16.30	17.44
Percentage of data removed as a result of the data trim	3.03	5.56	2.47	2.47	2.75	4.02
Percentage of data removed due to item loss	4.55	1.48	4.93	1.30	4.74	1.39
Total percentage of data removed	29.06	26.48	18.51	19.20	23.79	22.98
Percentage of data remaining	70.94	73.52	81.49	80.80	76.21	77.16

Using the trimmed data set, a participant mean response time was calculated for each condition for the SVT and a participant mean reading speed was calculated for each condition for the IST. Participant means were then used to calculate mean response times and reading speeds for each condition, see Table 4.5. for SVT data and Table 4.6. for IST data.

Table 4.5. Mean Responses Times (and Standard Deviations) in Milliseconds by Age, Plausibility and Inference Type

	Real-World			Counterfactual-World		
	Literal	**Coherence**	**Elaborative**	**Literal**	**Coherence**	**Elaborative**
Children	3275.69	3661.28	4637.31	4095.96	4555.27	6619.22
	(1693.87)	(533.67)	(709.26)	(776.98)	(940.75)	(708.48)
Adults	2039.09	2276.49	2519.43	2357.71	2739.52	3568.92
	(482.34)	(724.94)	(1097.67)	(694.27)	(606.52)	(763.27)

Table 4.6. Mean Reading Speed (and Standard Deviations) in Milliseconds by Age, Plausibility and Inference Type

	Real-World			Counterfactual-World		
	Literal	**Coherence**	**Elaborative**	**Literal**	**Coherence**	**Elaborative**
Children	5559.19	6608.03	6238.40	6874.22	7260.55	6227.43
	(1198.45)	(1993.48)	(1404.66)	(1809.46)	(2480.92)	(2312.92)
Adults	4807.67	6092.06	4158.54	4994.76	5500.46	4824.96.46
	(1735.47)	(5406.70)	(518.26)	(6006.78)	(1678.34)	(930.03)

For the SVT, inspection of skewness and kurtosis statistics revealed that for most of the conditions the data were normally distributed[13]. However, for the IST, a leptokurtic skew was observed for literal counterfactual-world, real-world coherence, real-world elaborative and counterfactual-world elaborative conditions for children and a leptokurtic skew was observed for literal real-world, literal counterfactual-world, real-world coherence, and real-world elaborative conditions for adults. ANOVA is argued to be

---

[13] z scores for skewness and kurtosis were less than 1.96, the level suggested by Field (2005) for sample sizes less than 200

robust enough to deal with skewed data (Field, 2013). The assumption of sphericity was met for the SVT, such that Mauchly's test of sphericity was not significant. For reading speeds, Mauchly's was violated for plausibility * inference type. Therefore, Huynh-Feldt statistics are reported for this analysis[14]. Homogeneity of variance was not violated. Two parametric ANOVAs were conducted to explore the effects of age, plausibility and inference type on response times on the SVT and reading speeds on the IST.

For the SVT, a 2 (Age: Children, Adults) * 2 (Plausibility: Real-World, Counterfactual-World) * 3 (Inference Type: Literal, Causal, Elaborative) mixed ANOVA was conducted. A main effect of plausibility was observed, such that real-world items were responded to significantly faster than counterfactual-world items ($F (1,28) = 104.88$, $p < .001$, $\eta p^2 = .79$). A main effect of inference type was observed ($F (2,56) = 79.61$, $p < .001$, $\eta p^2 = .74$). Three post-hoc pairwise comparisons were conducted to explore further. To control for a Type I error, a conservative alpha level of .01 was set in line with previous research (e.g. Cain et al., 2004; Calvo & Castillo, 2001a; 2001b). This alpha level was used for all post-hoc analyses in this thesis. There was found to be no significant difference between the response time to coherence items and literal items ($t (29) = 3.08$, $p = .01$). Elaborative items were found to be responded to significantly slower than literal items ($t (29) = 11.63$, $p < .001$) and coherence items ($t (29) = 9.48$, $p <.001$). A main effect of age was also found, with adults responding significantly faster than children ($F (1, 28) = 151.64$, $p < .001$, $\eta p^2 = .84$).

A significant interaction was observed between plausibility and age ($F (1, 28) = 11.94$, $p = .002$, $\eta p^2 = .30$). Four t-tests were conducted to explore further. Children were found to respond to real-world items ($t (28) = 7.38$, $p < .001$) and counterfactual-world items ($t (28) = 16.62$, $p < .001$) significantly slower than adults. Both adults and children were found to respond to counterfactual-world items significantly slower than real-world items (children: $t (14) = 4.23$, $p = .001$; adults: $t (14) = 11.47$, $p < .001$). An interaction between inference type and age was observed ($F (2, 56) = 13.86$, $p < .001$, $\eta p^2 = .33$). Nine t-tests

---

[14] Girden (1992) suggest that Huynh-Feldt, not the Greenhouse-Geisser correction, should be used when the estimates of sphericity are greater than 0.75, as in these cases the Greenhouse-Geisser correction is likely to be too conservative.

were conducted to explore further[15]. Compared to adults, children were found to respond significantly slower to literal items (t (23.32) = 9.37, p < .001), coherence items (t (28) = 7.25, p < .001) and elaborative items (t (28) = 11.64, p < .001). Adults were found to respond to coherence items just as fast as literal items (t (14) = 2.31, p = .04), whereas elaborative items were responded to significantly slower than coherence (t (14) = 3.88, p = .002) and literal items (t (14) = 5.94, p < .001). The same pattern was observed for children (literal vs coherence: t (14) = 2.15, p = .05; literal vs elaborative: t (14) = 10.61, p < .001; coherence vs elaborative t (14) = 9.10, p < .001). An interaction between plausibility and inference type was also observed[16]. Six post-hoc t-tests were conducted to explore further. Real-world coherence items were found to be responded to just as fast as real-world literal items (t (29) = 2.95, p = .01). Real-world elaborative items were found to be responded to significantly slower than real-world coherence items (t (29) = 5.56, p <.001) and literal items (t (29) = 4.30, p < .001). The same pattern was observed for counterfactual-world items (coherence vs literal: t (29) = 2.01, p = .05; elaborative vs coherence: t (29) = 6.34, p < .001; elaborative vs literal: t (29) = 7.67, p < .001). The interaction between plausibility, inference type and age was not significant (F (2, 56) = 0.54, p = .58, ηp^2 = .02).

For the IST, a 2 (Age: Children, Adults) * 2 (Plausibility: Real-World, Counterfactual-World) * 3 (Inference Type: Literal, Causal, Elaborative) mixed ANOVA was conducted. Although means suggest a trend for counterfactual-world critical texts (mean = 6011.71, SD = 1744.68) to be read slower than real-world critical texts (mean = 5620.23, SD = 1790.50), a significant main effect of plausibility was not observed (F (1, 28) = .96, p = .34, ηp^2 = .04). A main effect of inference type was observed (F (2, 56) = 3.71, p = .04, ηp^2 = .13). Post-hoc pairwise comparisons revealed that there was no difference between literal and elaborative reading speeds (t (29) = 0.69, p = .50). Coherence items were read slower than literal and elaborative critical texts. However, due to the correction applied this did not reach significance (p >.03 for both comparisons). A main effect of age was observed (F (1, 28) = 7.06, p =.01, ηp^2 = .23), with adults reading significantly faster than

---

[15] To explore change quantitatively, differences between adults and children for each inference type were explored. To explore change qualitatively, differences between coherence, elaborative and literal inference types were explored for adults and children separately.

[16] Since this thesis is concerned with changes in time-course, coherence, literal and elaborative items were compared for real-world conditions and counterfactual-world conditions.

children. None of the remaining interactions were significant ($p > .30$ for all remaining interactions – see Appendix 4.2).

In sum, patterns on the SVT suggest that coherence inferences are generated whilst reading whereas elaborative inferences are not, such that response times for coherence items were just as fast as response times to literal items, whereas response times for elaborative items were significantly slower than response times to both coherence and elaborative items. Visual inspection of means on the IST also supports this assumption, such that reading speeds were slower for coherence items compared to literal and elaborative items, with no difference between literal and elaborative items. However, these comparisons did not reach significance. Counterfactual-world items were found to be responded to significantly slower than real-world items on the SVT, however, this effect was not observed on the IST. Whilst an interaction was observed between plausibility and age on the SVT, when explored further results revealed that both adults and children responded to counterfactual-world items slower than real-world items, although children responded slower than adults. An interaction between inference type and age was also observed on the SVT, however, again the same patterns were observed for adults and children, such that both groups responded slower to elaborative items compared to coherence and literal items, with no difference in response times to coherence and elaborative items. Children responded slower to adults in all conditions. These results suggest that coherence inferences are generated online whereas elaborative inferences are not by adults and children, although this process may be more taxing for children compared to adults. An interaction was observed between plausibility and inference type on the SVT but not the IST. However, similar to above, when the data was explored the patterns suggested were the same as those suggested by the main effects on the IST – coherence inferences are generated online and elaborative inferences offline, regardless of the plausibility of the information being processed. Therefore, whilst interactions were observed on the SVT but not the IST, suggested inferential time-course patterns were similar across tasks. Overall, the IST appears to be a valid measure of the inferential time-course, such that similar patterns were observed on the IST and SVT.

4.2.2.3. Assessing Inferential Skill

For both the SVT and IST, the number of errors made by each participant was calculated for each condition. From this, the mean number of errors for each condition was calculated, as can be seen in Table 4.7.

Inspection of the skewness and kurtosis statistics revealed that for many of the conditions the data were normally distributed. However, the adult real-world and counterfactual-world literal IST data were positively skewed with a leptokurtic distribution. This is due to very low error rates, with many adults often making no errors at all in these conditions. Mauchly's Test of Sphericity was not significant. However, homogeneity of variance was violated such that the largest standard deviation was more than four times larger than the smallest standard deviation. Due to the skewed data and violation of homogeneity of variance both parametric and non-parametric ANOVAs were conducted[17].

---

[17] Both parametric and non-parametric tests resulted in the same patterns. Therefore, only parametric results are reported.

Table 4.7. Mean Number of Errors (and Standard Deviations) by Age, Plausibility, Inference Type and Task

	Children						Adults					
	Real-World			Counterfactual-World			Real-World			Counterfactual-World		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Sentence Verification Task	0.93 (0.96)	1.53 (1.19)	1.73 (1.28)	1.13 (0.92)	2.67 (1.40)	2.13 (2.00)	0.07 (0.26)	0.93 (0.80)	0.80 (0.68)	0.07 (0.26)	0.87 (0.83)	1.20 (1.08)
Image Selection Task	1.87 (2.33)	1.73 (1.62)	1.93 (1.79)	2.13 (2.07)	2.73 (1.33)	2.13 (1.89)	0.13 (0.35)	0.73 (0.88)	0.80 (0.78)	0.20 (0.41)	0.87 (0.74)	1.40 (1.12)

To explore the effects of age and task on the patterns of real-world and counterfactual-world inference generation a 2 (Age: children, adults) * 2 (Task: SVT, IST) * 2 (Plausibility: real-world, counterfactual-world) * 3 (Inference Type: literal, causal, elaborative) mixed factorial ANOVA was conducted. A main effect of plausibility was found (F (1, 56) = 11.16, p = .001, ηp^2 = .17), with more counterfactual-world than real-world errors being made. A main effect of inference type was observed (F (2, 112) = 5.38, p = .02, ηp^2 = .07). Three t-tests were conducted to explore further. There was found to be no difference between the number of literal and coherence errors made (t (59) = 2.31, p = .03), literal and elaborative errors made (t (59) = 2.44, p = .02), or coherence and elaborative errors made (t (59) < .001, p = 1.00). A main effect of age was observed (F (1, 56) = 26.80, p < .001, ηp^2 = .32), with children making significantly more errors than adults. There was no main effect of task (F (1,56) = 0.41, p = .53, ηp^2 = .01). None of the interactions were significant (p > .09 for all interactions – see Appendix 4.3.).

4.2.2.4. Error Analysis

To explore the interference experienced due to the activation of real-world knowledge when generating counterfactual-world inferences an error analysis was conducted. Given the high error rates across both the IST and SVT, an error analysis was also conducted on the real-world and literal items to see if there were any improvements that could be made to the stimuli.

4.2.2.4.1. Children's Data

80% of the errors made in the counterfactual-world inferential conditions were real-world interference errors - i.e. the selection of the real-world alternative instead of the counterfactual-world target. For instance, for item 4b, which evokes the inference *'the leaves are red,'* three children chose the green tree (real-world alternative) instead of the red tree (counterfactual-world target). The high percentage of real-world errors suggests that when drawing counterfactual-world inferences, real-world interference may occur.

The remaining 20% of counterfactual-world errors made by children are likely to be due to ambiguous stimuli, lack of understanding, or no inference/an erroneous inference generation process. For instance, in item 2a (see Figure 4.4.), the critical sentence stated that the river was filled with coffee, with participants expected to infer that the river would be hot. However, two children chose the warm river (Quad 1) instead of the hot river (Quad 2) – see Figure 4.4. This may be because coffee can be served both hot and warm,

or it could be that to the children both of the temperatures represented on the thermometers in Quad 1 and Quad 2 reflected 'hot coffee.' To remove this ambiguity, the critical sentence was changed and instead of simply referring to the liquid in the river, the critical sentence was modified to refer to 'the steaming liquid'. The picture with a warm thermometer was also edited so that thermometer reflected a cooler temperature. Since some of the errors thought to reflect real-world interference could also reflect ambiguity in the stimuli or a lack of understanding, several of these items were also edited to ensure that counterfactual-world errors are actually reflective of real-world interference (final stories can be found in Appendix 4.4).
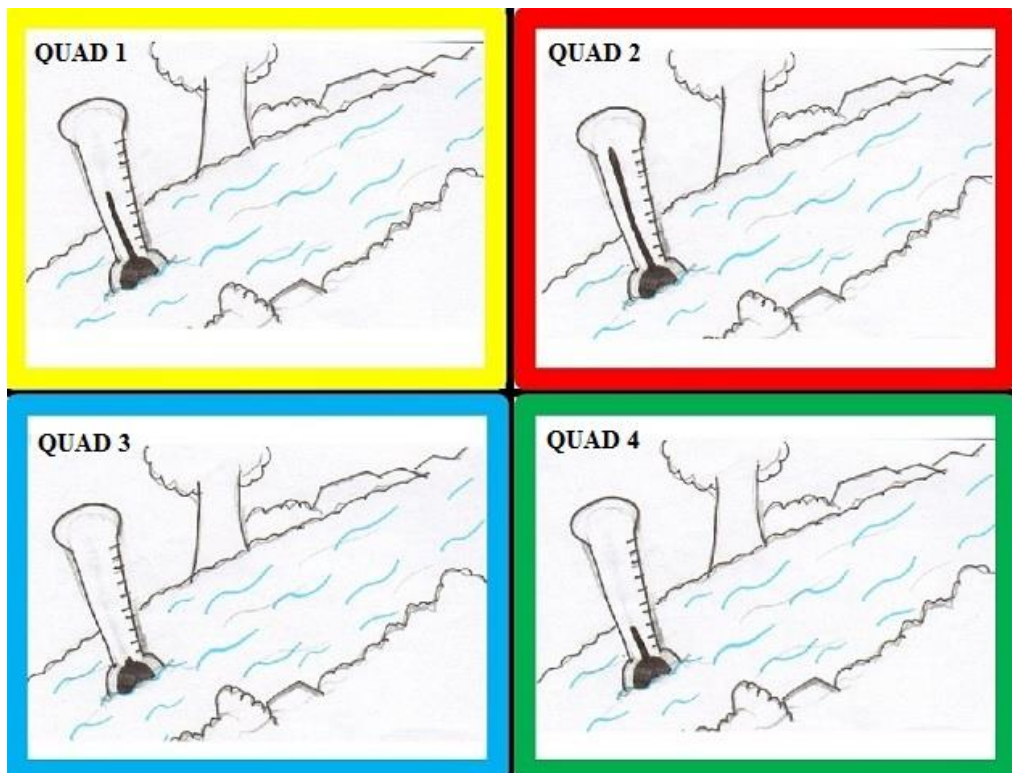
Figure 4.4. Test Item 2a

Real-world errors were split into three categories: (1) picture ambiguity, (2) text ambiguity, and (3) no inference drawn. An error was assigned to one of the ambiguity categories if there was a consistent error pattern - i.e. selection of the same (incorrect) picture. By analysing the critical text and picture selected, a decision was made as to where the ambiguity laid. For instance, item 9b states that the hippo crossed the thin ice (the intended inference was 'hippos are heavy') with many selecting the 'light hippo' picture. This suggests that the issue was text ambiguity. Therefore, this item was changed so the hippo was just standing on firm ground. If there was no consistent pattern of errors made, errors were assigned to the no inference category. 29% of errors were attributed to

picture ambiguity, 36% of errors were attributed to text ambiguity, and 29% of errors were attributed to no inference being drawn.

In the literal conditions, 100% of the errors made were likely to reflect the need for pronoun resolution. Only one of these errors was in the real-world literal condition (item 13a). For item 13a (for which the intended target is: 'Holly speaks to Ben') one child chose a dog saying hello instead of Holly saying hello. This may be because the critical sentence only said 'quietly she spoke…' This sentence was edited so that it was clear that Holly was the one speaking. The remaining errors were in the counterfactual-world literal conditions and the children consistently chose the real-world alternative. This suggests that when working memory is taxed (i.e. when the child needs to expend resources engaging in pronoun resolution) children may revert to a real-world baseline.

4.2.2.4.2. Adult's Data

91% of the counterfactual-world errors made reflected real-world interference. The remaining 9% of the counterfactual-world errors made could reflect guessing and thus may be indicative of either a response error or no inference being drawn. As with the child data, real-world errors were split into three categories: (1) picture ambiguity, (2) text ambiguity, and (3) no inference drawn. The same process of category assignment was used. 30% of the real-world errors made by adults were attributed to picture ambiguity, 40% to text ambiguity, and 30% to no inference being drawn. 100% of the literal errors made by adults could be response errors, since, typically, only one participant made the error.

The error analysis conducted shows that many of the errors made may reflect issues with the stimuli as opposed to the participants' inferential abilities. Subsequently, it was necessary to amend some of the ambiguous stimuli. Stage 2 presents the findings of the evaluation of the IST using the amended stimuli.

4.3. Stage 2 – Comparing Time-Course and Skill Patterns on the Sentence Verification Task and Image Selection Task Using Amended Stimuli

4.3.1. Method

4.3.1.1. Participants

30 adults (mean age = 25.63 years, SD = 8.05, range = 18-47 years; 18 females) and 30 children (mean age = 9.97 years, SD = 1.33, range = 9-10 years; 16 females) were recruited. Adults were recruited from Sheffield Hallam University undergraduate and postgraduate courses, Sheffield and the surrounding area. All undergraduates received research credits for their participation. Children were recruited from local schools and a sports group at a local leisure centre. All participants were native speakers of English and did not have a known learning difficulty. All participants had normal or corrected to normal vision and none were colour blind. None of the participants recruited for Stage 2 had taken part in Stage 1 of this study or Study 1 (Chapter 3).

4.3.1.2. Design

The design was the same as that employed in Stage 1.

4.3.1.3. Materials

4.3.1.3.1. Modified Image Selection Task

Whilst the general design of the IST remained the same as in Stage 1 the main character's names were changed from Holly and Ben to Lily and Joe, as the researcher was made aware of a children's TV show called Holly and Ben's Magic Kingdom. Several other changes, regarding pronoun resolution and stimuli ambiguity, were also made following the results of Stage 1. The changes that were made did not affect content with all stories maintaining their original plot. All 18 stories remained between 190 and 210 words long, with all stories in Version 1 having the same number of words as Version 2 (mean = 207.00, SD = 2.06). All 18 stories received a readability rating of grade 3 (8-9years) or less using Fry's readability formula (Version 1: mean = grade 1.96, SD = 0.52; Version

2: mean = grade 1.93, SD = 0.50)[18], thus, after modifications, stories were still suitable for the youngest participants (See Appendix 4.5).

The original IST was designed with 20 practice trials to familiarise participants with the response method. To ensure participants were not just comfortable with the response method but also the general design of the task a practice story was created. This story followed the same format as all of the experimental stories, such that it was 202 words in length and received a readability rating of less than grade 3 (see Appendix 4.6). Real-world literal and counterfactual-world literal critical sentences were used in the story. If, during the practice story, participants selected the incorrect picture, participants were asked to explain their decision. The researcher then explained why this was not the correct picture and which picture was correct. Very few participants selected the incorrect picture on this practice block.

Finally, it was decided that the comprehension questions should appear in blue text to allow participants to easily differentiate between the story text and the comprehension questions. No other changes were made.

4.3.1.3.2. Modified Sentence Verification Task

The stories, critical sentences and comprehension questions used in the modified SVT are the same as those used in the modified IST. The design and format of the modified SVT was the same as the SVT used in Stage 1.

4.3.1.4. Procedure

Consent was gained from adults and children in the same way as in Stage 1. Administration of the modified IST and modified SVT followed the same procedure as the administration of the original IST and SVT in Stage 1. However, the two main characters were introduced as Lily and Joe not Holly and Ben, participants were informed that the comprehension questions at the end of each story would appear in blue text, and participants were asked to complete a practice story before reading the first proper story.

---

[18] Readability of stories was measured using Fry's readability formula (Fry, 1968) instead of Hatcher's book grading formula. Fry's readability formula was used as it took into account not just length of word, but number of syllables which has been found to be more closely related to readability (e.g. Just & Carpenter, 1980). Therefore, Fry's formula is likely to provide a more subtle, accurate measure of readability.

The modified IST/SVT took adults on average 20 minutes and children 30 minutes to complete.

4.3.2. Results

4.3.2.1. Preliminary Analysis

Story effects were investigated since two versions of each story were created. Analysis was conducted to determine if one version of the stories was any more difficult than the other. No difference in response times was found between Set 1 and Set 2 for adults (t (17) = 0.39, p = 0.71) or children (t (17) = 0.58, p = 0.57). There was also no difference in reading speeds between Set 1 and Set 2 for adults (t (17) = 0.92, p = 0.37) or children (t (18) = 0.16, p = 0.88). Nor was there found to be any difference between the number of errors made on Set 1 and Set 2 for adults (t (17) = 0.41, p = 0.69) or children (t (17) = 0.25, p = 0.81). The data from both sets of stories were, therefore, collapsed.

4.3.2.2. Assessing Inferential Time-Course

The data were inspected prior to analysis and all incorrect responses were removed. A data trim was also conducted with all scores 2 standard deviations above or below the participant's grand mean removed and all scores 2 standard deviations above or below each items grand mean removed. Overall, for children, 17.46% of data were removed from the modified SVT, and 12.22% of data were removed from the modified IST. For adults, overall 7.78% of data were removed from the modified SVT, and 9.26% of data were removed from the modified IST – see Table 4.8. for a full breakdown of data removed. The exclusion rates observed are comparable to those found in similar inference generation studies with adults (e.g. Calvo & Castillo, 2001a; 2001b; Casteel, 1993) and children (e.g. Bowyer-Crane & Snowling, 2010; Casteel, 1993).

Table 4.8. Percentage of Data Removed

	Children		**Adults**		**Children and Adults Combined**	
	IST	**SVT**	**IST**	**SVT**	**IST**	**SVT**
Percentage of data removed due to incorrect responses	7.40	11.90	5.56	5.56	6.48	8.73
Percentage of data removed as a result of the data trim	4.82	5.56	3.70	2.22	4.29	3.89
Total percentage of data removed	12.22	17.46	9.26	7.78	10.74	12.62
Percentage of data remaining	87.78	82.54	90.74	92.22	89.26	87.38

Using the trimmed data set, a participant mean response time was calculated for each condition for the SVT and a participant mean reading speed was calculated for each condition for the IST. Participant means were then used to calculate a mean response time or reading speed for each condition, see Table 4.9. for SVT data and Table 4.10. for IST data.

Table 4.9. Mean Response Times (and Standard Deviations) in Milliseconds by Age, Plausibility and Inference Type

	Real-World			Counterfactual-World		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Children	3263.50	3186.39	4041.76	3977.23	4184.47	4786.16
	(398.13)	(321.58)	(505.34)	(546.08)	(1299.83)	(598.75)
Adults	2151.79	2130.42	2366.07	2566.03	2496.11	3015.37
	(464.58)	(641.32)	(362.93)	(739.77)	(562.18)	(1230.80)

Table 4.10. Mean Reading Speeds (and Standard Deviations) by Age, Plausibility, Inference Type, and Task

	Real-World			Counterfactual-World		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Children	5429.06	6267.50	5234.89	5859.03	6265.13	6065.90
	(1232.27)	(1605.89)	(1555.86)	(1683.58)	(1363.81)	(1655.78)
Adults	3437.28	4419.40	4023.27	3860.47	4697.72	4029.91
	(425.15)	(1269.26)	(1099.46)	(670.38)	(1076.07)	(801.85)

Inspection of the skewness and kurtosis statistics revealed that for most of the conditions the data were normally distributed. However, the distribution of real-world literal scores on the SVT for children was found to have a slight leptokurtic skew, however, this was not severe. Mauchly's test of sphericity was not significant for all conditions. Homogeneity of variance was not violated for reading speed data, but was violated, for response time data such that the largest standard deviation was more than four times higher than the smallest standard deviation. Since the sample sizes are equal it is argued that ANOVA is robust enough to deal with this violation (Field, 2013). Therefore, two parametric ANOVAs were conducted.

A 2 (Age: Children, Adults) * 2 (Plausibility: Real-World, Counterfactual-World) * 3 (Inference Type: Literal, Causal, Elaborative) mixed ANOVA was conducted on SVT

response times. A main effect of plausibility was observed (F (1,28) = 17.17, $p < .001$, $\eta p^2 = .38$), with real-world items (mean = 3068.22, SD = 988.40) found to be responded to significantly faster than counterfactual-world items (mean = 3989.43, SD = 1174.91). A main effect of inference type was observed (F (1, 56) = 16.10, $p < 001$, $\eta p^2 = .37$). Three post-hoc t-tests were conducted to explore further. Coherence items (mean = 3308.14, SD = 1007.48) were found to be responded to just as fast as literal items (mean = 2942.11, SD = 868.75; t (28) = .09, $p = .927$). Elaborative items (mean = 4336.22, SD = 1443.57) were found to be responded to significantly slower than coherence items (t (28) = 4.69, $p < .001$) and literal items (t (28) = 4.79, $p < .001$). A main effect of age was observed (F (1,28) = 116.93, $p < .001$, $\eta p^2 = .81$), with children responding significantly slower than adults. None of the interactions were significant ($p > .31$ for all interactions – see Appendix 4.7). Taken together, results suggest that coherence inferences are read whilst reading, whereas elaborative inferences are not, regardless of age group and plausibility of information. Counterfactual-world information does appear to be more taxing than real-world information, however, evidenced by the longer response times in the counterfactual-world condition compared to the real-world condition.

A 2 (Age: Children, Adults) * 2 (Plausibility: Real-World, Counterfactual-World) * 3 (Inference Type: Literal, Causal, Elaborative) mixed ANOVA was conducted on IST reading speeds. A main effect of age was found (F (1,28) = 21.22, $p < .001$, $\eta p^2 = 0.43$), with adults reading critical texts faster than children. A main effect of plausibility was found (F (1,28) = 12.15, $p = 0.002$, $\eta p^2 = 0.30$), with participants reading counterfactual-world items significantly slower than real-world items. A main effect of inference type was also found (F (2,56) = 12.70, $p < 0.001$, $\eta p^2 = 0.31$). Post-hoc pairwise comparisons found that coherence items were read significantly slower than both literal (t (29) = 4.64, $p < .001$) and elaborative items (t (29) = 3.73, $p = .001$). There was no difference between reading times of literal and elaborative items (t (29) = 1.24, $p = .23$). None of the interactions were significant ($p > .10$ for all interactions – see Appendix 4.8). Taken together, results suggest that coherence inferences are read whilst reading, whereas elaborative inferences are not, regardless of age group and plausibility of information. Counterfactual-world information does appear to be more taxing than real-world information, however, as evidenced by the longer reading speeds for all inference types in the counterfactual-world condition compared to the real-world condition.

4.3.2.3. Assessing Inferential Skill

For both versions of the task, the number of errors made by each participant was calculated for each condition. From this the mean number of errors for each condition was calculated, as can be seen in Table 4.11.

Inspection of the skewness and kurtosis statistics revealed that for many of the conditions the data were not normally distributed, such that the data were positively skewed with a leptokurtic distribution. This is due to the low error rates observed in some conditions. Mauchly's Test of Sphericity was significant for plausibility * inference type. Therefore, the Huynh-Feldt statistic is reported. Homogeneity of variance was also violated, with the highest standard deviation more than four times larger than the smallest standard deviation for both sets of data. Due to the skewed data and violation of homogeneity of variance both parametric and non-parametric ANOVAs were conducted[19].

---

[19] Parametric and nonparametric tests resulted in the same patterns. Therefore, only parametric results are reported below.

Table 4.11. Mean Number of Errors (and Standard Deviations) by Age, Plausibility, Inference Type, and Task

	Children						Adults					
	Real-World			Counterfactual-World			Real-World			Counterfactual-World		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Sentence Verification Task	0.26	2.14	2.66	0.40	3.06	4.66	0.14	0.14	0.40	0.26	1.06	1.60
	(0.35)	(0.33)	(1.35)	(0.41)	(0.74)	(1.11)	(0.26)	(0.26)	(0.41)	(0.35)	(0.74)	(0.86)
Image Selection Task	0.54	0.66	1.34	0.54	2.00	1.94	0.14	0.80	0.26	0.80	0.94	1.20
	(0.46)	(0.70)	(0.49)	(0.23)	(0.93)	(0.83)	(0.26)	(0.23)	(0.17)	(0.63)	(0.74)	(0.83)

To explore the effects of age, task, plausibility and inference type a 2 (Age: adults, children) * 2 (Task: SVT, IST) * 2 (Plausibility: real-world, counterfactual-world) * 3 (Inference Type: literal, causal, elaborative) mixed ANOVA was conducted on errors made on SVT and IST. A main effect of plausibility was found ($F (1,56) = 23.51, p < .001$, $\eta p^2 = .30$), with more counterfactual-world errors than real-world errors being made. A main effect of inference type was found ($F (2,112) = 28.88, p < .001, \eta p^2 = .34$). Post-hoc pairwise comparisons found that more coherence errors *(t (29) = 6.27, p < .001)* and elaborative errors ($t (29) = 7.18, p < .001$) than literal errors were made. There was found to be no significant difference between the number of coherence and elaborative errors made ($t (29) = 1.58, p = .06$). A main effect of age was observed ($F (1,56) = 34.27, p < .001, \eta p^2 = .38$), with children making more errors than adults. A main effect of task was also observed ($F (1,56) = 7.84, p = .007, \eta p^2 = .12$), with more errors made on the SVT than the IST.

A significant interaction between inference type and age was observed ($F (2,112) = 10.45, p < .001, \eta p^2 = .16$). Nine post-hoc t-tests were conducted to explore this interaction further[20]. Children were found to make significantly more coherence errors ($t (58) = 4.00, p < .001$) and elaborative errors ($t (58) = 4.44, p < .001$) than adults. However, there was found to be no significant difference between the number of literal errors made by adults and children ($t (58) = 0.63, p = .53$). Adults were found to make a similar number of literal, coherence and elaborative errors ($p > .01$ for all comparisons). Children were found to make significantly fewer literal errors than coherence errors ($t (29) = 5.43, p < .001$) and elaborative errors ($t (29) = 5.65, p < .001$). There was no difference in the number of coherence and elaborative errors made by children ($t (29) = 1.79, p = .08$). This suggests children find inferential, but not literal processing more difficult than adults. Similarly, whilst adults find inferential processing no more difficult than literal processing, children appear to find inferential processing more difficult than literal processing.

A significant interaction was observed between plausibility and inference type ($F (1.96, 109.58) = 3.88, p = .024, \eta p^2 = .07$). Nine post-hoc t-tests were conducted to explore this

---

[20] To explore change quantitatively, differences between adults and children for each inference type were explored. To explore change qualitatively, differences between coherence, elaborative and literal inference types were explored for adults and children separately.

interaction further [21]. Significantly fewer real-world literal errors than real-world coherence errors (t (59) = 2.82, p = .006) and real-world elaborative errors (t (59) = 3.83, p < .001) were made. Similarly, significantly fewer counterfactual-world literal errors than counterfactual-world coherence errors (t (59) = 5.44, p < .001) and counterfactual-world elaborative errors (t (59) = 5.48, p < .001) were made. Participants were found to make significantly more counterfactual-world coherence errors than real-world coherence errors (t (59) = 2.91, p = .005) and significantly more counterfactual-world elaborative errors than real-world elaborative errors (t (59) = 4.13, p < .001). There was no significant difference between the number of literal errors made in either the real-world or counterfactual-world conditions (p > .10 for all comparisons). This suggests, regardless of plausibility, adults and children find inferential processing more difficult than literal processing. Similarly, whilst counterfactual-world inferential processing appears to be more difficult than real-world inferential processing, literal processing appears to be just as easy when processing both real-world and counterfactual-world information.

A significant interaction was observed between inference type and task (F (2,112) = 7.66, p = .001, ηp^2 = .12). Nine post-hoc t-tests were conducted to explore further. Significantly fewer literal errors than coherence errors (SVT: t (29) = 4.75, p < .001; IST: t (29) = 2.83, p = .008) and elaborative errors (SVT: t (29) = 5.43, p <. 001; IST: t (29) = 2.79, p = .009) were made on the SVT and the IST. Significantly more elaborative errors were made on the SVT compared to the IST (t (58) = 2.83, p = .006). No other comparisons were significant (p > .03 for all comparisons). Taken together, results suggest that elaborative inference generation appears to be more difficult on the SVT than the IST, whereas there is no difference for coherence and literal processing.

A significant interaction was observed between age and task (F (1,56) = 10.95, p = .002, ηp^2 = .16). Four post-hoc t-tests were conducted to explore further. Children were found to make significantly more errors than adults on the SVT (t (28) = 6.20, p < .001), whereas there was no significant difference between the number of errors made by adults and children on the IST (t (28) = 1.90, p = .07). Adults were found to make a similar number of errors on the IST and SVT (t (28) = 0.43, p = .67), whereas children were found to

---

[21] Since this thesis is concerned with changes in time-course, coherence, literal and elaborative items were compared for real-world conditions and counterfactual-world conditions.

make significantly more errors on the SVT compared to the IST (t (28) = 3.70, p = .001). Taken together, results suggest that children find inferential processing easier on the IST compared to the SVT, such that they perform at adult-like levels on the IST.

A significant interaction was observed between inference type, age and task (F (2,112) = 3.42, p = .036, ηp^2 = .06). Twelve post-hoc t-tests were conducted to explore further. Children were found to make significantly more elaborative errors on the SVT than the IST (t (28) = 3.57, p = .001). Compared to adults, children were found to make significantly more coherence errors (t (28) = 4.63, p < .001) and elaborative errors (t (28) = 4.71, p < .001) on the SVT. Children were found to make significantly fewer literal than coherence errors (t (14) = 5.41, p < .001) and elaborative errors (t (14) = 6.07, p < .001) on the SVT. The difference between literal errors and coherence errors (t (14) = 2.86, p = .01) and elaborative errors (t (14) = 2.96, p = .01) also approached significance on the IST. No other comparisons were significant (p > .01 for all comparisons). None of the remaining interactions were significant (p > .01 for all interactions – see Appendix 4.9). Taken together results suggest that children, compared to adults, display difficulty on the SVT more so than the IST when processing inferential items. Children appear to find the generation of elaborative items particularly difficult on the SVT.

The processing of literal information appears to be no different for real-world and counterfactual-world information. However, counterfactual-world inferential processing appears to be more difficult than real-world inferential processing. It appears that children perform at adult-like levels on the IST, such that there is no difference in performance for coherence, elaborative or literal items, regardless of age, when assessed using the IST. Importantly, children appear to find inferential processing more difficult on the SVT compared to IST, such that they display particular difficulty with inferential processing and specifically elaborative processing when assessed using the SVT. This appears to be the case for both real-world and counterfactual-world information.

4.3.2.4. Error Analysis

4.3.2.4.1. Children's Error Patterns

Errors made in the counterfactual-world condition were categorised into three conditions: (1) real-world interference, (2) literal processing, and (3) random guessing. 64% of counterfactual-world errors fell into the real-world interference category, such that the participant chose the real-world alternative instead of the correct counterfactual-world target. 18% of counterfactual-world errors fell into the literal processing condition. The remaining 18% of errors fell into the random guess condition.

The real-world errors were categorised into four categories[22]: (1) literal processing, (2) lack of knowledge, (3) picture ambiguity, and (4) no inference/response error. An error was categorised as indicative of literal processing if the picture depicting information explicitly stated in the critical was selected. 43% of the errors made in the real-world condition reflected literal processing. When completing the IST some participants would make comments suggesting a lack of knowledge – e.g. 'What colour is clay?' An error was categorised as indicative of a lack of knowledge if such a comment had been made. 28.57% of errors made could potentially reflect a lack of knowledge. 14% of the real-world errors made were categorised as picture ambiguity. For instance, for item 7a, one child chose the picture of Joe saying hello to the whale, instead of the picture of the whale calling. The image used lines to represent sonar, however, this may have been too complex for the children. This item was modified; such that this item was changed to 'Suddenly Joe heard a loud Moo! Joe turned around to see a cow walking towards him.' with the target picture depicting a cow mooing. Finally, 29% of the real-world errors were categorised as response errors - i.e. accidentally choosing the button next to the correct one – or no inference being drawn resulting in a guess, due to no consistent error pattern being observed.

60% of the literal errors made are likely to reflect misinterpretation of the picture. However, this was widely spread across items suggesting there is no issue with any particular picture. 20% of the literal errors made reflected a real-world error. As suggested in Stage 1, this may be indicative of children reverting to a real-world baseline when their working memory is taxed and/or they are unable to make a decision. Finally, 20% of the errors made could be indicative of response errors.

---

[22] These categories are different to those in Stage 1 as clearer patterns were revealed.

75% of the errors made in the counterfactual-world condition were categorised as real-world interference errors. The remaining 25% of errors demonstrated no clear pattern, thus are likely to reflect guessing or response errors. Errors made in the real-world and literal conditions, were rare and widely spread meaning error analysis could not be conducted on these data.

Overall, the error analysis conducted suggests that a) the modified stimuli are more suitable than the original stimuli, such that ambiguity and the need for pronoun resolution appears to have been reduced; and b) real-world interference is likely to be a major cause of difficulty, for both adults and children, when generating counterfactual-world inferences.

4.4. Discussion

The Image Selection Task (IST) is a novel paradigm specially created for this thesis to assess inference generation abilities during adolescence. This study sought to evaluate the effectiveness of the IST as a measure of inference generation by comparing the results of the IST with the results of a timed Sentence Verification Task (SVT) comprising the same stimuli. In Stage 1, patterns on the IST largely mirrored those on the SVT. However, counterfactual-world items appeared to be more cognitively demanding than real-world items on the SVT but not the IST. The high error rates and error analysis conducted suggested that amendments to the stimuli were necessary. In Stage 2, consistent with the predictions made, time-course patterns on the IST largely mirrored those on the SVT. However, in conflict with the predictions made, there were some differences when exploring the skill (number of errors) data. Specifically, children appeared to find the SVT more difficult than the IST, experiencing particular difficulty when processing elaborative inferences. These differences may be due to the pictorial versus textual presentation of test items. As discussed below, this does not minimise the validity of the IST as a measure of inference generation.

4.4.1. The Image Selection Task as a Measure of the Inferential Time-Course

The Constructionist Theory suggests that those real-world inferences needed for coherence are generated *whilst* reading (online), whereas elaborative inferences are

generated with strategic effort *after* reading (offline; Graesser et al., 1994). Research exploring counterfactual-world inference generation, suggests that children generate counterfactual-world coherence inferences online, whereas adults generate all counterfactual-world inferences offline (Bowyer-Crane & Snowling, 2010; Graesser et al., 1998). Real-world patterns on both the IST and SVT were consistent with the Constructionist Theory. However, patterns on both the IST and SVT, suggest that both children and adults generate those counterfactual-world inferences needed for coherence whilst reading. As will be discussed further in Chapter 7, the difference between the findings of the current research and previous research may be due methodological differences in stimuli used.

In line with previous research, it was predicted that adults and children would respond to counterfactual-world items significantly slower than real-world items on the SVT and read counterfactual-world items significantly slower than real-world items on the IST, indicating that the processing of counterfactual-world information is more demanding than the processing of real-world information (e.g. Ferguson, 2012). This was not the pattern of results found in Stage 1, such that whilst means suggested counterfactual-world items were read slower, there was no significant difference between real-world and counterfactual-world reading speeds on the IST. However, in Stage 2, after amendments to the stimuli, counterfactual-world items were responded to significantly slower than real-world items on both tasks. This suggests that coherence inferences are generated whilst reading, regardless of plausibility, however this process appears to be more cognitively demanding for counterfactual-world, compared to real-world information. In sum, inferential time-course patterns were the same on both the IST and SVT. This suggests the IST is a valid measure of the inferential time-course in adults and children, for real-world and counterfactual-world information.

4.4.2. The Image Selection Task as a Measure of Inference Generation Skill

Previous research suggests that the generation of elaborative inferences is more difficult than the generation of coherence inferences (e.g. Bowyer-Crane & Snowling, 2010; Cain & Oakhill, 1999; Cain et al., 2001). The interaction between inference type and age, revealed that this was not the case in the current study, however, such that on both tasks a similar number of coherence and elaborative errors were made. Whilst there was no difference within the tasks, significantly more elaborative errors were made on the SVT

145

than IST. The first possible interpretation of these results is the stimuli used in the current research did not evoke the necessary inference types. However, stimuli were carefully designed to evoke target inference types, with the results of Chapter 3 suggesting that coherence inference-evoking sentences evoke coherence inferences and elaborative inference-evoking sentences evoke elaborative inference generation. The second interpretation is that the difference in difficulty observed in previous research is due to differences in the demands of coherence and elaborate inference-evoking stimuli not present in the current research. As discussed in more detail in Chapter 7, previous research typically explores knowledge-based elaborative inferences, but text-based coherence inferences. Third, it is possible that the use of pictures encouraged participants completing the IST to create a pictorial mental model. Given that the static properties assessed in the elaborative condition were largely visual, elaborative inference generation, particularly, on the IST may have been aided by the construction of a multi-modal mental representation. In sum, it appears that the generation of knowledge-based elaborative inferences is no more difficult than the generation of knowledge-based coherence inferences, particularly when elaborative static properties are assessed using pictorial stimuli.

Previous research shows that inference generation skill increases with age (e.g. Barnes et al., 1996; Casteel, 1993; Oakhill et al., 2003). However, whilst children were found to make more errors than adults on the SVT, adults and children were found to make a similar number of errors on the IST. This was due to children making significantly fewer errors on the IST compared to the SVT, whereas adults made a similar number of errors on both tasks. Previous research finding that inference generation skill increases with age, converges to suggest the ability to generate inferences unprompted, improves with age, as opposed to basic inference generation skill – i.e. young children are capable of generating the same inferences as adults, but do not do so unless prompted (Omanson et al., 1978; Paris, et al., 1977). Since demands are likely reduced in the IST compared to the SVT, children may have had more resources available to focus on inference generation – specifically elaborative inference generation - explaining the superior performance observed. The current study thus highlights the necessity of non-verbal methods of assessment.

Consistent with previous research, the error analysis conducted suggested that real-world interference was the primary cause of inference generation failure when generating

counterfactual-world inferences. Research suggests that counterfactual-world text processing becomes more difficult with age (Ceci et al., 1981; Dorfman, 1988). It was, therefore, predicted that counterfactual inferential skill would be significantly worse for adults compared to children on both the IST and SVT. However, this was not the case, such that both groups made more counterfactual-world errors than real-world errors on the IST and SVT. As discussed in more detail in Chapter 7, this may be due to methodological differences in the stimuli used to assess the counterfactual-world inference generation abilities of adults and CYP in previous research.

4.4.3. Conclusion

The current study was conducted to explore the validity of the IST, a novel paradigm designed to assess inference generation abilities. After some modifications, the IST resulted in the same inferential time-course patterns as the SVT. Whilst some differences in number of errors were observed between patterns on the IST and SVT, these differences may reflect differences in task demands, such that the IST promotes the construction of a more enriched mental model more so than the SVT. Similarly, the IST reduces task demands allowing for a purer measure of inference generation skill, resulting in children performing significantly better on the IST than the SVT. The IST, then, also appears to be a valid measure of inference generation skill. Subsequently, the IST was used to assess inference generation abilities in this thesis.

Part 3: Using the IST to Explore Inference Generation during Adolescence – A Series of Cross-Sectional Studies

CHAPTER 5 METHODOLOGY FOR STUDIES 3-5

5.1. Introduction

The aim of this research is to explore real-world and counterfactual-world inference generation abilities during adolescence. Part 1 of this thesis reviewed current literature in the area. Part 2 focused on the design and evaluation of an inference generation task suitable for the assessment of both real-world and counterfactual-world inference generation abilities during adolescence. This resulted in the creation of the Image Selection Task (IST). Part 3 of this thesis utilises the IST to explore how inference generation abilities develop during adolescence and the skills and processes underpinning this development. The purpose of this Chapter is to describe, evaluate, and justify the research methods used in Studies 3, 4 and 5 (Chapters 6, 7 and 8, respectively).

5.2. Design

When studying age-related changes three designs can be employed: 1) cross-sectional 2) longitudinal, and 3) sequential. A longitudinal or sequential design would have been preferable since repeated assessment of the same participants allows for the assessment of changes over time. However, typically, research exploring reading comprehension and related skills employ time intervals of no less than one year (e.g. Kim, Wagner, & Lopez, 2012; Seigneuric & Ehrlich, 2005; Verhoeven & van Leeuwe, 2008). Due to the time constraints of this PhD, neither approach was feasible. A cross-sectional design was thus employed in this research. Cross-sectional designs assess and compare several age groups. Cross-sectional studies are popular as all of the necessary data can be collected relatively quickly. However, since each child is only tested once, firm conclusions regarding individual development *over* time cannot be made. Although, attempts should be made to control for confounding variables, differences between age groups could be due to discrepancies in the educational, economical, or socio-cultural experiences as opposed to age-related experiences (e.g. Salthouse, 2010). For instance, 11 year olds could show superior performance compared to 9 year olds, however, this could be due to the 11 year olds receiving a higher quality education than the 9 year olds rather than an improvement in said skill between the ages of 9 and 11 years. Every attempt was made to control for these variables in the current study as discussed in Section 5.3.

5.3. Participant Recruitment

5.3.1. Selection of Age Groups

Four age groups were selected to take part in this research: (1) 9-10 year olds, (2) 11-12 year olds, (3) 13-14 year olds, and (4) 18 years and over. When exploring development and possible stages of change it is essential that the age-range chosen reflects the point in development where the change is expected to occur (Harris, 2008). Whilst many Primary School CYP appear to have typical literacy abilities, government statistics suggest that some of these CYP begin to struggle in Secondary School (Department for Education, 2015a; 2015b). Additionally, research shows that whilst adults struggle to generate counterfactual-world coherence inferences online, CYP aged 9-10years do not (e.g. Bowyer-Crane & Snowling, 2010; Graesser et al., 1998). This finding suggests a conceptual shift in the processing of counterfactual-world inferential abilities, with this period of change occurring between 9-10 years and adulthood. This thus set the upper and lower age-range of participants for the research. 14-16 years is the age at which CYP are preparing for their GCSEs and other equivalent exams. As such, schools can be reluctant to release CYP from their studies during this time. Fortunately, research suggests little development in those key skills explored in this thesis during 14-16 years (e.g. Romine & Reynolds, 2005). Therefore, it was decided to assess 9-10 year olds, 11-12 year olds 13-14 year olds and adults (18 years and over).

5.3.2. Sample Selection

Sample size (50 participants per group) was determined by Harris' (1985) rule of thumb such that for every predictor variable ten participants are needed[23]. Sample size in this study is thus higher than that used in similar developmental studies (e.g. Bowyer-Crane & Snowling, 2010).

This thesis is concerned with the development of inferential abilities in typical developers; therefore, recruiting a sample reflective of the typical UK school-age population was important. Research suggests that socioeconomic status is predictive of

---

[23] Originally working memory was a predictor variable not control variable.

literacy abilities, such that those from a lower socioeconomic background may be disadvantaged (Jefferson, Gibbons, Rentz, Carvalho, Manly, Bennet & Jones, 2011; Rowe & Goldin-Meadow, 2009; Zhang, Tardif, Hong Li, Liu, McBride-Chang, Liang & Zhang, 2013). There is also a strong link between teaching practices and literacy abilities (Bui & Fagan, 2013). To gain a representative sample it was important to recruit participants from a range of socioeconomic backgrounds and educational institutions, which, despite all following the National Curriculum, may implement different teaching strategies. Therefore, every mainstream Secondary School in the South Yorkshire area (84; and one school in Derbyshire[24]) was invited to take part in the research. There was a response rate of 11.76%, with ten Secondary Schools agreeing to take part. Five Year 7s and five Year 9s were selected from each school. Every mainstream Primary School in Barnsley was also invited to take part in the research (80). There was a response rate of 10%, with eight Primary Schools agreeing to take part. 5 Year 5s were recruited from the majority of Primary Schools, ten Year 5s were recruited from one Primary School and eleven Year 5s were recruited from another Primary School. Table 5.1 presents a summary of characteristics for each school. Participants were recruited from a range of areas, thus it is likely that a diverse sample, with regards to socioeconomic backgrounds and teaching strategies, was obtained. Variation in teaching strategy, in particular may not have been achieved if all participants were recruited from the same school. Teachers or year heads (for secondary schools) were all given criteria to help guide their selection of children and young people. The researcher was not involved in pupil selection after this point, with selecting participants in line with this criteria the responsibility of the school. The criteria given to schools along with a rationale is outlined below.

Table 5.1. Summary of School Information

---

[24] This was due to researcher error such that a school in Derby was confused with a school in Sheffield

	School	Area	No. and Age of participants recruited	Percentage of Pupils Eligible for Pupil Premium
Primary Schools	1	Barnsley	11 Year 5s	70.1%
	2	Sheffield	10 Year 5s	10%
	3	Barnsley	5 Year 5s	5.8%
	4	Sheffield	5 Year 5s	55.9%
	5	Sheffield	5 Year 5s	34.5%
	6	Sheffield	5 Year 5s	16.4%
	7	Sheffield	5 Year 5s	37.2%
	8	Sheffield	5 Year 5s	49.7%
Secondary Schools	1	Sheffield	5 Year 7s 5 Year 9s	27.6%
	2	Sheffield	5 Year 7s 5 Year 9s	52.0%
	3	Rotherham	5 Year 7s 5 Year 9s	31.5%
	4	Rotherham	5 Year 7s 5 Year 9s	23.4%
	5	Rotherham	5 Year 7s 5 Year 9s	23.8%
	6	Rotherham	5 Year 7s 5 Year 9s	38.8%
	7	Doncaster	5 Year 7s 5 Year 9s	18.3%
	8	Doncaster	5 Year 7s 5 Year 9s	45.4%
	9	Barnsley	5 Year 7s 5 Year 9s	33.9%
	10	Derby	5 Year 7s 5 Year 9s	34.7%

A variety of other factors have been found to impact upon literacy development, which have guided selection criteria. *Criteria 1:* Children must be typical developers, such that they do not have a known learning difficulty and/or exceptional reading abilities. Those with known learning difficulties may fall within the typical range for literacy abilities, but may be employing qualitatively different strategies to those without learning difficulties (Graham & Bellert, 2009; Ricketts, Jones, Happe, & Charman, 2013). *Criteria 2:* Children should be selected from a range of family backgrounds, ethnicities and both genders. Research has found correlations between literacy abilities and family background, ethnicity and gender (Denton, Wolters, York, Swanson, Kulesz, & Francis, R2014; Lindo, 2014). *Criteria 3:* Children should be native speakers of English. research shows that children with English as an additional language (EAL) display atypical literacy development (e.g. Burgoyne, Whiteley, & Hutchinson, 2011; 2013). *Criteria 4:* Children should not be colour blind or have any other (non-corrected) visual impairments. The IST requires participants to use coloured buttons to make their selection. To sum, schools

were asked to select a sample of children who varied in regards to literacy abilities (but did not have a known learning difficulty or exceptional literacy abilities), family background, ethnicity, and gender, were native English speakers, and were not colour blind.

Adults were recruited from Sheffield Hallam University. Research suggests that university students are not representative of the adult population, such that compared to the general population, this group have higher literacy skills (e.g. Kuczera, Field & Windisch, 2016). Whilst a representative adult sample would have been desirable, with regards to availability a student sample was most convenient to recruit. All adults were native English speakers, did not have a known learning difficulty and were not colour blind.

To conclude, to ensure a representative sample of typical developers was obtained multiple schools from different socio-economic backgrounds utilising different teaching methods were recruited. Within each school, CYP with a range of literacy abilities, ethnicities and genders were selected. All participants spoke English as a first language, had no known learning difficulties and none were colour-blind.

5.3.3. Ethical Considerations

Ethical approval for this project was gained from the SHU ethics committee, with these and the BPS ethical guidelines being adhered to throughout (see Appendix 5.1). A three-tiered method of obtaining CYP consent was employed: initially consent was obtained from the school, then the CYP's parent, and finally the CYP (see Appendix 5.2 for information sheets, consent forms, and debrief forms for all age groups). All schools were approached in the same way. An email was sent to all schools, briefly explaining the study and asking if they would be interested in participating. If schools responded a meeting was set-up (typically with the head teacher or head of year) to discuss the research. If the school was happy to participate at the end of the meeting, they were given an information pack containing details of the research and tasks to be undertaken, a school consent form to be signed by the head teacher, and parental consent forms. Schools sent out and collected the parental consent forms for the researcher. Once informed, written consent from the school and parents was obtained a date for data collection was confirmed with all school staff who would be affected. During the first data collection session written,

informed consent was gained from each CYP. In every session following, verbal consent was obtained. See Figure 5.1. for a summary of the CYP recruitment process.
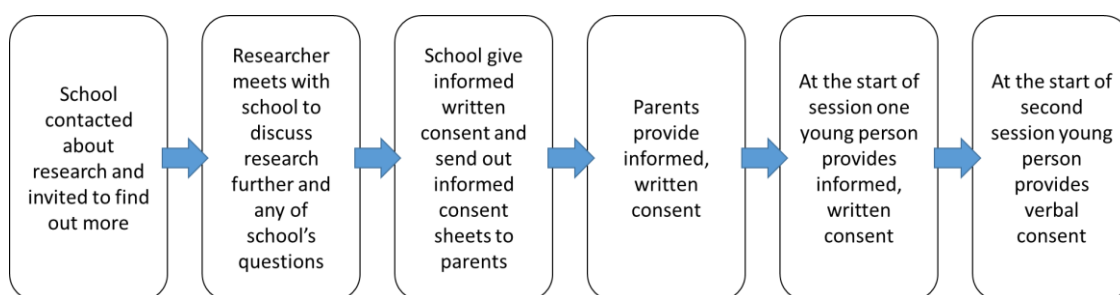
Figure 5.1. Flow diagram of the CYP recruitment procedure

Adults were recruited through an advert on the Sheffield Hallam University psychology research participation scheme. Adults were invited to discuss the research further and, if happy to take part, a time and date was arranged. Before the initial data collection session adults were given an information sheet followed by a consent form. Informed, written consent was gained from all adults.

Regardless of age, the information sheet and consent form was read aloud to all participants. Each task was also explained to all participants before being undertaken, participants then had to confirm that they were happy to complete the task. Whilst positive encouragement was given to all participants (e.g. "you're doing well" "keep trying your best"), participants (nor, in the case of CYP, parents or schools) were not advised of their individual performance on any task.

5.4. Cognitive Assessment of Inference Generation, Underpinning Skills and Processes, and Higher-Order Literacy Skills

5.4.1. Introduction

Part 3 of the thesis is concerned with the exploration of inference generation abilities during adolescence. Three research studies were conducted: 1) exploration of the role of inference generation in reading comprehension and text production during adolescence, 2) the development of inference generation abilities during adolescence, 3) the development and role of those cognitive skills underpinning inference generation during adolescence. The data for all studies were collected at the same time point (although this comprised two sessions). The following sections address issues concerning the

administration of a large battery of cognitive tasks, in addition to describing, evaluating, and justifying the cognitive test battery used.

5.4.2. Administration of the Test Battery

The tasks were presented in the same order to all participants so as not to advantage or disadvantage any participant. The order of task administration was designed to minimise fatigue and boredom, such that similar tasks were not presented one after another and participants were assessed over two sessions. The first session was an individual session and the second a small group session (for some adults, two individual sessions were held). In addition to reducing boredom, the splitting of tasks across two sessions was also undertaken to reduce carry-over effects between tasks utilising the same information – e.g. inference generation and knowledge tasks. See Appendix 5.3 for order of tasks.

5.4.3. Measures

Each participant completed all assessments discussed in this section. Table 5.1. provides a summary of the measures used.

Table 5.1. Summary of Measures

Task	Assesses	Data collected used in..
Image Selection Task	Inference generation	**Studies 3, 4, and 5**
WASI Matrix Reasoning	Non-verbal reasoning	**Studies 3, 4, and 5**
WASI Vocabulary Definitions	Verbal reasoning	**Studies 3, 4, and 5**
BAS Backwards Digit Span	Working Memory	**Studies 3, 4 and 5**
York Assessment of Reading Comprehension	Reading Comprehension Ability	**Study 3**
Narrative Production Task	Text Production Skills	**Study 3**
Timed Sentence Verification Task	Amount of knowledge Accessibility of knowledge	**Study 5**
Reasoning from False Premises Task	Strength of Belief Biases	**Study 5**
Rule Shift Card Task	Inhibition	**Study 5**
York Assessment of Reading Comprehension – single word reading	Word Reading Ability	

5.4.3.1. Inference Generation

As detailed in Chapters 3 and 4, a new task was developed to assess inference generation abilities in this study: Image Selection Task (IST). Participants were presented with a series of stories, in which critical texts followed by 2*2 picture selection tasks, were embedded. Participants select the picture that best matches the text they have just read. Critical texts evoke either real-world coherence inferential processing, real-world elaborative inferential processing, real-world literal processing, counterfactual-world coherence inferential processing, counterfactual-world elaborative inferential processing, or counterfactual-world literal processing. Number of errors and reading speed (in milliseconds) were recorded.

5.4.3.2. Control Measures

5.4.3.2.1. Non-Verbal and Verbal Reasoning

Consistent with previous developmental research, this study sought to control for intelligence (e.g., Cain & Oakhill, 1999; 2006a; 2011). The Matrix Reasoning Subscale of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was used to assess non-verbal reasoning. In this task, participants were presented with a pattern with a piece missing. Participants must select the missing piece from a choice of five pictures. The Vocabulary Definitions Subscale of the WASI was used to assess verbal reasoning. Participants hear a word. They then define it. These subscales were selected since the WASI was designed for easy and efficient assessment of an individual's intellectual functioning. Subsequently, when combined, these two subscales are able to provide a measure of full scale IQ, vital given the relatively large test battery and time constraints. The WASI is also suitable for the assessment of $6 - 89$ year olds. Specifically, the matrix reasoning subscale was chosen over the block design subscale since it is untimed and does not disadvantage any of the younger participants. Combined, the matrix reasoning and vocabulary definitions subscales can be easily and efficiently administered to a wide range of age groups to provide a full-scale measure of IQ.

5.4.3.2.2. Working Memory

Working memory has been implicated in the reading comprehension process and is highlighted as central to the inference generation process throughout Chapter 1 (Dixon, Lefevre, & Twilley, 1988; Engle, Cantor, & Carullo, 1992; Goff, Pratt, Ong, 2005;

Masson & Miller, 1983; Stothard & Hulme, 1992; Singer, Andrusiak, Reisdorf, & Black, 1992; Singer & Richot, 1996; Whitney, Ritchie, & Clark, 1991). Baddley and Hitch developed the Working Memory Model in 1974 which suggests that working memory comprises several components: (1) Central Executive, (2) Visuo-Spatial Sketchpad, (3) Phonological Loop, (4) Phonological Store, and (5) Articulatory Control Process. Three types of working memory are commonly discussed: general, verbal and visual (spatial). The importance of working memory in the inferential and comprehension processes has been found to vary as a function of the specific aspect of working memory being assessed (e.g. Cain & Oakhill, 2006a; De Beni, Palladino, Pazzaglia, & Cornoldi, 1998; Kershaw & Schatschneider, 2012; Yuill, Oakhill, & Parkin 1989). Verbal working memory has been most strongly linked with inference generation and reading comprehension (e.g. Baddeley, Logie, Nimmo-Smith, & Brereton, 1985). This is unsurprising given that successful completion of verbal working memory tasks is often dependent on sentence comprehension skills (e.g. Daneman & Carpenter, 1980; Siegel & Ryan, 1989). Establishing the relationship between verbal working memory and inference generation and comprehension is, therefore, difficult with results likely to be biased and multicollinearity high. The role of general working memory has been implicated in reading comprehension and text production, such that whilst the reader may rely on several skills to integrate information and create a coherent mental representation, when working memory capacity is limited all of these skills cannot be utilised (Kintsch & van Dijk, 1978; Oakhill & Cain, 2011; van den Broek, 2012). Subsequently, working memory constraints exert a general restriction across the comprehension and inference generation processes that is not limited to verbal processing only. A measure of general working memory capacity, specifically one distinct from comprehension skill, was thus required.

The Backwards Digit Span Task from the British Ability Scales II was chosen to assess working memory in the current research (Elliott, Smith, & McCulloch, 1996). In this task, participants hear a series of digits, which increase in length on each trial. Participants must recall the digits in reverse order. Accuracy is recorded. The backwards digit span task is argued to provide a general measure of working memory capacity, such that the need to maintain and manipulate information places high demands on the attentional element of working memory (Gathercole, 2008). Moreover, it is not dependent on sentence comprehension skills (Case, Kurland & Goldberg, 1982) and is quick and easy to administer.

5.4.3.2.3. Single-Word-Reading

In the current research, the York Assessment of Reading for Comprehension (YARC) Primary and Secondary was used to assess single-word reading (Snowling, Stothard, Clarke, Bowyer-Crane, Harrington, Nation, Truelove, & Hulme, 2009; 2010). See Section 5.4.3.3.1. for justification and description.

5.4.3.3. Higher-Order Literacy Measures

5.4.3.3.1. Reading Comprehension

A range of standardised measures exist to assess reading comprehension ability in CYP. In the current research, the York Assessment of Reading for Comprehension (YARC) Primary and Secondary was used to assess reading ability (Snowling et al., 2009; 2010). Participants are presented with a page of words and read the words they know. Accuracy is recorded, providing a measure of single-word-reading ability. Passage readability is then selected based on the CYP's single-word-reading ability. The child reads the passage aloud (silently in secondary version). Time taken to read the text is recorded providing a measure of reading fluency. Participants are then asked a series of literal and inferential questions. Participants respond verbally. Accuracy is recorded providing a measure of reading comprehension. The YARC employs an open-ended question format, allowing for the qualitative analysis of answers. Possible causes of comprehension failure can thus be explored. The YARC also has other advantages.

First, many other tools used to assess reading place heavy demands on decoding skills (Sponner et al., 2004). For instance, in the Neale Analysis of Reading Ability (NARA; Neale, 1999; Neale, Christophers, & Whetton, 1989) participants begin reading the same passage. Additional passages are then read until a certain number of reading errors are made. Assessment of reading comprehension can thus be a time-consuming process. Conversely, for those with weak decoding abilities the testing will stop very quickly. However, the YARC only requires two short passages to be read and these are determined by the participant's performance on the single-word-reading task. Moreover, for those with particularly weak word-reading abilities supplementary passages are provided. These passages are read aloud and errors corrected. Therefore, weak decoding ability does not impair comprehension performance. Second, by combining the Primary and Secondary YARC sets a relatively wide age-range can be assessed (4-16 years). Although

the YARC Secondary is not designed for the assessment of adult reading comprehension abilities, it was decided that this task would also be used to assess adults in the current study to ensure consistency and thus comparability. The YARC Secondary was piloted on adults (n=5) and ceiling effects were not observed on either the single-word-reading or passage reading tasks.

5.4.3.3.2. Text Production

A narrative production task from picture prompts was used to assess text production skills. Participants were given 10 minutes to write a story based on six black and white line drawings, using the same characters as those in the IST. Picture prompts were used since these have been successfully used with different age groups in oral narrative production tasks (e.g. Shapiro & Hudson, 1991) and written language tasks (e.g. Cain, 2003). Asking participants to write without such prompts can disadvantage those with limited experience or story exposure. To assess the quality of the text productions the Wechsler Objective Language Dimensions analytic scoring system was used (Wechsler, 1996). This system comprises subscales assessing ideas and development, organisation, unity and coherence, vocabulary, sentence structure and variety, grammar and usage, and capitalisation and punctuation.

5.4.3.4. Predictor Variables
5.4.3.4.1. Amount and Accessibility of Knowledge

Amount of general knowledge has been found to correlate with inference generation skill, however, even those with the highest levels of general knowledge cannot generate the target inference if they cannot readily access the specific target knowledge (e.g. Cromley & Azevedo, 2007; Barnes et al., 1996; Tarchi, 2010; 2012). Subsequently, a common cause of inference generation failure is insufficient knowledge and/or insufficient access to this knowledge (Barnes et al., 1996; Kintsch, 1994). It also follows that participants will not experience real-world interference when generating a counterfactual-world inference if they do not possess and/or cannot readily access said real-world knowledge. Previous research exploring the relationship between amount of knowledge and inferential abilities has employed general topic knowledge tests, multiple-choice questionnaires, and free recall tasks to assess amount of knowledge (e.g. Barnes et al., 1996; Cain, et al., 2001). Timed responses on free recall tasks have been used to assess accessibility of knowledge (e.g. Barnes et al., 1996; Cain et al., 2001). In this thesis, a

timed sentence verification task was used to assess both amount and accessibility of knowledge for several reasons.

It was essential that the knowledge-base assessed was directly associated with each item in the Image Selection Task. The administration of three tasks (to assess inference generation, amount of knowledge, accessibility of knowledge) tapping the exact same knowledge-base could, however, cause problems if administered in the same session, such that one task could prime responses on another. It was approximated that, for most participants, data collection would take place over two sessions. Consequently, one task suitable for the assessment of both amount of knowledge and accessibility of knowledge was needed so that inference generation could be assessed in Session 1 and then knowledge in Session 2.

Both multiple-choice and free recall tasks can be designed to tap a specific knowledge-base, with response times being recorded, making them suitable for the simultaneous assessment of both amount and accessibility of the target knowledge. As discussed in Chapter 2, text-based multiple-choice tasks place additional processing demands on participants that tax children's limited resources more so than adults'. In the researcher's experience, CYP, even when asked a very specific question can go off topic, thus assessment of the target knowledge may not have been obtained if a free recall task was used. Some have minimised this problem by first teaching a knowledge-base (e.g. Cain et al., 2001). However, for counterfactual-world items, this would mean that the counterfactual-world information was added to the knowledge-base calling into question its 'counterfactual nature.' Therefore, neither a multiple-choice or free recall task was suitable for the current research.

A timed sentence verification task can be designed to assess a specific knowledge-base. Since sentences are presented one at a time processing costs are reduced compared to a multiple-choice task, making this task suitable for the assessment of a range of age groups. A 72-item timed sentence verification task was created to assess amount and accessibility of knowledge. 36 items required a true response and were directly associated with the knowledge-base used in the IST. The remaining 36 items were lexically and syntactically similar fillers, which required a false response. These items were not related to the knowledge-base used to develop the IST. The task was piloted on adults (n=5) and children aged 9-11 years (n=5). The instructions were found to be clear and

161

understandable. Floor effects were not observed, however, some adults did perform at ceiling on amount of knowledge. However, given that the IST was designed based on knowledge learnt during Key Stages 2 and 3 this was expected.

5.4.3.4.2. Inhibition

Due to associative mechanisms of knowledge activation and thus the presence of related, but irrelevant information, successful real-world inference generation has been found to be underpinned by inhibitory control, such that the reader must create a clear workspace, if inferences are to be generated efficiently (Barnes et al., 2004; Cain, 2006; Harnishfeger, 1995). Successful counterfactual-world inference generation is also argued to be, at least in part, dependent on an individual's ability to inhibit a familiar response (i.e. real-world information) and respond in a new way (i.e. counterfactual-world inference; Graesser et al., 1998). The Rule-Shift-Cards task, Go/No-Go task and the Stroop task assess the ability to suppress an unwanted response or information (Gomez, Ratcliff, & Perea, 2007; Salthouse & Meinz, 1995). However, the Go/No-Go task is dependent on motor skills, and is thus a measure of response inhibition rather than interference suppression. Conversely, in the Rule-Shift-Cards task the conflict between the familiar response and the new rule must be resolved before responding. There is not merely a simple motor response, but a decision between which rule to attend to and which rule to ignore. Moreover, unlike the Stroop task the Rule-Shift-Cards task is not dependent on reading abilities. As such, the Rule-Shift-Cards task is best suited to tap the inhibitory processing associated with successful counterfactual-world inference generation. Therefore, the Rule-Shift-Cards task from the Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie & Evans, 2003) was used to assess inhibition abilities. In this task, participants see a series of cards. As each card is presented participants must respond 'true' or 'false' depending on the rule given. Participants see the cards again, but must respond according to a new rule. Errors are recorded. The Rule-Shift-Cards task was chosen because it assesses ability to inhibit a familiar pattern of responding and respond in a new way. Specifically, the rule-shift-cards task provides a measure of intrusion errors, such that information that was once relevant is no longer relevant, but continues to 'intrude' on processing. This is similar the role of real-world knowledge when processing counterfactual-world information.

162

Whilst the Rule-Shift-Cards Task is designed for the assessment of adults (16 years and over), the Play Card test in the Behavioural Assessment of the Dysexecutive Syndrome in Children (BADS-C) has the same format and same rules. Therefore, since raw scores as opposed to standard scores or profile scores were used, the Rule-Shift-Cards Task from the BADS was used with all participants to allow for comparability. The Rule-Shift-Cards task was piloted on children aged 9-11 years (n=5). All participants understood the task and did not display ceiling or floor effects. Finally, the Rule-Shift-Cards Task is easily and efficiently administered.

5.4.3.4.3. Belief Bias

As discussed in Section 1.3.3. belief biases may guide suppression during the inference generation process. In line with previous research exploring belief biases, a reasoning from false premises task was used to assess belief biases in the current research (e.g. Markovits & Schroyens, 2007; Moutier, et al., 2006). Five items presented a true premise and true conclusion (control), five items presented a true premise and false conclusion (filler a), five items presented a false premise and true conclusion (experimental), and five items presented a false premise and false conclusion (filler b), see Table 5.2. for example stimuli. Stimuli were acquired through personal communication with Claudine Bowyer-Crane. Participants must decide if the conclusion is true or false based on the preceding premise. Typically, measures of probability ratings and accuracy are obtained. It is assumed that a belief bias is present if either (a) conclusions in the experimental condition are rated has having lower probabilities than conclusions in the control condition or (b) more errors are made in the experimental condition than the control condition. It is argued that errors made on reasoning from false premises tasks can be attributed to weak inhibitory processes rather than poor logic (Houdé, 2000; Moutier, Angeard & Houdé, 2002). Consequently, a belief bias may be held, but not result in an error if inhibitory skills are strong. A measure of response time was also recorded in the current research. In accordance with the commonly accepted assumption that information which is more cognitively taxing takes longer to process (Donders, 1968;1969) it stands that stronger belief biases that cause more interference and thus require stronger inhibitory processes will result in longer response times. By calculating a difference score (control − experimental), strength of belief bias was gained whilst controlling for differences in reading speed and general processing speed. The task was piloted on adults

(n=5) and children aged 9-11 years (n=5), with all participants finding the instructions clear, the task easy to complete, and no ceiling or floor effects being observed.

Table 5.2. Example Reasoning from False Premises Stimuli

Condition	Premise	Conclusion
True Premise, True Conclusion	All fish live in water, a trout is a fish.	A trout lives in water.
True Premise False Conclusion	All fish live in water, a trout is a fish.	A trout lives in trees.
False Premise, True Conclusion	All fish live in trees, a trout is a fish.	A trout lives in trees.
False Premise, False Conclusion	All fish live in trees, a trout is a fish.	A trout lives in water.

5.4.4. Test Scores

When standardised measures are employed three types of score are available: (1) age-equivalent score, (2) standardised score, and (3) raw score. This research was concerned with assessing change, as such, neither age-equivalent or standard scores were appropriate. Age equivalent scores are not on an interval scale and thus the difference between 9:2 and 10:2 is rarely the same as the difference between 11:2 and 12:2 since rates of development differ. Similarly, standardised scores cannot determine change as they are linked to age-related expectations - whilst an 11-year-old may demonstrate significantly superior performance on a task compared to a 9-year-old, if both are performing as expected for their age group they will receive the same score. Therefore, on all tasks (except reading comprehension, for which reading comprehension age was used), raw scores were used as they provide detail of actual level of attainment and thus changes with age can be observed.

5.5. Summary

This thesis is concerned with exploration of real-world and counterfactual-world inferential abilities during adolescence, the cognitive skills and mechanisms underpinning this development, and the role of these skills in reading comprehension and text production. Before running the studies, careful consideration was given to the methodology used. A cross-sectional design was employed to capture the age-range at which development is occurring. A representative CYP sample was recruited by inviting a large number of schools to take part in the research and providing specific selection

criteria. Ethical considerations were made to ensure all participants were able to provide informed consent. The test battery was also carefully compiled, with age and capabilities of participants considered. Furthermore, tasks were ordered to reduce fatigue and tasks were selected to result in efficient data collection. Table 5.3. provides a summary of the test battery. Finally, raw scores were used to assess changes with age.

Table 5.3. Summary of Test Battery

Task	Purpose
The Image Selection Task	Specially designed for this thesis to assess inference generation abilities
Matrix Reasoning and Vocabulary Subscales from the WASI	Selected for their ability to efficiently provide a full measure of IQ
Backwards Digit Span Task	Selected to provide a general measure of working memory
The York Assessment of Reading Comprehension (Primary and Secondary)	Selected for the efficient assessment of comprehension abilities in a wide age range
A Written Narrative from Picture Prompt Task	Selected to support those with limited experiences of story knowledge
A Timed Sentence Verification Task	Selected to assess amount and accessibility of knowledge at the same time
The Rule-Shift-Cards Task	Selected to assess the inhibitory control implicated in suppressing a known way of processing for an alternative method of processes
A Reasoning from False Premises Task	Response time is measured to provide insight into the strength of belief biases, even when successfully overcome

CHAPTER 6 THE ROLE OF INFERENTIAL ABILITIES IN HIGHER-ORDER LITERACY SKILLS DURING ADOLESCENCE

6.1. Introduction

A literate individual is characterised by not only the ability to read and spell words, but the ability to extract the intended meaning from a text and produce a coherent piece of writing (e.g. Bain, 2006). Therefore, this study explores the role of inferential skill in reading comprehension and text production in 9-10 year olds, 11-12 year olds, 13-14 year olds, and adults.

6.1.1. Inference Generation and Reading Comprehension

Successful reading comprehension is characterised by the construction of an integrated and coherent mental model in which two or more textual propositions are connected using the reader's background knowledge (e.g. Kintsch, 1988; van den Broek et al., 2005; van den Broek et al., 1999). Background knowledge may also be used to elaborate on textual propositions, with information regarding possible outcomes, instruments used, and physical attributes of characters and objects also being integrated into the model (Graesser et al., 1994; Kaup et al., 2007). Inference generation is thus central to many models of proficient reading comprehension (Kintsch, 1988; van den Broek et al., 2005; Zwaan, 2003). Inference generation has also been implicated in the development of reading comprehension in CYP (e.g. Cain & Oakhill, 2012; Cain et al., 2004; Verhoeven & van Leeuwe, 2008). For instance, Cain and Oakhill (2012) found that inference generation skill at age 8-9years was a longitudinal predictor of reading comprehension skill age 10-11 years. Similarly, many CYP identified as poor comprehenders are characterised by weak inferential skills (e.g. Cain & Oakhill, 2006a), with interventions designed to develop inference generation skills improving both inference generation and comprehension skills (e.g. Dewitz, Carr & Patberg 1987; McGee & Johnson, 2003). Inference generation abilities thus appear to underpin reading comprehension in CYP, rather than being a consequence of reading comprehension or an incidental correlate.

UK government statistics suggest that some CYP, who were developing typically during Primary School (i.e. achieving age-related expectations), begin to struggle to comprehend

the texts they are presented with in Secondary School (Department for Education, 2015a; 2015b). This is worrying given that the importance of comprehension skills to academic success increases with age as the focus moves from 'learning-to-read' to 'reading-to-learn' (Chall 1996; Snow et al., 1999). In recent years, researchers have identified the need to improve adolescent comprehension (see Cassidy et al., 2010). This has sparked research into the exploration of the relationships between reading comprehension and inference generation during adolescence. Specifically, Barth et al. (2015) found that inferential skill was a predictor of reading comprehension in both adequate and struggling adolescents aged 11-18 years, accounting for 2-5% of unique variance. Similarly, Cromley and Azevedo (2007) found that inference generation made a significant contribution of 19.1% to the reading comprehension abilities of adolescents aged 14 years, after controlling for word-reading, background knowledge, strategy, and vocabulary. Previous research thus implicates inference generation in the comprehension process of adolescents, however, several questions remain unanswered.

First, the predictive utility of inference generation in the comprehension process of adolescents appears to be inconsistent. Cromley and Azevedo (2007) found that inference generation accounted for around four times more variance in reading comprehension skill than Barth et al. (2015). This discrepancy may be due to the tools used to assess inference generation. Barth et al. (2015) used inference generation measures in which the knowledge needed to successfully generate the target inference was presented in the text. Conversely, Cromley and Azevedo (2007) used an inference generation measure which was dependent on the reader activating their own background knowledge – more typical of the natural reading process (Barnes et al., 1996; Kintsch, 1993). As such, Barth et al. (2015) may have underestimated the role of inference generation in reading comprehension skill due to the use of an inference generation measure that did not fully assess all aspects of the inference generation process.

Second, the potentially fluctuating role of inference generation in the reading comprehension process during adolescence has not yet been explored. Cromley and Azevedo (2007) only explored the relationship between inference generation and reading comprehension in one age group. Whilst Barth et al. used a wide age-range (11-18years), instead of creating separate models for each age group, age was controlled for, meaning the developmental trajectory of the role of inference generation abilities is unclear. Third, there are several types of inference (see Section 1.1.1.). The research discussed above has

focused on coherence inferences. However, elaborative inferences result in the construction of a more enriched mental model and thus deeper comprehension (e.g. Graesser et al., 1994; Kintsch, 1994). It is possible then that elaborative inferences play a key role in adolescent comprehension, given the shift in importance from explicit to implicit understanding (e.g. Leach et al., 2003). Fourth, inferences can be generated from information that violates real-world beliefs, these are known as counterfactual-world inferences. The generation of counterfactual-world inferences is thought to be more cognitively demanding than the generation of real-world inferences (e.g. Dorfman, 1989; Dorfman & Brewer, 1988; Graesser et al., 1998). However, the specific role of counterfactual-world inference generation abilities in the reading comprehension process has not been explored. In sum, further research is needed to explore the nature of the relationship between reading comprehension and inference generation during adolescence, specifically, the individual and potentially unique role of different inference types needs to be explored.

6.1.2. Inference Generation and Text Production

Text production is a multi-faceted process, resulting from the dynamic interaction of several skills, with proficient mental model construction and thus inference generation central to the text production process (Hayes, 1996; 2012). However, the role of inference generation in text production quality has not been explicitly explored in CYP. The knowledge-telling model, originally proposed by Bereiter and Scardamalia (1987) and updated by Hayes (2012), conceptualises developments in textual coherence, organisation and elaboration. Three distinct stages of writing are proposed: knowledge-telling, knowledge-structuring, and knowledge-transforming. The knowledge-telling stage reflects the writing of CYP aged 5-9 years, who are typically found to produce a series of unconnected statements about a particular topic. The writing process relies primarily on long-term memory and transcription skills. The knowledge-structuring stage was added by Hayes (2012) based on research conducted with CYP aged 10-13 years. Hayes found that during this time, CYP produced more ideas than younger writers, such that initial ideas were elaborated on. Some ideas were linked using the writer's background knowledge. This is supported by research showing that at around 9 years of age, connectives such as 'because' can be seen in CYP's writing (Fayol, 1991). At this age, and into adulthood, the use of a connective has been found to be determined by the

strength of relationship between the actions, events or states in two adjoining phrases or sentences (Fayol & Lete, 1987). Both coherence and elaborative inference generation skills may thus become important to text production during this stage as the writer begins to link, build relationships and elaborate on their ideas using their background knowledge.

In the knowledge-transforming stage, ideas are selected, elaborated on, and then organised to meet the audience's needs. The construction of more organised and coherent texts appears to stem from prolonged periods of planning and revising. Chanquoy, Foulin, and Fayol (1990) found that adults paused for significantly longer periods than CYP before beginning writing, suggesting that adults spend longer planning and revising than CYP. Inferential processes are thought to be essential to both planning and revision. Specifically, planning requires the integration and elaboration of ideas, as such, in many cases inferences must be generated (Keys, 1999). Similarly, the process of revision requires the writer to read back the text they have written to determine coherence, MacArthur (2012) thus suggests that those skills essential for reading comprehension are also essential to the revision process. The transition between the three stages reflects a growing need for coherence, integration, organisation and elaboration of ideas. Inference generation may thus play a potentially increasing role in text production quality during the adolescent years. However, the explicit role of inference generation in text production during adolescence is yet to be established.

6.1.3. The Current Study

Inference generation is implicated in the process of both reading comprehension and text production during adolescence. However, potential age-related changes in the role of inference generation abilities in reading comprehension and text production during adolescence have not been fully explored. The unique contribution of different inference types has also been overlooked, with research either failing to specify the type of inference explored or focusing on real-world coherence inferences. The current study thus sought to establish the role of inference generation skill in reading comprehension skill and text production quality in Year 5 (9-10 year olds), Year 7 (11-12 year olds), Year 9 (13-14 year olds) and adults (18 years and over). It was predicted that inference generation skill would make a unique and significant contribution to reading comprehension skill and text production quality in all age groups, even after controlling for IQ, word-reading skill and working memory capacity.

The study also sought to explore the unique role of real-world coherence, real-world elaborative, counterfactual-world coherence, and counterfactual-world elaborative inference generation skill. Coherence inferences are essential for basic understanding, whereas elaborative inferences result in a deeper understanding (Graesser et al., 1994). It was thus predicted that the role of coherence inference generation abilities in reading comprehension would decrease with age, regardless of plausibility. Conversely, the role of elaborative inference generation abilities was predicted to increase with age. For text production, research suggests that both textual coherence and content (number of ideas) improve with age (Berman & Slobin, 1994; Stein & Trabasso, 1981; Wigglesworth, 1997). Subsequently, the predictive utility of both coherence and elaborative inference generation abilities was predicted to increase with age.

6.2. Method

6.1.1. Participants

51 CYP from Year 5 (9-10yers), 50 CYP from Year 7 (11-12years), 50 CYP from Year 9 (13-14years), and 56 adults (18years and over) were recruited. Table 6.1. provides a summary of demographics for all age groups. All participants met the selection criteria – see Chapter 5. None of these participants had participated in any previous research.

Table 6.1. Number of females and mean age (in years) for Year 5, Year 7, Year 9 and Adults

	Year 5	Year 7	Year 9	Adults
N	51	50	50	56
Number of females	31	28	26	50
Age (SD, Range)	10:03	12:01	14:01	19:03
	(0:06,	(0:04,	(0:05,	(1:02,
	9:03-10:10)	11:06-12:07)	13:01–14:09)	18:02-23:11)

6.1.2. Design

A cross-sectional design with four age groups: Year 5, Year 7, Year 9, and adults (18 years and over) was employed to explore the predictive utility of real-world coherence, real-world elaborative, counterfactual-world coherence and counterfactual-world elaborative inference skill (number of errors) in reading comprehension skill and text production quality. The control variables were intelligence (non-verbal and verbal), single word reading ability, and working memory capacity. The predictor variables were real-world coherence skill, real-world elaborative skill, counterfactual-world coherence skill

and counterfactual-world elaborative inference skill[25]. The outcome variables were reading comprehension age and text production quality.

6.1.3. Materials

6.1.3.1.   Inference Generation

The Image Selection Task (IST) was used to assess inference generation – see Section 4.3.1.3.1. for a description of the task used. Real-world coherence inference generation skill, real-world elaborative inference generation skill, counterfactual-world coherence inference generation skill and counterfactual-world elaborative inference generation skill were explored. The mean number of errors score for each condition was used in analysis.

6.1.3.2.   Reading Comprehension

The York Assessment of Reading for Comprehension (YARC; Snowling et al., 2009; 2010). Primary and Secondary was used to assess reading ability. Participants were presented with a page of words and asked to read the words they knew. Passage readability was then selected based on the participant's single-word-reading ability. Participants read the passages aloud, without any help from the researcher. Participants were then asked a series of literal and inferential questions to provide a measure of comprehension abilities. Participants responded verbally. Correct answers received a score of one and incorrect answers a score of zero. A total comprehension score was calculated by summing all correct responses. Since participants read texts of different levels, a reading comprehension age score was calculated to explore age-related changes.

6.2.3.1.Text Production

A narrative production task from picture prompts was used to assess text production skills. Participants were given 10 minutes to write a story based on six black and white line drawings. To assess the quality of the text productions the Wechsler Objective Language

---

[25] Time-Course (speed) data was also used as predictor variables in this study and study 5 reported in Chapter 8. However, this analysis did not result in significant findings and/or clear patterns. Therefore, it is not included in this thesis.

Dimensions analytic scoring system was used (Wechsler, 1996). This system comprised six subscales assessing ideas and development, organisation, unity and coherence, vocabulary, sentence structure and variety, grammar and usage, and capitalisation and punctuation. Each subscale received a score of between one and four, with one indicating poor quality and four excellent quality. A total quality score was calculated by summing scores on all subscales.

6.2.3.2. Control Measures

IQ, word-reading skill and working memory capacity were controlled for in the current study as these skills have been implicated in reading comprehension and text production skill (Carver, 1990; Dixon, Lefevre, & Twilley, 1988; Goff et al., 2005; Masson & Miller, 1983; Nation, Clarke, & Snowling, 2002; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000; Singer, Andrusiak, Reisdorf, & Black, 1992; Singer & Richot, 1996; Stanovich, 1993; Stanovich, Cunningham, & Freeman, 1984; Whitney, Ritchie, & Clark, 1991).

6.2.3.2.1. Intelligence Measures

The Matrix Reasoning Subscale of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was used to assess non-verbal reasoning. Participants were presented with patterns with one piece missing and asked to select the picture that completed the pattern. Participants scored one for a correct answer. The Vocabulary Definitions Subscale of the WASI was used to assess verbal reasoning. Participants heard a word and were then asked to define it. Participants scored two for a full answer and one for a partially correct answer. Raw scores are reported.

6.2.3.2.2. Working Memory Measure

The Backwards Digit Span Task from the British Ability Scales II (Elliot et al., 1996) was used to assess working memory capacity. Participants heard a series of digits, which increased in length. Participants were asked to recall the digits in reverse order. Working memory capacity was indexed by the number of digit strings recalled correctly – see Section 5.4.3.2.2. for more detail. An accuracy score was used in this study.

6.2.3.2.3. Word-Reading Measure

The York Assessment of Reading Comprehension (Primary and Secondary) was used to assess single-word-reading ability. Participants were presented with a page of 60

(Primary) or 70 (Secondary) words and asked to read as many as possible. Word-reading skill was indexed by the number of words read correctly. An accuracy score was used in this study.

6.2.4. Procedure

All tasks were administered as part of a larger test battery used to obtain data for Studies 3, 4, and 5. Written consent was gained from all CYP using a three-tiered approach as described in Section 5.3.3. Written consent was gained from adults. Participants were assessed individually in a quiet area of their school or university. Before the test battery was administered the researcher spent 10-15minutes talking to the participant to make them feel comfortable. Tests were administered across two sessions – see Appendix 5.3 for order of tasks. The first session lasted approximately 30-40 minutes for adults and 40-60 minutes for CYP. The second session lasted approximately 30-40 minutes for adults and 40-60 minutes for CYP. Participants were briefed on each task before starting it and reminded of their right to withdraw from the study or task. At the end of the final session participants were thanked and debriefed, with a debrief letter being sent out to the parents of all CYP.

6.3. Results

6.3.3. Age-Related Changes in Reading Comprehension, Text Production, IQ, Single Word Reading Skill and Working Memory Capacity

Age-related changes in reading comprehension skill, text production quality, IQ, word-reading skill and working memory were explored using ANOVAs. Descriptives for inference generation skill can be found in Table 6.2., for completeness. However, a full exploration of age-related changes in inference generation skill is not presented until Chapter 7.

6.3.3.1.Preliminary Analysis

Inspection of the data revealed that the assumption of homogeneity of variance was met for all variables, however, the data were skewed for matrix reasoning, vocabulary, and word-reading skill. Outliers were left in for all measures as they were thought to reflect true scores. Across all age groups, this led to one outlier being included for matrix reasoning and vocabulary and two outliers being included for word-reading skill. Non-

parametric tests were conducted for matrix reasoning, vocabulary, and word-reading skill. Means and standard deviations for word-reading skill, matrix reasoning, vocabulary, working memory capacity, reading comprehension skill, and text quality are presented in Table 6.2. for each year group. Table 6.2. shows that most skills appear to improve with age. A series of one-way independent ANOVAs were conducted to explore age-related changes in reading comprehension, text production, matrix reasoning, vocabulary, word-reading, and working memory capacity further.

Table 6.2. Descriptive Statistics (Means and Standard Deviations) for Single-Word-Reading, Vocabulary, Matrix Reasoning, IQ, Decoding, Working Memory Capacity, Real-World Coherence Inference Errors, Real-World Elaborative Inference Errors, Counterfactual-World Inference Errors, Counterfactual-World Elaborative Inference Errors, Reading Age and Text Production Quality for All Age Groups.

	Year 5 Mean (SD)	Year 7 Mean (SD)	Year 9 Mean (SD)	Adults Mean (SD)	Minimum Score Possible	Maximum Score Possible
Vocabulary Raw Score	23.2 (5.26)	30.16 (5.56)	35.91 (5.60)	40.68 (6.82)	0	Age 9-11: 64 Age 12-16: 72 Age 17-89: 80
Matrix Reasoning Raw Score	18.33 (5.84)	22.89 (3.86)	21.93 (5.56)	25.84 (4.88)	0	Age 9-11: 32 Age 12-44: 35
IQ Raw Score	41.41 (8.88)	53.04 (7.58)	57.85 (8.76)	66.52 (9.74)	0	Age 9-11: 96 Age 12-16: 107 Age 17-89: 115
Single Word Reading	44.77 (7.47)	54.56 (6.44)	57.88 (5.87)	63.91 (4.79)	0	70
Working Memory Capacity	14.35 (3.48)	16.60 (3.61)	17.85 (4.06)	19.16 (4.89)	0	30
Reading Comprehension Skill	107.82 (13.50)	165.10 (29.89)	149.77 (29.88)	179.57 (15.57)	0	Age 5-11: 144 Age: 11 and over: 192
Text Production Quality	9.42 (2.65)	11.30 (2.76)	11.63 (3.66)	15.38 (2.90)	6	24
Overall Inference Generation Accuracy	8.61 (3.13)	6.71 (3.34)	6.42 (3.23)	5.47 (2.94)	0	24
Real-World Coherence Inference Generation Accuracy	1.91 (0.97)	1.47 (1.24)	1.75 (1.36)	1.18 (1.25)	0	6
Real-World Elaborative Inference Generation Accuracy	2.14 (1.54)	1.94 (1.52)	1.25 (1.54)	1.43 (1.46)	0	6
Counterfactual-World Coherence Accuracy	2.51 (1.58)	2.00 (1.48)	2.33 (1.52)	1.57 (1.51)	0	6
Counterfactual-World Elaborative Accuracy	2.05 (1.77)	1.29 (1.47)	1.08 (1.32)	1.29 (1.45)	0	6

6.3.3.2.The Development of Reading Comprehension

A significant main effect of age was found for reading comprehension age (F (3,196) = 53.87, $p < .001$, $\eta p^2 = .06$). Six independent t-tests were conducted to explore further. Year 5 were found to have a significantly lower reading comprehension age than all other age groups (Y5 vs Y7: t (99) = 8.38, $p < .001$; Y5 vs Y9: t (99) = 6.67, $p < .001$; Y5 vs adults: t (105) = 12.48, $p < .001$). Year 7 were found to have a lower reading comprehension age than adults, although due to the correction applied this was not significant (t (104) = 2.48, $p = .02$). Year 7 were found to have a higher reading comprehension age than Year 9, due to the correction applied this difference was not significant (t (98) = 2.41, $p = .02$). Year 9 were found to have a significantly lower reading comprehension age than adults (t (104) = 5.74, $p < .001$). Due to the limited progress between Year 7 and Year 9, the number of weak comprehenders in Year 5, Year 7 and Year 9 was determined. In this thesis, a weak comprehender is defined as an individual with a reading comprehension age more than a year less than their chronological age. The percentage of weak comprehenders in each group can be found in Table 6.3. As can be seen from this table, Year 9 is comprised of a higher percentage of weak comprehenders than all other groups. This may explain the limited progress observed between Year 7 and Year 9.

Table 6.3. Percentage of Weak Comprehenders in Year 5, Year 7 and Year 9

	Year 5	Year 7	Year 9
Percentage of Weak Comprehenders	27	14	57

6.3.3.3.The Development of Text Production Quality

A significant main effect of age was observed for text production quality (F (3,196) = 27.75, $p < .001$, $\eta p^2 = .36$). Six independent t-tests were conducted to explore further. Year 5 were found to have significantly lower text production quality scores than all other age groups (Y5 vs Y7: t (99) = 2.85, $p = .005$; Y5 vs Y9: t (99) = 3.05, $p = .003$, Y5 vs Adults: t (105) = 8.86, $p < .001$). Year 7 and Year 9 were both found to have significantly lower text production quality scores than adults (Y7 vs adults: t (104) = 6.25, $p < .001$; Y9 vs adults: t (104) = 5.22, $p < .001$). However, there was found to be no significant

difference between the quality of text produced by Year 7 and by Year 9 (t (98) = .047, p = .64).

6.3.3.4. The Development of IQ

An independent measures non-parametric ANOVA was conducted to explore the effect of age group on overall IQ – see Appendix 6.1 for individual analysis for vocabulary and matrix reasoning. A main effect of age was found for overall IQ score ($\chi2$ (3, N = 207) = 105.11, p < .001). Six non-parametric independent t-tests were conducted to explore further. Year 5 were found to have lower composite IQ scores than Year 7 (U = 345.00, p < .001, N = 101), Year 9 (U = 167.50, p < .001, N = 101), and adults (U = 91.50, p < .001, N = 106). Whilst Year 7 were found to have a lower composite IQ score than Year 9, due to the correction applied this difference did not reach significance (U = 499.50, p = .01, N = 100). Adults were found to have higher composite IQ scores than Year 7s (U = 301.50, p < .001, N = 106) and Year 9s (U = 456.00, p < .001, N = 106).

6.3.3.5. The Development of Single-Word-Reading Skill

A non-parametric independent measures ANOVA was conducted to explore the effect of age on single-word-reading ability. A main effect of age was found ($\chi2$ (3, N = 207) = 115.51, p < .001). Six non-parametric post-hoc independent t-tests were conducted to explore further. Year 5 were found to read fewer words correctly than Year 7 (U = 331.50, p < .001, N = 101), Year 9 (U = 141.50, p < .001, N = 101), and adults (U = 43.50, p < .001, N = 106). Year 7 were found to read fewer words correctly than Year 9, however, due to the correction applied this failed to reach significance (U = 490.50, p = .011, N = 100). Year 7s (U = 250.50, p < .001, N = 106) and Year 9s (U = 354.50, p < .001, N = 106) were found to read fewer words correctly than adults.

6.3.1.6. The Development of Working Memory Capacity

An independent measures ANOVA was conducted to explore the effect of age on working memory capacity. A significant main effect of age was found (F (3,196) = 12.96, p < .001, ηp^2 = 0.18). Six independent t-tests were conducted to explore further. Year 5 were found to have a smaller working memory capacity than Year 7 (t (99) = 2.69, p = .008), Year 9 (t (99) = 3.83, p < .001) and adults (t (105) = 6.08, p <.001). There was found to be no significant difference between the working memory capacity of Year 7 and Year 9 (t (98) = 1.33, p = .18). However, Year 7 were found to have a smaller working memory

capacity than adults (t (204) = 3.13, p = .002). There was found to be no significant difference between the working memory capacity of Year 9 and adults (t (204) = 1.46, p = .15).

In sum, reading comprehension skill, text production quality, overall IQ score, single word reading skill appear to improve throughout adolescence, despite a period of limited progress between Year 7 and Year 9. Conversely, for working memory capacity, developments are observed between Year 5 and Year 7, with adult-like levels being achieved between Year 7 and Year 9. Taken together, for most variables, limited progress is observed between Year 7 and Year 9.

6.3.4. *The Relationships between Reading Comprehension, Text Production and Inference Generation Abilities during Adolescence*

This analysis seeks to explore the role of inference generation skill (number of errors) in reading comprehension and text production. The role of inference generation skill in reading comprehension and text production skill was analysed using multiple hierarchical regression analyses. The order of variables was selected a priori. Composite IQ scores, single-word-reading scores and working memory capacity scores were entered in Step 1 to account for relationships between reading comprehension, text production, IQ, word-reading skill and working memory capacity. Individual contributions made by each control component are not reported as this was not the aim of the study, however, they can be found in Appendix 6.2. Overall IST performance, real-world coherence inferential skill, real-world elaborative inferential skill, counterfactual-world coherence inferential skill, and counterfactual-world elaborative inferential skill were each entered, in turn, in Step 2.

To explore patterns of development and potential age-related changes in the contribution of inference generation abilities to text production and reading comprehension, separate regression analyses were conducted for each age group, as opposed to controlling for age.

6.3.4.1.Preliminary Analysis

Data screening revealed that all correlations were below .8, all VIF values were below 10 and all tolerance values were greater than .1. For most analyses the Durbin-Watson statistic was between 1 and 3, suggesting the assumption of independence was met. However, for the regression model exploring the role of real-world elaborative inference generation skill in text production quality of adults, the Durbin-Watson statistic was 0.97. Consequently, the conclusions made from this model must be made with caution. For all analyses, assumptions of homoscedasticity and linearity were met. Residual statistics were examined for extreme cases[26]. For most analyses, there were never more than three cases (out of 50) with standardised residual statistics higher than +/- 2 or more than one standardised residual greater than +/-2.5. However, the model exploring the independent contribution of counterfactual-world coherence inference generation in the reading comprehension skill of adults resulted in four participants with standardised residuals above two. The accuracy of this model must be viewed with caution. For most analyses, residual errors in the model were fairly normally distributed. For some models there was slight skew, however, this was never severe. Means were calculated for each group for performance on each task. As can be seen from Table 6.2. there were no ceiling or floor effects.

6.3.4.2. The Relationship between Inference Generation Abilities and Reading Comprehension Skill

To explore the relationship between inference generation abilities and reading comprehension skill, a hierarchical regression analysis was conducted for each group with reading comprehension skill as the outcome variable. Control variables were entered in Step 1 and total Inference Generation Score in Step 2. Total inference generation score was calculated by summing errors across all four inference conditions. Results can be found in Table 6.4.

---

[26] Field (2005) suggests that 95% of cases should have standardised residuals within +/-2 and 99% of cases should have standardised residuals within +/-2.5.

Table 6.4. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2)
With Reading Comprehension Age as the Dependent Variable and Image Selection Task
Score as Predictor, Controlling for IQ, Word-Reading Ability and Working Memory
Capacity

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Word-Reading Skill, Composite IQ, Working Memory Capacity	.121	.251	.108	.063
2	Image Selection Task (Inference)	.457*	.334*	.270*	.226*

*$p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

Inference generation abilities were found to make a unique contribution to reading comprehension age in all age groups, even after controlling for IQ, word-reading skill and working memory capacity. A positive relationship was observed, such that stronger inference generation skill was related to superior reading comprehension skill. The predictive utility of inference generation in reading comprehension skill appears to decrease with age, such that inference generation abilities account for 45.7%, 33.4%, 27.0%, and 22.6% of unique variance in Year 5, Year 7, Year 9, and adults, respectively. Fisher's r-to-z transformation found that none of the comparisons of coefficients were significant, however (*p > .20 for all comparisons*). Therefore, whilst visual inspection of the predictive utility of inference generation skill in reading comprehension suggests a decline, this is not significant.

This study aimed to explore the potentially differing contributions of four inference types in reading comprehension skill: real-world coherence, real-world elaborative, counterfactual-world coherence and counterfactual-world elaborative. Therefore, for each age group, four hierarchical regression analyses were conducted with reading comprehension as the outcome variable. To explore the independent contribution of the four inference types, inference skill for each type was entered in Step 2 sequentially. These analyses are summarised in Table 6.5.

Table 6.5. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Reading Comprehension Skill as the Dependent Variable and All Inference Conditions as Predictor Variables, Controlling for Word-Reading, IQ, and working memory capacity

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Word-Reading Skill, Composite IQ, Working Memory Capacity	.121	.251	.108	.063
2	Real-World Coherence Accuracy	.373***	.235**	.166*	.094*
2	Real-World Elaborative Accuracy	.011	.009	.137*	.109*
2	Counterfactual-World Coherence Accuracy	.007	.116	.002	.010
2	Counterfactual-World Elaborative Accuracy	.064	.033	.000	.036

*$p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

A positive relationship was observed between real-world coherence inference skill and reading comprehension, such that stronger real-world coherence inference generation skill was related to superior reading comprehension skill. The predictive utility of real-world coherence inference skill in reading comprehension skill appears to decrease with age, accounting for 37.3%, 23.5%, 16.6%, and 9.4% unique variance in Year 5, Year 7, Year 9s and adults, respectively, after controlling for IQ, word-reading skill and working memory capacity. Elaborative inference generation skill explains a significant amount of unique variance in reading comprehension skill in Year 9 (13.7%) and adults (10.9%), after controlling for IQ, word-reading skill and working memory capacity. A positive relationship was observed, such that stronger real-world elaborative inference generation skill was related to superior reading comprehension skill. The amount of unique variance in reading comprehension skill accounted for by counterfactual-world coherence inference skill approaches significance for Year 7 ($p = .051$). Neither counterfactual-world coherence nor counterfactual-world elaborative skill made a unique contribution to reading skill for any other age groups. Fisher's r-to-z transformation was used to explore differences in the unique contribution made by the different inference types further. None of the comparisons of coefficients were significant (*p > .15 for all comparisons*). This suggests that whilst the pattern of results may suggest the predictive utility of real-world coherence inference skill declines throughout adolescence, whereas real-world elaborative inference skill increases until Year 9, this is not significant.

6.3.4.3. The Relationship between Inference Generation Abilities and Text Production Quality

To explore the relationship between inference generation and text production quality a hierarchical regression analysis was conducted for each group with text production quality as the outcome variable. Control variables were entered in Step 1 and total Inference Generation Score in Step 2. Results can be found in Table 6.6.

Table 6.6. Summary of Fixed-Order Hierarchical Multiple Regression Analyses ($\Delta R2$) With Text Production Quality Score as the Dependent Variable and Image Selection Task Score as Predictor, Controlling for IQ, Word-Reading Ability and Working Memory Capacity

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Word-Reading Skill, Composite IQ, Working Memory Capacity	.459***	.291**	.165	.056
2	Image Selection Task (Inference)	.080	.185*	.215*	.240**

$*p \leq .05 ** p \leq .01 *** p \leq .001$

Inference generation skill did not account for a significant amount of variance in the text production quality of Year 5. However, results suggest that the predictive utility of inference generation skill in text production quality increases with age, with inference generation skill accounting for 18.5%, 21.5%, and 24.0% unique variance in text production quality in Year 7, Year 9, and adults, respectively, after controlling for IQ, word-reading skill and working memory capacity. Conversely, the predictive utility of the control variables appears to decline with age, with word-reading skill, IQ and working memory capacity collectively accounting for 45.9%, 29.1%, 16.5% and 5.6% unique variance in text production quality of Year 5, Year 7, Year 9 and adults, respectively. Fisher's r-to-z transformation was used to explore these patterns further. None of the comparisons of coefficients were significant (*p > .03 for all comparisons*). This suggests that whilst the predictive utility of inference skill in text production quality appears to increase throughout adolescence, whereas the predictive utility of the control variables declines, this is not significant.

This study aimed to explore the potentially differing contributions of four inference types in text production quality: real-world coherence, real-world elaborative, counterfactual-world coherence and counterfactual-world elaborative. For each age group, four hierarchical regression analyses were conducted with text production quality as the

outcome variable. To explore the independent contribution of the four inference types, inference skill scores for each inference type were entered in Step 2 sequentially. These analyses are summarised in Table 6.7.

Table 6.7. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Text Production Quality as the Dependent Variable and All Inference Conditions as Predictor Variables, Controlling for Word-Reading, IQ, and working memory capacity

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Word-Reading Skill, Composite IQ, Working Memory Capacity	.459***	.291**	.165	.056
2	Real-World Coherence Accuracy	.008	.048	.084	.225***
2	Real-World Elaborative Accuracy	.000	.129**	.121*	.069*
2	Counterfactual-World Coherence Accuracy	.071*	.051	.021	.012
2	Counterfactual-World Elaborative Accuracy	.001	.000	.003	.003

*p ≤.05 ** p ≤.01 *** p ≤.001

The predictive utility of real-world coherence inference skill in text production quality was found to increase with age, after controlling for word-reading skill, IQ and working memory capacity. A positive relationship was observed, such that stronger real-world coherence inference generation skill was related to superior text production quality. Specifically, inference generation explained 4.8% unique variance in text production in Year 7. However, this was not significant. Inference generation accounted for 8.4% unique variance in Year 9, however this contribution only approached significance (p = .059). Finally, inference generation accounted for a significant 22.5% significant unique variance in adults. Real-world elaborative inference skill was found to make a unique contribution to text production quality in Year 7 (12.9), Year 9 (12.1%), and adults (6.9%), respectively. A positive relationship was observed, such that stronger real-world elaborative inference generation skill was related to superior text production quality. For the most part, counterfactual-world inferential skill did not make a unique contribution to text production quality. However, counterfactual-world coherence skill was a unique predictor of text production quality in Year 5, accounting for 7.1% unique variance, after controlling for word-reading skill, IQ and working memory capacity. A positive relationship was observed, such that stronger counterfactual-world coherence inference generation skill was related to superior text production quality. Fisher's r-to-z transformation was used to explore these differences further. None of the comparisons of coefficients were significant (*p > .14 for all comparisons*).

184

6.4 Discussion

This study sought to establish the role of inference generation in reading comprehension skill and text production quality in Year 5, Year 7, Year 9 and adults. The results of this study extend previous knowledge in several ways. First, whilst reading comprehension skill and text production quality were found to improve from Year 5 to adulthood, progress between Year 7 and Year 9 was limited. Second, inference generation skill explained unique variance in text production quality from Year 7 onwards, after controlling for IQ, word-reading skill, and working memory capacity. Finally, the role of inference generation in reading comprehension skill and text production quality during adolescence varied as a function of inference type.

6.4.1. Age-Related Changes in Reading Comprehension Skill and Text Production Quality

Reading comprehension skill and text production quality were both found to improve from Year 5 to adulthood. However, there was no improvement in text production quality between Year 7 and Year 9 and reading comprehension skill decreased during this time (due to the correction applied this difference did not reach significance). Taken together, results suggest that Year 7 to Year 9 is a period of limited literacy progress for some CYP. This is consistent with UK government statistics which suggest around one third of CYP do not make the progress expected during the Secondary School years (Department for Education, 2015a).

6.4.2. The Developing Role of Inference Generation Skill in Reading Comprehension Skill

Inference generation was found to make a significant unique contribution to reading comprehension skill throughout adolescence, accounting for 45.7%, 33.4%, 27.0%, and 22.6% unique variance in Year 5, Year 7, Year 9 and adults, respectively, after controlling for IQ, word-reading skill, and working memory capacity. There is a discrepancy in previous research regarding the predictive utility of inference generation during the Secondary School years. Barth et al. (2015) found that inference generation accounted for only 2-5% unique variance in reading comprehension skill. Conversely, Cromley and Azevedo (2007) found that inference generation accounted for a much larger 19.1% unique variance in the reading comprehension skill of 14-15year olds. The findings of the

current study support Cromley and Azevedo (2007), such that inference generation at all ages was found to account for substantial variance in reading comprehension skill (22.6%-45.7%). Like Cromley and Azevedo (2007), the current study used an inference generation measure which was dependent on the reader activating their own background knowledge, as opposed to a measure that presented the necessary knowledge in the text. Subsequently, it appears that when an inference generation measure is used that does not minimise the contribution of those processes central to successful inference generation (e.g. knowledge activation), inference generation makes a considerable contribution to reading comprehension skill during adolescence.

The amount of variance in reading comprehension skill accounted for by inference generation in the Cromley and Azevedo (2007) study appears to be less than the amount of variance explained in the current study, however. This discrepancy may be due to the wide range of age groups assessed in the current study. Cromley and Azevedo used a sample of 14-15 year olds only. This group would fall between Year 9 and adults in the current study. When the results of Cromley and Azevedo are considered against just Year 9 (27.0%) and adults (22.6%), predictive utility explained is more comparable.

The current study extends previous research in a second way, such that visual inspection of regression coefficients, suggests the role of inference generation skill in reading comprehension declines with age between 9-10 years and adulthood. However, these differences did not reach significance. This effect had not previously been observed and is surprising given the prominent role of inference generation in skilled models of reading comprehension (e.g. Kintsch in 1988; van den Broek et al., 2005; Zwaan. 2003). Given that this decline was not statistically significant, further research is needed to establish the changing role of inference generation in the reading comprehension process during adolescence.

The current study was also unique in nature, such that the independent contribution of four different inference types was investigated. Visual inspection of regression coefficients suggests that the predictive utility of real-world coherence inference generation in reading comprehension declines with age. However, this was not significant. The converse was observed for real-world elaborative inferential skill, such that it did not make a significant unique contribution to reading comprehension skill until Year 9. However, again, changes were not significant. Despite not being significant the increases

and declines in predictive utility do fit with existing research. Real-world coherence inferences are essential to basic understanding and the maintenance of coherence, with real-world coherence skill found to improve with age (Barnes et al., 1996; Kendeou et al., 2008; Oakhill et al., 2003). The decline in importance may thus be due to real-world coherence inferential skill becoming sufficient for basic understanding. Since the cognitive demand incurred when generating coherence inferences has been found to decrease with age, with adult-like performance achieved around 12-14 years, Year 9 and adults may have more resources available for the generation of the more cognitively demanding elaborative inferences (e.g. Casteel, 1993). This would explain why elaborative inferences do not make a significant unique contribution to comprehension skill until Year 9. It was found that counterfactual-world inferences did not make a unique contribution to reading comprehension skill at any age. However, given that fantasy texts were not used to assess comprehension this is not surprising. In sum, inference generation appears central to reading comprehension during the Secondary School years, particularly when the text requires knowledge activation, with coherence and elaborative inference skill sharing unique relationships with reading comprehension.

6.4.3. The Developing Role of Inference Generation in Text Production Quality

Previous research had not explicitly explored the role of inference generation in text production quality. The current study thus extends previous research by showing that inference generation makes a unique contribution to text production quality between 11-12 years and adulthood. In contrast to the role of inference generation skill in reading comprehension process, the predictive utility of inference generation skill in text production quality was found to increase with age. Whilst a significant contribution was not made in Year 5, inference generation skill accounted for 18.5%, 21.5%, and 24.0% unique variance in Years 7, 9 and adults respectively, after controlling for IQ, word-reading skill, and working memory capacity. However, whilst a visual inspection of regression coefficients suggests an increase in predictive utility of inference generation skill, this was not significant, thus further research exploring the potentially changing role of inference generation in text production is needed.

Real-world coherence inferences did not make a unique contribution to text production quality until Year 9. Conversely, real-world elaborative inferences only made a unique contribution to text production quality in Years 7 and 9. The knowledge-telling model

suggests that writing moves through three distinct stages: 1) knowledge-telling, 2) knowledge-structuring, 3) knowledge-transforming (Bereiter and Scardamalia, 1987; Hayes 2012). The knowledge-telling stage occurs before 10 years of age and is characterised by the activation of ideas, although these may not be linked or well developed. It follows, then, that in Year 5 neither elaborative or coherence inference generation skill would make a unique contribution to text production quality since ideas are neither elaborated on nor linked together. The knowledge-structuring stage is characterised by the elaboration of ideas, with some links being made. CYP are thought to go through this stage between 10-13 years. During this period, the results of the current study suggest that elaborative inference generation skill makes a larger unique contribution to text production quality than coherence inference generation skill. This fits with CYP's growing ability to elaborate on ideas during this stage. Conversely, whilst some links may be made, which may require coherence inference skills, these links have been found to be relatively limited during this stage. The knowledge-transforming stage, typical of CYP aged 13 years and over, is characterised by a coherent representation, with the reader in mind, created through extended periods of planning and revision. It thus follows that real-world coherence inferences, central for linking ideas and maintaining coherence, do not make a significant unique contribution until Year 9 (13-14 years). Progression through the stages of the knowledge-telling model thus appears to coincide with changes in the role of the different inference types in text production quality. This highlights the possible pivotal role of inference generation skill in the text production process throughout adolescence.

Whilst real-world inferences did not make a significant unique contribution to text production quality in Year 5, counterfactual-world coherence inference generation skill did, accounting for 7.1% unique variance, after controlling for IQ, word-reading skill, and working memory capacity. This is an interesting finding that requires further exploration to be fully understood. For instance, real-world and counterfactual-world inference generation may be underpinned by different processes. Belief biases for example may underpin counterfactual-world inference generation whereas the activation of real-world knowledge may be central to real-world inference generation. If so, it is important to understand how those skills underpinning counterfactual-world inference generation operate during text production in Year 5 and how this differs to the older age groups.

Answering these questions would provide a deeper insight into the knowledge-telling stage of writing development.

6.4.4. Conclusions

This study has identified a continued role of inference generation skill in both reading comprehension and text production between the ages of 9-10 years and adulthood. Real-world coherence and elaborative inferences were found to share unique relationships with reading comprehension and text production. Specifically, the results of this study highlight the growing role of elaborative inference generation skills in the reading process as coherence inference generation skills become proficient. Although non-significant, the results of this study also support Hayes (2012) model of text production development, highlighting the growing role of coherence inference generation skill.

CHAPTER 7 A CROSS-SECTIONAL STUDY EXPLORING THE DEVELOPMENT OF INFERENCE GENERATION SKILL AND INFERENTIAL TIME-COURSE IN YEAR 5s, YEAR 7s, YEAR 9s AND ADULTS

7.1. Introduction

Research has implicated a role for inference generation skill in reading comprehension across the lifespan (Cain & Oakhill, 2012; Kintsch, 1988; Verhoeven & van Leeuwe, 2008). Exploring the development of inference generation abilities during adolescence may thus inform effective educational practice. The cross-sectional study reported here was designed to investigate age-related changes in real-world and counterfactual-world inference generation skill and changes in the inferential time-course in Year 5 (9-10years), Year 7 (11-12years), Year 9 (13-14years) and adults.

7.1.1. Real-World Inference Generation across the Lifespan

Research finds that whilst very young children have the ability to generate inferences from a narrative, they do not do this spontaneously (e.g. Omanson et al., 1978; Paris et al., 1977). With age, however, the ability to generate inferences whilst reading without prompting appears to improve, leading to increases in inferential skill (Ackerman, 1986, 1988; Barnes et al., 1996; Kendeou et al., 2008). Understanding the development of those processes underpinning whether an inference is drawn spontaneously whilst reading or not is thus important. Research exploring the inferential time-course in adults converges to support the Constructionist Theory (Graesser et al., 1994) which suggests that coherence inferences are generated whilst reading (i.e. online), whereas elaborative inferences are drawn after reading (i.e. offline). Research utilising online methodologies to explore the inferential processes of CYP suggests that, despite generating fewer inferences than adults do, children and adolescents engage in a qualitatively similar process to adults, such that they, too, generate coherence inferences online and elaborative inferences offline (Bowyer-Crane & Snowling, 2010; Casteel, 1993; Casteel & Simpson, 1991). Similarly, typically, all age groups find the generation of elaborative inferences more difficult than the generation of coherence inferences (e.g. Bowyer-Crane & Snowling, 2010; Cain & Oakhill, 1999; Cain et al., 2001). Therefore, whilst inference

generation skill improves from childhood to adulthood, the processes underpinning inference generation appear to remain the same.

The research discussed above may not reflect the current UK picture, however. Many of the studies above have relatively small sample sizes, with participants recruited from predominantly white suburban schools. Often, participants are described as 'average' readers, such that they are reading at the level expected for their age. Research and UK government statistics suggest that a number of CYP, who do not have special educational needs nor have been identified as poor comprehenders, are still struggling to comprehend the texts presented to them and produce coherent texts (Department for Education, 2015a; 2015b). These CYP may be processing text in a different way to their peers who are achieving or exceeding age-related expectations. However, they are unlikely to have been represented in the studies above. Whilst some have explored inference generation abilities at several ages, many of the studies discussed above compare just two groups, typically an adult and child group (9-10 years or younger), thus little is known about the developmental trajectory of inferential abilities *during* adolescence. Of the research that does exist, results appear inconsistent, with some suggesting that inference generation abilities develop throughout adolescence (e.g. Barnes et al., 1996) whereas others find that inference generation abilities reach adult-like levels by 13-14 years of age (e.g. Casteel, 1993). Further research is needed to advance understanding of age-related changes in inference generation abilities during adolescence in a sample reflective of the current UK school-aged population.

7.1.2. Counterfactual-World Inference Generation across the Lifespan

Research exploring counterfactual-world processing suggests that counterfactual-world information is more difficult to process than real-world information, due to the need to resolve conflict experienced due to the activation of related, but contradictory real-world knowledge – i.e. real-world interference (e.g. Alter et al., 2007; Byrne, 2002; 2007; Evans, 2006; Evans et al., 2005). However, Ceci et al. (1981) suggest that this is not always the case for CYP. Ceci et al. explored the real-world and counterfactual-world recall ability of CYP aged 7 years and 10 years. For both groups, Ceci et al. found no difference between the recall accuracy of real-world and counterfactual-world information after immediately hearing a story. However, when stories were recalled after a three-week delay, Ceci et al. found that significantly more errors were made when recalling

counterfactual-world information compared to real-world information, particularly for the 10 year olds. Specifically, the CYP failed to correctly recall counterfactual-world information about well-known characters, such that participants recalled the real, rather than counterfactual, traits of the characters. This suggests that errors were due to real-world interference.

Given that 10 year olds displayed more real-world interference when recalling counterfactual-world stories than 7 year olds, Ceci et al. concluded that when processing counterfactual-world texts, the amount of real-world interference faced increases with age. This is supported by Dorfman and colleagues (Dorfman, 1989; Dorfman & Brewer, 1988). They presented participants with fables, in which key information ran counter to existing beliefs (e.g. the hare wins the race instead of the tortoise), then asked participants to generate a thematic inference regarding the point of the fable, thus requiring participants to put aside existing beliefs. Superior performance was observed by 7-8 year olds, compared to 9-10 year olds, 12-13 year olds and adults. Dorfman and colleagues asked participants to do this immediately after hearing the story, however. The difficulty associated with generating counterfactual-world inferences is thus surprising given that Ceci et al. found that real-world interference was not present in immediate recall for any age group. The difference observed may be due to the methodological variations below.

Text can be processed at three levels, increasing in difficulty: surface, propositional and situational (Van Dijk and Kintsch, 1983). Gordon et al. (2000) suggest that participants will adopt the easiest strategy possible when completing tasks. Therefore, participants in Ceci et al.'s immediate recall condition may have only created a surface-level representation. A surface-level representation of the text is a verbatim record stored in short-term memory, retention is limited. Consequently, participants in the three-week delay condition would not be able to maintain a surface-level representation for this period. If a text is to be stored for later retrieval, a more robust situational model is needed. Unlike surface-level representations, situational-level representations are thought to be structured according to the reader's real-world knowledge (e.g. Kintsch, 1988). The construction of a surface-level representation, compared to a situational-level representation, may thus explain why significantly less real-world interference was observed in the immediate recall condition compared to the delayed recall condition in Ceci et al.'s (1981) study. Inferences are also thought to be generated primarily at the situational-level (Graesser et al., 1994; Schmalhofer et al., 2002). The task used by

Dorfman and colleagues requires a thematic inference to be generated meaning successful completion is dependent on the reader constructing a situational-level representation of the text. Taken together, it appears that real-world interference is present when completing tasks requiring the construction of situational-level representations, but not surface-level representations.

The recall task and thematic inference task employed by Ceci et al. (1981) and Dorfman and colleagues (Dorfman, 1989; Dorfman & Brewer, 1988) are offline measures of reading comprehension and inference generation. Bowyer-Crane and Snowling (2010) used a timed sentence verification task (SVT) to explore real-world and counterfactual-world inferential abilities in children aged 9-10 years. They found that accuracy of inference generation for both coherence and elaborative inferences did not differ in the real-world and a fairy story conditions. Consistent with Ceci et al. (1981), this suggests that, compared to real-world inference generation, children do not find it more difficult to generate inferences based on counterfactual information. Moreover, Bowyer-Crane and Snowling found that in both the real-world and fairy story condition, participants responded to coherence items significantly faster than elaborative items, suggesting coherence inferences are drawn online, regardless of the plausibility of the information. However, responses in the real-world condition were significantly faster than in the fairy story condition, suggesting children may find the process of counterfactual inference generation slightly more taxing than real-world inference generation.

Despite also employing a SVT, Graesser et al. (1998) found that adults displayed an accuracy advantage for real-world rather than counterfactual-world inference generation. Moreover, all adults took much longer to make the true or false decision when confronted with counterfactual-world inference items compared to real-world inference items. Graesser et al. thus concluded that, due to the presence of real-world interference, adults find the generation of all counterfactual inferences cognitively taxing and thus generate them offline. The discrepancy in online-offline patterns between Graesser et al. (1998) and Bowyer-Crane and Snowling (2010) suggests a qualitative shift in the processing of counterfactual-world information between 9-10 years and adulthood. The differences between adults and children could also be due to methodological differences between the two studies, however.

The texts used by Bowyer-Crane and Snowling (2010) were 114-144 words long, with readability ratings of 7:4years or less, whereas the texts used by Graesser et al. (1998) were 609-763 words long, with readability ratings ranging from approximately 20-40 years. Whereas Bowyer-Crane and Snowling used fairy tales in the counterfactual-world condition, Graesser et al. used texts that violated the laws of time and space – information that is far more implausible (Byrne, 2007; Zwaan et al., 1995). Additionally, the SVT followed the critical text after an almost immediate time interval (500ms) in the Bowyer-Crane and Snowling study. Conversely, the SVT was presented at the end of the text in the Graesser et al. study, meaning that the time interval between critical text and test item was longer than that employed by Bowyer-Crane and Snowling and inconsistent across texts. Difficulty associated with processing counterfactual-world information has been found to increase as length of text, complexity of text, implausibility of information, and time interval increase (Byrne, 2007; Markman, 1979; Otero & Campanario, 1990; Singer, 1994; Vosniadou et al., 1988). This may thus explain the difficulty faced by adults compared to CYP. Similarly, research shows that for real-world information, coherence inferences are generated online whereas elaborative inferences are not (e.g. Calvo & Castillo, 1996; 1998; 2001a; 2001b). This is also the pattern of results observed by Bowyer-Crane and Snowling (2010) for counterfactual-world information. However, whilst Bowyer-Crane and Snowling (2010) specify the types of inferences explored, Graesser et al. (1998) do not. It is thus possible that a) the inferences explored by Graesser et al. were predominantly elaborative inferences, b) the lack of discrimination between coherence and elaborative inferences resulted in skewed results, c) due to the differing time intervals, the task may not be sensitive to differences between coherence and elaborative inferences, since there is adequate time for both types of inference to be generated before the test item is presented. In sum, it is unclear if adults would display as much difficulty generating counterfactual-world inferences if given shorter, simpler texts comprising more plausible counterfactual-world information, with exploration and comparison of coherence and elaborative inferences. Due to limited research, it is thus unclear if the differences observed between adults and children are due to a conceptual shift in the processing of counterfactual-world information or methodological discrepancies.

7.1.3. The Current Study

The current study sought to establish age-related changes in real-world and counterfactual-world inferential abilities in Year 5, Year 7, Year 9, and adults. There was a specific focus on exploring whether the differences observed between the counterfactual-world inference generation processing difficulties of adults and children were due to a conceptual shift or methodological differences. Research suggests that real-world coherence inferences are drawn online, regardless of age whereas real-world elaborative inferences are drawn offline, regardless of age (Casteel, 1993; Poynor & Morris, 2003). The same pattern of results has been found for counterfactual-world inferences for children (Bowyer-Crane & Snowling, 2010). However, research suggests that adults draw all counterfactual-world inferences offline (e.g. Graesser et al., 1998). Therefore, two critical texts were embedded into one of eighteen stories, reflecting one of six conditions (1) real-world coherence, (2) real-world elaborative, (3) real-world literal, (4) counterfactual-world coherence, (5) counterfactual-world elaborative, and (6) counterfactual-world literal. Critical texts were then followed by a forced-choice picture-selection task, with one picture depicting the target inference (or target information in literal conditions). In the coherence condition, causal inferences were explored and in the elaborative condition, static inferences were explored. - see Chapter 3. Reading speeds of critical texts and number of errors on the forced-choice picture selection task were recorded to provide measures of the inferential time-course and inference generation skill, respectively.

7.1.3.1. Predictions

The current study predicts quantitative, but not qualitative changes in both real-world and counterfactual-world inference generation abilities during adolescence, proposing that qualitative changes in counterfactual-world processing are due to methodological discrepancies as opposed to a conceptual shift in processing. Inference generation skill and speed is predicted to improve, across all conditions, during adolescence for real-world inferences but decline for counterfactual-world inferences. Adult-like performance is predicted to be achieved for both skill and speed by Year 9. Qualitative changes during adolescence are not predicted, such that regardless of age and plausibility of information, coherence inferences are predicted to be generated with more success than elaborative inferences and coherences inferences are predicted to be generated online and elaborative

inferences offline. Due to the need to resolve the conflict experienced due to the activation of related, but contradictory real-world knowledge, the counterfactual-world process is predicted to differ to the real-world inference generation process for all age groups, such that additional processing costs are predicted when generating counterfactual-world inferences compared to real-world inferences. Therefore, it is predicted that all groups will read counterfactual-world critical texts slower than real-world critical texts.

7.2. Method

7.2.1. Participants

Those who participated in Study 3 were also included in this study.

7.2.2. Design

A 4 (Age: Year 5, Year 7, Year 9, adults) * 3 (Inference Type: Coherence, Elaborative, Literal) * 2 (Plausibility: Real-World, Counterfactual-world) cross-sectional design was employed to explore age-related changes in real-world and counterfactual-world inferential abilities during adolescence. The dependent variables were reading speed of critical texts resulting in a correct response (milliseconds) and number of errors made on forced-choice picture-selection task.

7.2.3. Materials

7.2.3.1. Inference Generation Measure

The Image Selection Task (IST) from Study 3 was used to explore inference generation skill (number of errors) and the inferential time-course (reading speeds) – See Chapters 3 and 4.

7.2.4. Procedure

The IST was administered as part of a large test battery used to obtain data for Studies 3, 4, and 5 – see Section 6.2.4 for the procedure.

7.3. Results

7.3.1. Assessing Inferential Skill

The data were inspected prior to analysis. Since two versions of each story were created, analysis was conducted to determine if one version of the stories was more difficult than the other. There was found to be no difference between the number of errors made on Version 1 and Version 2 for Year 5 (t (50) = 0.76, p = .45), Year 7 (t (49) = 0.13, p = .90), Year 9 (t (49) = 1.01, p = .33) or adults (t (55) = 0.29, p = .78). The data were collapsed across both versions.

Age-related changes in inference generation abilities were established using a mixed factorial ANOVA with number of errors as the dependent variable. The independent variables were inference type, plausibility and age. To calculate the mean number of errors for each condition, the number of errors for each participant in each condition was first calculated. These totals were then used to create a group mean for each condition - see Table 7.1. A visual summary of means can also be found in Figure 7.1.

Inspection of the skewness and kurtosis statistics revealed that for all six conditions the data were normally distributed. Homogeneity of variance was not violated. Mauchly's Test of Sphericity was significant for inference type. A parametric ANOVA was conducted to explore the effects of age, plausibility and inference type on skill (number of errors), with Huynh-Feldt statistics[27] reported for all analyses involving inference.

Table 7.1. Mean number of errors (and Standard Deviations) all conditions for all age groups (max number of errors = 6)

	Real-world			**Counterfactual-world**		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Year 5	0.63 (0.69)	1.91 (0.97)	2.14 (1.54)	0.54 (0.91)	2.51 (1.58)	2.05 (1.77)
Year 7	0.41 (0.56)	1.47 (1.24)	1.94 (1.52)	0.62 (0.85)	2.00 (1.48)	1.29 (1.47)
Year 9	0.33 (0.48)	1.75 (1.36)	1.25 (1.54)	0.29 (0.46)	2.33 (1.52)	1.08 (1.32)
Adults	0.41 (0.57)	1.18 (1.25)	1.43 (1.46)	0.25 (0.58)	1.57 (1.51)	1.29 (1.45)

---

[27] Girden (1992) suggests that Huynh-Feldt, not the Greenhouse-Geisser correction, should be used when the estimates of sphericity are greater than 0.75, as in these cases the Greenhouse-Geisser correction is likely to be too conservative.
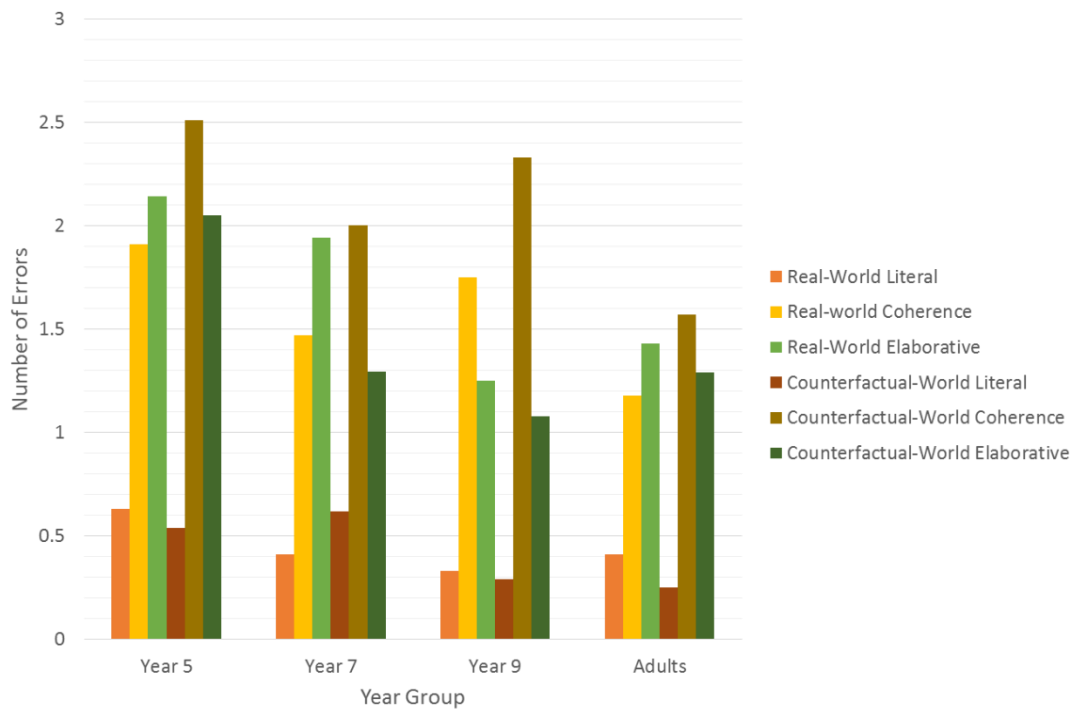
Figure 7.1. Mean number of errors (and Standard Deviations) for all six inference conditions for all age groups

A 4 (Age: Year 5, Year 7, Year 9, adults) * 3 (Inference Type: Literal, Coherence, Elaborative) * 2 (Plausibility: Real-World, Counterfactual-World) mixed factorial ANOVA was conducted. There was a main effect of age (F (3, 203) = 8.37, $p < .001$, ηp^2 = 0.14). Post-hoc comparisons were conducted to explore this main effect further. As illustrated in Figure 7.2. Year 5 were found to make more errors than Year 7, however, due to the correction applied this did not reach significance (t (99) = 2.44, $p = .02$). Year 5 were found to make significantly more errors than Year 9 (t (99) = 2.95, $p = .004$), and adults (t (105) = 4.94, $p < .001$). Whilst means suggest a steady increase in inference generation skill, none of the remaining comparisons were significant ($p > .043$ for all remaining comparisons - see Appendix 7.1).
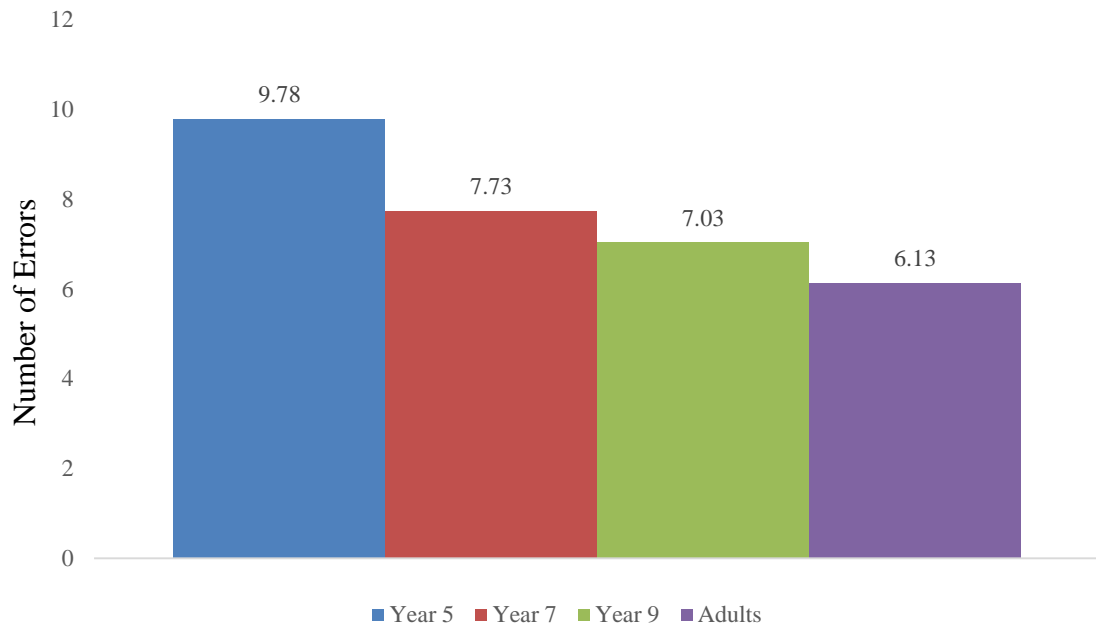
Figure 7.2. Bar Chart to show Main effect of age Across All Six Inference Conditions

There was a main effect of inference type (F (1.71, 362.08) = 123.89, p <.001, ηp^2 = 0.45). Post-hoc comparisons show that significantly fewer literal errors than coherence (t (206) = 16.16, p < .001) or elaborative errors were made (t (206) = 13.89, p < .001). Error rates were found to be higher for coherence items compared to elaborative items, however, due to the correction applied this difference did not reach significance (t (206) = 2.52, p = .013). There was no main effect of plausibility (F (1, 203) = .78, p = .38, ηp^2 = 0.01). There was a significant interaction between inference type and age (F (5.14, 362.08) = 3.01, p = .011, ηp^2 = 0.06) – see Table 7.2. for post-hoc analyses and Figure 7.3. for visual depiction. Year 5 and Year 9 were found to make significantly more coherence errors than adults. This suggests that coherence inference generation skill develops between Year 5 and adulthood. However, this progress may be steady, with a period of limited development between Year 7 and Year 9. Year 5 were found to make significantly more elaborative errors than Year 9 and adults. This suggests elaborative inference generation skill develops during adolescence, with adult-like levels achieved around Year 9. All age groups made fewer literal errors than coherence or elaborative errors. There was no significant difference between the number of coherence and elaborative errors made for Year 5, Year 7 and adults. However, Year 9 were found to make significantly more coherence than elaborative errors.
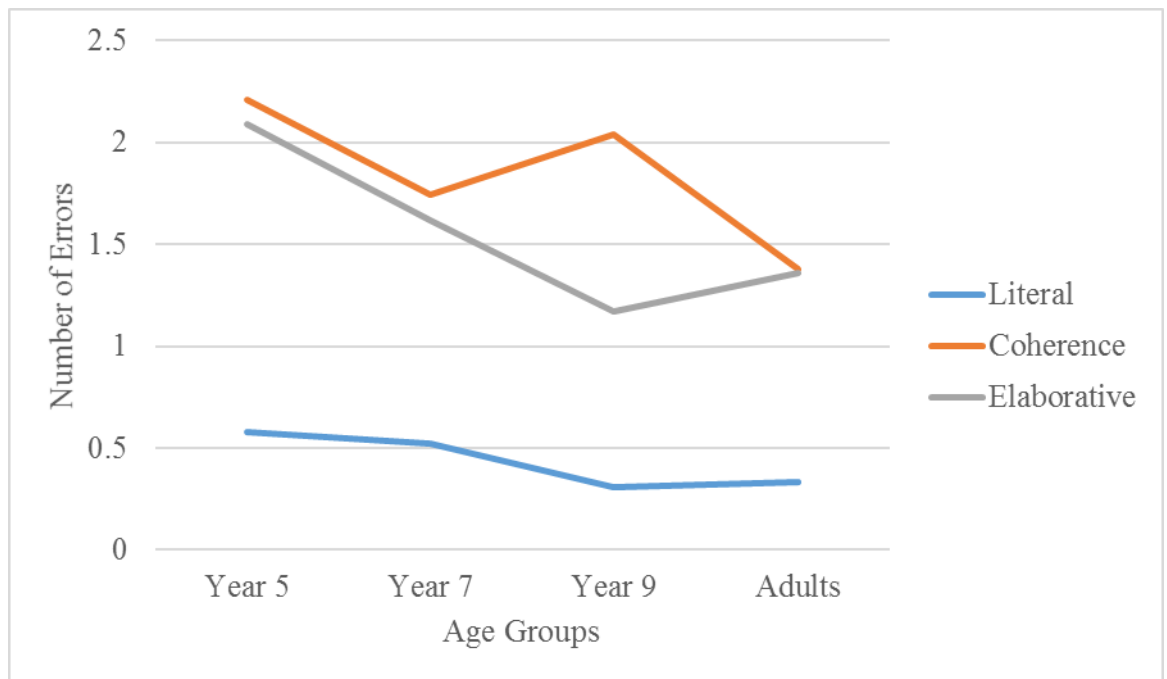
Figure 7.3. Line Graph showing the interaction between age and inference type

Table 7.2. Summary of comparisons conducted to explore the interaction between inference type and age group

Comparison	Mean (Standard Deviation)	t-value	Significance value
Year 5 Coherence vs Year 7 Coherence	2.21 (0.91) vs 1.74 (0.86)	2.31	.02
Year 5 Coherence vs Year 9 Coherence	2.21 (0.91) vs 2.04 (1.04)	0.68	.50
Year 5 Coherence vs Adults Coherence	2.21 (0.91) vs 1.38 (1.06)	4.13	.000**
Year 7 Coherence vs Year 9 Coherence	1.74 (0.86) vs 2.04 (1.04)	1.22	.23
Year 7 Coherence vs Adults Coherence	1.74 (0.86) vs 1.38 (1.06)	1.68	.10
Year 9 Coherence vs Adults Coherence	2.04 (1.04) vs 1.38 (1.06)	2.60	.01*
Year 5 Elaborative vs Year 7 Elaborative	2.09 (1.09) vs 1.62 (1.16)	1.85	.07
Year 5 Elaborative vs Year 9 Elaborative	2.09 (1.09) vs 1.17 (0.97)	3.48	.001**
Year 5 Elaborative vs Adults Elaborative	2.09 (1.09) vs 1.36 (1.07)	3.37	.000**
Year 7 Elaborative vs Year 9 Elaborative	1.62 (1.16) vs 1.17 (0.97)	1.57	.12
Year 7 Elaborative vs Adults Elaborative	1.62 (1.16) vs 1.36 (1.07)	1.09	.28
Year 9 Elaborative vs Adults Elaborative	1.17 (0.97) vs 1.36 (1.07)	0.75	.46
Year 5 Literal vs Year 7 Literal	0.58 (0.67) vs 0.52 (0.48)	0.49	.63
Year 5 Literal vs Year 9 Literal	0.58 (0.67) vs 0.31 (0.32)	1.84	.07
Year 5 Literal vs Adults Literal	0.58 (0.67) vs 0.33 (0.44)	2.24	.03
Year 7 Literal vs Year 9 Literal	0.52 (0.48) vs 0.31 (0.32)	1.78	.08
Year 7 Literal vs Adults Literal	0.52 (0.48) vs 0.33 (0.44)	1.86	.07
Year 9 Literal vs Adults Literal	0.31 (0.32) vs 0.33 (0.44)	0.18	.86
Year 5 Coherence vs Year 5 Elaborative	2.21 (0.91) vs 2.09 (1.09)	0.60	.55
Year 5 Literal vs Year 5 Coherence	0.58 (0.67) vs 2.21 (0.91)	11.15	.000**
Year 5 Literal vs Year 5 Elaborative	0.58 (0.67) vs 2.09 (1.09)	11.12	.000**
Year 7 Coherence vs Year 7 Elaborative	1.74 (0.86) vs 1.62 (1.16)	0.59	.56
Year 7 Literal vs Year 7 Coherence	0.52 (0.48) vs 1.74 (0.86)	7.36	.000**
Year 7 Literal vs Year 7 Elaborative	0.52 (0.48) vs (1.62 (1.16)	6.49	.000**
Year 9 Coherence vs Year 9 Elaborative	2.04 (1.04) vs 1.17 (0.97)	3.60	.002**
Year 9 Literal vs Year 9 Coherence	0.31 (0.32) vs 2.04 (1.04)	8.65	.000**
Year 9 Literal vs Year 9 Elaborative	0.31 (0.32) vs 1.17 (0.97)	4.17	.000**
Adult Coherence vs Adult Elaborative	1.38 (1.06) vs 1.36 (1.07)	0.09	.93
Adult Literal vs Adult Coherence	0.33 (0.44) vs 1.38 (1.06)	6.70	.000**
Adult Literal vs Adult Elaborative	0.33 (0.44) vs 1.36 (1.07)	7.66	.000**

*significant at .01 level ** significant at .001 level*

There was a significant interaction between plausibility and inference type (F (1.85, 383.70) = 8.63, $p < .001$, $\eta p^2 = 0.05$). Post-hoc contrasts show that significantly fewer literal than coherence or elaborative errors were made in the real-world condition (t (206) = 10.38, $p < .001$; t (206) = 10.14, $p < .001$, respectively) and counterfactual-world condition (t (206) = 12.13, $p < .001$; t (206) = 9.20, $p < .001$). There was no significant difference in the number of coherence and elaborative errors made in the real-world condition (t (205) = 1.20, $p = .23$). Conversely, significantly more coherence than elaborative errors were made in the counterfactual-world condition (t (205) = 3.76, $p < .001$). No other interactions were significant ($p > .70$ for all interactions - see Appendix 7.2.). This suggests that throughout adolescence coherence errors are no more difficult than elaborative errors when processing real-world information. However, when processing counterfactual-world information, the generation of coherence inferences appear to be more difficult than the generation of elaborative inferences.

7.3.1.1. Additional Analysis

In the above analyses, Year 9 displayed a qualitatively different pattern of inferential skill to the other three age groups. The Year 9 sample comprises a high number of CYP with reading comprehension levels below age-related expectations – see Chapter 6.3.3.2. It is possible then that the qualitatively different pattern observed in Year 9 is due to a sample which comprise predominantly weak comprehenders. To explore this further four new groups were created based on reading comprehension age: 7-9:11 years (n = 49), 10-13:11 years (n = 51), 14-15:11 years (n = 48), and 16 years+ (n=59). See Table 7.3. for a summary of means. To determine if the unique pattern observed in Year 9 is due to the number of weak comprehenders in that group, a mixed factorial ANOVA, with number of errors as the dependent variable was conducted to explore changes in plausibility and inference type as a function of reading comprehension age.

As with the analysis above, all assumptions for parametric testing were met, therefore, a 3*2*2 parametric ANOVA was conducted to explore the effects of age, plausibility and inference type on skill (number of errors), with Huynh-Feldt statistics reported due to Mauchly's test of Sphericity being violated for all inference conditions.

Table 7.3. Mean number of errors (and Standard Deviations) all six inference conditions for all Reading-Age groups

	Real-world			Counterfactual-world		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
7-9 Years	0.67 (0.58)	1.43 (0.93)	2.29 (1.93)	0.57 (0.81)	2.48 (1.54)	2.10 (1.61)
10-13 Years	0.58 (0.72)	1.50 (1.06)	1.17 (1.44)	0.33 (0.64)	1.50 (1.47)	1.33 (1.52)
14-15 Years	0.33 (0.53)	1.50 (1.38)	1.67 (1.47)	0.39 (0.69)	1.78 (1.64)	1.11 (1.30)
16 and Over	0.46 (0.58)	1.07 (1.38)	1.29 (1.36)	0.32 (0.67)	2.00 (1.22)	1.43 (1.71)

There was a main effect of inference type ($F (1.65, 352.56) = 66.54, p < .001, \eta p^2 = 0.39$). Post-hoc tests revealed that significantly fewer literal than coherence ($t (206) = 11.99, p < .001$) or elaborative errors ($t (206) = 11.11, p < .001$) were made. There was no difference in the number of coherence and elaborative errors made ($t (206) = 0.78, p = .44$). There was a main effect of age ($F (3, 203) = 3.58, p = .016, \eta p^2 = 0.09$). Post-hoc pairwise comparisons revealed that the 7-9:11 years group made significantly more errors than 10-13:11 years group (t (98) = 2.82, $p = .006$), 14-15:11 years group ($t (95) = 2.72, p = .008$), and 16 years+ group ($t (106) = 2.78, p = .006$). None of the remaining comparisons were significant ($p > .71$ for all remaining comparisons – see Appendix 7.3). There was no main effect of plausibility ($F (1,206) = 1.13, p = .26, \eta p^2 = 0.01$). There was, however, a significant interaction between plausibility and inference type ($F (1.86, 396.65) = 6.38, p = .003, \eta p^2 = 0.06$). There was no difference between the number of literal and elaborative errors made in real-world and counterfactual-world conditions ($F (1, 205) < .001, p = 1.00, \eta p^2 < .001$). However, significantly more coherence than elaborative errors were made in the counterfactual-world condition compared to the real-world condition ($F (1,205) = 47.38, p = .009, \eta p^2 = 0.06$). None of the remaining interactions were significant ($p > .20$ for all interactions – See Appendix 7.4). The interaction between age and inference type present in the original analysis was not present in this additional analysis. This suggests that the difficulty generating coherence inferences displayed by Year 9 was not due to the high proportion of poor comprehenders in this group. The possible reasons for Year 9 to display a qualitatively different error pattern to all other are groups are discussed in Section 8.4.3.

7.3.2. Error Analysis

Analysis was conducted on the errors made by all age groups to explore the role of real-world interference when processing counterfactual-world information. Real-world

interference is argued to be present if the participant selects the real-world alternative instead of the counterfactual-world alternative. The forced-choice picture-selection task was designed so that errors caused by literal processing could also be identified, such that one of the pictures depicted the literal state of events. Remaining errors were categorised as unexplained errors, such that the cause of inferential failure could not be determined by the IST. Tables 7.4. and 7.5. summarise the types of errors made when processing real-world inferences and counterfactual-world inferences, respectively.

Table 7.4. Mean Percentage of Error Types made in Real-World Conditions

	Coherence		Elaborative	
	Literal Processing	Unexplained Error	Literal Processing	Unexplained Error
Year 5	64.00	36.00	13.16	86.84
Year 7	81.81	18.18	6.67	93.33
Year 9	63.64	36.36	21.43	78.57
Adults	66.67	33.33	23.08	76.92

Table 7.5. Mean Percentage of Error Types made in Counterfactual-World Conditions

	Coherence			Elaborative		
	Real-World Interference	Literal Processing	Unexplained Error	Real-World Interference	Literal Processing	Unexplained Error
Year 5	72.72	23.64	3.64	76.59	4.26	19.15
Year 7	68.18	22.72	9.09	76.00	0.00	24.00
Year 9	54.54	36.36	9.09	77.78	0.00	2.22
Adults	63.04	30.43	6.52	87.10	0.00	12.90

The general trends, presented in Table 7.4, suggest that for all age groups real-world coherence errors are largely attributed to literal processing, such that of the total errors made, between 64%-82% of these errors could be attributed to literal processing. Conversely, for all groups real-world elaborative errors are predominantly categorised as unexplained errors. For counterfactual-world processing, the means presented in Table 7.5. suggest that both coherence and elaborative errors are predominantly attributable to real-world interference.[28] In sum, when processing counterfactual-world information, real-world interference appears to be the primary cause of errors. However, when generating real-world inferences, the primary cause of inference generation failure varies as a function of inference type, such that the primary cause of coherence inference failure

---

[28] Statistical analysis was not conducted due to limited data.

appears to be literal errors, whereas the majority of elaborative errors appear to be response errors.

7.3.3. Reading Speed

The analysis of reading speeds is concerned with the inferential time-course of successful inference generation, therefore, all errors were removed. This resulted in five participants from Year 5, two participants from Year 7, and one participant from Year 9 losing more than 50% of their data. These participants were excluded from all subsequent analyses. Consistent with previous research, a data trim was also conducted, with all scores +/-2 standard deviations the participant's grand mean being removed and all scores +/-2 standard deviations each item's grand mean being removed (e.g. Calvo & Castillo, 2001a). If a participant had one or less available answer in any one condition, this was treated as a missing value and replaced with the said condition's mean. This data trim resulted in one Year 5 losing more than 50% of their reading speed data. This participant was removed. Whilst most outliers were evenly distributed across all items and participants, some items resulted in more than 50% of errors or outliers. For Year 5 and Year 7 items 2a, 8b, and 11a were excluded from all analyses. For Year 9, items 2a and 8b were excluded from all analyses. For adults, items 2a, 5b and 8b were excluded from all analyses. In total, 4.47% of Year 5 data, 5.06% of Year 7 data, 3.97% of Year 9 data, and 6.30% of adult data were lost due to the data trim conducted. Due to both errors and the data trim, for Year 5, 22.22% of reading speed data were removed. For Year 7, 19.66% of reading speed data were removed. For Year 9, 18.577% of reading speed data were removed. For adults, 19.49% of reading speed data were removed. Whilst the level of data lost due to errors appears quite high, it is consistent with data loss rates in similar studies recruiting child and adolescent samples (Bowyer-Crane & Snowling, 2010; Casteel, 1993).

Since two versions of each story were created, analysis was conducted to determine if one version of the stories was more taxing than the other. There was found to be no difference between the reading speeds of critical texts on Version 1 and Version 2 for Year 5 (t (18) = 0.91, p = .37), Year 7 (t (18) = 1.67, p = .10), Year 9 (t (18) = 1.66, p = .10) or adults (t (18) = .22, p = .83). The data were collapsed across both versions. For all groups a participant mean reading speed was calculated for each condition. These means were then

used to calculate group means, which can be seen in Table 7.6. A visual summary of means can also be found in Figure 7.4. Means suggest that, for all age groups, coherence-inference-evoking critical texts are read slower than elaborative-inference-evoking critical texts and literal texts. For all age groups, there appears to be no difference between reading speeds for elaborative-inference-evoking critical texts and literal critical texts.

Inspection of the skewness and kurtosis statistics revealed that for all conditions the data were normally distributed. Homogenity of variance was not violated. However, Mauchly's Test of Sphericity was violated for inference type. A parametric ANOVA was conducted to explore the effects of age, plausibility and inference type on skill (number of errors), with Huynh-Feldt statistics reported for all analyses involving inference[29].

Table 7.6. Mean Reading Speed of Critical Texts (and Standard Deviations (milliseconds)) for all conditions for all age groups

	Real-world			Counterfactual-world		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
Year 5	7070.09	7304.71	7036.09	7112.85	7444.58	7191.17
	(2028.27)	(2194.19)	(1740.39)	(1994.55)	(2175.78)	(2006.18)
Year 7	5888.93	6098.61	5897.20	6017.99	6227.19	6024.52
	(1231.80)	(1583.88)	(1247.59)	(1494.81)	(1759.01)	(1384.21)
Year 9	4463.16	4872.25	4464.30	4930.61	5186.79	4920.99
	(1133.20)	(1634.10)	(1270.45)	(1442.13)	(1626.67)	(1414.30)
Adults	4279.76	4459.70	4320.56	4475.10	4730.81	4553.81
	(986.42)	(1350.99)	(1144.27)	(1003.94)	(1119.25)	(1103.30)

---

[29] Girden (1992) suggest that Huynh-Feldt, not the Greenhouse-Geisser correction, should be used when the estimates of sphericity are greater than 0.75, as in these cases the Greenhouse-Geisser correction is likely to be too conservative.
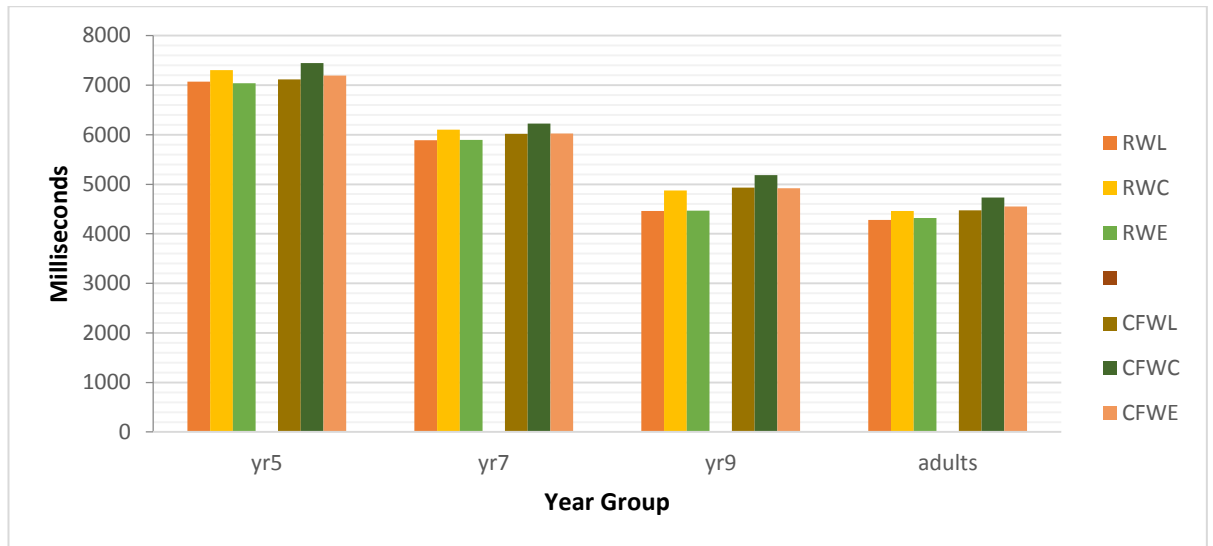
Figure 7.4. Year 5's Year 7's Year 9's and Adults Mean Reading Speed of Critical Texts (milliseconds) for All Six Inference Conditions

A 4 (Age: Year 5, Year 7, Year 9, adults) * 3 (Inference Type: Literal, Coherence, Elaborative) * 2 (Plausibility: Real-World, Counterfactual-World) mixed factorial ANOVA was conducted. There was found to be a main effect of plausibility (F (1,180) = 12.45, p = .001, ηp^2 = 0.07), such that counterfactual-world items (mean = 5734.70, SD = 1342.78) were found to be read significantly slower than real-world items (mean = 5512.95, SD = 1256.24). This suggests that the processing of counterfactual-world information is more taxing than real-world information.

A main effect of inference type was observed (F (1.92, 346.25) = 6.93, p = .001, ηp^2 = 0.04). Post-hoc pairwise comparisons revealed that there was no significant difference between the reading speeds of literal (mean = 5529.81, SD = 1300.04) and elaborative items (mean = 5551.08, SD = 1257.82; t (180) = 0.32, p = .75). Coherence items (mean = 5790.58, SD = 1527.65) were found to be read significantly slower than literal items (t (180) = 3.12, p = .002) and elaborative items, however (t (180) = 2.92, p = .004). This suggests that coherence inferences are generated online, whereas elaborative inferences are generated offline. A main effect of age was observed (F (1,180) = 44.73, p < .001, ηp^2 = 0.44). Post-hoc pairwise comparisons found that the reading speeds of all groups significantly differed (p < .001 - see Appendix 7.5) except for Year 9 and adults (t (93) = 1.31, p = .19). Adults (mean = 4469.95, SD = 1125.55) and Year 9 (mean = 4806.35, SD = 1440.91) had the fastest reading speeds, followed by Year 7 (mean = 6025.74, SD =

1451.45), and then Year 5 (mean = 7193.25, SD = 2012.99) who displayed the slowest reading speeds. This suggests inferential efficiency is achieved by Year 9. There were no significant interactions ($p > .01$ for all interactions - see Appendix 7.6) suggesting the inferential patterns were the same for both real-world and counterfactual-world information and across age groups[30].

7.4. Discussion

This study sought to establish age-related changes in real-world and counterfactual-world inference generation skill (number of errors) and speed (reading speed) in Year 5 (9-10 years), Year 7 (11-12 years), Year 9 (13-14 years) and adults (18 years and over). Unlike previous research, the current study used the same methodology with all age groups. The results of this study extend previous knowledge in several ways. First, quantitative changes in both real-world and counterfactual-world inference generation skill and speed were observed during adolescence. Both skill and speed improved during adolescence for real-world and counterfactual-world information, with adult-like performance observed for inference skill by Year 7 and inference speed by Year 9. This both conflicts with and supports predictions made, such that counterfactual-world skill and speed were predicted to decline during adolescence whereas real-world skill and speed were predicted to improve. Second, consistent with predictions, despite quantitative improvements during adolescence, qualitative changes in online-offline inferential time-course patterns were not observed. Regardless of the plausibility of information, all age groups read coherence-inference-evoking critical texts significantly slower than literal and elaborative evoking critical texts. This suggests coherence inferences are generated whilst reading whereas elaborative inferences are not, throughout adolescence. Previous differences, then, may thus be due to methodological discrepancies, as predicted. Third, qualitatively different error patterns were observed for Year 9 compared to all other groups. Years 5 and 7 and adults found the generation of coherence inferences no more difficult than the generation of elaborative inferences. Conversely, Year 9 made significantly more coherence errors than elaborative errors. This effect disappeared when groups were split by reading

---

[30] Additional analysis was conducted with groups split by reading comprehension age instead of chronological age. Patterns were the same when groups were split by reading age and chronological age – see Appendix 7.7.

comprehension age as opposed to chronological age. This suggests that Year 9 may be a special time in the development of inferential abilities and conflicts with predictions made, such that no qualitative changes were predicted. Fourth, consistent with predictions, for all age groups, the process of real-world and counterfactual-world inference generation does appear to differ. Counterfactual-world inference generation appears to have additional processing costs, compared to real-world inference generation, evidenced by slower reading speeds in all age groups. However, these costs appear to be effectively managed for counterfactual-world elaborative inference generation but not counterfactual-world coherence inference generation. Whilst not statistically significant, means suggest this effect is most prominent in Year 9. Analysis of errors suggests that difficulties faced when processing counterfactual-world information can be attributed to real-world interference. This suggests that throughout adolescence real-world interference is harder to inhibit whilst reading than when generating inferences with strategic effort. These findings are discussed in relation to previous research below.

7.4.1. The Developmental Trajectory of Real-World Inference Generation

Real-world inference generation skill and real-world inference generation speed both improved in all conditions during adolescence, with adult-like performance in skill achieved by Year 7 and speed by Year 9. For inference generation skill, different trajectories were observed for coherence and elaborative inferences. For coherence inference generation, adult-like levels are observed in Year 7 but not Years 5 and 9, with Years 5 and 9 making significantly more errors than adults. It is possible then, that coherence inference generation skill develops throughout adolescence, with some achieving adult-like levels in Year 7, but some not achieving proficiency until later. Alternatively, whilst not significant, visual inspection of mean number of errors made suggest a decline in performance between Year 7 and Year 9. This is consistent with the decline in reading comprehension observed between Year 7 and 9 in Chapter 6. It is possible, then, that for some, coherence inference generation skill declines between Year 7 and Year 9. However, this is a cross-sectional study, therefore, it is unclear if these results reflect true age-related changes. For elaborative inference generation skill, steady improvements were observed, with adult-like levels achieved between Year 7 and Year 9. Whilst quantitative gains in inference generation skill during adolescence are consistent with previous research, differing developmental trajectories for the two inference types had not been noted (Ackerman, 1986, 1988; Barnes et al., 1996). This is likely due to

previous research not distinguishing between the two inference types. This study thus highlights the importance of exploring the development of both inference types during adolescence. Current results also highlight slow progress after Year 7. There are several possible interpretations. First, significant improvements in inference generation skill may occur between Year 7 and adulthood. However, the Image Selection Task (IST) may not be sensitive enough to detect them. Whilst error rates were low, particularly in Year 9 and adulthood, ceiling effects were only observed for literal conditions, it is likely then that the task provided enough breadth to tap the inference generation skills of all age groups. Moreover, as discussed in Chapter 3, the stimuli and format in the IST were designed to be suitable for the assessment of inferential abilities in a range of age groups. It is thus unlikely that the limited development observed is due to the sensitivity of the IST. Second, it is possible that basic inference generation skills are achieved by Year 7. Research shows that even young CYP are able to generate the same inferences as adults when prompted (Omanson et al., 1978; Paris et al., 1977). It is possible then that basic inference generation skills are achieved early on, but successful application of inferences is constrained by decoding skills, text demands and/or ability to express the inference. This may explain why adult-like inference generation skills were observed at a younger age than in previous research – texts in the IST were designed to be readable by the youngest participants and expressive output demands were minimised by the use of a pictorial response method.

In addition to the main effect of age, an interaction between inference type and age was observed for skill (number of errors), such that whilst Years 5 and 7 and adults found elaborative inference generation no more difficult than coherence inference generation, Year 9 made significantly more coherence than elaborative errors. An interaction was also observed between plausibility and inference type for skill (number of errors). However, a three-way interaction between inference type, plausibility and age was not significant. Visual inspection of means suggests that, whilst not significant, coherence inferences are more difficult than elaborative inferences for Year 9 in the counterfactual-world condition only. Therefore, means suggest that for real-world information, despite quantitative gains in inference generation skill during adolescence, qualitative changes do not occur, such that for all age groups there was no difference in the number of coherence and elaborative errors made. This refutes previous research, which suggests that elaborative inferences are more difficult than coherence inferences for both children

and adults (e.g. Bowyer-Crane & Snowling, 2010; Cain & Oakhill, 1999; Cain et al., 2001). The difference between the findings of the current study and previous research may be due to the measure used. Some studies have assessed inferential skill using online measures only (e.g. Calvo & Castillo, 1996; 1998; 2001a). Whilst these measures provide information regarding those inferences that can be generated successfully whilst reading, research shows that participants, particularly CYP, are able to generate more inferences when prompted than spontaneously whilst reading (Omanson et al., 1978; Paris et al., 1977). Moreover, those inferences generated whilst reading are typically found to be coherence inferences, with elaborative inferences tending to be generated offline (e.g. Calvo & Castillo, 1996; 1998; 2001a; 2001b; Casteel, 1993; Casteel & Simpson, 1991). This thus explains the superiority of coherence inferential skill compared to elaborative inferential skill when online measures are used.

Some studies finding that elaborative inference generation is more difficult than coherence inference generation have used offline measures and prompts, however (Cain & Oakhill, 1999; Cain et al., 2001; Casteel & Simpson, 1991). Typically, in this previous research, whereas elaborative inferences require the integration of the text and the reader's background knowledge (i.e. knowledge-based inferences), coherence inferences can be generated successfully by integrating two premises in the text (i.e. text-based inferences). In the current study, both coherence and elaborative inferences were knowledge-based inferences. This suggests that the discrepancy between the two inference types in previous research may be due to differences in processing costs associated with the knowledge-based versus text-based inferences. This is supported by Cain et al., (2001) who used knowledge-based inference-evoking stimuli to assess both coherence and elaborative inferences and also found no significant difference between coherence and elaborative inferential skill.

Alternatively, differences between the current findings and previous research may be reflective of the pictorial response method used in the current study. Static inferences were explored in the elaborative condition, with a focus on properties that could be depicted pictorially e.g. colour and texture. The Immersed Experiencer Framework (IEF) suggests that mental representations are multi-modal, including not just verbal information, but other sensory information (Zwaan, 2003 – see Chapter 2.3.5). The use of pictorial test stimuli may thus have encouraged participants to create a pictorial situational mental model. Moreover, the IEF suggests that information is depicted in the

mental model using the modality it was experienced in, since static properties are more often experienced visually than verbally, these properties may thus be more easily and/or strongly represented pictorially than verbally (Zwaan, 2003). Compared to participants in previous research using text-based response methods, participants in the current study may have found the generation of elaborative inferences no more difficult than the generation of coherence inferences due to the nature of the pictorial situational model, such that both coherence and elaborative inferences were readily depicted.

For inference generation speed, as measured by reading speed, quantitative developments were observed until Year 9. Year 5 read critical texts significantly slower than all groups, Year 7 read critical texts significantly slower than Year 9 and adults, however, there was no difference between the reading speeds of Year 9 and adults. This suggests adult-like inference generation speeds are achieved by Year 9. This is consistent with previous research (e.g. Casteel, 1993). It is notable, though, that speed continues to improve after skill has reached adult-like levels. It is possible that whilst CYP may have the basic inference generation skills needed for success on the IST by Year 7, they are still unable to execute these skills as quickly as Year 9s and adults, such that they are constrained by slower access to the necessary knowledge, for example. In line with previous research, qualitative changes in inference speed were not observed, such that all age groups were found to read coherence-inference-evoking critical texts significantly slower than elaborative-inference-evoking and literal critical texts. There are two possible interpretations. First, it may be that coherence-inference-evoking critical texts were simply more taxing than elaborative-inference-evoking critical texts and literal critical texts. This is unlikely given the measures taken to ensure similar levels of complexity and readability across conditions – see Chapter 3. When exploring inference generation, research has shown that even if two sentences are equal in syllable length, if one sentence requires an inference for coherence to be maintained and the other does not, the sentence requiring an inference will take longer to read (e.g. Poynor & Morris, 2003). Therefore, an alternative interpretation is that the additional processing costs associated with coherence-inference-evoking critical texts are attributable to the generation of the coherence inference whilst reading the text. This interpretation is consistent with previous research, which finds that CYP and adults generate coherence inferences, but not elaborative inferences, whilst reading (e.g. Barnes et al., 1996; Calvo & Castillo, 1996; 1998; 2001a; 2001b; Casteel, 1993; Graesser et al., 1994).

In sum, quantitative changes in both inference generation skill, as measured by number of errors, and inference generation speed, as measured by reading speed, were observed during adolescence, with adult-like levels being achieved in Year 7 and Year 9, respectively. Different trajectories were observed for coherence and elaborative inferences, highlighting the distinctiveness of these inference types. Specifically, all age groups appear to generate coherence inferences whilst reading and elaborative inferences after reading. All age groups appear to find the generation of elaborative inferences no more difficult than the generation of coherence inferences, though. Therefore, the processes underpinning coherence and elaborative inferences, whilst distinct, do not appear to qualitatively change during adolescence. The speed at which inference generation skills can be executed may be constrained during adolescence, however.

7.4.2. The Developmental Trajectory of Counterfactual-World Inference Generation

Counterfactual inference generation skill and real-world inference generation speed both improved in all conditions during adolescence, with adult-like performance in skill achieved by Year 7 and speed by Year 9. For inference generation skill, possible qualitative changes were observed. There was no difference in number of coherence and elaborative errors made in Year 5, Year 7, or adults, however, Year 9 made more coherence than elaborative errors. For inference generation speed, patterns were consistent across all age groups, such that coherence inferences were generated whilst reading whereas elaborative inferences were not. The findings of this study thus conflict with previous research in several ways. First, previous research suggests that counterfactual-world inference generation abilities decline with age (Ceci et al., 1981; Dorfman, 1989; Dorfman & Brewer, 1988). However, in the current study, counterfactual-world patterns largely mirrored real-world patterns, such that quantitative *gains* in inference generation skill and speed were observed during adolescence, with adult-like performance being achieved by Year 7 and Year 9, respectively. Second, previous research suggests a conceptual shift in counterfactual-world processing during adolescence, such that whilst CYP may generate counterfactual-world coherence inferences online, adults generate all inferences offline (Bowyer-Crane & Snowling, 2010; Graesser et al., 1998). However, in the current study, reading speed patterns were consistent across all age groups, with coherence-inference-evoking critical texts read significantly slower than both literal and elaborative-inference-evoking critical texts by

all age groups, regardless of the plausibility of the information. This suggests, throughout adolescence, counterfactual-world coherence inferences are generated whilst reading, whereas elaborative inferences are generated after reading. In previous research there were methodological discrepancies regarding text length, complexity, plausibility, and time interval when exploring the counterfactual-world inference generation abilities of CYP and adults. The current study was unique, such that the same methodology was employed with all age groups. Discrepancies between the results of the current study and previous research thus suggest that the differences observed between CYP and adults in previous research are due to inconsistencies in stimuli and method as opposed to age-related changes and conceptual shifts in processing. The stimuli used in the current study were similar to those used by Bowyer-Crane and Snowling (2010) - short texts comprising fairy-tale-like implausible counterfactual-world information, with readability ages of 9 years or less. It is possible, then, that difficulty generating counterfactual-world coherence inferences whilst reading may arise for adults or all age groups, with longer, more complex texts, comprising implausible natural-law-violating information (e.g. Graesser et al., 1998).

Whilst the changes in the inferential time-course suggested by previous research were not observed in the current study, Year 9 did display a qualitatively different error pattern to all other age groups. Unlike the other groups, Year 9 found counterfactual-world coherence inference generation more difficult than counterfactual-world elaborative inference generation. One possible interpretation of these results is that Year 9 are struggling to generate inferences whilst reading. Previous research suggests that coherence inferences are generated with minimal effort whilst reading and result from largely unconscious processing. Conversely, elaborative inferences are thought to require strategic effort, with the reader more aware of their processing (Graesser et al., 1994; Long & Golding, 1993; Magliano et al., 1993; McKoon & Ratcliffe, 1992). Year 9 may thus be able to more actively suppress real-world interference when generating counterfactual-world elaborative inferences compared to counterfactual-world coherence inferences. However, this was a cross-sectional study, meaning the differences observed across groups may be due to individual differences within each sample and not age. Specifically, the Year 7 cohort was found to comprise stronger readers than the Year 9 cohort – See Section 6.3.3.2. Therefore, the unique inference generation pattern observed in Year 9 may be attributable to the sample used. This is supported by the additional

analysis reported. When groups were split according to reading comprehension age rather than chronological age, for all age groups, there was found to be no significant difference in the number of coherence and elaborative errors made. Further research exploring inference generation during Year 9 is needed, specifically exploring those skills and processes that may underpin the suppression of real-world interference whilst generating inferences.

It was predicted that for all age groups the process of real-world and counterfactual-world inference generation would differ. Despite similar online-offline patterns being observed for both real-world and counterfactual-world inference generation, counterfactual-world critical texts were found to be read significantly slower than real-world critical texts by all age groups. This suggests that counterfactual-world texts are more demanding than real-world texts. This is consistent with previous research (e.g. Byrne, 2002; 2007). Previous research suggests that this additional processing cost is due to the construction of two mental models, a real-world and counterfactual-world mental model, with the resolution of conflict experienced due to the activation of related, but contradictory real-world information, cognitively demanding (e.g. Ferguson, 2012). This is supported by the error analysis conducted, such that 54-87% of the counterfactual-world inference errors made could be attributed to real-world interference. Importantly, the extra time spent processing counterfactual-world inferences appears to be effective for Years 5 and 7 and adults, such that these groups made no more errors in the counterfactual-world conditions compared to real-world conditions. However, as discussed above, means suggest that Year 9 made more coherence than elaborative inference errors in the counterfactual-world condition but not the real-world condition. This suggests, despite additional processing time, Year 9s still appear to be displaying difficulty when generating counterfactual-world coherence inferences. Further research exploring why this is the case for Year 9 but no other age group is warranted.

In sum, whilst quantitative improvements in counterfactual-world inference generation skill and speed were observed during adolescence, developmental patterns of counterfactual-world inference generation whilst reading were found to be relatively consistent across age groups, such that coherence inferences appear to be generated online, whereas elaborative inferences appear to be generated offline. The finding that all groups were able to generate coherence inferences whilst reading, regardless of the plausibility of the information, directly conflicts with previous research. This may be due to

methodological discrepancies in previous research. Regardless of inference type, counterfactual-world critical texts were read significantly slower than real-world critical texts by all age groups. Consistent with previous research, this suggests that the processing of counterfactual-world information is more taxing than the processing of real-world information, with this difficulty being attributed to the resolution of real-world interference. For Years 5 and 7 and adults, the additional time spent processing counterfactual-world information appears to be effective, such that there is no significant difference in real-world and counterfactual-world skill levels. However, Year 9 may be experiencing difficulty generating counterfactual-world coherence inferences, despite the extra processing time. Future research is needed to explore why this is the case in Year 9 but at no other point during adolescence.

7.4.3. Conclusions

Inference generation skill, as measured by number of errors, and speed, as measured by reading speed, improved during adolescence, with adult-like levels achieved around Year 7 and Year 9, respectively. Whilst inference generation skill appears to reach adult-like levels early in adolescence it is possible that the successful application of these skills is constrained in some way – e.g. speed at which knowledge can be accessed. These factors (e.g. knowledge accessibility) may thus limit inference generation speed between Year 7 and Year 9, explaining why adult-like performance is achieved earlier for inferential skill than inferential speed. When explored separately, coherence and elaborative inference skills had unique trajectories. Coherence inference skill appears to develop throughout adolescence, with a period of limited progress between Year 7 and Year 9. Elaborative inference skill appears to develop steadily during adolescence, with adult-like levels achieved between Year 7 and Year 9. For real-world inference generation, despite quantitative gains in inference generation skill and speed, patterns tended to remain the same throughout adolescence, such that the generation of elaborative inferences was found to be no more difficult than the generation of coherence inferences and whilst coherence inferences were found to be generated online, elaborative inferences were found to be generated offline. For counterfactual-world inference generation, patterns of inferential time-course were consistent throughout adolescence, such that coherence inferences were generated online and elaborative inferences offline. However, throughout adolescence, the resolution of conflict experienced due to real-world interference, appears to result in the generation of counterfactual-world inferences being more cognitively

taxing than the generation of real-world inferences. For the most part, the extra processing time attributed to counterfactual-world information appears to be effective, however, this may not be the case for Year 9s when generating coherence inferences. Further research is needed to explore those factors constraining inference generation throughout adolescence and, in particular, in Year 9.

CHAPTER 8 PREDICTORS OF REAL-WORLD AND COUNTERFACTUAL-WOLRD INFERENTIAL ABILITIES DURING ADOLESCENCE

8.1. Introduction

Inference generation is a multi-faceted skill, with areas such as the superior frontal gyrus and dorsal lateral prefrontal cortex routinely activated when reading a text requiring an inference for comprehension (e.g. Ferstl et al., 2005; 2008 Jung-Beeman, 2005; Mason & Just, 2004; Virtue et al., 2006). Activation of these brain areas is related to skills and processes such as the selection of appropriate knowledge, semantic integration and detection of inconsistencies. It is essential to understand how those skills underpinning inference generation operate during adolescence as this may further illuminate the role of inference generation in higher-order literacy skills during adolescence. The cross-sectional study reported here investigates the potentially varying independent contributions of four key predictors: 1) amount of knowledge, 2) accessibility of knowledge, 3) inhibitory control, and 4) strength of belief bias in real-world and counterfactual-world inferential skill in Year 5 (9-10 years), Year 7 (11-12 years), Year 9 (13-14years) and adults (18years and over).

8.1.1. The Role of Knowledge in the Inference Generation Process

This thesis is primarily concerned with knowledge-based inferences due to their centrality in situational model construction (e.g. Graesser et al., 1994; Kintsch, 1993; 1994). Kintsch (1993) defines knowledge-based inferences as those that result from the activation and subsequent integration of a reader's background knowledge with the text they are reading. When generating knowledge-based inferences, research finds that, for both adults and CYP, those with high topic knowledge generate more inferences than those with low topic knowledge, regardless of IQ (Birch & Bloom, 2004). In this way, the specific background knowledge a reader possesses determines the meaning they can derive from a text (Graesser et al., 1994). For instance, if the reader does not know what a rod is used for they cannot infer that the girl is going fishing when reading sentence A. Furthermore, if a reader is an expert in finances and banking but naïve in the area of agriculture and fishing, when reading sentence A, the reader may infer that the girl is going to a monetary bank, rather than a riverbank. Coherence would be lost.

Sentence A: The girl picked up her rod and headed to the bank.

Whilst specific topic knowledge does not necessarily vary as a function of age – i.e. a CYP can be an expert in an area an adult knows little about – the amount of general knowledge possessed has been found to increase with age (Schroeders et al., 2015; Thorsen et al., 2014). Subsequently, the likelihood of an individual possessing the relevant knowledge needed to generate an inference on any given topic increases with age. Simply increasing the knowledge-base does not always result in associated improvements in inference generation skill, however. Barnes et al. (1996) found that, even after teaching the necessary knowledge, age-related differences in inference generation skill were still observed. This may be because inference generation is dependent upon not only possession of the necessary knowledge, but also ready access to said knowledge (Barnes et al., 1996).

Accessibility of knowledge refers to how quickly information can be retrieved (Glucksberg et al., 1993). It has been shown that certain brain areas become more active when individuals read sentences that require an inference (e.g. Chow et al., 2008; Kuperberg et al., 2006). This suggests that these areas are involved in inference generation. Skilled-comprehenders have been found to display less neuronal activation than less-skilled comprehenders when generating an inference, however (Prat & Just, 2010; Prat et al., 2007; 2011). This suggests that skilled-comprehenders engage in more efficient processing – one possible reason for more efficient processing could be that skilled-comprehenders are able to access knowledge more readily than less skilled-comprehenders.

Barnes et al. (1996) explored the relationship between accessibility of knowledge and inference generation in CYP aged 6 - 16 years. They found that the role of knowledge accessibility varied according to inference type. The importance of knowledge accessibility was found to increase for real-world elaborative inferences, but decline for real-world coherence inferences. Barnes et al. suggests that this is due to improvements in accessibility of knowledge, such that, with age, more information is readily available for elaborative inference generation. Barnes et al. argue that as comprehension monitoring improves with age, older participants are more likely to detect coherence breaks whilst reading. To maintain coherence, older participants are thus more likely to engage in a search for the necessary knowledge, regardless of the accessibility of this

knowledge. Therefore, for older participants, accessibility of knowledge becomes less important as participants identify more coherence gaps and actively strive to fill these gaps with the necessary knowledge. For younger participants, on the other hand, coherence breaks may not be detected, thus a conscious search of long-term memory may not be undertaken. As a result, the only inferences generated without prompting will be those for which knowledge is easily accessible.

When processing counterfactual-world information the automatic activation of large amounts of related, but contradictory, real-world knowledge is argued to result in real-world interference (Alter et al., 2007; Byrne, 2002; 2007; Evans, 2006). The resolution of the conflict experienced due to real-world interference is argued to be one of the primary difficulties faced when generating counterfactual-world inferences (Byrne, 2012; Ferguson & Sandford, 2008). It follows, then, that the more related real-world knowledge an individual possesses and the more easily accessible this knowledge is, the more real-world interference they will face when processing counterfactual-world information. Since both amount and accessibility of knowledge are found to increase with age, this would fit with previous research which finds that cognitive demands and difficulty associated with counterfactual-world inferential processing increase with age (Ceci et al., 1981; Dorfman, 1989; Dorfman & Brewer, 1988). In sum, both amount and accessibility of knowledge are implicated in the inference generation process, with improvements observed in both as CYP age. The role and importance of knowledge may vary as a function of both age and inference type, however. Developments in said skills may thus affect the inference generation process during adolescence.

8.1.2. The Role of Inhibitory Control in Inference Generation

The representation of semantic knowledge is thought to be associative, resulting in the activation of both related and contradictory knowledge when reading (e.g. van den Broek et al., 2005; van den Broek et al., 1996; van den Broek et al., 1999). Language processing takes place in working memory, with several skills and processes interacting (Cain & Oakhill, 2006a; De Beni et al., 1998; Kershaw & Schatschneider, 2012; Yuill et al., 1989). Working memory capacity, however, is limited (Baddeley, 1986; Just & Carpenter, 1992). Subsequently, the maintenance and integration of large amounts of knowledge would leave little processing capacity for the execution of those other cognitive processes, such as comprehension monitoring, needed to successfully generate an inference and

220

comprehend a text (Cain, 2006; Oakhill & Cain, 2011; van den Broek, 2012). The amount of knowledge activated, maintained, and integrated into the situational model whilst reading must be constrained in some way.

Inhibitory control has been implicated in the management of knowledge during the inferential process (e.g. Barnes et al., 2004; Cain, 2006; De Beni & Palladino, 2000). Research suggests that for comprehension to be successful, irrelevant and contradictory knowledge must be suppressed in the early stages of the inferential process (e.g. Cain, 2006; Harnishfeger, 1995). This creates a more manageable workspace, such that processing costs are reduced as only the most relevant information is processed (Barnes et al., 2004; Cain, 2006). Good inhibitory control thus appears central to efficient and accurate inference generation. This is supported by research showing that, compared to those with strong inhibitory skills, those with weak inhibitory control struggle to generate inferences and ultimately comprehend a text (Cain, 2006; Gernsbacher & Faust, 1991).

Inhibitory control has also been implicated in counterfactual-world processing, such that it is essential to the suppression of conflicting real-world knowledge (Houde & Moutier, 1996; 1999; Houde et al., 2001). Inhibitory control has been found to develop throughout adolescence and into adulthood (Bjorklund & Harnishfeger, 1990; Romine & Reynolds, 2005). Developmental improvements in counterfactual-world inference generation ability during adolescence would be expected. However, previous research suggests that compared to real-world inferential processing, difficulty associated with counterfactual-world inferential processing increases with age (e.g. Ceci et al., 1981; Dorfman, 1989; Dorfman & Brewer, 1988). The increasing processing difficulties associated with counterfactual-world information may be due to *when* and *how* in the inferential time-course inhibitory control is executed. The filtering model suggests that readers have a bias for typical information, such that they will attend to, process and retrieve this information more readily than atypical information (Graesser et al., 1998). Typical information is that which fits with pre-existing schemas and real-world knowledge. The filtering model suggests that in the early stages of text processing, much of the contradictory and irrelevant knowledge, that is automatically activated due to associative knowledge activation mechanisms, is suppressed.

Counterfactual-world information is, by nature, contradictory, as such, this information may be suppressed early in the inferential process. Subsequently, counterfactual-world

information would need to be reactivated if the counterfactual-world inference is to be successfully generated. Research suggests that in many cases the reader may not recognise that they need to reprocess the text since filtering information may also provide 'an illusion of comprehension' (Glenberg et al., 1982; Otero & Campanario, 1990; Vosniadou et al., 1988). An illusion of comprehension refers to a basic representation of the text in which inconsistent or difficult information is excluded, to create a version of the text the reader is able to understand. Strong inhibitory control in the early stages of text processing may thus hinder counterfactual-world inference generation, such that the necessary inference is not generated or the information needed is not readily activated, resulting in a more time-consuming process. This may thus explain the increasing difficulty associated with counterfactual-world processing from childhood to adulthood (e.g. Dorfman, 1989). However, the filtering model is based on the processing of counterfactual-world information in real-world contexts. There is a growing body of research showing that counterfactual-world information processed in a counterfactual-world context is not treated as contradictory and, therefore, is not suppressed (Nieuwland, 2012; 2013; 2015, Nieuwland & Berkum, 2006; Nieuwland & Martin, 2012). These studies have explored literal, rather than inferential text processing, however. It is thus unclear if, when presented in a counterfactual-world context, counterfactual-world inferences could be generated with minimal real-world interference. The role of inhibitory control in the processing of counterfactual-world information in counterfactual-world contexts requires further exploration.

8.1.3. The Role of Belief Biases in the Inferential Process

A belief bias is a cognitive preference to process information depending on real-world beliefs, rather than logic. The mental model theory proposes a three stage model of comprehension, with belief biases playing a central role in the final stage of inference validation (Johnson-Laird & Byrne 1991; 2002). For the inference to be fully encoded into the reader's mental model of the text, Johnson-Laird and Byrne (1991) suggest that a conscious, cognitively demanding search or construction of a model in which the inference is compatible takes place. To reduce processing costs, inferences consistent with real-world beliefs are thought to be automatically accepted. For real-world inferential processing, belief biases thus serve to reduce processing costs whilst reading.

Counterfactual-world inferences are inconsistent with real-world beliefs, however, meaning the mental model theory would predict that they are not automatically accepted. A conscious and cognitively demanding search and/or construction process must be executed to resolve this conflict. Since real-world knowledge is automatically activated, in addition to a counterfactual-world inference, an inconsistent and inaccurate real-world inference may also be generated (Kintsch, 1988; 1998; van den Broek et al., 1995; 1999; 2005). This effect is evident in the literature, such that errors, when processing counterfactual information, typically reflect conflicting real-world information (e.g. Ceci et al., 1981). The effects of belief biases when processing counterfactual-world information have been observed for adults and CYP such that, compared to real-world consistent information, reading speeds and response times for counterfactual-world information are longer and error rates are higher (Houde & Guichart, 2001; Moutier, et al., 2006; Bjorklund and Harnishfeger, 1990;).

If the counterfactual-world information is to be processed successfully, the interference caused by belief biases must be inhibited (Houde & Moutier, 1996; 1999; Houde et al., 2001). Since inhibitory control has been found to increase with age (Bjorklund and Harnishfeger, 1990; Romine & Reynolds, 2005), it is unsurprising that some have found that an individual's ability to inhibit belief biases also increases with age (e.g. Frank, 1996; 1997). However, it has also been found that children are able to overcome the effects of a belief bias with more ease than adolescents and adults (Mitchell et al., 1996). Mitchell et al. (1996) argue that CYP apply a 'seeing is believing rule' – i.e. the character must see an event to believe it and any other form of contradictory evidence is not believed even if the reader knows the real state of events. Conversely, adults are argued to adopt a more complex processing strategy, drawing upon their existing previous knowledge to consider those factors that may lead one to accept another's message. This leaves adults more vulnerable to belief biases. This may explain the possible conceptual shift in counterfactual-world inference generation processing observed between 9-10 years and adulthood (Bowyer-Crane & Snowling, 2010; Graesser et al., 1998).

When generating counterfactual-world inferences during the comprehension of fantasy and fictional texts the necessary counterfactual-world model is already provided. Recent research suggests that counterfactual-world information is compared to a counterfactual-world model rather than real-world model, when presented within a counterfactual-world context (Nieuwland, 2012; 2013; 2015, Nieuwland & Berkum, 2006; Nieuwland &

Martin, 2012). This suggests that readers may be able to put aside belief biases when a counterfactual-world context is provided. Moreover, it has been argued that when reading naturally readers do not validate their inferences whilst reading, as the cognitive effort needed to do this would disrupt the natural reading process (Markman & McCullen, 2003; 2005; 2007; McMullen, 1997; McMullen & Markman, 2000). Instead, to avoid comprehension disruption, all inferences are argued to be automatically accepted. At present, research predominantly explores the role of belief biases in counterfactual-world reasoning, rather than counterfactual-world inference generation. Subsequently, belief biases may not play such an influential role in the generation of counterfactual-world inferences when reading fantasy and fictional texts. Further research exploring the potentially changing role of belief biases in the inferential process during adolescence is needed.

8.1.4. The Current Study

Amount of knowledge, accessibility of knowledge, inhibitory control, and strength of belief bias appear to be some of the skills implicated in the inferential process during adolescence. Not only are these skills central to real-world and counterfactual-world inference generation, developments in these skills are thought to be advantageous for real-world inference generation, but detrimental to counterfactual-world inference generation, thus explaining the possible conceptual shift in processing suggested by the literature (e.g. Bowyer-Crane & Snowling, 2010; Graesser et al., 1998). However, previous research typically explores the role of these cognitive skills in isolation. Furthermore, the predictive utility of the cognitive components of inference generation is rarely explored developmentally. Little is known about which skill(s) is most important at a particular time, how importance and role changes with age, or how the role of these skills may change according to inference type. In addition, extant literature focuses on counterfactual-world reasoning, and/or counterfactual-world inference generation with no context. Therefore, little is known about how the four key cognitive skills operate when generating counterfactual-world inferences from fantasy or fiction texts. The current study sought to explore the role of amount of knowledge, accessibility of knowledge, strength of belief bias, and inhibitory control in real-world and counterfactual-world inference generation skill for four age groups: Year 5, Year 7, Year 9, and adults. This study aimed to (a) investigate the independent predictive utility of

amount of knowledge, accessibility of knowledge, strength of belief bias, and inhibitory control in real-world and counterfactual-world coherence and elaborative inference generation; and (b) establish age-related changes in the predictive utility of the four key predictors in real-world and counterfactual-world inferential processing in Year 5, Year 7, Year 9, and adults. Since this study is primarily exploratory, no predictions were made, other than all skills would be unique predictors of inference generation skill between Year 5 and adulthood.

8.2. Method

8.2.1. Participants

The same participants recruited for Study 3 were also included in this study – see Chapter 6.

8.2.2. Design

A cross-sectional design was employed with four age groups: Year 5 (9-10 years), Year 7 (11-12 years), Year 9 (13-14 years), and adults (18 years and over) to explore age-related differences in the predictive utility of four predictor variables: accessibility of knowledge, amount of knowledge, strength of belief bias, and inhibitory control. Working memory capacity, IQ score, and single-word-reading ability were controlled for due to the relationship these variables have been found to share with inference generation skill (Carver, 1990; Dixon et al., 1988; Goff et al., 2005; Masson & Miller, 1983; Singer et al., 1992; Singer & Richot, 1996; Whitney et al., 1991). The outcome variables were:

- Real-world Coherence Skill (number of errors)
- Real-World Elaborative Skill (number of errors)
- Counterfactual-World Coherence Skill (number of errors)
- Counterfactual-World Elaborative Skill (number of errors)[31]

---

[31] The IST provides measures of both the inferential time-course (reading speed) and inference generation skill (number of errors). Number of errors was used in the current study as the dependent variable instead of reading speed. Reading speed may be affected by decoding skill and other aspects of reading comprehension, which are not directly related to inference generation skill. To control for these factors a difference score (inference reading speed – literal reading speed) would have been needed. However, when calculated this resulted in some negative scores. Due to the regressional nature of this study these scores could not be used. Consequently, only skill was explored.

8.2.3. Materials

8.2.3.1. Intelligence, Working Memory, and Word-Reading Measures

IQ score (a composite of the vocabulary subscale and matrix reasoning subscale) measured using the WASI, working memory capacity measured using the BAS II backwards digit span task, and single-word-reading ability measured using the York Assessment of Reading Comprehension (Primary and Secondary) were used as control variables (Elliott et al., 1996; Snowling et al., 2009; 2010; Wechsler, 1999). Full details of these measures can be found in Chapters 5 and 6.

8.2.3.2. Inference Generation Measure

The Image Selection task used in Study 3 was used – see Chapters 3 and 4.

8.2.3.3. Amount and Accessibility of Knowledge Measure

A seventy-two-item timed sentence verification task was created to assess amount and accessibility of knowledge. Thirty-six items required a true response and were directly associated with the knowledge-base used in the Image Selection Task. Table 8.1 provides examples – see Appendix 8.1 for a list of stimuli. The remaining thirty-six items were lexically and syntactically similar fillers however, a false response was required. These items were related to the knowledge-base used in the IST but were not based on this knowledge-base (e.g. plants give birth to baby plants). Each sentence was seven syllables long and recorded by the researcher. Each sentence was presented orally once (to account for differences in reading skill) and visually in the middle of the screen until participants verified the sentence. Participants were instructed to press the tick button if they felt the sentence was true and the cross button if they felt the sentence was false, immediately after hearing the sentence. A button press was not accepted until the entire sentence had been presented orally. Sentences were presented in a random order for each participant. Accessibility of knowledge was indexed by response time in milliseconds. Amount of knowledge was indexed by accuracy. To familiarise the participant with the response method and ensure appropriate volume, the task began with five practice trials in which participants heard a sentence, which was syntactically and lexically similar to the experimental sentences, and had to respond true or false. If participants responded incorrectly, the researcher first ensured the participant could hear the sentences and then

explained the instructions again. The researcher was confident all participants understood the task and response method before the task, proper, began. A HP ProBook Laptop using E-Prime software was used to present stimuli and record response times (milliseconds) and accuracy.

Table 8.1. Example of Knowledge Stimuli

	Image Selection Task Item	**Knowledge Task Item**
Real-World Coherence Inference	Holly planted a tiny seed. A century later an oak tree stood in its place.	Most trees grow from tiny seeds
Counterfactual-World Coherence Inference	The giant egg began to crack loudly. Soon, instead of the egg there was a small bear cub.	Mammals give birth to live young

8.2.3.4. Belief Bias Measure

A twenty-item reasoning from true and false premises task was used to assess belief biases. To successfully overcome a belief bias, the participant must reason based on logic not beliefs. Experimental items required participants to reason from a false premise where the conclusion was true based on the premise presented, but not beliefs. Control items were used to ensure participants had basic reasoning abilities – i.e. could reason from true premises. Fillers presented both true and false premises, but the conclusion was always false based on the premise presented. Five items presented a true premise and true conclusion (control), five items presented a true premise and false conclusion (filler a), five items presented a false premise and true conclusion (experimental), and five items presented a false premise and false conclusion (filler b) – see Table 8.2. for example stimuli. See Appendix 8.2 for a full list of stimuli. The premise was presented in the top half of the screen in a white box with a blue outline. The conclusion was presented in a white box with a green outline in the bottom half of the screen. Participants were instructed to decide if the conclusion was true or false based on the preceding premise. Participants were asked to respond as quickly and as accurately as possible. Items were presented in a random order for each participant. Strength of belief bias was indexed by a response time difference score for correct items, to control for differences in motor control (experimental – control). To familiarise the participant with the response method,

the task began with four practice trials, one trial representing each condition. If a participant answered incorrectly, the researcher explained the task and instructions to participants and what the correct answer should have been. The researcher was confident all participants understood the task before the experiment, proper, began. A HP ProBook Laptop using E-Prime software was used to present stimuli and record response times (milliseconds) and accuracy.

Table 8.2. Example Reasoning from False Premises Stimuli

Condition	Premise	Conclusion
True Premise, True Conclusion	All fish live in water, a trout is a fish.	A trout lives in water.
True Premise False Conclusion	All fish live in water, a trout is a fish.	A trout lives in trees.
False Premise, True Conclusion	All fish live in trees, a trout is a fish.	A trout lives in trees.
False Premise, False Conclusion	All fish live in trees, a trout is a fish.	A trout lives in water.

8.2.3.5. Inhibition Measure

The Rule-Shift-Cards task from the Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 2003) was used to assess inhibition. Participants saw a series of cards. As each card was presented participants were instructed to respond 'true' or 'false' depending on the rule given –*say true if the card is black, false if the card is not black*. Participants then saw the cards again, but were instructed to respond according new a new rule –*say true if the card is even, false if the card is not even*. Errors were recorded by the researcher. Inhibition skill is indexed by the number of errors made.

8.2.4. Procedure

All of the tasks used in this study were administered as part of a larger test battery used to obtain data for Studies 3, 4, and 5. The procedure is the same as Study 3 – see Chapter 6.

8.3. Results

This study sought to explore the role of amount of knowledge, accessibility of knowledge, inhibitory control, and strength of belief biases in inference generation skill in Year 5,

Year 7, Year 9 and adults. First, age-related changes in each predictor variable were explored using ANOVAs. Then the role of these four predictors in inference generation skill was analysed using multiple hierarchical regression analyses. The order of variables was selected a priori. Composite IQ scores, single-word-reading scores and working memory capacity scores were entered in Step 1 to account for relationships between inference generation, IQ, word-reading skill and working memory capacity. Individual contributions made by each control component are not reported as this was not the aim of the study, however, they can be found in Appendix 8.3. To determine if a predictor made an independent contribution to inference generation skill, above and beyond the other predictors, three of the four predictors were entered in Step 2 and then the remaining predictor was entered in Step 3.

To explore potential age-related changes in the contribution of the four key predictors to inference generation skill, separate regression analyses were conducted for each age group, as opposed to controlling for age.

8.3.1. Preliminary Analysis

For strength of belief bias and accessibility of knowledge task data, all incorrect responses were removed as the dependent variable was response time. Consistent with previous research, a data trim was also conducted with all scores 2 standard deviations above or below each participant's grand mean being removed and all scores 2 standard deviations above or below each item's grand mean being removed (e.g. Calvo & Castillo, 2001a). In sum, for belief biases, 20% of Year 5 data, 16.80% of Year 7 data, 10% of Year 9 data, and 3.90% of adult data were lost due to errors and the trim conducted. These loss rates are in line with previous research (e.g. Moutier et al., 2006). Similarly, for accessibility of knowledge 10.57% of Year 5 data, 8.22% of Year 7, 6.00% of Year 9 data, and 8.13% of adult data were lost due to errors and the trim conducted. These loss rates are in line with previous research (e.g. Nelson & Kosslyn, 1975). After this trim all participants still had more than 50% of their data remaining.

8.3.2. Age-Related Changes in Accessibility of Knowledge, Amount of Knowledge, Strength of Belief Bias, and Inhibitory Control

8.3.2.1. Data Screening

Inspection of the data revealed that whilst the assumption of homogeneity of variance was met for most datasets, it was violated for the amount of knowledge dataset with the largest standard deviation being more than four times greater than the smallest standard deviation. The amount of knowledge data set was transformed using a log transformation, with the transformed data set used in all analyses. The data were skewed for all tasks. Moreover, whilst outliers were transformed for the accessibility of knowledge and strength of belief bias datasets, such that they were replaced with the highest or lowest data point within normal ranges, outliers were left in for all other measures as they were thought to reflect true scores. Across all age groups, this led to one outlier being included in IQ, vocabulary and amount of knowledge, and two outliers being included in single-word-reading. Consequently, non-parametric tests were conducted for all comparisons.

In Table 8.3., the means and standard deviations for IQ score (including vocabulary subscale and matrix reasoning subscale), single-word-reading skill, working memory capacity, inference generation skill (number of errors), amount of knowledge (total correct), accessibility of knowledge (response time), strength of belief bias (response time), and inhibitory control (number of errors) are presented as a function of year group. Table 8.3. shows that for all tasks performance appears to improve with age. A series of independent non-parametric ANOVAs were conducted to explore these differences further.

Table 8.3. Descriptive Statistics (Means and Standard Deviations) for Vocabulary, Matrix Reasoning, IQ, Single-Word-Reading, Accessibility of Knowledge, Amount of Knowledge, Strength of Belief Bias, Inhibition Skill and Working Memory Capacity for All Age Groups

	Year 5 Mean (SD)	Year 7 Mean (SD)	Year 9 Mean (SD)	Adults Mean (SD)	Minimum Score Possible	Maximum Score Possible
Amount of Knowledge (no. correct)	32.98 (2.77)	32.96 (2.54)	33.03 (2.69)	34.93 (1.44)	0	36
Accessibility of Knowledge (milliseconds)	1481.21 (839.29)	772.69 (232.81)	667.08 (373.61)	537.23 (308.81)	0	∞
Inhibitory Control (no. of errors)	6.41 (3.10)	5.47 (2.94)	4.18 (2.13)	3.07 (1.23)	0	20
Strength of Belief Biases (milliseconds)	2526.65 (1636.22)	1962.95 (1420.95)	1596.13 (1246.74)	1159.89 (952.34)	0	∞
Vocabulary Raw Score	23.2 (5.26)	30.16 (5.56)	35.91 (5.60)	40.68 (6.82)	0	Age 9-11: 64 Age 12-16: 72 Age 17-89: 80
Matrix Reasoning Raw Score	18.33 (5.84)	22.89 (3.86)	21.93 (5.56)	25.84 (4.88)	0	Age 9-11: 32 Age 12-44: 35
IQ Raw Score	41.41 (8.88)	53.04 (7.58)	57.85 (8.76)	66.52 (9.74)	0	Age 9-11: 96 Age 12-16: 107 Age 17-89: 115
Single Word Reading	44.77 (7.47)	54.56 (6.44)	57.88 (5.87)	63.91 (4.79)	0	70
Working Memory Capacity	14.35 (3.48)	16.60 (3.61)	17.85 (4.06)	19.16 (4.89)	0	30
Overall Inference Generation Errors	8.61 (3.13)	6.71 (3.34)	6.42 (3.23)	5.47 (2.94)	0	24
Real-World Coherence Inference Generation Errors	1.91 (0.97)	1.47 (1.24)	1.75 (1.36)	1.18 (1.25)	0	6
Real-World Elaborative Inference Generation Errors	2.14 (1.54)	1.94 (1.52)	1.25 (1.54)	1.43 (1.46)	0	6
Counterfactual-World Coherence Errors	2.51 (1.58)	2.00 (1.48)	2.33 (1.52)	1.57 (1.51)	0	6
Counterfactual-World Elaborative Errors	2.05 (1.77)	1.29 (1.47)	1.08 (1.32)	1.29 (1.45)	0	6

Minimum and Maximum scores refer to minimum and maximum possible scores

231

8.3.2.2. Accessibility of Knowledge

A non-parametric independent measures ANOVA was conducted to explore the effect of age on accessibility of knowledge. A main effect of age was found ($\chi2$ (3, $N = 207$) = 65.204, $p < .001$). Six independent non-parametric t-tests were conducted to explore further. Results suggest that Year 5 access knowledge significantly slower than Year 7 ($U = 403.00$, $p < .001$, $N = 101$), Year 9 ($U = 273.00$, $p < .001$, $N = 101$), and adults ($U = 333.00$, $p < .001$, $N = 107$). There was found to be no significant difference between Year 7 and Year 9 ($U = 528.00$, $p = .030$, $N = 100$) or Year 9 and adults ($U = 723.00$, $p = .088$, $N = 106$). Year 7 were found to access knowledge significantly slower than adults ($U = 680.00$, $p < .001$, $N = 107$). This suggests that after Year 7, development in accessibility of knowledge is gradual.

8.3.2.3. Amount of Knowledge

A non-parametric independent measures ANOVA was conducted to explore the effect of age on amount of knowledge. A significant main effect of age was observed ($\chi2$ (3, $N = 207$) = 31.38, $p < .001$). Six independent non-parametric t-tests were conducted to explore further. Adults were found to have significantly more knowledge than Year 5 ($U = 720.00$, $p < .001$, $N = 107$), Year 7 ($U = 559.50$, $p < .001$, $N = 106$), and Year 9 ($U = 497.00$, $p < .001$, $N = 106$). No other comparisons were significant ($p > .81$ for all remaining comparisons). This suggests that amount of knowledge does not vary from Year 5 to Year 9, however, adults do appear to have more knowledge than all younger groups[32].

8.3.2.4. Strength of Belief Bias

A non-parametric independent measures ANOVA was conducted to explore the effect of age on strength of belief bias. A significant main effect of age was found ($\chi2$ (3, $N = 207$) = 25.27, $p < .001$). Six non-parametric independent t-tests revealed that Year 5 responded significantly slower than Year 7 ($U = 796.00$, $p = .009$, $N = 101$), Year 9 ($U = 536.00$, p

---

[32] It is important to note that amount of knowledge was near ceiling levels for all groups, although gradual development was observed. The lack of knowledge acquisition with age is surprising since the knowledge-base necessary to generate inferences in the Image Selection Task was based on the Key Stage 2 and Key Stage 3 science curriculums. Subsequently, Year 5s would not have encountered the Key Stage 3 curriculum. However, the knowledge-base used was fairly general. For example, rather than focusing on the specifics of reproduction explored in Key Stage 3 (e.g. mitosis), the knowledge-base is focused on general ideas (e.g. mammals give birth to live young, reptiles lay eggs).

= .005, N = 101) and adults (U = 719.00, p < .001, N = 107). Year 7 were found to respond significantly slower than adults (U = 799.00, p = .002, N = 106). However, none of the remaining comparisons were significant (p > .15 for remaining comparisons). This suggests a gradual development in strength of belief biases from Year 7 to adulthood.

8.3.2.5. Inhibitory Control

A non-parametric independent measures ANOVA was conducted to explore the effect of age on inhibitory control. A significant main effect of age was found ($\chi2$ (3, N = 207) = 46.34, p < .001). There was found to be no significant difference between the number of errors made by Year 5 and Year 7 (U = 934.00, p = .113, N = 101). Year 5 were found to make significantly more inhibition errors than Year 9 (U = 482.50, p = .001, N = 101) and adults though (U = 435.00, p < .001, N = 107). There was found to be no significant difference between the number of inhibition errors made by Year 7 and Year 9 (U = 552.50, p = .052, N = 100). However, Year 7 were found to make significantly more inhibition errors than adults (U = 579.50, p < .001, N = 106). Year 9s were also found to make significantly more inhibition errors than adults (U = 612.50, p = .007, N = 106). This suggests that whilst inhibitory control develops from Year 5 to adulthood, development between Year 5 and Year 7 and Year 7 and Year 9 is minimal.

8.3.3. The Relationships between Amount of Knowledge, Accessibility of Knowledge, Inhibitory Control, Strength of Belief Biases and Inferential Skill

8.3.3.1. Data Screening

Data screening revealed that all correlations were below .8, all VIF values were below 10 and all Tolerance values were greater than .1. For all analyses the Durbin-Watson statistic was between 1 and 3, suggesting the assumption of 'lack of autocorrelation' was met. Residual statistics were examined for extreme cases. For all analyses, there were never more than three cases (out of 50) with standardised residual statistics higher than +/- 2. Moreover, there was never more than one standardised residual greater than +/-2.5. Therefore, it can be assumed that the regression models were accurate. For most analyses residual errors in the model were normally distributed. For some models there was slight skew, however, this was never severe.

For most analyses assumptions of homoscedasticity and linearity were met. However, there appears to be slight heteroscedasticity for the following models:

- Year 7 real-world coherence model

233

- Year 7 counterfactual-world coherence model

- Adult real-world coherence model

- Adult counterfactual-world coherence model

The validity of these models is questionable and caution must be taken when generalising these results beyond the current sample.

Means were calculated for each group for performance on each task. As can be seen from Table 8.3., ceiling effects were present for amount of knowledge. No other ceiling or floor effects were present.

8.3.3.2. The Predictive Utility of Amount of Knowledge, Accessibility of Knowledge, Inhibitory Control, and Strength of Belief Biases in the Inferential Process

This study aimed to explore the potentially changing unique contribution of each component in the inferential process throughout adolescence and into adulthood. Therefore, the individual contribution of each component, after controlling for all other components is explored below for each inference type and each age group. For all analyses composite IQ scores, working memory capacity scores and single-word-reading scores were entered in Step 1 to account for relationships between inference generation and IQ, working memory capacity and decoding. The results of composite analyses, combining all four predictor variables in Step 2, for each dependent variable can be found in Appendix 8.4[33]. Beta values for all variables for each regression model can be found in Appendix 8.5.

---

[33] Additional analysis was also ran, such that groups were split by reading comprehension age instead of chronological age. This was due to the high number of weak comprehenders in Year 9 – see Section 6.3.3.2. Results of this analysis are not reported as variance was largely explained by the control variables, with no clear relationships emerging. This is unsurprising given the relationships reported between reading comprehension, single-word-reading ability, working memory and IQ (Carver, 1990; Dixon, Lefevre, & Twilley, 1988; Masson & Miller, 1983; Singer, Andrusiak, Reisdorf, & Black, 1992; Singer & Richot, 1996; Whitney, Ritchie, & Clark, 1991).

8.3.3.2.1. Amount of Knowledge and Inferential Skill

To explore the independent contribution of amount of knowledge, accessibility of knowledge, inhibitory control, and strength of belief bias were entered at Step 2 and amount of knowledge at Step 3. Results can be found in Table 8.4.-8.7.

Table 8.4. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-World Coherence Inference Generation Skill as the Dependent Variable and Amount of Knowledge as the Predictor Variable, controlling for Word-Reading, IQ, and Working Memory Capacity, Accessibility of Knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.273**	.145	.106	.126
2	Accessibility of knowledge, inhibitory control, and strength of belief bias,	.095	.179*	.306**	.387***
3	Amount of Knowledge	.163***	.153**	.076*	.050*

* p < .05, ** p < .01, *** p < .001

Table 8.5. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-World Elaborative Inference Generation Skill as the Dependent Variable and Amount of Knowledge as the Predictor Variable, controlling for Word-Reading, IQ, and Working Memory Capacity, Accessibility of Knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.043	.182*	.047	.052
2	Accessibility of knowledge, inhibitory control, and strength of belief bias,	.269**	.134	.188	.437***
3	Amount of Knowledge	.020	.035	.335***	.010

* p < .05, ** p < .01, *** p < .001

Table 8.6. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Counterfactual-World Coherence Inference Generation Skill as the Dependent Variable and Amount of Knowledge as the Predictor Variable, controlling for Word-Reading, IQ, and Working Memory Capacity, Accessibility of Knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.053	.074*	.207	.030
2	Accessibility of knowledge, inhibitory control, and strength of belief bias,	.537***	.227*	.265**	.276***
3	Amount of Knowledge	.003	.020	.005	.009

* p < .05, ** p < .01, *** p < .001

Table 8.7. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Counterfactual-World Elaborative Inference Generation Skill as the Dependent

Variable and Amount of Knowledge as the Predictor Variable, controlling for Word-Reading, IQ, and Working Memory Capacity, Accessibility of Knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.114	.061	.081	.054
2	Accessibility of knowledge, inhibitory control, and strength of belief bias	.258**	.473***	.385***	.177*
3	Amount of Knowledge	.000	.004	.003	.012

* p < .05, ** p < .01, *** p < .001

Amount of knowledge appears central to real-world coherence inferential skill, sharing a negative relationship, such that as amount of knowledge increases number of real-world coherence inferential errors made decreases. Specifically, amount of knowledge was found to account for 16.3%, 15.3%, 7.6% and 5% of variance in Year 5, Year 7, Year 9 and adults, respectively, after controlling for accessibility of knowledge, inhibitory control, strength of belief bias, IQ, working memory capacity, and word-reading skill. Whilst a significant contribution was not made in Year 9, this was approaching significance (*p = .06*). Amount of knowledge also accounted for a significant 33.5% unique variance in real-world elaborative inferential skill in Year 9. The relationship was negative. Amount of knowledge did not make a unique contribution to any of the remaining variables (*p > .17*).

8.3.3.2.2. Accessibility of Knowledge and Inferential Skill

To explore the independent contribution of accessibility of knowledge, amount of knowledge, inhibitory control, and strength of belief bias were entered at Step 2 and accessibility of knowledge at Step 3. Results can be found in Tables 8.8-8.11.

Table 8.8. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-world Coherence Inference Generation Skill as the Dependent Variable and Accessibility of Knowledge as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.273**	.145	.106	.126*
2	Amount of knowledge, inhibitory control, and strength of belief bias	.235**	.309***	.195	.144
3	Accessibility of knowledge	.023	.023	.188**	.294***

* p < .05, ** p < .01, *** p < .001

Table 8.9. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-world Elaborative Inference Generation Skill as the Dependent Variable and Accessibility of Knowledge as the Predictor Variable, controlling for Word-Reading,

IQ, Working Memory Capacity, Amount of knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.043	.182*	.047	.052
2	Amount of knowledge, inhibitory control, and strength of belief bias	.029	.169*	.410***	.322***
3	Accessibility of knowledge	.260***	.000	.113*	.125***

* p < .05, ** p < .01, *** p < .001

Table 8.10. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Counterfactual-World Coherence Inference Generation Skill as the Dependent Variable and Accessibility of Knowledge as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.053	.074	.207	.030
2	Amount of knowledge, inhibitory control, and strength of belief bias	.510***	.223*	.239*	.274***
3	Accessibility of knowledge	.030	.024	.031	.010

* p < .05, ** p < .01, *** p < .001

Table 8.11. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Counterfactual-World Elaborative Inference Generation Skill as the Dependent Variable and Accessibility of Knowledge as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of knowledge, Inhibitory Control, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.114	.061	.081	.054
2	Amount of knowledge, inhibitory control, and strength of belief bias	.249**	.476***	.369**	.177*
3	Accessibility of knowledge	.008	.002	.019	.011

* p < .05, ** p < .01, *** p < .001

Accessibility of knowledge was a unique predictor of real-world coherence inferential skill in Year 9 and adults, accounting for 18.8% and 29.4% unique variance, respectively, after controlling for amount of knowledge, inhibitory control, strength of belief bias, IQ, working memory capacity, and word-reading skill. The positive relationships observed suggest that as time taken to access knowledge increases number of real-world coherence inferential errors made also increases. Accessibility of knowledge also appears central to real-world elaborative inferential skill accounting for 26.0%, 11.3% and 12.5% unique variance in Year 5, Year 9 and adults, after controlling for amount of knowledge, inhibitory control, strength of belief bias, IQ, working memory capacity, and word-reading skill. Again, the relationship was positive. Accessibility of knowledge did not make a unique contribution to any remaining variables (*p > .09*).

8.3.3.2.3. Inhibitory Control and Inferential Skill

To explore the independent contribution of inhibitory control, amount of knowledge, accessibility of knowledge, and strength of belief bias were entered at Step 2 and inhibitory control at Step 3. The results can be found in Tables 8.12.-8.15.

Table 8.12. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-world Coherence Inference Generation Skill as the Dependent Variable and Inhibitory Control as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of Knowledge, Accessibility of Knowledge, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.273**	.145	.106	.126
2	Amount of knowledge, accessibility of knowledge, and strength of belief bias	.258**	.313***	.376***	.434***
3	Inhibitory Control	.000	.019	.007	.003

* p < .05, ** p < .01, *** p < .001

Table 8.13. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-world Elaborative Inference Generation Skill as the Dependent Variable and Inhibitory Control as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of Knowledge, Accessibility of Knowledge, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.043	.182*	.047	.052
2	Amount of knowledge, accessibility of knowledge, and strength of belief bias	.267**	.048	.522***	.237**
3	Inhibitory Control	.022	.121*	.001	.210***

* p < .05, ** p < .01, *** p < .001

Table 8.14. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Counterfactual-World Coherence Inference Generation Skill as the Dependent Variable and Inhibitory Control as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of Knowledge, Accessibility of Knowledge, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ, Working Memory Capacity	.053	.074	.207	.030
2	Amount of knowledge, accessibility of knowledge, and strength of belief bias	.537***	.247**	.124	.277**
3	Inhibitory Control	.003	.000	.146**	.057

*p < .05, **p < .01, ***p < .001*

Table 8.15. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Counterfactual-World Elaborative Inference Generation Skill as the Dependent Variable and Inhibitory Control as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of Knowledge, Accessibility of Knowledge, and Strength of Belief Bias

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Word-Reading Skill, Composite IQ, Working Memory Capacity	.114	.061	.081	.054
2	Amount of knowledge, accessibility of knowledge, and strength of belief bias	.255**	.074	.094	.188*
3	Inhibitory Control	.002	.403***	.294***	.001

Inhibitory control appears to play a key role in real-world elaborative inferential skill in Year 7 and adulthood. After controlling for reading skill, IQ, working memory capacity, amount of knowledge, accessibility of knowledge, and strength of belief bias, inhibitory control accounted for 12.1% and 21.0% unique variance in real-world elaborative inferential skill for Year 7 and adults, respectively. Similarly, inhibitory control accounted for 40.3% and 29.4% unique variance in counterfactual-world elaborative inferential skill for Year 7 and Year 9, respectively. In addition, inhibitory control accounted for 14.6% unique variance in counterfactual-world coherence inferential skill in Year 9. Whilst inhibitory control was a positive predictor of real-world and counterfactual-world elaborative inference generation skill in Year 5 and Year 7, such that those with weaker inhibitory control made more inferential errors, the relationships were negative for Year 9 and adults, such that those with weaker inhibitory control made fewer inferential errors. Inhibitory control shared a positive relationship with counterfactual-world coherence inferential skill in Year 9. The relationship between inhibitory control and inference generation skill appears to be complex. Consequently, this relationship is explored further in the additional analysis below[34].

8.3.3.2.4. Belief Biases and Inferential Skill

---

[34] Additional analysis was conducted, such that Year 9s with very high error rates were identified and removed (n = 2). However, when these participants were removed, patterns did not change. Therefore, these results are not reported.

To explore the independent contribution of strength of belief bias, amount of knowledge, accessibility of knowledge, and inhibitory control were entered at Step 2 and strength of belief bias at Step 3. Results can be found in Tables – 8.16.-8.19.

Table 8.16. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-world Coherence Inference Generation Skill as the Dependent Variable and Strength of Belief Bias as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of knowledge, Accessibility of Knowledge, and Inhibitory Control

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ Working Memory Capacity	.273**	.145	.106	.126
2	Amount of knowledge, accessibility of knowledge, and inhibitory control	.257***	.316***	.375***	.420***
3	Strength of belief bias	.000	.016	.007	.017

* p < .05, ** p < .01, *** p < .001

Table 8.17. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Real-world Elaborative Inference Generation Skill as the Dependent Variable and Strength of Belief Bias as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of knowledge, Accessibility of Knowledge, and Inhibitory Control

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ Working Memory Capacity	.043	.182*	.047	.052
2	Amount of knowledge, accessibility of knowledge, and inhibitory control	.289**	.134	.493***	.447***
3	Strength of belief bias	.000	.035	.029	.001

* p < .05, ** p < .01, *** p < .001

Table 8.18. Summary of Fixed-Order Hierarchical Multiple Regression Analyses (ΔR2) With Counterfactual-World Coherence Inference Generation Skill as the Dependent Variable and Strength of Belief Bias as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of knowledge, Accessibility of Knowledge, and Inhibitory Control

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ Working Memory Capacity	.053	.074	.207	.030
2	Amount of knowledge, accessibility of knowledge, and inhibitory control	.006	.107	.207*	.117
3	Strength of belief bias	.534***	.140**	.063	.168***

* p < .05, ** p < .01, *** p < .001

Table 8.19. Summary of Fixed-Order Hierarchical Multiple Regression Analyses ($\Delta R2$) With Counterfactual-world Elaborative Inference Generation Skill as the Dependent Variable and Strength of Belief Bias as the Predictor Variable, controlling for Word-Reading, IQ, Working Memory Capacity, Amount of knowledge, Accessibility of Knowledge, and Inhibitory Control

Step	Variable	Year 5	Year 7	Year 9	Adults
1	Single Word Reading, Composite IQ Working Memory Capacity	.114	.061	.081	.054
2	Amount of knowledge, accessibility of knowledge, and inhibitory control	.008	.320***	.296*	.017
3	Strength of belief bias	.250***	.157***	.092*	.172**

* p < .05, ** p < .01, *** p < .001

Strength of belief biases appear central to both coherence and elaborative counterfactual-world inferential skill with positive relationships observed, such that the more time taken to respond on the belief bias task the more counterfactual-world errors made. Specifically, after controlling for word-reading, IQ, working memory capacity, amount of knowledge, accessibility of knowledge, and inhibitory control, strength of belief biases accounted for 53.4%, 14.0%, and 16.8% of variance in counterfactual-world coherence inference skill in Year 5, Year 7 and adults, respectively. Similarly, strength of belief biases accounted for 25.0%, 15.7%, 9.2% and 17.2% of variance in counterfactual-world elaborative inference skill in Year 5, Year 7, Year 9 and adults, respectively. Strength of belief biases therefore appears to play a central role in counterfactual-world inference generation processing during adolescence.

8.3.3.3. Additional Analysis – Exploring the Role of Inhibition

Floyd et al. (2012) highlighted the indirect effects of many skills on comprehension across the lifespan using multiple regression. However, as noted by Floyd et al. (2012) most multiple regression analyses fail to explore these indirect relationships and the factors that may mediate them. Floyd et al. argue that consideration of these relationships is likely to lead to a much richer and clearer insight into reading comprehension. Given that the direction of the relationship between inhibitory control and inference generation skill changed from positive to negative between Year 7 and Year 9, further analyses were conducted to explore the role of inhibition in elaborative inference generation in more detail. For each group, the median inhibition score was calculated and used to explore

individual differences in real-world elaborative inference generation skill and counterfactual-world elaborative inference generation skill[35].

8.3.3.3.1. Real-World Elaborative Inference Generation Skill

For Year 5, there is no significant difference in number of real-world elaborative errors made by those with good inhibition skills and those with weak inhibition skills (t (49) = 0.66, p = .51). For Year 7, there appears to be no significant difference in the number of real-world elaborative errors made by those with good inhibition skills and those with weak inhibition skills (t (48) = 1.91, p = .06). For Year 9, there was found to be no significant difference in the number of real-world elaborative errors made by those with good inhibition skills and those with weak inhibition skills (t (48) = 0.72, p = .48) [36]. For adults, significantly more real-world elaborative errors were found to be made by those with good inhibitory skills (mean = 1.88, SD = 0.96) compared to those with weak inhibitory skills (mean = 0.63, SD = 0.95; t (54) = 4.43, p < .001). Therefore, inhibitory control appears central to real-world elaborative inference generation skill in adults, with strong inhibitory skills resulting in weaker inference generation skills.

8.3.3.3.2. Counterfactual-World Elaborative Inference Generation Skill

For Year 5, there was no significant difference in number of counterfactual-world elaborative errors made by those with good inhibition skills and those with weak inhibition skills (t (49) = 0.71, p = .49). For Year 7, those with good inhibition skills were found to make significantly fewer errors (mean = 1.08, SD = 1.10) than those with weak inhibition skills (mean = 2.29, SD = 1.31; t (48) = 3.35, p = .002). For Year 9, those with weak inhibition skills were found to make significantly fewer errors (mean = 0.81, SD = 1.09) than those with good inhibition skills (mean = 2.14, SD = 1.03; t (48) = 3.95, p < .001) [37]. For adults, there was no significant difference between those with weak inhibition and good inhibition skills (t (54) = 0.21, p = .83). Therefore, inhibitory control appears to be central to counterfactual-world elaborative inference generation in Year 7

---

[35] Median split analysis was conducted with those Year 9s making a high number of inhibition errors removed (n = 2). These results are not reported as patterns were the same regardless of whether these participants were included or excluded.

[37] The same pattern was observed when those Year 9s identified has making a high number of inhibition errors were removed (t (46) = 3.76, p = .001).

and Year 9, however, its primary function appears to change during this time. The nature of this change is discussed below.

8.4. Discussion

This study sought to establish age-related changes in the cognitive underpinnings of real-world coherence, real-world elaborative, counterfactual-world coherence and counterfactual-world elaborative inference generation skill in Year 5 (9-10 years), Year 7 (11-12 years), Year 9 (13-14 years) and adults (18 years and over). The independent predictive utility of four key predictors (amount of knowledge, accessibility of knowledge, inhibitory control, and strength of belief bias) was explored. The data presented here extend knowledge in several ways. First, the cognitive underpinnings of real-world coherence, real-world elaborative, counterfactual-world coherence and counterfactual-world elaborative inference generation skill differed and changed with age. Specifically, knowledge was central to real-world coherence inference, with a shift in importance from amount of knowledge to accessibility of knowledge between Year 7 and Year 9. Knowledge and inhibitory control were central to real-world elaborative inferential skill, with inhibitory control shifting from sharing a positive relationship with elaborative inferential skill in Year 7 to a negative relationship in Year 9. Belief biases and inhibitory control were central to both counterfactual-world coherence and elaborative inference generation skill. The direction of the relationship between inhibitory control and counterfactual-world coherence and elaborative inferential skill was found to fluctuate. Taken together, findings suggest that coherence and elaborative inferences are distinct inference types with unique developmental trajectories.

8.4.1. The Cognitive Underpinnings of Real-World Coherence Inference Generation Skill

Real-world coherence inference generation skill appears to be underpinned by knowledge. Amount of knowledge made a significant unique contribution to real-world coherence inference generation skill in Year 5 and Year 7 accounting for 16.3% and 15.3% unique variance, respectively. Accessibility of knowledge made a significant unique contribution in Year 9 and Adults, accounting for 18.8% and 29.4% unique variance, respectively. Subsequently, whilst knowledge remains central to real-world coherence inference generation skill in all age groups, there appears to be a developmental shift between Year 7 and Year 9 in importance from amount of knowledge to accessibility of knowledge. These results support those implicating knowledge in the inferential process and build

upon previous research by showing how the role of knowledge changes during adolescence (Birch & Bloom, 2004; Prat & Just, 2010; Prat et al., 2007; 2011).

The current findings conflict with Barnes et al. (1996), one of the few studies to explore the role of knowledge in inferential processing developmentally, however. Barnes et al. (1996) found that the role of accessibility of knowledge in coherence inference generation declined with age. Real-world coherence inferences are argued to be drawn whilst reading since they are essential to maintaining coherence (see Graesser et al., 1994). Subsequently, Barnes et al. suggested that, due to increasing comprehension monitoring skills and identification of coherence breaks, accessibility of knowledge becomes less important to coherence inference generation skill with age, such that readers will strive to generate the inference and maintain coherence even if the necessary knowledge is not readily accessible. However, research suggests that in most cases coherence inferences are generated with minimal effort, particularly as the reader develops, such that inferential speed is found to increase with age (Magliano et al., 1993; Millis & Graesser, 1994). This is unsurprising given that research shows that both word-reading skill and reading fluency increase with age (Adlof, Catts, & Little, 2006; Carver, 1993; Tilstra, McMaster, van den Broek, Kendeou, & Rapp, 2009). For older readers to generate inferences without disrupting fluent reading, highly accessible knowledge would facilitate the efficient and relatively automatic generation of those inferences needed to maintain coherence *whilst* reading. For participants in Year 5 and Year 7, amount of knowledge was the strongest independent predictor of real-world coherence inferential skill. Year 5 and Year 7 may thus be more likely to engage in laborious search processes since reading is a slower and more demanding process at this age (Adlof et al., 2006; Carver, 1993; Tilstra et al., 2009).

An alternative explanation may be due to Barnes et al. (1996) priming the necessary knowledge to control for the possibility of differing amounts of knowledge across the age groups by, first, teaching participants the necessary knowledge-base. As a result, when generating coherence inferences in the Barnes et al. study the necessary knowledge, even if not readily accessible, may have been weakly activated, thus requiring minimal effort to be fully activated. Participants may have engaged in this additional processing, knowing that they would find the knowledge they would need. However, when reading naturally, knowledge is not primed in such a way, instead, the knowledge necessary for coherence inferences is thought to be activated through associative processing, (van den Broek et al., 1995; 1999; 2005). Therefore, if the necessary knowledge is not freely available, to prevent reading disruption, older readers may not engage in this potentially

time-consuming yet possibly fruitless task. Subsequently, accessibility of knowledge would be essential to successful real-world coherence inference generation for these older groups. The results of the current study may thus be more reflective of inferential processing during the natural reading process. However, it is important to note that near ceiling effects were observed on the amount of knowledge task. Therefore, the role of amount of knowledge may not be adequately represented in the current study, particularly in older participants where variance in scores within the groups is minimal. In sum, knowledge appears to be key to real-world coherence inferential skill, such that with age accessibility of knowledge becomes more important than mere possession of knowledge. This may be due to the increasing necessity to generate coherence inferences whilst reading, with only minimal effort.

8.4.2. The Cognitive Underpinnings of Real-World Elaborative Inference Generation Skill

For Year 5, accessibility of knowledge was found to make the strongest unique contribution to real-world elaborative inferential skill, accounting for 26% independent variance, after controlling for all other skills of interest in this thesis. Knowledge activation is argued to be associative, such that semantic knowledge is thought to be stored in networks, with those knowledge nodes with multiple or strong links to the target knowledge node also being activated (Kintsch, 1988; 1998). Therefore, when reading or performing a conscious search of long-term memory, improvements in knowledge activation could increase the likelihood of the knowledge necessary for elaborative inference generation being activated and the target inference being generated. Research suggests that knowledge activation increases during adolescence (Barnes et al., 1996). This is supported by the findings of the current study, such that accessibility of knowledge significantly increased from Year 5 to Year 7. Therefore, it is possible that even when given a prompt and time, Year 5 are still unable to retrieve the necessary knowledge. Subsequently, when generating real-world elaborative inferences, Year 5 may only do so with knowledge that is easily accessible if they are unable to quickly access the target knowledge.

The increase in knowledge accessibility may explain the central role of inhibitory control in Year 7. Inhibitory control was the only skill found to make an independent contribution to real-world elaborative inference generation skill in Year 7, accounting for 12.10% unique variance, such that those Year 7s with stronger inhibitory skills were more

successful when generating real-world elaborative inferences. Increasing knowledge accessibility means that compared to Year 5, when searching long-term memory, Year 7 may activate more knowledge automatically. Subsequently, some of the knowledge activated may be irrelevant. Inhibitory control would be needed to reduce processing costs by ensuring only the most relevant information is processed (e.g. Cain, 2006). This reduces the load on working memory, allowing for the execution of those other skills and processes necessary for inference generation (e.g. semantic integration).

There appears to be a shift in the primary function of inhibitory control in the elaborative inferential process between Year 7 and Year 9, such that, whilst inhibitory control shares a positive relationship with real-world elaborative inferential skill in Year 5 and Year 7, in Year 9 and adults a negative relationship is observed. This suggests that for the two older groups, those with stronger inhibitory control are less successful when generating real-world elaborative inferences. However, inhibitory control only makes a significant contribution to real-world elaborative inference generation skill in adults, accounting for 21% unique variance. There are several possible interpretations. First, inhibitory control has been implicated in the suppression of irrelevant information and consequent maintenance of a clear workspace whilst generating inferences (e.g. Cain, 2006). However, if this process was to become too efficient, with only knowledge with very high activation levels being maintained, inhibitory control could hinder elaborative inference generation. Research suggests that elaborative inferences are generated offline for a variety of reasons, one being the number of possibilities is too numerous to be fully explored whilst reading (e.g. Calvo & Castillo, 2001a; 2001b). Given that adults are able to access knowledge with more ease than CYP, the number of possibilities available when given time to consciously search long-term memory may be very high. Those with strong inhibitory control may quickly isolate one idea whereas those with weak inhibitory control maintain more possibilities. If, however, the one possibility selected by those with good inhibitory control is inconsistent with the target inference then accuracy will not be achieved. Alternatively, those with weak inhibitory control are more likely to have the target inference activated as a possibility. This is supported by the findings of the current study, such that accessibility of knowledge also made a significant independent contribution to the real-world elaborative inferential processing of adults, accounting for 12.5% unique variance. Similarly, inhibitory control was found to improve between Year 7 and adulthood and Year 9 and adulthood, but not between Year 7 and Year 9. This may explain the pattern of results observed, such that Year 9 are beginning to focus on just

one possibility, but do not (or only some participants) have the level of inhibitory control needed to do this successfully.

Second, the IST may not be suitable for assessment of inhibition associated with language processing (Barnes et al., 2004; Cain, 2006; Gernsbacher, 1993; Harnishfeger, 1995; Kintsch, 1988). Inhibition is thought to comprise a 'family' of skills (Friedman & Miyake, 2004). The three functions include: 1) prepotent response inhibition – the suppression of a dominant, often automatic response to the stimuli presented 2) response to distractor inhibition – suppression of irrelevant stimuli to focus on relevant stimuli, when both are present simultaneously 3) resistance to proactive interference – suppression of information that is no longer relevant. Prepotent response inhibition and response to distractor inhibition are related to incoming environmental information. Research exploring the relationship between prepotent response inhibition and response to distractor inhibition finds no difference between good and poor comprehenders suggesting these aspects of inhibitory control do not underpin the reading comprehension processes (Arrington, Kulesz, Francis, Fletcher & Barnes, 2014; Borella et al., 2014).

Due to the motor response method, the IST may assess prepotent response inhibition. The IST requires participants to select one picture from four. It is possible then that more impulsive responders do not consider all four possibilities when making their selection. The IST may thus be tapping this aspect of the individual as opposed to the inhibitory control associated with accurate inference generation. This may explain the variation in results observed.

Similarly, the rule-shift-cards task is dependent upon deliberate and conscious inhibitory processing (Wilson et al., 1996). The method of suppression in the comprehension process, however, is thought to be largely unconscious (e.g. Kintsch, 1988). The Construction-Integration theory, for example, suggests that the knowledge integrated into a situational model and the knowledge lost results from several iteration cycles, with knowledge nodes with high activation levels becoming stronger, and knowledge nodes with low activation levels becoming weaker and eventually being degraded or pruned. The rule-shift-cards task is unlikely to tap this implicit process of suppression. In addition, the rule-shift-cards task is not a pure measure of inhibitory control, such that other skills such as working memory and long-term memory are required for successful completion. Finally, there is little variation in the adult sample's inhibition scores. As a result, any dip in performance could result in a strong relationship being observed due to the regressional

nature of this study. It is important to note that all of these possibilities are extremely speculative. Further research using several measures of inhibition is needed to unpick the relationship between inhibition and inference generation skill.

For Year 9 amount of knowledge was the strongest independent predictor, accounting for 33.5% unique variance. It is possible that Year 9 are engaging in a laborious search processes to find the knowledge which matches the target inference presented at the time of test. Accessibility of knowledge was the second strongest predictor accounting for 11.3% unique variance. This suggests that whilst Year 9 engage in the laborious search processes needed for elaborative inference generation, this process is facilitated by readily accessible knowledge. However, these suggestions are made very tentatively. The unique pattern of skills underpinning inference generation in Year 9 coincides with a unique pattern of inferential skill also occurring in Year 9. This suggests that Year 9 may be a unique time in the processing of inference generation. This is discussed further below.

In sum, elaborative inference generation appears to be underpinned by accessibility of knowledge, however, as accessibility of knowledge increases, inhibitory control seems to become key. However, it is possible that the primary function of inhibitory control changes with age. Whereas in Year 7 it appears that inhibitory control is needed to inhibit irrelevant knowledge resulting from increasing accessibility of knowledge, in adults inhibitory control appears to constrain the search process, minimising the possibilities weakly activated at the time of test and/or the number of possibilities isolated when searching long-term memory. Similar to the role of knowledge in coherence inferential processing, the shift in function of inhibitory control may also occur between Year 7 and Year 9.

8.4.3. The Cognitive Underpinnings of Counterfactual-World Coherence Inferential Processing

Strength of belief bias was the only predictor to make a unique contribution to counterfactual-world coherence inferential skill in Year 5 and Year 7, accounting for 53.4% and 14.0% unique variance respectively, such that those with stronger belief biases made more counterfactual-world coherence errors. This suggests that when processing counterfactual-world coherence inferences, belief biases result in the activation of and preference for processing real-world knowledge, which serves to limit successful counterfactual-world inference generation. This is consistent with previous research

implicating belief biases in counterfactual-world processing (Houde & Guichart, 2001; Moutier et al., 2006; Richards & Sanderson, 1999). Extant research exploring belief biases had focused on counterfactual-world reasoning rather than counterfactual-world inference generation. Therefore, these findings extend previous research by highlighting the key role of belief biases in counterfactual-world coherence inferential processing. These results also extend previous research by showing that even when a counterfactual-world context is provided, counterfactual-world inferential processing still appears to be affected by belief biases. This conflicts with previous research exploring counterfactual-world literal processing which finds that whilst initially incoming text is compared against real-world beliefs, as the text progresses and a counterfactual-world mental model is constructed text is compared against the counterfactual-world model instead (e.g. Nieuwland, 2015). A tentative explanation is that belief biases may play a stronger role in inferential rather than literal processing as inferences are constructed by integrating one or more textual premise with existing knowledge. The reader thus creates the inference; to be accepted the reader evaluates the validity of the inference generated (Byrne, 2007). Belief biases may thus impede validation, a process not necessary when processing literal information.

In Year 9, strength of belief biases was no longer a significant independent predictor of counterfactual-world coherence inference skill. Instead, inhibitory control was found to be the strongest unique predictor, accounting for 14.6% independent variance, such that those Year 9s with stronger inhibitory control made fewer counterfactual-world coherence errors. Although not significant, strength of belief biases was the second strongest predictor of counterfactual-world coherence inference skill, accounting for 6.3% unique variance. Previous research suggests that for counterfactual-world processing to be successful, the real-world interference experienced must be inhibited (e.g. Moutier et al., 2006). This may thus explain the pivotal role of inhibitory control, above and beyond strength of belief biases, during Year 9. It is notable, then, that inhibitory control does not make a significant independent contribution to counterfactual-world coherence inference skill in Year 5 and Year 7. It is possible that the central role of inhibitory control in Year 9, but not Year 5 and Year 7, is due to a qualitative shift in the application of belief biases.

Mitchell et al. (1996) suggest that when processing a narrative, those aged 9-10 years apply a 'seeing is believing rule', such that processing is based on a character's actual experience of an event. Conversely, adults are believed to adopt a more complex

processing strategy, such that they consider other real-world alternatives that may affect the character's actions. Therefore, CYP may more readily accept the information in the text, even when it is contradictory to what they already know, whereas adults will try to find a real-world alternative that can explain the text. For example, when reading *"Every day the boy ate chocolate cake all day long and every day the boy grew skinnier and skinner"* all readers may activate background knowledge relating to food intake changing body size. This may then lead to the inference: *eating chocolate cake makes the boy skinny*. Given that the text provides no additional information, CYP may accept this inference because it is based on the character's experiences. However, adults may question the inference generated, and draw upon other knowledge sources to generate an alternative real-world inference that is consistent with the text – e.g. *the boy has bulimia*. For the correct counterfactual-world inference to be drawn, the adult must inhibit the real-world interference. Mitchell et al. (1996) only used a child (9-10 years) and adult group, thus the application of belief biases during adolescence was unknown. However, given the central role of inhibitory control in Year 9 but not the younger groups, current results suggest that Year 9 patterns are consistent with the more complex application of belief biases observed by the adults in the Mitchell et al. study. The findings of the current study, therefore, extend previous research by suggesting that the qualitative shift in the application of belief biases occurs between Year 7 and Year 9.

By adulthood, inhibitory control no longer makes a unique contribution to counterfactual-world coherence inference skill, instead strength of belief biases was the only predictor to make a significant unique contribution, accounting for 16.8% independent variance. The difference in patterns between Year 9 and adults may be due to improvements in inhibitory control. For adults, inhibitory control may be sufficient, such that adults can effectively inhibit the real-world interference experienced by the more complex application of belief biases. In sum, belief biases appear central to counterfactual-world coherence inferential skill across the lifespan. However, there appears to be a qualitative shift in the way belief biases are applied from Year 7 to Year 9, such that older readers apply a more complex strategy.

8.4.4. The Cognitive Underpinnings of Counterfactual-World Elaborative Inferential Processing

Strength of belief biases was the only predictor to make a unique contribution to counterfactual-world elaborative inferential skill in Year 5, accounting for 25.0%

independent variance, such that stronger belief biases resulted in more errors. This is consistent with the patterns observed for counterfactual-world coherence inferential processing. For Year 7, strength of belief biases was still a unique predictor of counterfactual-world elaborative inferential skill, accounting for 15.7% independent variance. However, inhibitory control made the strongest unique contribution, accounting for 40.3% unique variance. Those with good inhibitory control were more successful when generating counterfactual-world elaborative inferences than those with weak inhibitory control. The increasing importance of inhibitory control from Year 5 to Year 7 is surprising, given that inhibitory control did not increase during this time, and strength of belief biases decreased. However, the inhibitory control measure used assessed inhibitory skill (number of errors), not inhibitory speed. Research suggests that inhibitory speed significantly increases between the ages of nine years and twelve years (Bjorklund and Harnishfeger, 1990; Romine & Reynolds, 2005). Therefore, it is possible that those Year 7s who were able to inhibit more accurately may also inhibit information more efficiently. Subsequently, these Year 7s would be able to more quickly inhibit the real-world interference caused by belief biases, thus aiding counterfactual-world elaborative inference generation.

In Year 9, inhibitory control remained the strongest unique predictor of counterfactual-world elaborative inference skill, accounting for 29.4% independent variance. Strength of belief biases also continued to make an independent contribution, accounting for 17.2% unique variance. However, the direction of relationship between inhibitory control and counterfactual-world elaborative inferential skill was negative, such that those with good inhibitory control made more counterfactual-world elaborative errors. There are two possible interpretations. First, elaborative inferences, are typically found to be generated offline, such that the knowledge needed to generate the target inference is typically not readily accessible – i.e. minimal activation levels (Graesser et al., 1994; Kintsch, 1988). Research suggests that irrelevant and/or contradictory information is suppressed in the initial stages of inferential processing to reduce processing costs (e.g. Cain, 2006). Given its counterfactual-world nature and low activation levels, the knowledge necessary for the generation of counterfactual-world elaborative inferences is thus likely to be suppressed in the early stages of inferential processing. Strong inhibitory control would thus hinder counterfactual-world elaborative inference generation. Conversely, those with weak inhibitory control may maintain the counterfactual-world information, thus aiding later

inference generation. However, it is not easy to explain why this would occur in Year 9, but not Year 5 and 7.

Second, given that Chapter 7 suggests counterfactual-world elaborative inferences were generated offline, these inferences may have been processed with strategic effort. Year 9 is a critical period in the development of reasoning and meta-cognitive skills, such that adolescents begin to question and evaluate the possibilities of information presented to them (Swanson, 1990). Developing meta-cognition may hinder the conscious search of long-term memory, such that Year 9 may consciously evaluate the plausibility of the information, and due to belief biases, those with stronger inhibitory control may inhibit information that is inconsistent with real-world beliefs. This was observed by the researcher during the task, such that Year 9s made comments such as "but that can't be right because…" This is consistent with the qualitative shift in the application of belief biases discussed above. Given that the negative relationship between inhibitory control and counterfactual-world elaborative inferential skill is not observed until Year 9, the latter possibility is most likely. However, both suggestions are extremely tentative.

By adulthood, inhibitory control is still a negative predictor of counterfactual-world elaborative inferential skill. However, it does not make a significant independent contribution, despite inhibitory control improving from Year 9 to adulthood. For adults, strength of belief biases was the only predictor to make an independent contribution to counterfactual-world elaborative inferential processing, accounting for 17.2% unique variance. The declining role of inhibitory control may be due to adults becoming more flexible, such that further developments in abstract thinking allow adults to more readily accept counterfactual-world information (Flavell, 1971). However, the continued independent contribution of strength of belief biases, suggests that counterfactual-world processing continues to be governed by a preference for real-world information. In sum, belief biases appear central to counterfactual-world elaborative inferential processing from Year 5 to adulthood. Inhibitory control appears to be important also, particularly during the Secondary School years. The function of inhibitory control during this time appears to change, however. Whilst in Year 7 inhibitory control inhibits real-world interference, in Year 9 inhibitory control may inhibit the activation of knowledge deemed implausible during a conscious search of long-term memory. Further research is needed, given that, due to time constraints, it was only possible to use one measure to explore each component of inference generation. For inhibitory control, in particular, it would be interesting to explore the relationship between inhibitory control and inference generation

using several measures sensitive to both skill and speed and conscious and unconscious processing.

8.4.5. Conclusions

This study highlights the unique developmental trajectories of four inference types during adolescence: real-world coherence, real-world elaborative, counterfactual-world coherence, and counterfactual-world elaborative. Specifically, knowledge appears central to real-world coherence inferences with a shift from mere possession of knowledge to accessibility of knowledge as the necessity to generate these inferences automatically whilst reading increases. Real-world elaborative inference generation, on the other hand, appears to be underpinned by developments in knowledge accessibility and inhibitory control, such that as more knowledge is readily accessible stronger inhibitory control is executed. However, in some instances, this control may be too strong. Belief biases appear central to both coherence and elaborative counterfactual-world inference generation across adolescence. However, there may be a conceptual shift in the application of belief biases between Year 7 and Year 9. This study also highlights Year 9 as a special time in inference generation development, with the possibility of qualitative shifts, largely driven by inhibitory control, occurring during this time.

CHAPTER 9 GENERAL DISCUSSION

This thesis sought to explore age-related changes in real-world and counterfactual-world inference generation abilities during adolescence. To assess this wide age-range, whilst also effectively and efficiently assessing both inferential skill and time-course along with possible real-world interference when processing counterfactual-world information, a new measure of inference generation was created, the Image Selection Task (IST). The development and evaluation of the IST is detailed in Chapters 3 and 4. To promote a natural reading process, the IST embeds critical sentences into short fantasy stories. Participants move through the text at their own speed, with the time taken to read critical texts recorded thus providing a measure of the inferential time-course. Each critical text is followed by a forced-choice picture-selection task which allows for a measure of skill (number of errors) to be obtained. The use of diagnostic fillers also allows for conclusions to be drawn regarding the cause of inference generation failure. The IST was found to be a valid measure of inference generation abilities. Due to the use of a pictorial response method, which minimises additional translational demands, the IST also appears to provide a purer measure of CYP's inference generation abilities than existing measures utilising verbal response methods.

The IST was used with Year 5, Year 7, Year 9 and adults in three studies to investigate:

1. The role of real-world and counterfactual-world inference generation abilities in reading comprehension and text production during adolescence (Chapter 6)
2. Age-related changes in real-world and counterfactual-world inference generation abilities during adolescence (Chapter 7).
3. The development and potentially changing role of those skills and processes underpinning real-world and counterfactual-world inference generation abilities during adolescence (Chapter 8).

This chapter begins by summarising the key findings of this thesis and then relating these to previous research. The educational implications of this research are considered. This is followed by a discussion of research limitations. Finally, directions for future research and conclusions are provided.

9.1. Summary of Findings

9.1.1. How does the role of inference generation in higher-order literacy skills change during adolescence?

One aim of this thesis was to explore the role of inference generation skill in reading comprehension and text production processes during adolescence. Inference generation was implicated in both reading comprehension and text production during adolescence. Therefore, this research is unique in nature such that it explicitly establishes a relationship between inference generation and text production quality during adolescence. Second, whilst not significant, visual inspection of regression coefficients suggests that, during adolescence, the predictive utility of overall inference generation skill declines for reading comprehension but increases for text production. This research thus highlights the unique developmental trajectories of these higher-order literacy skills during adolescence. Third, whilst not significant, visual inspection of regression coefficients for each inference type suggests that real-world coherence inference skill plays a growing role in text production quality during adolescence, such that predictive utility appears to increase. Conversely, the role of real-world coherence inference skill in reading comprehension appears to decline during adolescence. The opposite pattern is observed for real-world elaborative inference generation. The role of specific inference types in the reading comprehension and text production process is discussed in more detail below.

9.1.1.1. Inference Generation and Reading Comprehension

Reading comprehension skill was found to improve throughout adolescence, however, a period of limited development was observed between Year 7 and Year 9. A shift in importance of inference type to the reading comprehension process was also observed in Year 9. Specifically, real-world coherence inference generation skill was found to be a unique significant predictor of inference generation in all age groups. This is consistent with skilled models of reading comprehension and inference generation that highlight the importance of maintaining coherence whilst reading (e.g. Kintsch, 1988; 1998). However, elaborative inference generation skill did not make a significant contribution to reading comprehension skill until Year 9. This suggests a possible shift in the way coherence inferences and elaborative inferences are processed when reading at 13-14 years (Year 9). Coherence inferences are essential for basic understanding of a text. However, it appears that as the CYP progresses a more enriched mental model is needed for the deep level of understanding expected. This is consistent with the researcher's classroom experience, such that texts not only become more challenging with regards to syntactic complexity

and vocabulary, but also the depth of understanding expected. Year 9 was also the point where inferential processing speed (measured by reading speed) achieved adult-like levels. Taken together, results could suggest that as inferential processing becomes more efficient the reader is able to elaborate on the text in more detail.

Reading comprehension progress was not observed between Year 7 and Year 9, however, potentially suggesting that the increasing importance of elaborative inference skill may be difficult to for Year 9s to manage. However, by adulthood, elaborative inference generation was the strongest predictor of reading comprehension skill and reading comprehension skill had significantly improved between Year 9 and adults. This suggests adults are better able to deal with the demands associated with elaborative inference generation as part of the reading process. The findings of Chapter 8, suggest that this may be due to increased knowledge accessibility and, as discussed below, the possibility that some adults are generating elaborative inferences whilst reading. The research presented here thus adds to the current literature by showing that developments in reading comprehension during adolescence may be fuelled, at least in part, by the more efficient generation of elaborative inferences and specifically the use of more accessible knowledge, with this shift occurring around Year 9. These findings also demonstrate the distinct role of coherence and elaborative inference generation skills in reading comprehension, thus highlighting the need for further research to distinguish between the two inference types.

9.1.1.2. Inference Generation and Text Production

Text production quality was found to improve throughout adolescence, however, a period of limited development was observed between Year 7 and Year 9. The results of Chapter 6 also suggested that there is a shift in importance of inference type during Year 9, such that coherence inference generation becomes more important than elaborative inference generation. Specifically, the role of real-world coherence inference generation skill in text production quality was found to increase with age, with the unique contribution of real-world coherence skill approaching significance in Year 9 and reaching significance in adulthood. Elaborative inference generation skill made a significant contribution in Years 7, 9 and adults, but, predictive utility was found to decline with age (although this decline was not significant). This pattern of results is consistent with the knowledge-telling model (Bereiter and Scardamalia, 1987; Hayes 2012) such that whilst the first stage (knowledge-telling) is characterised by the generation of ideas with developments due to elaboration,

the second and third stages are characterised by improvements in coherence and the organisation of text. As discussed below, developments in coherence inference generation were characterised by a shift in the function of knowledge, such that accessibility of knowledge became more important than amount of knowledge. It is possible then, that developing writing is constrained at first by amount of knowledge (i.e. the CYP does not have the knowledge necessary to make a link between one or more ideas) and then accessibility of knowledge (i.e. the CYP only makes links between ideas when the knowledge needed to do so is readily available).

In Year 5 and Year 7 the control variables (working memory capacity, IQ, decoding) accounted for substantial amounts of variance in text production quality (45.9% and 29.1% respectively). However, as the predictive utility of inferential skill increased, the predictive utility of the control variables declined. This pattern is consistent with the Simple View of Writing (SVW; Juel, 1988). According to the SVW, text production quality develops as low-level transcription skills become automatic and working memory capacity increases, since this allows for the execution of higher-level ideation skills – i.e. mental model construction. The research presented here thus supports the SVW and further extends this by explicitly highlighting the growing importance of coherence inference generation skill as low-level transcription skills become automatic. To sum, the application of coherence inference generation skills to the text production process appears to be mediated, at first by transcription skills, then amount of knowledge and then the accessibility of this knowledge.

9.1.2. How do inference generation abilities change during adolescence?

A further aim of this thesis was to explore changes in inferential skill (number of errors) and the inferential time-course (speed) during adolescence. Inference generation development appears to be largely quantitative during adolescence, with proficiency achieved between Year 7 and Year 9 for both real-world and counterfactual-world information. However, there also appears to be a period of qualitatively different processing during Year 9, such that a qualitatively different error pattern was observed for this group compared to all other age groups.

Inference generation skill appears to develop between Year 5 and Year 7, at which point adult-like proficiency is achieved for both real-world and counterfactual-world inference generation. The lack of development in real-world inference generation skill between

Year 7 and adulthood is surprising and whilst consistent with some research (e.g. Casteel, 1993) conflicts with others (e.g. Barnes et al., 1996). The results of this thesis may be explained by use of the IST compared to a text-based assessment tool. A pictorial response method was chosen to minimise those verbal demands associated with the translational process. It was hoped that this would provide a purer, less confounded measure of CYP's inference generation abilities. It is possible then, that CYP have the same inference generation skills as adults by 11-12 years, but these skills are constrained by the verbal skills needed to translate text or construct an answer. This interpretation is supported by the findings of Chapter 4, such that despite the use of the same methodology, CYP performed worse on the sentence verification task (which used a verbal response method, thus requiring decoding and translational processes) compared to the IST (which used a pictorial response method, thus minimising decoding and translational processes) whereas there was no difference for adults. This has implications for the classroom, such that the implementation of the most effective intervention is dependent on successful identification of the CYP's difficulty(s). Therefore, the use of measures in which inference generation can be isolated from translation and verbal output processes may be needed. This finding also provides an avenue for further research, such that it could be fruitful to explore the translational and verbal skills that are constraining the successful application of the inference generation process between Year 7 and adulthood.

Previous research suggests counterfactual-world inference generation becomes more difficult with age, with a conceptual shift in the processing of counterfactual-world inferences occurring between 9-10 years and adulthood (Bowyer-Crane & Snowling, 2010; Ceci et al., 1981; Dorfman, 1989; Dorfman & Brewer, 1988; Graesser et al., 1998). The findings of this thesis thus conflict with previous research. This discrepancy may be due to the tasks used. For instance, in the Ceci et al. (1981) study older participants only displayed counterfactual-world errors in their recall after a one-week delay. Research suggests that as a CYP ages they are more likely to store information by fitting it into pre-existing knowledge structures (Barnes et al., 1996; Chi, 1978). Typically, these are real-world knowledge structures. In the current thesis, participants' inference generation skill was assessed after an almost immediate delay (500ms). Therefore, storage and recall was not necessary. Real-world interference may thus be higher when storing and recalling information than when reading, particularly for older CYP.

Qualitative developments during adolescence were not observed for the inferential time-course. All age groups were found to generate coherence inferences online and

elaborative inferences offline, regardless of the plausibility of information. This was evidenced by significantly longer reading speeds for real-world coherence inference-evoking critical texts compared to real-world literal critical texts. There was no significant difference between reading speeds for real-world elaborative inference-evoking critical texts and real-world literal critical texts. This finding is consistent with previous research (Calvo et al., 2006; Casteel, 1993; Ford & Milosky, 2008; Graesser et al., 1994; Prat et al., 2011; Virtue et al., 2006; 2008). These findings also serve to extend the Constructionist Theory by suggesting coherence inferences will be generated online, by all age groups, even when reading fantasy texts and generating counterfactual-world inferences. This finding conflicts with previous research which suggests a qualitative shift in the processing of counterfactual-world information between 9-10 years and adulthood, though (Bowyer-Crane & Snowling, 2010; Graesser et al., 1998). The notion of a conceptual shift was not based on direct comparison of CYP and adults, however, but discrepancies in findings between studies exploring the counterfactual-world inferential processing of adults and CYP. Moreover, previous research exploring counterfactual-world text processing has utilised offline measures of inference generation, explored only one age group, and/or not been explicit about inference type (e.g. Bowyer-Crane & Snowling, 2010; Ceci et al., 1981; Dorfman, 1989; Dorfman & Brewer, 1988; Graesser et al., 1998). The differences between adults and CYP observed in the previous literature thus appear to be due to methodological differences, such that this thesis has shown when the same methodology is used to assess CYP and adults, all groups display qualitatively similar patterns.

In line with previous research, for all age groups, inference generation was found to be more difficult than literal processing, potentially resulting from a different process (e.g. Chow et al., 2008; Jin et al., 2009; Kim et al., 2012; Prat et al., 2011; Virtue et al., 2006; 2008). Qualitative changes in inference generation skill during adolescence were observed, such that a unique error pattern was observed in Year 9. Specifically, Years 5 and 7 and adults found the generation of elaborative inferences no more difficult than the generation of coherence inferences. Previous research has found that the generation of elaborative inferences is more difficult than the generation of coherence inferences (e.g. Bowyer-Crane & Snowling, 2010; Cain & Oakhill, 1999). There are several possible explanations for this discrepancy in findings. First, elaborative inferences are typically found to be generated offline (e.g. Calvo et al., 2006). However, many of the aforementioned studies have used online measures of inference generation, thus only

those inferences generated whilst reading are measured. Second, inferences can be either text-based or knowledge-based, with knowledge-based inferences more difficult (Kintsch, 1994). Many of the aforementioned studies have explored text-based coherence inferences but knowledge-based elaborative inferences. Consistent with the findings of this thesis, those exploring both knowledge-based coherence and knowledge-based elaborative inferences find no difference in skill levels (e.g. Cain et al., 2001). Third, the use of pictures may have encouraged the construction of a highly enriched multi-modal situational model. The static inferences explored in this study often reflected visual features – e.g. colour. The construction of highly enriched visual situational models may thus facilitate elaborative, specifically static, inference generation. If this is the case, the results of this thesis support those theories suggesting that mental models are multi-modal as opposed to merely propositional (e.g. Zwaan, 2003).

Year 9 found coherence inference generation more difficult than elaborative inference generation. Although not significant, visual analysis of means suggests that this effect is primarily due to Year 9 experiencing more difficulty with coherence inference generation compared to elaborative inference generation in the counterfactual-world condition. Positioning Year 9 as a unique time in development is supported by the additional analysis conducted in Chapter 7 such that when data were analysed according to reading comprehension age, and Year 9s were dispersed across groups, all groups made a similar amount of coherence and elaborative errors, suggesting the pattern observed is due to chronological age not reading age. Year 9 also displayed a different pattern to all other age groups when those skills underpinning counterfactual-world inference generation were explored (Chapter 8), such that inhibitory control, not strength of belief biases was key. Taken together, it is possible that, compared to the other age groups, for some reason Year 9 are facing more real-world interference when processing counterfactual-world coherence inferences compared to counterfactual-world elaborative inferences. As a result, suppression of this conflicting information becomes key to success. Several reasons for the possible qualitative shift in counterfactual-world processing during this time are discussed in Section 9.1.3.3.

For the most part patterns were the same for real-world and counterfactual-world processing, however, counterfactual-world information was processed significantly slower than real-world information by all groups. This finding was predicted, such that previous research suggests the processing of counterfactual-world information is more taxing than the processing of real-world information for all age groups (e.g. Bowyer-

Crane & Snowling, 2010; Ferguson, 2012; Graesser et al., 1998). Previous research suggests that the additional processing time is spent resolving conflict experienced due to real-world interference – the automatic activation of related but contradictory real-world knowledge. This is supported by the error analysis conducted in Chapter 7, such that for all age groups and both inference types, the primary cause of counterfactual-world inference generation failure was attributed to real-world interference. The additional time spent processing counterfactual-world information appears to be effective in most cases, such that counterfactual-world inference generation was not found to be any more difficult than real-world inference generation, except for counterfactual-world coherence inferences in Year 9, as discussed above. Both real-world and counterfactual-world inference generation development during adolescence, then, is largely quantitative in nature. However, some qualitatively different processing may be occurring during Year 9, such that counterfactual-world coherence processing during this time appears to cause some difficulty.

9.1.3. How do the cognitive underpinnings of successful inference generation change during adolescence?

The role of amount of knowledge, accessibility of knowledge, inhibitory control and strength of belief biases in inference generation skill during adolescence was also explored in this thesis. This research builds upon the findings of existing studies by showing that, during adolescence, each inference type appears to be underpinned by different skills at different times. Specifically, real-world coherence inference generation appears to be underpinned by knowledge throughout adolescence, with a shift in importance from amount of knowledge to accessibility of knowledge around Year 9. Knowledge and inhibitory control appear central to elaborative inference generation whereas strength of belief biases appear to underpin counterfactual-world inference generation. Each inference type is discussed in more detail below.

9.1.3.1. Real-World Coherence Inference Generation

The IST was carefully designed for this thesis such that fillers were designed to reflect common reasons for inference generation failure, allowing for a meaningful analysis of errors to be conducted. For all groups, real-world coherence errors were primarily attributed to literal processing (Chapter 7) – i.e. the participant not integrating the ideas

within the critical text using their background knowledge. This is supported by the results of Chapter 8, which implicate the integration of background knowledge in the successful generation of real-world coherence inferences, although there appears to be a shift in the role of knowledge processes. Whereas amount of knowledge was found to be the strongest unique predictor of real-world coherence inference generation skill in Years 5 and 7, accessibility of knowledge was found to be the strongest unique predictor of real-world coherence inference generation skill in Year 9 and adults. This shift may be due to *when* in the reading process coherence inferences are generated by different age groups.

The Simple View of Reading (SVR) suggests that the application of inferential processes (and other higher-order skills falling under the umbrella of linguistic comprehension) is constrained by decoding skills in children, such that it is not until decoding is automatic that inferential processes can operate whilst reading. The results of this research and previous studies, however, suggest that even young children are able to generate coherence inferences whilst reading, despite decoding skills not being automatic (Omanson et al., 1978; Paris et al., 1977). It is possible then that whilst not automatic, decoding skills are proficient enough by 9-10 years for the simultaneous execution of inferential skills whilst reading. Alternatively, online comprehension may be slowed by limited decoding skills, but not constrained completely. Instead, a two-stage more conscious reading comprehension process may take place where text is decoded and then higher-order processes are executed. Years 5 and 7 may engage in a conscious search for the necessary knowledge, once the text has been decoded, explaining why possession not accessibility of knowledge is key. This is also consistent with the researcher's data collection observations and classroom observations. Children will often appear to pause and consider what they have read once they have successfully decoded.

For Year 9, on the other hand, decoding skills may be automatic. This cannot be determined by the current study as the single-word-reading task employed assessed word-reading skill not speed. Proficient models of reading comprehension, that assume decoding proficiency, suggest that, due to associative mechanisms, knowledge is automatically activated whilst reading (Kintsch, 1988; van den Broek et al., 1995; 1999; 2005; Zwaan, 2003). Barnes et al. (1996) suggest that the amount of knowledge that is readily accessible via associative mechanisms improves with age as knowledge networks become better organised. This is supported by the findings of the current thesis such that accessibility of knowledge was found to increase with age. Working memory resources are thus no longer primarily attributed to conscious decoding and large amounts of

associated knowledge are also readily available. Inferential processes may be executed *whilst* reading, as opposed to the two-step process discussed above. A conscious search of long-term memory would disrupt the fluid reading process, explaining the increasing importance of accessibility of knowledge over amount of knowledge. The reading speed measure used in the current research does not distinguish between pauses and reading. Therefore, it remains unclear when Year 5 and Year 7 are engaging in inferential processing. An eye tracking study, which could map pauses, fixations and regressions to determine reading processes, would be needed to explore this further.

9.1.3.2. Real-World Elaborative Inference Generation

For Year 5, accessibility of knowledge was the strongest predictor of elaborative inference generation. The Constructionist Theory suggests that elaborative inferences are generated offline, when prompted, after a conscious search of long-term memory, as the knowledge needed for the generation of these inferences is neither readily accessible nor strongly activated whilst reading (Graesser et al., 1994). Knowledge is thought to be stored in networks (Kintsch, 1988; 1998). Developments in knowledge networks are thought to reflect not only the addition of new knowledge, but the reorganisation of knowledge, resulting in more efficient knowledge activation via associative and conscious search mechanisms (Barnes et al., 1996). Year 5 were slower to activate knowledge than all other groups, suggesting their knowledge networks may not be as well organised as the other groups. Therefore, even when given a prompt and time to search long-term memory, Year 5 may still experience difficulty efficiently activating the necessary knowledge. Year 5s may thus only generate elaborative inferences with information that is easily accessible. This is consistent with the results of Barnes et al. (1996) who found that knowledge that was readily accessible was twice as likely to be used when generating an inference than knowledge that was more difficult to access.

Typically, knowledge inhibited is that with low activation levels due to weak or minimal connections with other nodes in the currently activated mental model of the text (Kintsch, 1988). When reading, knowledge necessary for coherence inferences is likely to have strong and numerous connections with currently activated nodes, but this is not the case for elaborative inferences. Consequently, for coherence inference generation there is likely to be one strong overriding inference, with multiple connections. Conversely, for elaborative inferences, several competing possibilities, with similar activation levels may be accessed when the reader performs a conscious search of long-term memory.

Maintenance and integration of all possibilities would be cognitively demanding and possibly overwhelm the reader's limited working memory capacity and result in inference generation failure (Cain & Oakhill, 2006a; De Beni et al., 1998; Kershaw & Schatschneider, 2012). The reader must thus select one possibility, inhibiting all others. The number of competing possibilities will be determined by how readily accessible the reader's knowledge is. Knowledge accessibility improved from Year 5 to Year 7, Year 7 may thus access more representations than Year 5, explaining the central role of inhibitory control in the elaborative inferential process of Year 7s. This also explains the increase in inference generation skill and inference generation speed observed between Year 5 and Year 7 – i.e. developments in inference generation abilities during this time are driven by developments and interactions between accessibility of knowledge and inhibitory control.

In opposition to the positive relationship observed in Year 7, a negative relationship was observed between elaborative inference skill and inhibitory control in adults. This suggests that for adults those with stronger inhibitory control make more elaborative inference errors. There are several possible, highly speculative, reasons for this relationship. Although not detectable in the current findings, some adults may have been generating some elaborative inferences online[38]. A function of inhibitory control in online inference generation is to suppress the automatic activation of irrelevant and/or contradictory knowledge, thus creating a clear workspace (e.g. Cain, 2006; Harnishfeger, 1995). Elaborative inferences are argued to be generated offline as there are often multiple plausible possibilities (e.g. Albrecht et al., 1995). Whilst adults may be able to maintain some of these possibilities whilst reading, maintenance of all possibilities could still overload working memory and thus result in comprehension failure. Adults with strong inhibitory control may thus strongly inhibit all possibilities, except the one (or two) that is most strongly activated. However, if this does not match the target picture, the IST will suggest an error has been made. Those adults with weak inhibitory control, however, may also be those with a more limited working memory capacity (the rule-shift-cards task is not a pure measure of inhibition, such that success is dependent on other skills, including working memory capacity) and thus this group may generate elaborative inferences

---

[38] Data in the current study may have been skewed by those adults not generating elaborative inference online. This was the case in the Long et al. (1994) study. Their findings suggested that elaborative inferences were generated offline, until they split their sample by working memory capacity. The high working memory capacity group were found to generate elaborative errors whilst reading, whereas the low working memory capacity group were not.

offline, allowing for more conscious control and consideration of several possibilities. This interpretation fits with the developmental patterns of inference skill and speed observed between Year 7 and adulthood, such that whilst skill did not improve, speed did.

Alternatively, the negative relationship observed may be attributable to the inference generation measure used in the current research. Inhibition is thought to comprise a 'family' of skills (Friedman & Miyake, 2004). The three functions include: 1) prepotent response inhibition – the suppression of a dominant, often automatic response to the stimuli presented 2) response to distractor inhibition – suppression of irrelevant stimuli to focus on relevant stimuli, when both are present simultaneously 3) resistance to proactive interference – suppression of information that is no longer relevant. Resistance to proactive interference is the type of inhibitory control most often implicated in the reading comprehension process (e.g. Borella et al., 2014). However, due to the motor response method used, success on the IST may not be dependent on this type of inhibitory control. Similarly, it is possible that the rule-shift-cards task is not tapping the type on inhibitory control associated with reading comprehension. The rule-shift-cards task is dependent upon deliberate and conscious inhibitory processing (Wilson et al., 1996). The inhibitory control likely to operate during reading comprehension is largely unconscious (Cain, 2006; Harnishfeger, 1995; Kintsch, 1988). This interpretation is consistent with the lack of relationship observed between inhibitory control and coherence inference generation skill in the current research, such that coherence inference generation is thought to be automatic and largely unconscious. All of the possibilities discussed above are highly speculative, research exploring the role of inhibitory control using several measures would be needed to further clarify results.

For Year 9, amount of knowledge was the strongest predictor of elaborative inference generation skill. Year 9 were also found to display a different error pattern to all other groups, such that they made significantly fewer elaborative errors than coherence errors. The superior performance and strong role of amount of knowledge, followed by accessibility of knowledge, suggests that some Year 9s may be engaging in a laborious search process to find the necessary knowledge, even when this knowledge is not readily accessible. Year 9 may be the only group to engage in this additional processing due to their point in cognitive development. Whilst Year 7 are just entering Piaget's Formal Operational Stage, Year 9 are deep within it (Dulit, 1972). Those within this stage are more likely to engage in speculation and possibility (e.g. consideration of multiple alternatives and hypotheses) and to search for additional information that will confirm

their new way of thinking. Unlike Years 5 and 7, Year 9 may thus have considered several plausible real-world elaborative possibilities, judging these against the four pictures, before selecting the target. Whilst adults are also thought to be in the Formal Operational Stage, research shows that whilst conflict and questioning (e.g. challenging of ideas people and perspectives) is frequent during adolescence, by adulthood these behaviours begin to reduce (e.g. Steinberg & Morris, 2001). Unlike Year 9, adults, then, may select a possibility and stick with it, instead of considering all possibilities. This suggests that Year 9 may be a unique stage in development, such that qualitatively different processes are engaged.

9.1.3.3. Counterfactual-World Inference Generation

For Year 5, Year 7 and adults, for both coherence and elaborative inference skill, strength of belief biases was the strongest predictor, such that those more efficient at overcoming belief biases made fewer counterfactual-world errors. When text is read, proficient models of reading comprehension suggest that related knowledge from long-term memory is automatically activated via memory-based processes (e.g. Kintsch, 1988; 1998). For counterfactual-world critical texts, the real-world knowledge is consistent with beliefs but not the text. The reader must thus suppress their belief biases to generate the correct inference. The results of this thesis suggest that whilst belief biases may operate whilst reading the additional time spent resolving the conflict experienced is effective, such that no significant difference was observed between real-world and counterfactual-world skill. Additionally, all age groups were able to generate counterfactual-world inferences whilst reading when they were necessary for coherence. For all age groups, constructionist processes must thus operate, potentially reducing plausibility constraints, resulting in the inhibition of real-world knowledge and successful generation of the counterfactual-world inference.

Inhibitory control was also found to play a role in counterfactual-world inference generation. Inhibitory control was the second strongest predictor of counterfactual-world elaborative inference generation in Year 7, after belief biases, such that those Year 7s with stronger inhibitory control generated counterfactual-world elaborative inferences with more success. Given the discussion above, it is likely that inhibitory control plays a role in inhibiting the real-world interference caused by belief biases. Whilst inhibitory control was also a significant predictor in Year 9, the relationship was negative. A negative relationship was also observed between inhibitory control and real-world elaborative inference generation skill. As discussed above, the negative role of inhibitory

control in elaborative inference generation could be due to difficulty weighing up several options if all but one possibility is quickly suppressed. Alternatively, as discussed above, the pattern of results observed could be an artefact of the measures used.

It is also possible that the shift in the function of inhibitory control reflects a genuine shift in the role of inhibitory control in counterfactual-world processing. Previous research suggests a shift in counterfactual-world processing around the ages of 12-14 years, underpinned by developments in inhibitory control (Raefetseder et al., 2010; 2013). It is suggested that older CYP and adults apply a more complex counterfactual-world processing strategy, such that they consider two worlds. Younger CYP are thought to apply a simpler basic processing strategy, such that only one world is created and considered. When reading, the construction of only one model – a counterfactual-world model – would be beneficial such that real-world interference would be minimal. In conflict with previous research, the results of Chapter 7 suggest that inhibitory control does *not* improve from Year 7 to Year 9 (Bjorklund and Harnishfeger, 1990; Romine & Reynolds, 2005). However, previous research suggests that improvements in inhibitory control are largely due to increasing efficiency during adolescence. The inhibitory control measure used in the current research tapped accuracy not speed. It is possible then that those Year 9s with superior inhibitory accuracy in the current research are also those Year 9s who are able to more quickly inhibit information. If the shift in processing strategy is attributable to increasing inhibitory efficiency then Year 9 may be facing more interference than the other age groups, explaining the negative relationship observed. This interpretation is highly speculative. Further research would be needed to explore the role of inhibitory control, using several measures, in real-world and counterfactual-world inference generation skills during adolescence.

Finally, it is important to note that critical texts were presented within a fantasy world which provided a plausible context for the counterfactual-world critical texts. Previous research exploring counterfactual-world text processing within a counterfactual-world context finds accommodation after an initial delay which is associated with the setting up of a counterfactual-world mental model (e.g. Ferguson, 2012; Ferguson & Sanford, 2008, Hald et al., 2007). Real-world critical texts were also presented within a fantasy context, yet real-world information was processed faster than counterfactual-world information. Unlike previous research in which the counterfactual-world critical text is primed by the counterfactual-world context, the fantasy contexts used in the IST were designed to reflect the natural reading process such that the critical texts in the current thesis were not primed

by the story context. Consistent with Warren et al. (2008) the results of this thesis suggest that, for all age groups, likelihood predicts processing difficulty in counterfactual-world contexts, not plausibility per se. Real-world information would be more likely given that even counterfactual-world mental models are grounded by real-world knowledge. It is argued that readers create counterfactual-worlds by applying the nearest possible world constraint – i.e. only minimal changes are made to real-world beliefs (Rafetseder et al., 2010). Taken together, it is likely that readers experience less interference when processing counterfactual-world information in counterfactual-world contexts once the counterfactual-world context is familiar, constraints are set and counterfactual-world information is no longer unexpected. This would explain how readers are able to engage so fully with fantasy and fiction texts, becoming lost in the fantasy world. Further research would be needed to explore how these worlds are built over time.

9.2. Educational Applications

Through the researcher's work in schools she became aware of the vast use of strategies to promote comprehension in the classroom and small groups. However, many of these strategies appeared to be things teachers thought would work, rather than the result of evidenced-based practice or the design of an individualised intervention following a specific assessment. As discussed below, the findings of this thesis can be used to support the limited evidence-base exploring comprehension and text composition programmes and strategies for struggling adolescents and the design of assessments which could lead to individualised intervention for those most in need of support.

9.2.1. General Classroom and Small-Group Strategies

Gregg Brooks provides an overview of the literacy one-to-one, small group s and whole class strategies and programmes available to schools in the UK in his fifth edition of 'What works for children and young people with literacy difficulties?' (Brooks, 2016). However, whilst government statistics highlight the literacy needs of secondary school CYP (Department for Education, 2015a), compared to the number of interventions targeting word reading and spelling in primary and secondary school CYP, the number of interventions targeting comprehension and text composition are limited. Moreover, a review of the interventions reported by Brooks suggests that of those interventions claiming to support the development of comprehension, for many this is 'achieved' through a quiz at the end of each passage. However, in line with the systematic review conducted by Paul and Clarke (2016), those interventions found effective at improving

comprehension explicitly teach comprehension strategies. For example, Brooks highlights inference training as an intervention for secondary school CYP. Similarly, Brooks highlights 'Self Regulated Strategy Development' (SRSD) as an intervention to improve writing skill in CYP in Years 6 and 7. The programme promotes planning, monitoring and evaluation, with planning supported by real-world visits to different areas. Further support for this intervention comes from Torgerson, Torgerson, Ainsworth, Buckley, Heaps, Hewitt, and Mitchell (2014) who, using a randomised control trial, found SRSD had a positive effective on the writing skill of the intervention group, in comparison to the control group, after 20 weeks. Furthermore, Brooks highlights the efficacy of reciprocal reading, an intervention that can be used at the small group level to teach and promote skills in the following comprehension strategies: predicting, clarifying, questioning and summarising. Although Brooks presents this as an intervention for Years 5 to 6, through practical work in schools, the researcher is aware of this intervention being used more widely across the primary school years and in some secondary schools. It appears, then, that to promote comprehension abilities, one must directly target and teach comprehension strategies, with the teaching of inference generation one such skill to target. The results of the current thesis thus support this body of research and the classroom practice the researcher has observed. However, in line with previous research as discussed below, this thesis may be able to extend and elaborate on the specific teaching of inference generation to small groups and whole classes.

Given that coherence inferences were found to be generated online, regardless of plausibility, strategies should be used to promote the generation of these inferences whilst reading. The findings of the current thesis suggest that possession of and, particularly during the Secondary School years, ready access to the necessary knowledge-base is essential to the successful generation of a coherence inference. This is in line with research finding that training students to activate their background knowledge when reading can improve their comprehension (e.g. Berkley, Mastropieri, & Scruggs, 2011). Educational practitioners should thus encourage the activation of the knowledge necessary for these inferences before and whilst reading. This will be particularly important for those groups who may not possess the necessary knowledge themselves – e.g. EAL learners, pupils from deprived socio-economic areas who may have limited experiences. *Before reading activities* could include speaking and listening activities such as brain-storming ideas across the class, small group or paired discussion, and multimedia stimuli to prompt discussion and encourage learning of ideas and vocabulary. *During*

reading activities could include think-aloud modelling by an adult highlighting how background knowledge is activated and used to link two or more ideas (e.g. "this bit of the text reminds me of…"), questioning of texts whilst reading, or marked up texts with different symbols to encourage the activation of specific knowledge.

Elaborative inferences were found to be generated offline. Therefore, elaborative inference generation questions and activities should be reserved for after reading. The successful generation of these inferences is dependent on a solid mental model for use in later comprehension tasks. A solid mental model can be encouraged using the techniques and activities discussed above. Elaborative inference generation was found to be dependent on both accessibility of knowledge and inhibitory control during Secondary School. Therefore, tasks to promote activation and consideration of elaborative knowledge should be promoted. Activities could include drawing or writing about a setting or character (e.g. wanted posters), role-play and hot seating (e.g. based on what you know about X what would he say if…), comparative questions. This is supported by Tarchi (2015) who found that an intervention designed to promote knowledge activation during reading improved not only inference generation abilities but also reading comprehension skill in CYP aged 12-13 years. The findings of the current research suggest that the ability to consider multiple options becomes difficult for CYP in Secondary School, such that in Year 9 a negative relationship is observed between inhibitory control and elaborative inference generation skill. It is possible that adults inhibit knowledge too strongly, restricting their ability to weigh up and consider multiple possibilities. The use of debate activities may thus be beneficial in encouraging CYP to consider several options. Given that the role of elaborative inference generation in reading comprehension increased with age, the promotion of elaborative inferences is likely to be particularly key in the Secondary School years. This is supported by Vaughn, Klingner, Swanson, Boardman, Roberts, Mohammed, et al. (2011) who also suggests that for older adolescents support should focus on developing an enriched elaborated mental model. Many of the activities described above would fit within Reciprical Reading. For example, role play and hot-seating activities fit within predicting, questioning and clarifying strategies.

For counterfactual-world texts, belief biases were found to be key. It is suggested that the anomaly effects observed for counterfactual-world critical texts compared to real-world critical texts can be attributed to expectancy as both were presented in counterfactual-world contexts. Therefore, to minimise interference and associated cognitive demands it

may be beneficial to aid CYP in creating an initial counterfactual-world mental model using other sources in which constraints of impossibility are set. This is supported by the work of Filik & Colleagues who found that counterfactual-world models can be added to the reader's knowledge-base (Filik, 2008; Filik & Leuthold, 2008; 2013). Fantasy and fiction texts are commonly chosen for pleasure (Clark & Foster, 2005; Clark & Rumbold, 2006). Given the correlation between reading for pleasure and reading comprehension skill (Mahiri, 2003) and academic achievement (Moje, Overby, Tysvaer, & Morris, K. 2008) supporting engagement with fantasy and fiction texts may particularly aid those who have disengaged with reading. This may encourage these CYP to read for pleasure and ultimately improve their reading comprehension skill.

Developments in text production were underpinned by increasing coherence, with the role of coherence inferences increasing with age (although this increase was not significant). CYP's writing may thus be improved by providing scaffolds and prompts that help the CYP to maintain coherence – e.g. planning sheets, writing models, and vocabulary mats with connectives. This thus fits with the SRSD writing intervention discussed above. CYP may be encouraged to retell stories and sequence events throughout the school day to develop their oral organisation skills. Given that amount and accessibility of knowledge are the key predictors of coherence inference generation skill during adolescence, helping CYP to develop and discuss their content and ideas before writing may also be useful. This could take the form a pre-learning sessions, vocabulary mats, and pictorial or textual writing stimuli.

Finally, throughout the school day and across the curriculum in both Primary and Secondary Schools learning activities often require some form of reading and writing – e.g. worksheets, books, the internet, written questions, essays, posters. For those with literacy difficulties it is important that: a) reading comprehension and text production are promoted through explicit instruction and intervention when necessary; and b) opportunities for learning from alternate sources are provided. Tovani (2001) suggests that some CYP with reading comprehension difficulties may be 'fake reading' such that they gather information from the teacher and class discussion. Making use of these strengths is key to ensuring these CYP have opportunities to learn in a way that they are good at, maintaining their self-esteem and motivation to learn. This is supported by research finding that improving CYP's views of reading alongside a comprehension intervention leads to additional gains in comprehension skill, above any beyond those observed for a comprehension intervention alone (Berkley et al., 2011).

9.2.2. Identification of Literacy Difficulties and the Design of Specific Interventions

In a review of reading comprehension interventions for Secondary School CYP, Paul and Clarke (2016) found that those interventions targeting phonics through computer-based programmes were of limited effectiveness. However, interventions which focused specifically on comprehension strategies were found to be more effective. Designing interventions to target inference generation may thus result in effective interventions. The results of Chapter 6 suggested no progress in reading comprehension or text production between Year 7 and Year 9. Additionally, the Year 9 sample comprised higher percentage of poor comprehenders[39] (57%) than Year 5 (27%) and Year 7 (14%)[40]. These differences could be due to the cross-sectional design employed. It is possible that when in Year 7, the current Year 9 cohort had weaker skills than the current Year 7 cohort. However, as discussed in Chapter 5, the Year 7 and Year 9 cohorts were recruited from the same schools using the same selection criteria. It is possible then that findings of the current research reflect a genuine period of limited progress. This is supported by UK government statistics which suggest around one third of CYP do not make the literacy progress expected during the Secondary School years (Department for Education, 2015a). In the current thesis, word-reading skill was found to increase between Year 7 and Year 9, although due to the correction applied the difference between the two groups failed to reach significance. Reading comprehension text was determined by single-word-reading score. Therefore, typically, Year 9s read a higher-level comprehension passage than Year 7s. As discussed above, it is possible that whilst basic inference generation skills are achieved by Year 7, ability to apply these skills is mediated by complexity of text. Therefore, whilst Year 9 have basic inferential skills, they may be unable to apply these effectively when text demands increase. Alternatively, given that reading comprehension measures are largely dependent on literal understanding until Secondary School it is possible that the texts read by Year 7 tapped primarily literal understanding, whereas the texts read by Year 9 taxed more implicit understanding (Cain & Oakhill, 2006b). Therefore, some Year 9s may have a literacy difficulty, related to inference generation abilities, that has not been identified, thus limiting the progress they can make. Further, longitudinal, research is needed. What is clear, at present, however, is that more in-depth

---

[39] defined in this research as those with a reading comprehension age 12 months or less than their chronological age
[40] Due to YARC only providing reading ages up to 16 years the percentage of adults who may be classified as poor comprehenders could not be determined.

and accurate assessments of inference generation abilities are needed if specific intervention in Year 9 is to be put in place.

Based on the findings of the current research, to inform effective identification and intervention an inference generation measure must be able to:

- *Assess both coherence and elaborative inference types.* The two inference types were found to be distinct, underpinned by different processes and play different roles in reading comprehension and text production.

- *Assess both the inference generation time-course and skill.* Coherence inferences were found to be generated online. Research suggests that the ability to generate these inferences spontaneously whilst reading improves with age (e.g. Barnes, et al., 1996; Casteel, 1993; Oakhill et al., 2003). Therefore, some with limited comprehension skills may be capable of generating coherence inferences when given a prompt but not whilst reading. The intervention would need to target inferential efficiency.

- *Assess inference generation skills independent of expressive language skills.* The development of inference generation abilities observed in this study was consistent with Casteel and colleagues (Casteel, 1993; Casteel & Simpson, 1991), such that adult-like inferential skill was observed around 11-12 years and adult-like inferential efficiency around 13-14years. Barnes et al., (1996) found that inference generation abilities continued to develop until at least 16 years. Barnes et al. used a question-answer assessment method whereas Casteel and colleagues and the current study used measures of inference generation that were not dependent on expressive language skills. The discrepancy between the studies suggests that whilst CYP may be able to generate inferences at an adult-like level by 11-12 years, their ability to articulate the inference may continue to develop until adulthood. To determine if inference generation and mental model construction skills or articulation of ideas should be the target of the intervention the assessment tool used needs to explore these skills independently.

- *Provide diagnostic information.* The error analyses conducted in this thesis suggests that for all ages inference failure can result from several different faulty process. It is important to know not just which inference types the CYP struggles with, but also why. This will allow for specific and targeted intervention to be created.

A measure similar to the IST with passages varying in complexity may be beneficial in the classroom. In addition, educational practitioners may want to consider assessing knowledge and inhibitory control given that these cognitive components were found to be central to inference generation skill in the current study.

9.3. Limitations

9.3.1. Sample Age

Four age groups were recruited for this thesis: 9-10 year olds, 11-12 year olds, 13-14 year olds and adults. There is a large gap between 14 years and 18 years. Recruiting a group of 15-16 year olds was considered. However, in the researcher's experience schools can be reluctant to release Year 11 pupils due to coursework, revision and exam pressures. Research exploring adolescent development shows that many of the key changes take place between 11-14years, with development in many executive functions plateauing after 14 years (see Romine & Reynolds, 2005 for a meta-analysis). Therefore, it is possible that development between 14 years and 18 years would be minimal. However, further research would be needed to explore this.

9.3.2. Sampling Strategy

Primary and secondary school participants were recruited from several schools in the South Yorkshire and Derby area. Participants were recruited from several schools to obtain a representative sample. However, whilst recruiting participants from a range of schools may mean participants received a range of teaching styles and strategies, this method also has several limitations. First, there may have been selection bias on the parts of the schools, such that the responsibility of pupil selection was given to the schools. Therefore, despite being given a list of criteria asking for participants who ranged in reading abilities, schools may have selected pupils with good reading abilities to present themselves in a positive light. Alternatively, schools may have selected participants who presented challenges in the classroom to allow some respite for the child and/or teacher. To gain a more representative sample, it may thus have been more appropriate to work with just one or two schools, randomly selecting participants or working with a particular class. However, the researcher was aware of the pressures placed on schools and did not want to place additional strains on schools which could have possibly resulted in school drop-out. Second, the list of criteria schools were provided with asked for participants who did not have a known learning difficulty and/or did not have exceptional abilities.

The full range of abilities in the current UK schooling system is thus not fully represented by this research. Further research, then, should aim to seek a truly representative sample by recruiting participants from a small number of schools and then randomly selecting participants from this pool, regardless of learning needs.

9.3.3. Design

A cross-sectional design was used in this thesis. Differences observed between the age groups may be due to differences in group characteristics rather than developmental changes. Every effort was made to ensure similar group characteristics, such that all schools were given the same selection criteria. As highlighted above, the unique pattern observed in Year 9 may be due to sample characteristics rather than age-related changes. Due to time constraints, a longitudinal study was not possible, but should be considered in future research.

9.3.4. The Measures

Development between Year 7 and Year 9 was limited. This could reflect the measures used, such that they were not robust enough to detect the subtle changes during this time. Some of the issues related to robustness will be discussed below. However, it is important to note that development was limited on both researcher-designed and standardised tasks. The limited progress observed may thus reflect the sample rather than the measures used.

9.3.3.1. The IST

The IST explored the processing of coherence and elaborative inferences. However, for many items, generation of the inference was more complex than the label suggests, such that processing was dependent on more than just successful coherence or elaborative inference generation. For instance, many coherence inferences also required referential processing due to the need for pronoun resolution. For example, for a coherence inference to be generated when reading the critical text below the participant must first infer that 'it' refers to the penguin.

One penguin jumped into the cold ocean. A few seconds later it was at their feet.

However, research shows that these inferences are often mastered by young children (e.g. Ackerman, 1986). However, has highlighted in Chapter 4, for some children pronoun resolution was still an area of difficulty, such that they experienced difficulty when processing literal items that required pronoun resolution. It is possible then, that some of

the processing difficulty and errors made when processing coherence inferences can be attributed to difficulties with other types of inference. The need to generate two inferences to comprehend a text is not uncommon in real texts, however, and was often necessary to prevent the creation of contrived and stilted target texts and stories.

A successful elaborative inference on the IST could be defined as the enrichment of the text by adding knowledge about the character's or object's traits, knowledge or beliefs. Therefore, when reading the text "the boy was eating grapes," one reader could infer that the grapes are green the other could infer the grapes are black. In both cases, the readers have successfully generated an elaborative inference – i.e. they have added knowledge about the object's traits. However, if the target item is a picture of green grapes, only the first reader would be scored as successfully generating the inference. Therefore, whilst precautions were taken to avoid a correct inference being scored as incorrect (e.g., including grapes of three colours you would not typically see – light blue, orange, and pink), it is possible that some participants generated an elaborative inference whist reading, but not one that matched the target item, thus performing 'incorrectly' on the IST. This thus highlights the value of open-ended tasks when exploring elaborative inference generation.

Error rates in all groups were low. This may be due to the simplicity of the texts used – stories were designed to be readable by CYP aged 9 years or less. Inference generation skill may thus continue to develop after 11-12 years. Due to discrepancies in previous research it was important that the same methodology was used for all age groups in the current thesis to allow for direct comparison. Future research exploring inference generation skills using texts of varying complexity would greatly build upon the inference generation development literature, highlighting possible constraining factors.

Previous research shows that the ability to generate coherence inferences spontaneously whilst reading improves with age (e.g. Barnes et al., 1996; Casteel, 1993; Oakhill et al., 2003). Reading speeds in this thesis suggest that Year 5 were able to generate coherence inferences whilst reading. However, the researcher's observation of the Year 5 participants and Year 5s when reading naturally in the classroom suggests that Year 5 may not generate inferences whilst reading. Year 5s will often read a sentence(s) pause and then continue. It may be that the inference is generated during the pause not whilst reading. When reading naturally Year 5s may not generate all coherence inferences whilst reading, therefore, full understanding of text may not be gained. This would explain the

significantly lower reading comprehension skill observed in Year 5, compared to all other age groups. Research utilising eye-tracking methods would shed further light onto the online process of inference generation, such that fixations and pauses can be detected.

Errors were explored using error analysis. However, errors were excluded from reading speed data. Consequently, it is unclear if those coherence inferences generated incorrectly were still drawn whilst reading. Further research exploring the process underpinning inference generation failure is needed.

Care was taken to design pictures that reflected the idea intended by the researcher – see Chapter 3 – however, the validation of the images could have been improved. Each picture was piloted in isolation, however, in the image selection task four pictures were presented simultaneously. Therefore, the processing associated with viewing an image in isolation during the validation task may not have reflected the processing of viewing four images simultaneously during the image selection task. For instance, processing the four pictures simultaneously may have been more difficult than processing the pictures in isolation, such that interference or overload occurred. Alternatively, processing the four pictures simultaneously may have been easier than processing the pictures in isolation, such that comparison allowed for clearer distinction of differences between pictures. For example, in the piloting phase, participants were able to correctly label thermometers displaying extremes (e.g., boiling hot, freezing cold). However, there was more variation when labelling thermometers displaying temperatures in between these extremes, such that cooler temperatures were typically labelled as cold and warmer temperatures as hot. However, if all four pictures were presented simultaneously it may have been possible for the participants to more easily distinguish between hot and cold.

A mixed methods approach could also have been used, such that ratings could have been accompanied by an interview in which participants' explanations for their decisions were given. This would have allowed a further insight into which elements of the picture supported interpretation and alternatively, which elements of the picture were difficult to interpret or ambiguous. This would have guided picture development and may have led to clearer pictures being created that were representative of the participants' views not just the researcher's. This may have resulted in the need for fewer amendments to the pictures once in the image selection task. Similarly, some of the children viewed plausible pictures as implausible as they would be unlikely to witness the events depicted. This information was gathered through the children's vocalisations. Asking all participants to

talk through their plausibility ratings may thus have been useful in determining whether all participants viewed the images this way, and any other possible factors which influence interpretations of plausibility. This information could have then guided or added to future interpretations of performance patterns on the IST, providing insight into the processing that occurs when presented with images varying in plausibility.

This thesis was unique in nature such that it sought to explicitly explore the role of inference generation in text production quality. Inference generation was assessed during the reading comprehension process. Due to time constraints and an already large test battery, the inclusion of another measure of inference generation whilst writing was not feasible. This means, however, the role of inference generation in the writing process may differ to the one suggested. It is clear from the current research, that inference generation skills during reading are in some way borne out when writing. Further research would be needed to explore inference generation when writing. To this researcher's knowledge, there is not a measure of inference generation whilst writing, although the use of think-aloud methodologies may be useful a starting point.

9.3.3.2. Inhibitory Control

As discussed above, the varying role of inhibitory control in the real-world and counterfactual-world elaborative inference generation processes could be due to developmental shifts or the measure used (rule-shift-cards task) – i.e. not tapping the implicit inhibitory control associated with language comprehension (e.g. Kintsch, 1988). Similarly, research shows that improvements in inhibitory control are primarily attributed to improvements in efficiency (Romine & Reynolds, 2005), however, developments in inhibitory efficiency were not tapped by the current measure. Future research utilising several measures of inhibitory control is needed.

9.3.3.3. Knowledge

Amount of knowledge possessed by all groups was high. This was unexpected given that the knowledge-base used to create the critical texts was taken from the Key Stage 3 science curriculum. However, to ensure all critical texts were readable by CYP aged 9 years or less, the concepts were relatively simple. For instance, for reproduction the Key Stage 3 curriculum explores topics such as mitosis whereas the critical texts were based on simple concepts of life cycles – e.g. eggs hatching and seeds growing. Further research using texts based on more complex knowledge-bases is needed.

9.4. Future Research

This thesis has explored typical development, exploration of this area was necessary given the few studies exploring inference generation during adolescence and inconsistencies between these studies. Future research is now needed to explore poor comprehenders to determine how the developmental trajectory of poor comprehenders differs to that mapped out in this thesis. Specifically, research exploring poor comprehender profiles around Year 7 and Year 9 could prove fruitful in supporting those CYP who do not make the expected literacy progress during Secondary School. The use of a longitudinal methodology to explore the profiles of those CYP who appear to be developing typically during Primary School but then do not make expected progress during Secondary School could be fruitful and provide a useful insight into the cause of the limited progress made by some CYP during the Secondary School years – i.e. do these CYP have weak inference generation abilities that remain unidentified until Secondary School or do these CYP have difficulty applying their inference generation skills as text demands increase?

Previous research exploring counterfactual-world inference generation was inconsistent with regards to methodology, stimuli and findings. This thesis took steps in answering some of the questions surrounding counterfactual-world inference generation, particularly during adolescence. For instance, counterfactual-world processing was found to be more taxing than real-world processing due to the need to resolve the conflict associated with the real-world interference caused by belief biases. All groups appear to be able to generate counterfactual-world coherence inferences whilst reading, around Year 9. However, there is a possible qualitative shift in the process underpinning counterfactual-world inference generation, with inhibitory control potentially playing a key role in this. Counterfactual-world errors were primarily attributable to real-world interference. However, what is not clear from the current research is why participants were successful in some instances but not others. Exploration of why some participants were unable to successfully inhibit the real-world interference for some items but not others is needed. It is possible that success is due to strength of belief biases associated with particular concepts. To this researcher's knowledge this is the first piece of research to explicitly implicate belief biases in the counterfactual-world inferential process. Further research exploring belief biases in inference generation, its specific function and how this process is both similar and dissimilar to the role of belief biases in counterfactual-world reasoning is thus warranted.

Whilst in the current study all age groups were found to generate counterfactual-world inferences online, Graesser et al. (1998) suggested adults generate counterfactual-world inferences offline. However, there are a number of methodological differences, between their research and the current thesis. It may be that when both adults and children are presented with more complex and implausible texts counterfactual-world inferences are generated offline. Further research, using texts varying in complexity and plausibility would be needed to explore this.

Finally, since beginning this thesis there have been changes in the National Curriculum and assessment tools and procedures (Department for Education, 2014). Exploration of how these changes may impact upon teaching and necessity of inference generation at different levels is warranted. Specifically, the Primary School reading curriculum and assessments now place more emphasis on inferential understanding. Given that almost 50% of CYP did not achieve age-related expectations in the 2016 reading SATs, a drop of almost 30% from 2015, it appears that those CYP previously achieving age-related expectations in Primary School but then struggling to make expected progress in Secondary School had inferential difficulties throughout their school careers, but assessment tools did not assess this (Department of Education, 2016). The high percentage of CYP failing to achieve age-related expectations in their reading SATs thus highlights the need for much more research into the role of inference generation abilities in reading comprehension. Moreover, a 'world without levels' has freed schools up to use more informative and in-depth measures of reading comprehension (Department for Education, 2015c). Further research exploring the design of reading comprehension measures, consistent with the criteria laid out in the new National Curriculum and the findings of previous research is needed.

9.5. Conclusions

Government statistics suggest that around one third of CYP in the UK are not making the expected literacy progress during Secondary School, despite less than one fifth of CYP entering Secondary School with the reading and writing skills expected (Department for Education, 2015). Given the central role of inference generation to skilled models of reading comprehension, this thesis sought to explore the development of inference generation abilities during adolescence, with the goal of informing effective educational practice. Findings show that there is both quantitative and qualitative change in inference generation skill (measured by number of errors) and time-course (measured by reading speed). All age groups appear to generate coherence inferences online and elaborative

inferences offline, regardless of the plausibility of the information, although skill in doing so improves until Year 7 and speed in doing so improves until Year 9. Several qualitative differences were also observed in Year 9, such that Year 9 found the generation of coherence inferences more difficult than the generation of elaborative inferences, in contrast to all other age groups, inhibitory control, not belief biases, was also the strongest predictor of counterfactual-world coherence inference generation skill, and Year 9 displayed a negative relationship between inhibitory control and elaborative inference generation skill. In addition, shifts in the importance of inference types to the reading comprehension and text production processes also emerged around Year 9. Year 9, also comprised higher percentage of weak comprehenders than all other groups. Year 9, then, appears to be a distinct time in inference generation development. With regards to supporting literacy development in the classroom, this thesis has highlighted the role of inference generation abilities in both reading comprehension and text production during adolescence. The need to isolate basic inference generation skills from translational and expressive language skills to identify the specific area of need(s) is also promoted by the findings of this thesis. Similarly, the findings of this thesis highlight the distinctiveness of coherence and elaborative inference types, calling for assessments and interventions to recognise this difference, particularly in the Secondary School years when elaborative inference generation may be more important to the reading comprehension process than coherence inference generation. Some of those factors (knowledge and inhibitory control) that may constrain the successful application of basic inference generation abilities (achieved around Year 7) during the Secondary School years have also been identified by this thesis, with this thesis providing recommendations to aid teachers in supporting CYP overcoming these barriers. Year 9 has been highlighted as a period of special development that requires much further research, the role of inhibitory control during this time should be a specific focus. This thesis has also extended the Constructionist Theory by showing that, for all age groups, regardless of plausibility and information those inferences necessary for coherence are generated whilst reading. However, due to belief biases, real-world interference is still likely, meaning if teachers want to engage learners with fantasy and fiction texts they may first need to support the CYP in creating a counterfactual-world model.

REFERENCES

Ackerman, B. P. (1986). Referential and Causal Coherence in the Story Comprehension of Children and Adults. *Journal of Experimental Child Psychology, 41,* 336-366.

Ackerman, B. P. (1988). Reason Inferences in the Story Comprehension of Children and Adults. *Child Development, 59(5),* 1426-1442.

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities, 43,* 332 – 345.

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19,* 933–958.

Albrecht, J. E., O'Brien, E. J., Mason, R. A., & Myers, J. L. (1995). The Role of Perspective in the Accessibility of Goals During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(2),* 364-372.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning. *Journal of Experimental Psychology: General,136(4),*569-576.

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior Knowledge Activation and the Comprehension of Compatible and Incompatible Text. *Reading Research Quarterly, 20(4),* 420-436.

AQA (2013, June 30th). English Language June 2013 Marking Scheme. Retrieved from http://www.aqa.org.uk/subjects/english/gcse/english-language-4705/past-papers-and-mark-schemes.

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading*, *18*(5), 325-346.

Babayigit, S., & Stainthorp, R. (2010). Component processes of early reading, spelling, and narrative writing skills in Turkish: A longitudinal study. *Reading and Writing: An Interdisciplinary Journal, 23,* 539-568.

Babayigit, S., & Stainthorp, R. (2011). Modelling the Relationships Between Cognitive-Linguistic Skills and Literacy Skills: New Insights From a Transparent Orthography. *Journal of Educational Psychology, 103*(1), 169-189.

Baddeley, A. D. (1986). *Working memory (Oxford Psychology Series, Vol. II).* Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation (Vol. 8, pp. 47-90).* New York: Academic Press.

Baddeley, A. D., Logie, R, Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory & Language, 24*, 119-131.

Bain, R. B. (2006). Rounding Up Unusual Suspects: Facing the Authority Hidden in the History Classroom. *Teachers College Record, 108*(10), 2080-2114.

Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. In Graesser, A. C., Kassler, M. A., Kreuz, R. J., & McLain-Allen, B. (1998). Verification of Statements about Story Words That Deviate from Normal Conceptions of Time: What Is True about Einstein's Dreams? *Cognitive Psychology, 35,* 246-301.

Baker, S., Gersten, R., & Graham, S. (2003). Teaching expressive writing to students with learning disabilities: Research-based applications and examples. *Journal of Learning Disabilities, 36,* 109–123.

Baker, S., Gertsen, R., & Scanlon, D. (2002). Procedural facilitators and cognitive strategies: Tools for unraveling the mysteries of comprehension and the writing process, and for providing meaningful access to the general curriculum. *Learning Disabilities Research and Practice, 17,* 65–77.

Barch, D.M., Braver, T.S., Sabb, F.W., & Noll, D.C. (2000). Anterior cingulate and the monitoring of response conflict: evidence from an fMRI study of overt verb generation. *Journal of Cognitive Neuroscience, 12,* 298–309.

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology, 61,* 216–241.

Barnes, M. A., Faulkner, H., Wilkinson, M., & Dennis, M. (2004). Meaning construction and integration in children with hydrocephalus. *Brain and Language, 89,* 47–56.

Barth, A. E., Barnes, M., Francis, D., Vaughn, S., & York, M. (2015). Inferential processing among adequate and struggling adolescent comprehenders and relations to reading comprehension. *Reading and writing*,*28*(5), 587-609.

Bartlett, F. (1932). *Remembering.* In Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive psychology, 9*(1), 77-110.

BBC - Bitesize. (2012 October 30th). *BBC - Bitesize.* Retrieved from http://www.bbc.co.uk/bitesize/

Beeman, M., Bowden, E.M., & Gernsbacher, M.A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. *Brain Language, 71,* 310–336.

Bereiter, C., & Scardamalia, M. (1987). *The Psychology of Written Composition.* Hillsdale, NJ: Erlbaum.

Berkeley, S., Mastropieri, M. A., & Scruggs, T. E. (2011). Reading comprehension strategy instruction and attribution retraining for secondary students with learning and other mild disabilities. *Journal of Learning Disabilities, 44(1),* 18–32.

Berman, R. A., & Slobin, D. I. (1994). Narrative structure. *Relating events in narrative: A crosslinguistic developmental study*, 39-84.

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in Grades 1, 3, 5, and 7. *Journal of Educational Psychology, 102,* 635–651.

Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10,* 48–71.

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 & Cognition, 18,* 65-71.

Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy?: The effect of task demands and motivational factors on scanning strategy in visual search. *Journal of Vision, 9*(3), 7-7.

Borella, E., & de Ribaupierre, A. (2014). The role of working memory, inhibition, and processing speed in text comprehension in children. *Learning and Individual Differences, 34,* 86-92.

Borella, E., & de Ribaupierre, A. (2014). The role of working memory, inhibition, and processing speed in text comprehension in children. *Learning and Individual Differences*, *34*, 86-92.

Bowyer-Crane, C., & Snowling, M. J. (2010). Turning frogs into princes: can children make inferences from fairy tales? *Reading & Writing, 23,* 19-29.

Brooks, G. (2016). *What works for children and young people with literacy difficulties?* (Fifth Edition). Dyslexia-SpLD Trust

Brookshire, J., Scharff, L. F., & Moses, L. E. (2002). The influence of illustrations on children's book preferences and comprehension. *Reading Psychology, 23,* 323–339.

Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2011). The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual, English-speaking peers. *British Journal of Educational Psychology, 81,* 344-354.

Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2013). The role of background knowledge in text comprehension for children learning English as an additional language. *Journal of Research in Reading, 36*(2), 132-148.

Burkart, J. M., & Rueth, K. (2013). Preschool Children Fail Primate Prosocial Game Because of Attentional Task Demands. *PLoS ONE, 8*(7), e68440. doi:10.1371/journal.pone.0068440

Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Science, 6*(10), 426-431.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality.* MIT Press.

Byrne, R. M. J. (2007). Precis of the rational imagination: How people create alternatives to reality. *Behavioural and Brain Sciences, 30,* 439-480.

Bui, Y. N., & Fagan, Y. M. (2013). The effects of an integrated reading comprehension strategy: A culturally responsive teaching approach for fifth-grade students' reading comprehension. *Preventing School Failure: Alternative Education for Children and Youth*, *57*(2), 59-69.

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology, 21,* 335–351.

Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory*, *14*(5), 553-569.

Cain, K. (2007). Deriving word meanings from context: Does explanation facilitate contextual analysis? *Journal of Research in Reading, 30,* 347–359.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal, 11,* 489–503.

Cain, K., & Oakhill, J. (2006a). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76,* 683-696.

Cain, K., & Oakhill, J. (2006b). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76,* 697–708.

Cain, K., & Oakhill, J. (2011). Matthew Effects in Young Readers: Reading Comprehension and Reading Experience Aid Vocabulary Development. *Journal of Learning Disabilities, 44*(5), 431-443.

Cain, K., & Oakhill, J. (2012). Reading comprehension development from seven to fourteen years: Implications for assessment. *Measuring up: Advances in how we assess reading ability*, 59-76.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and the relation to knowledge. *Memory & Cognition, 29,* 850–859.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42.

Cain, K., Oakhill, J., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child and Language, 30,* 681–694.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96,* 671–681.

Calvo, M. G. (2000). The time course of predictive inferences depends on contextual constraints. *Language and Cognitive Processes, 15,* 293–319.

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. (1998). Predictive inferences take time to develop. *Psychological Research, 61,* 249-260.

Calvo, M. G., & Castillo, M. D. (2001a). Selective interpretation in anxiety: Uncertainty for threatening events. *Cognition and Emotion, 15,* 299–320.

Calvo, M. G. & Castillo, M. D. (2001b). Bias in Predictive Inferences During Reading. *Discourse Processes, 32*(1), 43-71.

Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-line predictive inferences in reading: Processing time during vs. after the priming context. *Memory and Cognition, 27,* 834–843.

Calvo, M. G., Castillo, M. D., & Schmalhofer, F. (2006). Strategic influence on the time course of predictive inferences in reading. *Memory & Cognition,34*(1), 68-77.

Campion, N. (2004). Predictive inferences are represented as hypothetical facts. *Journal of Memory and Language*, *50*(2), 149-164.

Campion, N., & Rossi, J. P. (2001). Associative and causal constraints in the process of generating predictive inferences. *Discourse Processes*, *31*(3), 263-291.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical education*, *42*(12), 1150-1152.

Carlson, S. E., Seipel, B., & McMaster, K. (2014). Development of a new reading comprehension assessment: Identifying comprehension differences among readers. *Learning and Individual Difference, 32,* 40-53.

Carver, R. P. (1990). Intelligence and reading ability in grades 2-12. *Intelligence, 14,* 449-455.

Carver, R. P. (1992). What do standardized tests of reading comprehension measure in terms of efficiency, accuracy, and rate? *Reading Research Quarterly, 27,* 346-359.

Carver, R. P. (1993). Merging the simple view of reading with rauding theory. *Journal of Reading Behavior, 25,* 439–455.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of experimental child psychology*, *33*(3), 386-404.

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, 14,* 116–129.

Catts, H.W., Hogan, H.C., & Adlof, S.M. (2005). Developmental changes in reading and reading disabilities. In H. Catts & A. Kamhi (Eds.), *Connections between language and reading disabilities (pp. 2025–2040).* Mahwah, NJ: Erlbaum.

Ceci, S. J., Caves, R. D., & Howe, M. J. A. (1981). Children's long-term memory for information that is incongruous with their prior knowledge. *British Journal of Psychology, 72*(4), 443-450.

Chall, J. S. (1996). *Learning to read: The great debate (2nd ed.).* New York: McGraw–Hill.

Chandrasekharan, S., & Nersessian, N. J. (2007). Counterfactuals in science and engineering. *Behavioral and Brain Sciences*, *30*(5-6), 454-455.

Chanquoy, L., Foulin, J. N., & Fayol, M. (1990). Temporal management of short text writing by children and adults. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition.*

Chen, R. S., & Vellutino, F. R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research, 29,* 1–24.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18,* 80-98.

Chow, H. M., Kaup, B., Raabe, M., & Greenlee, M. W. (2008). Evidence of fronto-temporal interactions for strategic inference processes during language comprehension. *NeuroImage, 40,* 940-954.

Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology, 54,* 74–89.

Clark, C., & Foster, A. (2005). Children's and Young People's Reading Habits and Preferences: The who, what, why, where and when. *National Literacy Trust.*

Clark, C., & Rumbold, K. (2006). Reading for Pleasure: A Research Overview. *National Literacy Trust.*

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological review*, *100*(4), 589.

Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental neuropsychology*, *31*(1), 103-128.

Connelly, V., Dockrell, J. E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology, 25,* 99-107.

Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available? A memory-based view of text processing. *Discourse Processes, 26,* 109–129.

Cook, A. E., Limber, J. E., O'Brien, E. J. (2001). Situation-Based Context and the Availability of Predictive Inferences. *Journal of Memory and Language, 44,* 220-234.

Curtis, M. E., & Longo, A. M. (1996). Reversing Reading Failure in Adolescents with Behavioral Disorders. *Paper presented at the Annual Council for Exceptio.*

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and interventions (pp. 113–136).* Mahwah, NJ: Erlbaum.

Cragg, L. & Nation, K. (2006). Exploring Written Narrative in Children with Poor Reading Comprehension. *Educational Psychology: An International Journal of Experimental Educational Psychology, 26(1)*, 55-72.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and indirect mediationmodel of reading comprehension. Journal of Educational Psychology, 99, 311 – 325.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. Journal of Educational Psychology, 102(3), 687.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934–945.

Currie, G. (1998). Pretence, pretending and metarepresenting. *Mind and Language, 13,* 35–55.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277–299.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & 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.

De Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 51A(2),* 305–320.

Denton, C. A., Wolters, C. A., York, M. J., Swanson, E., Kulesz, P. A., & Francis, D. J. (2015). Adolescents' use of reading comprehension strategies: Differences related to reading proficiency, grade level, and gender. *Learning and Individual Differences*, *37*, 81-95.

Department for Education (2013 October 30th). The national curriculum. Retrieved from https://www.gov.uk/national-curriculum

Department for Education (2014 October 30th). National Curriculum. Retrieved from https://www.gov.uk/government/collections/national-curriculum

Department for Education. (2015a October 30th). Revised GCSE and equivalent results in England: 2014 to 2015. Retrieved from https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2014-to-2015

Department for Education. (2015b October 30th). National curriculum assessments: key stage 2, 2015 (revised). Retrieved from https://www.gov.uk/government/statistics/national-curriculum-assessments-at-key-stage-2-2015-revised

Department for Education (2015c October 30th). Commission on Assessment Without Levels: final report. Retrieved from https://www.gov.uk/government/publications/commission-on-assessment-without-levels-final-report

Department for Education (2016 September 1st). National curriculum assessments at key stage 2: 2016 (revised). Retrieved from https://www.gov.uk/government/statistics/announcements/national-curriculum-assessments-at-key-stage-2-2016-revised

Dewitz, P., Carr, E. M., & Patberg, K. P. (1987). Effects of Inference Training on Comprehension and Comprehension Monitoring. *Reading Research Quarterly, 22 (1),* 99-121.

DFES (2006). *Five year strategy for children and learners: Maintaining the excellent progress strategy paper.* Department for Education and Skills. London, UK: TSO.

Dixon, P., Lefevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80,* 465-472.

Donders, F. C. (1969). On the speed of mental processes. *Acta psychologica*, *30*, 412-431.

Dorfman, M. H. (1989). Understanding the points of fables: A developmental study. In Williams, J. P., Brown, L. G., Silberstein, A. K, & decani, J. S. (1994). An Instructional Program in Comprehension of Narrative Themes for Adolescents. *Learning Disability Quarterly, 17,* 205-221.

Dorfman, M. H., & Brewer, W. F. (1988). Understanding the points of fables. In Williams, J. P., Brown, L. G., Silberstein, A. K, & decani, J. S. (1994). An Instructional Program in Comprehension of Narrative Themes for Adolescents. *Learning Disability Quarterly, 17,* 205-221.

Dubeck, L. W., Bruce, M. H., Schmuckler, J. S., Moshier, S. E., & Boss, J. E. (1990). Science fiction aids science teaching. *The Physics Teacher, 316*-318.

Duffy, S. A., & Rayner, K. (1990). Eye movements and anaphor resolution: effects of antecedent typicality and distance. *Language and Speech, 33,* 103-119.

Dulit, E. (1972). Adolescent thinking ala Piaget: The formal stage. *Journal of Youth and Adolescence*, *1*(4), 281-301.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Edinburgh Associative Thesaurus (EAT, 2013 October 30th). Edinburgh Associated Thesaurus. Retreived from http://www.eat.rl.ac.uk/

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, *22*(1), 75-87.

Elliott, C. D., Smith, P., & McCulloch, K. (1996). *British Ability Scales (BAS II): Full Age Range*. NFER-Nelson.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18,* 972-992.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com.

Ericcson, K. A, & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245.

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: extension and evaluation. *Psychonomic Bulletin Review, 13(3),* 378-395.

Evans, J. S. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality and conditionals: a critique of the model theory of Johnson-Laird & Byrne (2002). *Psychological Review, 112,* 1040-1052.

Fayol, M. (1991). From sentence production to text production: Investigating fundamental processes. *European Journal of Psychology of Education*, *6*(2), 101-119.

Fayol, M., & Lété, B. (1987). Ponctuation et connecteurs: Une approche textuelle et génétique. *European Journal of Psychology of Education*, *2*(1), 57-71.

Ferguson, H. J. (2012). Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. *The Quarterly Journal of Experimental Psychology, 65 (5),* 939-961.

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language, 58,* 609-626.

Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research, 1236,* 113-125.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A metaanalysis of neuroimaging studies on text comprehension. *Human Brain Mapping, 29,* 581–593.

Ferstl, E.C., Rinck, M., & von Cramon, D.Y., 2005. Emotional and temporal aspects of situation model processing during text comprehension: an event-related fMRI study. Journal of Cognition. *Neuroscience, 17,* 724–739.

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event related fMRI study. *Cognitive Brain Research, 11,* 325-340.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* Sage.

Filik, R. (2008). Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition, 106,* 1038 – 1046.

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from N400. *Psychophysiology, 45,* 554-558.

Filik, R., & Leuthold, H. (2013). The role of character-based knowledge in online narrative comprehension: Evidence from eye movements and ERPs. *Brain Research, 1506,* 94-104.

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 981-995.

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes, 22,* 225-246.

Flavell, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology*, *2*(4), 421-453.

Fletcher, J. M., & Vaughn, S. (2009). Response to Intervention: Preventing and Remediating Academic Difficulties. *Child Development Perspectives, 3* (1), 30-37.

Floyd, R. Meisinger, E. Gregg, N. & Keith, T (2012). An Explanation of Reading Comprehension Across Development Using Models From Cattell-Horn-Carroll Theory: Support for Integrative Models of Reading. *Psychology in Schools, 49(8),* 725-743.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Ford, J., & Milosky, L. M. (2003). Inferring Emotional Reactions in Social Situations: Differences in Children with Language Impairment. *Journal of Speech, Language, and Hearing Research, 46 (1),* 21-30.

Ford, J. A., & Milosky, L. M. (2008). Inference generation during discourse and its relation to social competence: An online investigation of abilities of children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *51*(2), 367-380.

Freebody, P., & Anderson, R.C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior, 15(3),* 19–39.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of experimental psychology: General*, *133*(1), 101.

Friese, U., Rutschmann, R., Raabe, M., Schmalhofer, F. (2008). Neural Indicators of Inference Processes in Text Comprehension: An Event-related Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience, 20* (11), 2110-2124.

Fry, E. (1968). A readability formula that saves time. *Journal of reading*,*11*(7), 513-578.

Garrod, S., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 250.

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6,* 514–523.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4,* 294–298.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 245–262.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107-119.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59,* 601-613.

Girden, E. R. (1992). *ANOVA: Repeated measures* (No. 84). Sage.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. Journal of Experimental Psychology: General, 118(1), 13.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31,* 129–151.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10,* 597-602.

Goff, D. A., Pratt, C.,& Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills, and components of reading decoding in a normal sample. *Reading and Writing, 18,* 583 – 616.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389.

Gordon, P. C., Hendrick, R., & Foster, K. L. (2000). Language comprehension and probe-list memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 766.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Process and intervention (pp. 1 – 13).* Mahwah, NJ: Erlbaum.

Gough, P. B., & Tunmer,W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6 – 10.

Graesser, A. C., Kassler, M. A., Kreuz, R. J., & McLain-Allen, B. (1998). Verification of Statements about Story Words That Deviate from Normal Conceptions of Time: What Is True about Einstein's Dreams? *Cognitive Psychology, 35*, 246-301.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395.

Graham, L., & Bellert, A. (2009). Reading comprehension difficulties experienced by students with learning disabilities. *Australian Journal of Learning Disabilities, 10 (2),* 71-78.

Green, M. C., & Brock, T. C. (2000). The role of transportation in persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79,* 701-721.

Gyselinck, V., & Tardieu, H. (1999). The role of illustrations in text comprehension: What, when, for whom, and why? In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading (pp. 195–218).* Mahwah, NJ: Lawrence Erlbaum.

Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research, 1146,* 210-218.

Hand, B., Wallace, C. W., & Yang, E. M. (2004). Using a science writing heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: Quantitative and qualitative aspects. *International Journal of Science Education*, *26*(2), 131-149.

Harris, R. J. (1985). *A primer of multivariate statistics* (2nd ed.). New York: Academic Press.

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the Writing, Knowledge, and Motivation of Struggling Young Writers: Effects of Self-Regulated Strategy Development with and without Peer Support. *American Educational Research Journal, 43(2),* 295-340.

Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development,* 58(1, Serial No. 231).

Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition (pp. 176–199).* San Diego: Academic Press.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In Levy, M. & Ransdell, S. (Eds.). *The science of writing: Theories, methods, individual differences, and applications.* Erbaum: Mahwah, NJ.

Hayes, J. R. (2012). My Past and Present as Writing Researcher and Thoughts About the Future of Writing Research. In Berninger, V. W. (Eds). *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology.* Psychology Press: Taylor & Francis Group, Hove, East Sussex.

Hayes, J. R., & Chenoweth, N. (2006). Is working memory involved in the transcribing and editing of texts? In Abbott, R. D, Berninger, V. W., & Fayol, M. (2010). Longitudinal Relationships of Levels of Language in Writing and Between Writing and Reading in grades 1 to 7. *Journal of Educational Psychology, 102(2),* 281-298.

Hayes, J. R., & Flower, L. S. (1996). Writing Research and the Writer. *American Psychologist, 41(10),* 1106-1113.

Hinojosa, J. A., Carretié, L., & Valcárcel, M. A. Méndez--Bértolo, C., & Pozo, MA (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cognitive, Affective & Behavioral Neuroscience*, *9*, 173-89.

Hinze, S. R., Bunting, M. F., & Pellegrino, J. W. (2009). Strategy selection for cognitive skill acquisition depends on task demands and working memory capacity. *Learning and Individual Differences, 19,* 590-595.

Hochberg, J. (1972). The representation of things and people. *Art, perception, and reality*, 47-94.

Hoover, W., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127 – 160.

Houde, O., & Guichart, E (2001). negative priming effect after inhibition of number/length interference in a Piaget-like task. *Developmental Science, 4,* 119-123.

Houde, O., & Moutier, S. (1996). Deductive reasoning and experimental inhibition training: The case of the matching bias. In Moutier, S., Plagne-Cayeux, S., Melot, A., & Houde, O. (2006). Syllogistic reasoning and belief-bias inhibition in school children: evidence from a negative priming paradigm. *Developmental Science, 9(2),* 166-172.

Houde, O., & Moutier, S. (1999). Deductive reasoning and experimental inhibition training: The case of the matching bias. New data and reply to Girotto. In Moutier, S., Plagne-Cayeux, S., Melot, A., & Houde, O. (2006). Syllogistic reasoning and belief-bias inhibition in school children: evidence from a negative priming paradigm. *Developmental Science, 9(2),* 166-172.

Houde, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B., & Tzourio-Mazoyer, N., (2001). Access to deductive logic depends on a right ventromedial prefrontal area devoted to emotion and feeling: evidence from a training paradigm. *NeuroImage, 14,* 1486-1492.

Howell, D. C. (1987). *Statistical methods for psychology*. Cengage Learning.

Huang, Y. T., & Gordon, P. C. (2011). Distinguishing the time course of lexical and discourse processes through context, coreference, and quantified expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 966.

Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, *47*(1), 72-87.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10), 1489-1506.

Jarrold, C.R., Carruthers, P., Smith, P.K., & Boucher, J. (1994). Pretend play: is it metarepresentational? *Mind and Language, 9,* 445–68.

Jefferson, A. L., Gibbons, L. E., Rentz, D. M., Carvalho, J. O., Manly, J., Bennett, D. A., & Jones, R. N. (2011). A life course model of cognitive activities, socioeconomic status, education, reading ability, and cognition. *Journal of the American Geriatrics Society*, *59*(8), 1403-1411.

Jin, H., Liu, H., Mo, L., Fang, S., Zhang, J. X., & Lin, C. (2009). Involvement of the left inferior frontal gyrus in predictive inference making. *International Journal of Psychophysiology, 71,* 142-148.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. In Byrne, R. M. J. (2007). Precis of The Rational Imagination: How People Create Alternatives to Reality. *Behavioural and Brain Sciences, 30,* 439-480.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 646-678.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437–447.

Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. *Trends in Cognitive Science, 9,* 512–518.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,*122-149.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology*, *60*(7), 976-990.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32,* 446-463.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behaviour, 23,* 115-126.

Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage independent items. *Scientific Studies of Reading, 10,* 363–380.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281–300.

Kellogg, R. T. (1999). Components of working memory in text production. In Torrance, M., & Jeffery, G. C. (Eds). *The cognitive demands of writing: Processing capacity and working memory in text production*. Amsterdam: Amsterdam University Press.

Kendeou, P., Bohn-Gettler, C., & Fulton, S. (2011). What we have been missing: The role of goals in reading comprehension. In M. T. McCrudden, J. Magliano, & G. Schraw (Eds.), *Relevance instructions and goal focusing in text learning (pp. 375–394).* Charlotte, NC: Information Age Publishing.

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*(3), 259-272.

Kershaw, S. & Schatschneider, C. (2012). A latent variable approach to the simple view of reading. *Reading and Writing, 25,* 433-464.

Keys, C. W. (1999). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science education*,*83*(2), 115-130.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805-825.

Kim, Y., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of Experimental Child Psychology, 113,* 93-111.

Kim, S., Yoon, M., Kim, W., Lee, S., & Kang, E. (2012). Neural Correlates of Bridging Inferences and Coherence Processing. *Journal of Psycholinguistic Research, 41,* 311-321.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95,* 163–182.

Kintsch, W. (1993). Information accretion and reduction in text processing: inferences. *Discourse Processes, 16,* 193-202.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, MA: Cambridge University Press.

Kintsch, W. (2012). Psychological models of reading comprehension and their implications for assessment. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability (pp. 21–38).* Lanham, MD: Rowman & Littlefield Education.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and language*, *29*(2), 133-159.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Klin, C., Guzmán, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language, 40,* 593–604.

Kuczera, M., Field, S., & Windisch, H. C. (2016). *Building Skills for All: A Review of England: Policy insights from the survey of adult skills.* OECD

Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., & Perner, J. (2013). Processing counterfactual and hypothetical conditions: An fMRI investigation. NeuroImage, 72, 265-271.

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95,* 211–224.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94,* 778–784.

Lindo, E. J. (2014). Family background as a predictor of reading comprehension performance: An examination of the contributions of human, financial, and social capital. *Learning and Individual Differences, 32,* 287-293.

Long, D. L., & Golding, J. M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes, 16,* 55-73.

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). The generation of goal related inferences during narrative comprehension. *Journal of Memory and Language, 5,* 634-647.

Long, D. L., Oppy, B. J., Seely, M. R. (1994). Individual Differences in the Time Course of Inferential Processing. *Journal of Experimental Psychology: Memory and Cognition, 20(6),* 1456-1470.

Long, D. L., Seely, M. R., & Oppy, B. J. (1996). The availability of causal information during reading. *Discourse Processes, 22,* 145–170.

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The Development of Nonverbal Working Memory and Executive Control Processes in Adolescence. *Child Development, 76 (3),* 697-712.

MacArthur, C. A. (2012). Evaluation and revision. *Past, present, and future contributions of cognitive writing research to cognitive psychology*, 461.

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.

Magorian, M. (1981). *Good night Mr. Tom.* New York: Harper.

Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology*, *47*(5), 315-335.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88,* 49–63.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50,* 643-655.

Markman, K. D. & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review, 7,* 244–67.

Markman, K. D. & McMullen, M. N. (2005). Reflective and evaluative modes of mental simulation. In Mandel, D. R, Hilton, D. J. & Catellani, P. (2005). *The psychology of counterfactual thinking (pp. 77–93).* Routledge.

Markman, K. D., & McMullen, M. N. (2007). Counterfactuals need not be comparative: The case of "As if." In Byrne, R. M. J. (2007). Precis of The Rational Imagination: How People Create Alternatives to Reality. *Behavioural and Brain Sciences, 30,* 439-480.

Markovits, H., & Schroyens, W. (2007). A curious belief-bias effect: Reasoning with false premises and inhibition of real-life information. *Experimental Psychology*, *54*(1), 38-43.

Martinussen, R., & Mackenzie, G. (2015). Reading comprehension in adolescents with ADHD: Exploring the poor comprehender profile and individual differences in vocabulary and executive functions. *Research in developmental disabilities*, *38*, 329-337.

Mason, R. A., & Just, M. A. (2004). How the brain processes causal inferences in text: A theoretical account of generation and integration component processes utilizing both cerebral hemispheres. *Psychological Science, 15,* 1–7.

Masson M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology, 75,* 314-318.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8,* 299-325.

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology, 86,* 256-266.

McDaniel, M. A., & Waddill, P. J. (1994). The mnemonic benefit of pictures in text: Selective enrichment for differentially skilled readers. In W. Schnotz & R. W. Kulhavy (Eds.), *Comprehension of graphics (pp. 165–184).* Amsterdam, The Netherlands: North Holland.

McGee, A., & Johnson, H. (2003). The Effect of Inference Training on Skilled and Less Skilled Comprehenders. *Educational Psychology, 23 (1),* 49-59.

McGonigle, B., & Chalmers, M. (1984). The selective impact of question form and input mode on the symbolic distance effect in children. *Journal of Experimental Child Psychology*, *37*(3), 525-554.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 82–91.

McKoon, G., & Ratcliff, R. (1989). Semantic association and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 326–338.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99,* 440-466.

McMullen, M. N. (1997). Affective contrast and assimilation in counterfactual thinking. *Journal of Experimental Social Psychology, 33,* 77–100.

McMullen, M. N. & Markman, K. D. (2000). Downward counterfactuals and motivation: The wake-up call and the Pangloss effect. *Personality and Social Psychology Bulletin, 26,* 575–84.

McMullen, M. N. & Markman, K. D. (2002). Affective impact of close counterfactuals: Implications of possible futures for possible pasts. *Journal of Experimental Social Psychology, 38*, 64–70.

McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 465.

Meneghetti, C., Carretti, B., & De Beni, R. (2006). Components of reading comprehension and scholastic achievement. *Language and Individual Differences, 16,* 291-301.

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.

Mitchell, P., Robinson, E. J., Issacs, J. E., Nye, R. M. (1996). Contamination in reasoning about false belief: an instance of realist bias in adults but not children. *Cognition, 59,* 1-21.

Moje, E., Overby, M., Tysvaer, N., & Morris, K. (2008). The complex world of adolescent literacy: Myths, motivations, and mysteries. *Harvard Educational Review*, *78*(1), 107-154.

Moutier, S., Angeard, N., & Houde, O. (2002). Deductive reasoning and matching-bias inhibition training: Evidence from a debiasing paradigm. *Thinking & reasoning*, *8*(3), 205-224.

Moutier, S., Plagne-Cayeux, S., Melot, A., & Houde, O. (2006). Syllogistic reasoning and belief-bias inhibition in school children: evidence from a negative priming paradigm. *Developmental Science, 9(2),* 166-172.

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology, 91,* 488–496.

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language & Communication Disorders*, *41*(3), 335-354.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology, 72,* 549–560.

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.

National Center for Education Statistics (NCES). (2009). *The Nations Report Card: Reading 2009 (NCES 2010-458).* Institute of Education Sciences (IES). Washington, DC: U.S. Department of Education.

National Center for Education Statistics (NCES). (2011). *The Nations Report Card: Science 2009 (NCES 2011-451).* IES. Washington, DC: U.S. Department of Education.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: The National Institute of Child Health and Human Development.

Neale, M. D. (1989). Neale Analysis of Reading Ability, revised British edition (British adaptation and standardisation by Una Christophers and Chris Whetton). *NFER-Nelson Publishing, Windsor, UK.*

Neale, M. D. (1999). *Neale analysis of reading ability: Manual.* ACER Press, Australian Council for Educational Research Limited, 19 Prospect Hill Road, Camberwell, Melbourne, Victoria 3124, Australia..

Nelson, K. E., & Kosslyn, S. M. (1975). Semantic retrieval in children and adults. *Developmental Psychology*, *11*(6), 807.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.

Nieuwland, M. S. (2012). Establishing propositional truth-value in counterfactual and real-world contexts during sentence comprehension: Differential sensitivity of the left and right inferior frontal gyri. *Neuroimage, 59,* 3433-3440.

Nieuwland, M. S. (2013). "If a lion could speak…": Online sensitivity to propositional truth-value of unrealistic counterfactual-world sentences. *Journal of Memory and Language, 68,* 54-67.

Nieuwland, M. S. (2015). The Truth Before and After: Brain Potentials Reveal Automatic Activation of Event Knowledge during Sentence Comprehension. Journal of Cognitive Neuroscience, 1-14.

Nieuwland, M. S., & Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition, 122,* 102-109.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18,* 1098-1111.

Oakhill, J., & Cain, K. (2011). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16,* 91–121.

Oakhill, J., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18,* 443–468.

Omanson, R. C., Warren, W. H., & Trabasso, T. (1978). Goals, inferential comprehension, and recall of stories by children∗. *Discourse Processes*,*1*(4), 337-354.

Otero, J. C., & Campanario, J. M. (1990). Comprehension evaluation and regulation in learning from science texts. *Journal of research in science teaching*, *27*(5), 447-460.

Paris, S. G., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology, 8,* 217–227.

Paris, S. G., Lindauer, B. K., & Cox, G. L. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension (pp. 141–183).* Hillsdale, NJ: Lawrence Erlbaum.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research*, *42*(1), 107-123.

Paul, S. S., & Clarke, P. J. (2016). A systemic review of reading interventions for secondary school students. *International Journal of Educational Research, 79,* 116-127.

Peeck, J. (1994). Retention of pictorial and verbal content of a text with illustrations. *Journal of Educational Psychology, 66,* 880–888.

Peracchi, K. A., & O'Brien, E. J. (2004). Character profiles and the activation of predictive inferences. *Memory & Cognition*, *32*(7), 1044-1052.

Perfetti, C. A. (1998). Two basic questions about reading and learning to read. In P. Reitsma, & L. Verhoeven (Eds.), *Problems and interventions in literacy development (pp. 15–48).* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Perfetti, C. A., & Adlof, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability (pp. 3–20).* Lanham, MD: Rowman & Littlefield Education.

Perfetti, C. A., Marron, M. A., & Foltz, P. W. (1996). Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention (pp. 137 – 166).* Mahwah, NJ: Erlbaum.

Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology, 105,* 243–255.

Poynor, D. V., & Morris, R. K. (2003). Inferred goals in narratives: Evidence from self-paced reading, recall, and eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 3.

Prat, C. S., & Just, M. A. (2010). Exploring the Neural Dynamics Underpinning Individual Differences in Sentence Comprehension. *Cerebral Cortex, 21,* 1747-1760.

Prat, C. S., Mason, R. A., & Just, M. A. (2011). Individual differences in the neural basis of causal inferencing. *Brain and Language, 116,* 1–13.

Pretorius. (2000). Inference generation in the reading of expository texts by university students. In Oakhill, J., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18,* 443–468.

Rader, A. W., & Sloutsky, V. M. (2002). Processing of Logically Valid and Logically Invalid Conditional Inferences in Discourse Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(1),* 59-68.

Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of ''nearest possible world''. *Child Development, 81,* 376–389.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology, 114,* 389-404.

Ransdell, S., & Levy, M. C. (1994). Writing as Process and Product: The Impact of Tool, Genre, Audience Knowledge, and Writer Expertise. *Computers in Human Behaviour, 10 (4),* 511-527.

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11(4),* 289–312.

Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading*, *16*(3), 187-217.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparison to other models. *Behavioural and Brain Sciences, 26,* 445-526.

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.

Richardson, J. T. (1980). *Mental imagery and human memory*. St. Martin's Press.

Ricketts, J., Jones, C. R., Happé, F., & Charman, T. (2013). Reading comprehension in autism spectrum disorders: The role of oral language and social functioning. *Journal of Autism and Developmental Disorders*, *43*(4), 807-816.

Riggs, K. J. & Peterson, D. M. (2000) Counterfactual thinking in preschool children: Mental state and causal inferences. In Mitchell, P. & Riggs, K. J. *Children's reasoning and the mind.* Psychology Press.

Robertson, D. A., Gernsbacher, M. A., Guidotti, S. J., Robertson, R. R., Irwin, W., Mock, B. J., & Campana, M. E. (2000). Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. *Psychological science*, *11*(3), 255-260.

Robinson, E.J., & Beck, S. (2000). What is difficult about counterfactual reasoning? In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind (pp. 101–119).* Hove, UK: Psychology Press.

Roese, N. J., Sanna, L. J., & Galinsky, A. D. (2005). The mechanics of imagination: Automaticity and control in counterfactual thinking. *The new unconscious*, 138-170.

Romine, C. B., & Reynolds, C. R. (2005). A Model of the Development of Frontal Lobe Functioning: Findings From a Meta-Analysis. *Applied Neuropsychology, 12 (4),* 190-201.

Rose, J. (2006). *Independent review of the teaching of early reading: Final report. Department for education and skills.* London, UK: TSO.

Rosinski, R. R., Pellegrino, J. W., & Siegel, A., W. (1977). Developmental Changes in Semantic Processing of Pictures and Words. *Journal of Experimental Child Psychology, 23,* 282-291.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science,323*(5916), 951-953.

Royer, J. M. (2001). Developing reading and listening comprehension tests based on the Sentence Verification Technique (SVT). *Journal of Adolescent & Adult Literacy*, *45*(1), 30-41.

Ryan, T. A., & Schwartz, C. B. (1956). Speed of perception as a function of mode of representation. *The American journal of psychology*, *69*(1), 60-69.

Sabatini, J. P., Albro, E. R., & O'Reilly, T. (Eds.). (2012). *Measuring up: Advances in how we assess reading ability.* Lanham, MD: Rowman & Littlefield Education.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The Nation's Report Card: Writing 2007 (NCES 2008–468).* Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, *24*(5), 563.

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *50*(6), P297-P306.

Samuels, S. J. (1988). Decoding and automaticity: Helping poor readers become automatic at word recognition. *The Reading Teacher*, *41*(8), 756-760.

Samuels, S. J., & Flor, R. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly, 13,* 107–122.

Savage, R. S., & Wolforth, J. (2007). An additive simple view of reading describes the performance of good and poor readers in higher education. *Exceptionality Education Canada, 17(1),* 243–268.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. *Toward a general theory of expertise: Prospects and limits*, 172-194.

Schacht, A., & Sommer, W. (2009). Emotions in word and face processing: early and late cortical responses. *Brain and cognition*, *69*(3), 538-550.

Schlochtermeier, L. H., Kuchinke, L., Pehrs, C., Urton, K., Kappelhoff, H., & Jacobs, A. M. (2013). Emotional Picture and Word Processing: An fMRI Study on Effects of Stimulus Complexity. *PLoS ONE, 8 (2),* e55619. doi:10.1371/journal.pone.0055619

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A Unified Model for Predictive and Bridging Inferences. *Discourse Processes, 33 (2),* 105-132.

Schmidt, C. R., & Paris, S. G. (1978). Operativity and reversibility in children's understanding of pictorial sequences. *Child Development*, 1219-1222.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13,* 141-156.

Schroeders, U., Schipolowski, S., & Wilhelm, O. (2015). Age-related changes in the mean and covariance structure of fluid and crystallized intelligence in childhood and adolescence. *Intelligence*, *48*, 15-29.

Seger, C.A., Desmond, J.E., Glover, G.H., & Gabrieli, J.D.E. (2000). Functional magnetic resonance imaging evidence for right hemisphere involvement in processing unusual semantic relationships. *Neuropsychology 14,* 361–369.

Seigneuric, A., & Ehrlich, M.-F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing, 18(7–9),* 617–656.

Seigneuric, A., Ehrlich, M.-F., Oakhill, J.V., & Yuill, N.M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing, 13,* 81–103.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*, 40–59.

Shanahan, T., & Shanahan, C. (2012).What is disciplinary literacy and why does it matter? *Topics in Language Disorders, 32,* 1–12.

Shapiro, E. S., Fritschmann, N. S., Thomas, L. B., Hughes, C. L., & McDougal, J. (2014). Concurrent and Predictive Validity of Reading Retell as a Brief Measure of Reading Comprehension for Narrative Text. *Reading Psychology*, *35*(7), 644-665.

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental psychology*, *27*(6), 960.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child development*, 973-980.

Singer, M. (1990). Psychology of Language. In Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. *Psychological Review, 101(3),* 371-395.

Singer, M. (1994). Discourse inference processes. In Graesser, A. C., Kassler, M. A., Kreuz, R. J., & McLain-Allen, B. (1998). Verification of Statements about Story Words That

Deviate from Normal Conceptions of Time: What Is True about Einstein's Dreams? *Cognitive Psychology, 35,* 246-301.

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. *Memory & Cognition, 20,* 539-548.

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behaviour, 22,* 437-448.

Singer, J. B., & Flavell, J. H. (1981). Development of knowledge about communication: Children's evaluations of explicitly ambiguous messages. *Child Development*, 1211-1215.

Singer, M., & Ritchot, K. F.M. (1996). The role of working memory capacity and knowledge access in text inference processing. *Memory & Cognition, 24(6),* 733-743.

Snodgrass, J. G., & McCullough, B. (1986). The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 147.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*(2), 174.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 516-536.

Snow, C. E., Scarborough, H. S., & Burns, M. S. (1999). What speech-language pathologists need to know about early reading. *Topics in Language Disorders, 20,* 48–58.

Snowling, M.J., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. & Hulme, C. (2009). York Assessment of Reading for Comprehension: Passage Reading. GL Assessment.

Snowling, M.J., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. & Hulme, C. (2010). York Assessment of Reading for Comprehension: Passage Reading - Secondary Version. GL Assessment.

Spear-Swerling, L., & Sternberg, R. J. (1994). The road not taken: An integrative theoretical model of reading disability. *Journal of Learning Disabilities, 27,* 91–103.

Spooner, A. L., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability?. *British Journal of Educational Psychology*, *74*(2), 187-204.

Stanovich, K. E. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. In H. Reese (Ed.), *Advances in child development and behavior (Vol. 24,pp. 133–180).* San Diego, CA: Academic Press.

Stanovich, K. E., Cunningham, A. E., & Freeman, D. J. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly, 19,* 278-303.

Steinberg, L., & Morris, A. S. (2001). Adolescent development. *Journal of Cognitive Education and Psychology*, *2*(1), 55-87.

Stein, N. L., & Trabasso, T. (1981). What's in a story: An approach to comprehension and instruction. *Center for the Study of Reading Technical Report; no. 200.*

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38,* 934–947.

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. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of educational psychology*, *82*(2), 306.

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology, 63 (2),* 358-385.

Székely, A., D'amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621-633.

Tamm, L., Menon, V., & Reiss, A. L. (2002). Maturation of brain function associated with response inhibition. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(10), 1231-1238.

Tarchi, C. (2010). Reading comprehension of informative texts in secondary school: A focus on direct and indirect effects of reader's prior knowledge. *Learning and Individual Differences, 20,* 415–420.

Tarchi, C. (2012). La comprensione del testo espositivo. L'interazione tra conoscenze precedenti del lettore e contenuto del testo. *Giornale Italiano di Psicologia, 1,* 91–114.

Tarchi, C. (2015). Fostering reading comprehension of expository texts through the activation of readers' prior knowledge and inference-making skills. *International Journal of Educational Research, 72,* 80-88.

Theios, J., & Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: reading, naming, and visual and conceptual comparisons. *Psychological review*, *96*(1), 5.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792-14797.

Thorsen, C., Gustafsson, J. E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology*, *84*(4), 556-570.

Thurlow, R., & van den Broek, P. (1997). Automaticity and inference generation during reading comprehension. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 165-181.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory and Cognition, 16,* 283-298.

Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading, 32,* 383–401.

Torgerson, D., Torgerson, C., Ainsworth, H., Buckley, H., Heaps, C., Hewitt, C. and Mitchell, N. (2014) *Improving Writing Quality. Evaluation Report and Executive Summary.* London: Education Endowment Foundation.

Torrance, M., & Galbraith, D. (2006). The Processing Demands of Writing. In MacArthur, C., Graham, S., & Fitzgerlad, J. (Eds.). (2006). *Handbook of Writing Research.* New York: Guilford Publications.

Torrens, D. (1999). Individual differences and the belief bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking & Reasoning*,*5*(1), 1-28.

Tovani, C. (2001). I Read It, But I Don't Get It: Strategies for Adolescent Readers. Journal of Adolescent & Adult Literacy, 45(1), 78-78.

Tyler, L. K., & Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1493), 1037-1054.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

van den Broek, P. W. (2012). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability (pp. 39–58).* Lanham, MD: Rowman & Littlefield Education.

van den Broek, P., Lorch, R. F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29,* 1081–1087.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39,* 299–316.

van den Broek, P., Risden, K., & Husebye-Hartman, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in text comprehension (pp. 353–373)*. Hillsdale, NJ: Erlbaum.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading (pp. 71–98)*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Van Dijk, T. A., Kintsch, W., & Van Dijk, T. A. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Reet, J., Pinkham, A. M., & Lillard, A. S. (2007). The development of the counterfactual imagination. In Byrne, R. M. J. (2007). Precis of The Rational Imagination: How People Create Alternatives to Reality. *Behavioural and Brain Sciences, 30,* 439-480.

Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S., et al. (2011). Efficacy of collaborative strategic reading with middle school students. *American Educational Research Journal, 48(4),* 938–964.

Verhoeven, L. van Leeuwe, J. (2008). Prediction of the Development of Reading Comprehension: A Longitudinal Study. *Applied Cognitive Psychology, 22,* 407-423.

Virtue, S., Parrish, T., Jung-Beeman, M. (2008). Inferences during Story Comprehension: Cortical Recruitment Affected by Predictability of Events and Working Memory Capacity. *Journal of Cognitive Neuroscience, 20 (12),* 2274-2284.

Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Jung-Beeman, M. (2006). Neural activity of inferences during story comprehension. *Brain Research, 1084,* 104-114.

Virtue, S., & Joss, L. M. (2012). Hemispheric processing of inferences during text comprehension: The role of consistency and task difficulty. *Laterality: Asymmetries of Body, Brain and Cognition, 17 (5),* 549-564.

Virtue, S., van den Broek, P., 2005. Hemispheric processing of anaphoric inferences: the activation of multiple antecedents. *Brain Language, 93,* 327–337.

Vosniadou, S., Pearson, P. D., & Rogers, T. (1988). What causes children's failures to detect inconsistencies in text? Representation versus comparison difficulties. *Journal of educational psychology*, *80*(1), 27.

Wagner, A.D., Parâe-Blagoev, E.J., Clark, J., Poldrack, R.A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338.

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about plausible and impossible events. *Journal of Experimental Pyschology: Learning, Memory, and Cognition, 34,* 1001-1010.

Wechsler, D. (1996). *Wechsler objective language dimensions*. London: The Psychological Corporation.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI).* Texas, USA: Harcourt Assessment.

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes, 14,* 133-146.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*(1), 85-106.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. (1996).*Behavioural assessment of the dysexecutive syndrome*. Thames Valley Test Company.

Wittrock, M.C., Marks, C., & Doctorow, M. (1975). Reading as a generative process. *Journal of Educational Psychology, 67,* 484–489.

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of Online Reading Paradigms: Eye Tracking, Moving-Window and Maze. *Journal of Psycholinguistic Research, 41,* 105-128.

Wright, M. J. (2005). Saliency predicts change detection in pictures of natural scenes. *Spatial Vision, 18 (4),* 413-430.

Wright, K. P, & Berch, D. B. (1992). Developmental Changes in Memorial Comparisons: The Effects of Stimulus Presentation Mode. *Journal of Experimental Child Psychology, 53,* 219-236.

Yuill, N.M., Oakhill, J.V., & Parkin, A.J. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology, 80,* 351–361.

Zhang, J., Feng, C., Fox, P. T., Gao, J., & Tan, L. H. (2004). Is left inferior frontal gyrus a general mechanism for selection? *NeuroImage, 23,* 596-603.

Zhang, Y., Tardif, T., Shu, H., Li, H., Liu, H., McBride-Chang, C., ... & Zhang, Z. (2013). Phonological skills and vocabulary knowledge mediate socioeconomic status effects in predicting reading outcomes for Chinese children. *Developmental psychology*, *49*(4), 665.

Zhu, Z., Zhang, J. X., Wang, S., Xiao, Z., Haung, J., & Chen, H. (2009). Involvement of left inferior frontal gyrus in sentence-level semantic integration. *NeuroImage, 47,* 756-763.

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82,* 51-59.

Zirinsky, D., & Rau, S.A. (2001). A classroom of teenaged readers: Nurturing reading processes in senior high English. In Hughes-Hassell, S. & Rodge, P. (2007). The

Leisure reading habits of urban adolescents. *Journal of Adolescent & Adult Literacy, 51(1),* 22-33.

Zwaan, R. A. (1996). Processing Narrative Time Shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(5),* 1196-1207.

Zwaan, R. A. (2003). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of learning and motivation*, *44*, 35-62.

Zwaan, R. A., & Singer, M. (2003). Text comprehension. *Handbook of discourse processes*, 83-121.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. *Psychological Science, 6,* 292-297.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1.

Zwaan, R. A., & Truitt, T. P. (2000). Inhibition of smoking-related information in smokers and nonsmokers. *Experimental and Clinical Psychopharmacology*, *8*(2), 192.

Zwaan, R. A., & Yaxley, R. H. (2003a). Spatial iconicity affects semantic relatedness judgments. *Psychonomic bulletin & review*, *10*(4), 954-958.

Zwaan, R. A., & Yaxley, R. H. (2003b). Hemispheric differences in semantic-relatedness judgments. *Cognition*, *87*(3), B79-B86.

APPENDICES

Appendix 3.1 – Readability levels for all stories

Story	Readability Rating
Counterfactual-World	
Story 1 – Version 1	8.732
Story 1 – Version 2	8.409
Story 2 – Version 1	8.951
Story 2 – Version 2	8.977
Story 3 – Version 1	8.732
Story 3 – Version 2	8.758
Story 4 – Version 1	8.732
Story 4 – Version 2	8.706
Story 5 – Version 1	8.758
Story 5 – Version 2	8.732
Story 6 – Version 1	8.706
Story 6 – Version 2	8.68
Real-World	
Story 7 – Version 1	8.435
Story 7 – Version 2	8.977
Story 8 – Version 1	8.435
Story 8 – Version 2	8.977
Story 9 – Version 1	8.487
Story 9 – Version 2	8.732
Story 10 – Version 1	8.977
Story 10 – Version 2	8.977
Story 11 – Version 1	8.951
Story 11 – Version 2	8.68
Story 12 – Version 1	8.706
Story 12 – Version 2	8.977
Literal	
Story 13 – Version 1	8.977
Story 13 – Version 2	8.977
Story 14 – Version 1	8.977
Story 14 – Version 2	8.977
Story 15 – Version 1	8.706
Story 15 – Version 2	8.977
Story 16 – Version 1	8.977
Story 16 – Version 2	8.732
Story 17 – Version 1	8.977
Story 17 – Version 2	8.977
Story 18 – Version 1	8.977
Story 18 – Version 2	8.732

Theme	Real-World Stimuli	Counterfactual-World Stimuli	Literal Stimuli
Coherence Inferences			
Reproduction	Holly planted a tiny seed. A century later an oak tree stood in its place. **The seed grew into a tree**	The giant egg began to crack loudly. Soon, instead of the egg there was a small bear cub. **The bear hatched from the egg**	The men were clones. They all looked the same because the men had been grown in giant test tubes. **The men had been grown in giant test tubes**
Transport	One penguin jumped into the cold ocean. A few seconds later it was at their feet. **The penguin swam to the children**	Arms spread wide he jumped off the ledge. A few seconds later he was on the other side. **The man flew across the hole**	The large fish ran across the green grass. He left a trail of gunky slime as he went. **The fish ran across the grass**
Communication	Ben heard a long deep sound. Ben turned around to see a whale swimming next to him. **The whale made the sound.**	"Hello," came a voice from below. They looked down to see a cat looking at them. **The cat talked to them**	Holly did not want to wake the **huge** fish up. So when **Holly** spoke to Ben she whispered. **Holly spoke to Ben**
Structure	Ben fell to the ground. He landed on his leg. There was a crack and Ben cried out in pain. **Ben's leg broke**	The tree was ten feet away. They span around again. The same tree was now two feet away. **The tree moved closer to Holly and Ben**	These plants did not suck up water from their roots. The plants sucked up water from a huge straw. **Plants suck up water through a straw**
Animal Attacks	However, a shark was approaching. It attacked. Blood began to spill out of Holly's arm.	The hungry tiger pounced. Almost instantly Ben's face was covered in warm sticky spit.	"ARGH!!!" Holly cried out in pain. A snake had bitten her ankle.

	The shark bit Holly's arm	The tiger licked Ben's face	The venom was very toxic. **The snake bit Holly**
Density	They cautiously crept on to the thin ice. Crack! Crack! Ice cold water swept all over them. **The children fell through the ice**	They slowly stepped onto the murky lake. They crossed the lake without so much as wet feet. **The children walked on water**	They ran back to the spaceship along the hard road. They made it back in no time at all. **The children ran on the road**

Elaborative Inferences

Colour	The gigantic clay statue shone in the sun light. It must have been at least ten feet high. **The statue was a burnt orange colour**	The tiny bear cub was quite odd. He was not covered in fur. He was covered in cheese. **The bear is yellow**	Every blade of grass was exactly the same. Every blade of grass was quite short and green. **Grass is green**
Texture	Luckily a snake skin floated past Holly. She took it and made a bag out of it. **The bag is scaly**	They were shaped like human hands. But, they were like a large bird's wing in every other way. **The man's hand is covered in feathers**	Surprisingly the fish was huge. The fish was also covered in pink and purple fur. **The fish is covered in fur**
Temperature	Holly and Ben spotted a little white house. The tiny round house was made out of snow. **The house is cold.**	The fluid in the stream was brown and steaming. "Watch out Holly, it is coffee!" said Ben. **The river is hot**	Holly licked the ice-lolly. She screamed. The ice-lolly was so hot it had burnt her tongue**. Ice is hot**
Weight	The hippo looked at Holly and Ben. As the hippo walked it's giant feet cracked the ground. **Hippos are heavy**	Holly and Ben looked up. The giant grey leaves above their heads were made of solid stone. **The leaves are heavy**	The rhino was very very big. It had huge feet and was very very heavy. **The rhino is heavy**
Rigidity	Ben grabbed the steel shield. He used it to protect them from the huge rocks.	The trees were odd. The tree trunks were not made of wood. They were made out	The people in this place wore tops made from elastic. Elastic is quite stretchy.

	The shield was hard and inflexible	of huge pipe cleaners. Tree trunks are bendy	Elastic is stretchy
Colour	The caring Blibs gave Ben a special medicine. It was made from mashed up bananas. The medicine was yellow.	The trees were odd. The leaves were covered in tiny rubies that glinted in the sunlight. The leaves were red	There was a large herd of very odd zebras. The zebras were covered in huge pink stripes. The zebra had pink stripes

Appendix 3.3 – Critical Texts and Test Sentences used in Critical Text Validation Booklet

Critical Text	Test Sentence
Although it glided through the water the humongous hippo was extremely heavy.	Hippos on this planet are very light.
He dove off the cliff. Arms spread wide he soared through the air and soon landed on the other cliff.	The friendly man had flown across the large gap.
They stepped onto the lake. Soon they were on the other side, without so much as wet feet.	The children had walked across the water.
The creature's gigantic hairy paw had been cut by the sharp metal spikes beneath it.	The sharp metal tickled the creatures paw.
Surprisingly the gigantic fish was covered in soft orange and violet fur.	Fish on this planet are covered in fur.
The hideous creature had toxic green blood leaking from a cut in its huge left paw.	The enormous creature had red thick blood.
Ben and Holly were trapped in a tiny cage. The cold bars were made of solid steel.	Steel is very soft.
The bear cub was odd. It wasn't covered in fur. It was covered in banana skins.	The bears on this planet are yellow.
Quietly Holly spoke to Ben, careful not to wake the sleeping creature behind them.	Quietly Holly spoke to Ben.
The beautiful granite statue towered above all of the other buildings.	The statue was black.
However, the shark was approaching. It pounced. Blood began to spill out of Holly's arm.	The evil shark had bitten Holly's arm.
The hairy brown elephant stood in front of them, using its trunk to suck up the snow.	Elephants are heavy.
The monkeys had very odd looking faces; in the place of their nose was a green snake.	The small monkey's nose was covered in fur.
Ben fell to the ground. He landed on his leg. There was a crack and Ben cried out in pain.	The bone in Ben's leg had broken.
Luckily unlike those on Earth the rocks on this planet were very soft and spongy.	The rocks on this planet are very hard.
Instead of sucking water up through their roots, these plants sucked up water through a straw.	The flowers suck up water through a straw.
The egg began to crack loudly. Soon, instead of the egg there was a small furry bear cub.	The tiny bear cub had hatched from the egg.
All of the inhabitants of this planet were exactly the same. They were all clones.	The inhabitants were all the same.

The whale made a low deep groan. Suddenly many more whales appeared around him.	The whale called to the other whales.
The tiny animal had been shot. A bright orange substance was oozing from the wound.	The gigantic creature had bright red blood.
Instead of sparkling blue water, a thick stream of coffee was flowing through the river.	The rivers on this planet are hot.
The large tree was on their right. They turned around again, and saw the same tree on their left!	The trees had moved.
A great magnificent eagle soared high above their heads full of both grace and power.	The large eagle was covered in scales.
The fish ran across the green grass, leaving a trail of pink sticky slime as he went.	The wet fish ran across the grass.
A snake skin washed up on one of the islands. Holly quickly turned it into a hat.	The hat was scaly.
"Don't worry, it will grow back," said Cain. "Part of the plant lives on under the soil."	Plants do not have roots below the soil.
Instead of leaves all of the trees seemed to be growing tiny cans of coca cola.	Leaves on the trees on this planet are mostly red.
His tail had been cut off, but instead of a wound, there was a new smaller tail.	The dog's tail had grown back.
Enormous leaves were rustling above their heads, enormous leaves made of solid lead.	The leaves on this planet are very light.
They cautiously tiptoed across the thin ice. Crack! Ice cold water engulfed their bodies.	The children had fallen through the thin ice.
Very slowly Holly and Ben looked up. The grey leaves high above them were made of rock.	The leaves on this planet are heavy.
Splash. The heavy marble rocks quickly sank to the bottom of the very murky lake.	The rocks on this planet sink in water.
The giraffe had a huge neck which it was using to get leaves from the top of the trees.	The giraffes on this planet have short necks.
"Hello," came a voice from behind. They turned around to see a cat looking at them.	The cats on this planet can talk.
Holly took a bite of the pear. Instantly she fell to the ground clutching her belly.	The pears on this planet are good for you.
There was no way out. Just like everything else, their small cage was made of solid concrete.	Concrete is hard and inflexible.
Holly planted a tiny seed. A century later an oak tree stood in its place.	The seed had grown into a huge oak tree.

One penguin jumped into the ice cold ocean. Moments later it appeared at their feet.	The penguin had swum to the children.
"ARGH!!!" Holly cried out. A snake had bit her ankle. The venom was extremely deadly.	The evil snake bit Holly's pale ankle.
Whilst shaped like a human hand, it was like a bird's wing in every other way possible.	The mans hands were covered in feathers.
Holly fell forward and landed in something warm sticky and red. She screamed loudly.	She had landed in a pool of ketchup.
The zebras trotting ahead of them had bright yellow stripes all over their bodies.	All of the zebras had bright yellow stripes.
Instantly she screamed in pain. The ice-lolly was so hot it burnt her tongue when she licked it.	The ice lollies on this planet are hot.
Even every tree trunk was exactly the same. They were all perfect cylinders.	The tree trunks are all rectangular shaped.
Surprisingly, instead of wood, the tree trunks seemed to be made out of plastic rulers.	Trees trunks on this planet are flexible.
The enormous field ahead of them was filled to the brim with beautiful poppies.	Poppies are red.
However, they were trapped. There was no way out. The thick walls of rock were completely solid.	The rocks on this planet were really soft.
The whole planet was covered in thick white snow. Snow so deep it came up to their elbows.	Snow is very cold.

Appendix 3.4 – Critical Text ratings for each item

Critical Sentence Accuracy Ratings (%) - Children

Item	Percentage of time item was rated correctly
1a	90%
1b	90%
2a	90%
2b	80%
3a	100%
3b	80%
4a	80%
4b	100%
5a	90%
5b	90%
6a	100%
6b	80%
7a	90%
7b	90%
8a	80%
8b	80%
9a	90%
9b	80%
10a	90%
10b	90%
11a	100%
11b	80%
12a	100%
12b	80%
13a	80%
13b	**50%**
14a	80%
14b	80%
15a	80%
15b	**40%**
16a	90%

16b	80%
17a	80%
17b	90%
18a	80%
18b	100%

Critical Sentence Accuracy Ratings - Adults

Item	Percentage of time item was rated correctly
1a	80%
1b	80%
2a	90%
2b	90%
3a	80%
3b	90%
4a	100%
4b	80%
5a	100%
5b	90%
6a	80%
6b	90%
7a	80%
7b	80%
8a	90%
8b	80%
9a	100%
9b	100%
10a	90%
10b	100%
11a	100%
11b	90%
12a	90%
12b	90%
13a	80%
13b	**80%**
14a	90%
14b	80%
15a	100%
15b	**90%**

16a	90%
16b	100%
17a	80%
17b	80%
18a	90%
18b	100%

Appendix 3.5 – Table of Picture Accuracy Percentages for each Item

Pictorial Items Correctly identified (%) by children

Item	Target Picture	Filler 1	Filler 2	Filler 3
1a	100%	100%	100%	100%
1b	100%	100%	100%	100%
2a	100%	100%	100%	80%
2b	100%	80%	100%	100%
3a	100%	100%	100%	100%
3b	80%	100%	100%	100%
4a	**40%**	100%	100%	100%
4b	100%	100%	100%	100%
5a	100%	100%	100%	100%
5b	100%	100%	100%	100%
6a	100%	100%	100%	100%
6b	80%	80%	100%	80%
7a	80%	100%	80%	100%
7b	100%	100%	100%	100%
8a	100%	100%	100%	80%
8b	100%	100%	100%	100%
9a	100%	100%	100%	100%
9b	100%	100%	100%	100%
10a	100%	100%	80%	100%
10b	**60%**	**60%**	**20%**	**0%**
11a	100%	100%	100%	100%
11b	100%	100%	100%	100%
12a	80%	100%	100%	80%
12b	100%	80%	100%	80%
13a	100%	100%	100%	100%
13b	100%	100%	100%	100%
14a	**20%**	80%	**0%**	**60%**
14b	80%	100%	80%	100%
15a	100%	100%	100%	100%
15b	100%	100%	100%	100%
16a	100%	100%	100%	80%
16b	100%	100%	100%	100%
17a	100%	100%	100%	100%
17b	100%	100%	100%	100%
18a	100%	100%	80%	80%
18b	100%	100%	100%	100%

Pictorial Items Correctly identified (%) by adults

Item	Target Picture	Filler 1	Filler 2	Filler 3
1a	100%	100%	100%	100%
1b	100%	80%	100%	100%
2a	100%	100%	100%	100%
2b	100%	100%	100%	100%
3a	80%	80%	80%	80%
3b	80%	100%	100%	100%
4a	80%	100%	100%	100%
4b	100%	100%	100%	100%
5a	80%	100%	100%	100%
5b	100%	100%	100%	100%
6a	100%	100%	100%	80%
6b	80%	80%	80%	100%
7a	100%	100%	100%	100%
7b	100%	100%	100%	100%
8a	100%	100%	100%	80%
8b	80%	80%	100%	100%
9a	100%	100%	100%	80%
9b	100%	100%	100%	100%
10a	100%	100%	100%	100%
10b	80%	**60%**	**60%**	**20%**
11a	100%	100%	100%	100%
11b	100%	100%	100%	100%
12a	100%	100%	80%	100%
12b	100%	100%	80%	100%
13a	100%	100%	100%	100%
13b	80%	100%	100%	100%
14a	**20%**	**20%**	80%	**60%**
14b	80%	100%	100%	100%
15a	100%	100%	100%	100%
15b	100%	100%	100%	100%
16a	100%	100%	100%	80%
16b	100%	100%	100%	100%
17a	100%	100%	80%	100%
17b	100%	100%	100%	100%
18a	100%	100%	100%	80%
18b	100%	100%	100%	100%

Appendix 3.6 – Table of Mean Difficulty Ratings for Each Picture

Mean difficulty rating for each item for children

Item	Target Picture	Filler 1	Filler 2	Filler 3
1a	1.2 (0.45)	1.2 (0.45)	1.2 (0.45)	1.2 (0.45)
1b	1.2 (0.45)	1.2 (0.45)	1.4 (0.55)	1.4 (0.55)
2a	1.2 (0.5)	2 (0.82)	1.4 (0.50)	1.75 (0.96)
2b	1.6 (0.57)	1.25 (0.50)	2 (0.82)	2 (0.82)
3a	1.6 (0.55)	1.4 (0.55)	1.4 (0.55)	1.4 (0.55)
3b	1.5 (0.58)	1.4 (1.00)	1.4 (0.58)	1.2 (0.50)
4a	2 (1.41)	1.6 (0.71)	1.8 (1.41)	1.4 (0.55)
4b	1.2 (0.45)	1.2 (0.45)	1.4 (0.55)	1.4 (0.55)
5a	2 (0.71)	1.2 (0.45)	1.8 (0.45)	1.4 (0.55)
5b	1.4 (0.55)	1.4 (0.55)	1.6 (0.89)	1.6 (0.55)
6a	1.2 (0.45)	1.2 (0.45)	1.4 (0.55)	1.2 (0.45)
6b	1.5 (0.71)	1.25 (0.71)	1.4 (0.71)	1.5 (0.71)
7a	1.75 (0.5)	1.2 (0.45)	1 (0.00)	1.4 (0.55)
7b	1.4 (0.55)	1.2 (0.45)	1.4 (0.55)	1.4 (0.55)
8a	1.4 (0.55)	1.2 (0.45)	1.4 (0.55)	1.4 (0.55)
8b	1.6 (0.55)	1.6 (0.55)	2 (1.25)	2 (1.25)
9a	1.6 (0.55)	1.2 (0.45)	1.4 (0.55)	1.6 (0.55)
9b	1.6 (0.89)	1.4 (0.55)	1.4 (0.55)	1.6 (0.55)
10a	1.2 (0.50)	1.2 (0.50)	1.5 (0.58)	1.4 (0.50)
10b	1.67 (0.58)	1.67 (0.58)	2 (0.00)	1.5 (0.58)
11a	1.6 (0.55)	1.6 (0.89)	1.4 (0.55)	1.4 (0.55)
11b	1.4 (0.55)	1.4 (0.55)	1.6 (0.89)	1.6 (0.55)
12a	1.5 (0.58)	1.4 (0.71)	1.25 (0.71)	1.25 (0.71)
12b	1.2 (0.58)	1.5 (0.58)	1.8 (1.53)	1.67 (1.15)
13a	1.2 (0.45)	1.2 (0.45)	1.2 (0.45)	1.4 (0.55)
13b	1.4 (0.55)	1.2 (0.45)	1.2 (0.45)	1.4 (0.55)
14a	2 (1.26)	1.2 (0.50)	2 (0.82)	1.4 (0.50)
14b	2 (1.00)	1.5 (0.58)	1.6 (1.00)	1.75 (1.15)
15a	1.2 (0.45)	1.4 (0.55)	1.4 (0.55)	1.4 (0.55)
15b	1.4 (0.55)	1.2 (0.45)	1.6 (0.55)	1.4 (0.55)
16a	1.4 (0.58)	1.2 (0.58)	1.4 (0.58)	1 (0.00)
16b	1.6 (0.55)	1.2 (0.45)	1.4 (0.55)	1.2 (0.45)
17a	1 (0.00)	1.2 (0.50)	1.2 (0.50)	1.2 (0.50)
17b	1.6 (0.55)	1.4 (0.55)	1.6 (0.55)	1.2 (0.45)
18a	1.4 (0.50)	1.2 (0.45)	1.6 (0.50)	1.25 (0.50)
18b	1.2 (0.50)	1.2 (0.50)	1.25 (0.50)	1.25 (0.50)

Mean difficulty rating for each item for adults

Item	Target Picture	Filler 1	Filler 2	Filler 3
1a	1.00 (0.00)	1.40 (0.55)	1.20 (0.45)	1.00 (0.00)
1b	1.20 (0.45)	1.25 (0.50)	1.40 (0.55)	1.60 (0.55)
2a	1.20 (0.45)	1.80 (1.79)	1.40 (0.55)	1.60 (0.55)
2b	1.20 (0.45)	1.20 (0.45)	1.60 (0.55)	1.80 (0.84)
3a	**2.33 (1.53)**	1.75 (0.96)	1.75 (0.5)	1.75 (0.96)
3b	2.00 (1.00)	1.40 (0.55)	1.80 (0.84)	1.00 (0.00)
4a	2.00 (0.82)	1.20 (0.45)	1.60 (0.55)	1.80 (0.96)
4b	1.20 (0.45)	1.20 (0.45)	1.40 (0.55)	1.40 (0.55)
5a	1.50 (0.58)	1.00 (0.00)	2.00 (1.23)	1.20 (0.45)
5b	1.40 (0.55)	1.60 (0.89)	1.60 (0.89)	1.80 (0.84)
6a	1.20 (0.45)	1.00 (0.00)	1.40 (0.55)	1.00 (0.00)
6b	1.50 (0.58)	2.00 (1.41)	1.50 (0.58)	2.00 (1.25)
7a	1.60 (0.50)	1.60 (0.55)	1.40 (0.58)	1.80 (0.84)
7b	1.60 (0.55)	1.40 (0.55)	1.60 (0.55)	1.20 (0.45)
8a	1.20 (0.45)	1.60 (0.89)	1.40 (0.55)	2.00 (0.82)
8b	1.25 (0.50)	1.75 (0.96)	2.00 (1.25)	1.80 (0.84)
9a	2.00 (0.71)	1.80 (0.84)	1.40 (0.55)	1.20 (0.45)
9b	1.20 (0.45)	1.60 (0.55)	1.40 (0.55)	1.60 (0.89)
10a	1.40 (0.55)	1.00 (0.00)	1.60 (0.55)	1.20 (0.45)
10b	1.75 (1.16)	1.67 (1.16)	1.33 (0.00)	2.00 (1.00)
11a	1.40 (0.55)	1.80 (0.84)	1.40 (0.55)	1.40 (0.55)
11b	1.40 (0.55)	1.80 (1.30)	1.80 (0.84)	1.80 (0.84)
12a	1.20 (0.45)	1.80 (1.30)	1.50 (0.58)	1.80 (0.45)
12b	2.00 (1.73)	1.40 (1.52)	1.40 (0.55)	1.40 (0.55)
13a	1.20 (0.45)	1.20 (0.45)	1.20 (0.45)	1.20 (0.45)
13b	1.75 (0.50)	1.60 (0.89)	1.80 (0.84)	1.40 (0.55)
14a	1.67 (0.58)	1.50 (2.12)	1.75 (0.96)	1.75 (0.96)
14b	**2.20 (1.30)**	1.60 (0.89)	1.60 (0.55)	1.20 (0.45)
15a	1.20 (0.45)	1.00 (0.00)	1.40 (0.55)	1.60 (0.55)
15b	2.00 (1.23)	1.20 (0.45)	1.20 (0.45)	1.60 (0.89)
16a	1.60 (0.55)	1.80 (1.10)	1.80 (0.84)	2.00 (1.41)
16b	1.60 (0.55)	1.60 (0.89)	1.20 (0.45)	1.20 (0.45)
17a	1.00 (0.00)	1.40 (0.55)	1.50 (0.58)	1.40 (0.55)
17b	1.60 (0.55)	1.60 (0.89)	1.60 (0.89)	1.20 (0.45)
18a	1.00 (0.00)	1.20 (0.45)	1.60 (0.55)	1.25 (0.50)
18b	1.80 (0.84)	1.00 (0.00)	1.20 (0.45)	1.20 (0.45)

Appendix 3.7 – Table of Mean Plausibility Ratings for Each Picture

Mean plausibility rating for each item - Children

Item	Target Picture
1a	1 (0)
1b	2 (0)
2a	1 (0)
2b	1.6 (0.55)
3a	1.8 (0.45)
3b	1.67 (0.58)
4a	1.4 (0.55)
4b	1.6 (0.55)
5a	1.8 (0.45)
5b	1.2 (0.45)
6a	1.2 (0.45)
6b	1.33 (0.58)
7a	3.75 (0.50)
7b	3.6 (0.89)
8a	**2.8 (0.45)**
8b	3 (0.71)
9a	3 (0.71)
9b	3.4 (0.55)
10a	3.6 (0.55)
10b	3.4 (0.55)
11a	**2.8 (1.10)**
11b	3.4 (0.55)
12a	3.4 (0.89)
12b	3 (0.71)
13a	3.8 (0.45)
13b	1.25 (0.43)
14a	3.6 (0.55)
14b	1.25 (0.50)
15a	3.4 (0.55)
15b	1 (0.00)
16a	**2.8 (0.45)**
16b	1.8 (0.45)
17a	3.8 (0.45)
17b	1.4 (0.55)
18a	**2.8 (0.45)**
18b	2 (0.00)

Mean plausibility rating for each item - Adults

Item	Target Picture
1a	1.20 (0.00)
1b	1.80 (0.00)
2a	1.20 (0.00)
2b	1.80 (0.84)
3a	1.67 (0.58)
3b	1.25 (0.50)
4a	1.50 (1.00)
4b	1.20 (0.45)
5a	1.25 (0.50)
5b	1.20 (0.45)
6a	1.00 (0.00)
6b	2.00 (1.16)
7a	3.80 (0.45)
7b	3.20 (1.10)
8a	3.40 (0.89)
8b	3.75 (0.50)
9a	3.40 (0.89)
9b	3.80 (0.45)
10a	3.60 (0.55)
10b	3.00 (0.82)
11a	3.40 (0.55)
11b	3.00 (1.25)
12a	3.40 (1.34)
12b	3.40 (0.89)
13a	3.80 (0.45)
13b	1.25 (0.43)
14a	3.25 (0.96)
14b	1.00 (0.00)
15a	3.80 (0.45)
15b	1.00 (0.00)
16a	3.20 (0.84)
16b	1.40 (0.55)
17a	3.80 (0.45)
17b	1.60 (0.55)
18a	3.40 (0.55)
18b	1.20 (0.45)

Appendix 4.1 – Readability Levels for All Stories

Story	Readability Rating
Counterfactual-World	
Story 1 – Version 1	8.19
Story 1 – Version 2	8.164
Story 2 – Version 1	7.57
Story 2 – Version 2	7.815
Story 3 – Version 1	6.98
Story 3 – Version 2	8.513
Story 4 – Version 1	8.435
Story 4 – Version 2	7.919
Story 5 – Version 1	8.732
Story 5 – Version 2	8.216
Story 6 – Version 1	7.893
Story 6 – Version 2	7.57
Real-World	
Story 7 – Version 1	7.57
Story 7 – Version 2	7.893
Story 8 – Version 1	8.383
Story 8 – Version 2	7.596
Story 9 – Version 1	7.028
Story 9 – Version 2	7.325
Story 10 – Version 1	7.351
Story 10 – Version 2	7.839
Story 11 – Version 1	7.325
Story 11 – Version 2	7.325
Story 12 – Version 1	7.893
Story 12 – Version 2	7.054
Literal	
Story 13 – Version 1	8.164
Story 13 – Version 2	8.138
Story 14 – Version 1	8.435
Story 14 – Version 2	8.435
Story 15 – Version 1	7.596
Story 15 – Version 2	7.622
Story 16 – Version 1	8.164
Story 16 – Version 2	7.841
Story 17 – Version 1	7.325
Story 17 – Version 2	6.434
Story 18 – Version 1	6.976
Story 18 – Version 2	7.596

Appendix 4.2 – Interactions for IST Reading Speed (Stage 1)

Plausibility * Age: F (1, 28) = .56, p = .46, ηp^2 = .02

Inference * Age: F (2, 56) = .32, p = .73, ηp^2 = .02

Plausibility * Inference Type: F (2, 56) = .56, p = .57, ηp^2 = .02

Plausibility * Inference Type * Age: F (2, 56) = 1.24, p = .30, ηp^2 = .05

Appendix 4.3 – Interactions for Accuracy (Stage 1)

Plausibility * Age: F (1, 56) =3.04, p = .09, ηp^2 = .05

Plausibility * Task: F (1, 56) = 0.02, p = .88, ηp^2 < .001

Plausibility * Age * Task: F (1, 56) = 0.30, p = .59, ηp^2 = .01

Inference Type * Age: F (2, 112) = 0.76, p = .47, ηp^2 = .01

Inference Type * Task: F (2, 112) = 0.56, p = .57, ηp^2 = .01

Inference Type * Age * Task: F (2, 112) = 0.41, p = .66, ηp^2 = .01

Plausibility * Inference Type: F (2, 112) = 0.87, p = .42, ηp^2 = .02

Plausibility * Inference Type * Age: F (2, 112) = 2.36, p = .10, ηp^2 = .04

Plausibility * Inference Type * Task: F (2, 112) = 0.03, p = .97, ηp^2 < .001

Plausibility * Inference Type * Age * Task: F (2, 112) = 0.04, p = .96, ηp^2 = .001

Age * Task: F (1, 56) = 0.41, p = .53, ηp^2 = .01

Appendix 4.4 – Final Stories, Critical Texts and Pictorial Stimuli Used

Counterfactual-World Stories

There are six counterfactual-world stories, these stories contain one counterfactual-world causal inference evoking critical sentence and one counterfactual-world elaborative inference evoking critical sentence. Causal evoking inference items can be identified by a '(c)' at the end of the sentence and elaborative evoking inference items can be identified be an '(e).'

Lily and Joe's spaceship landed at the bottom of a cliff. Suddenly the ground shook.

The weight of the spaceship had made the cliff fall down. Huge rocks were crashing down all around Lily and Joe.

Eventually, the giant rocks stopped falling. Lily and Joe climbed out of their ship. They looked around.

A massive egg had fallen to the ground with the rocks! Slowly Lily and Joe made their way over to the egg.

The giant egg began to crack loudly. Soon, instead of the egg there was a small bear cub. (c) *RED*

The tiny bear cub had hatched from the egg
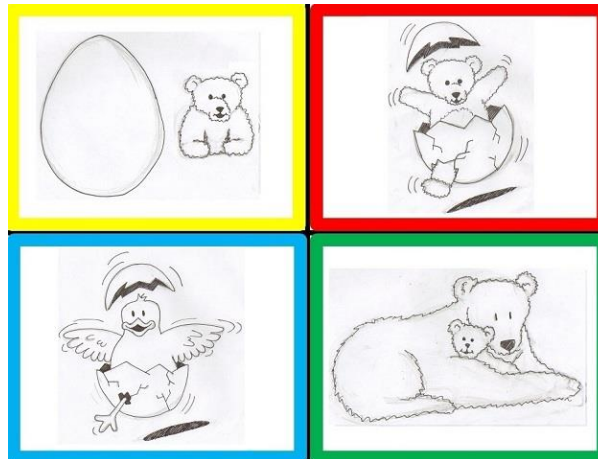
Lily ran over to the tiny bear cub. He was very upset.

"He must have fallen from the cliff. We have to take him back to his mum," said Joe. So Lilly and Joe started to climb the cliff.

The cliff was very steep. Climbing it was hard work. To make it worse it was quite a cold dark day.

When the sun came out Lily and Joe noticed something weird.

The tiny bear cub was quite odd. He was not covered in fur. He was covered in cheese. (e) *RED*

The bear cub was a yellowish colour
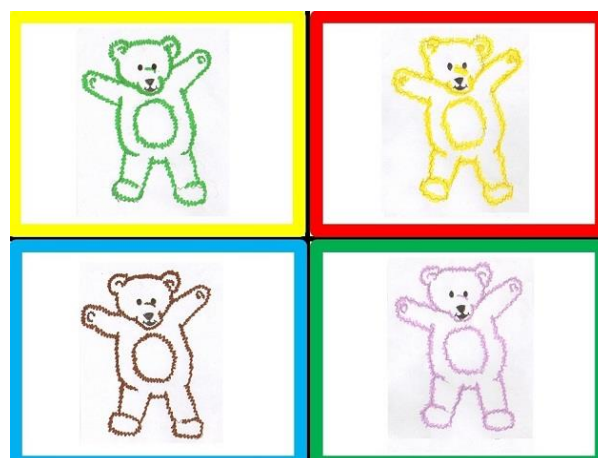
After a few hours Lily and Joe came to a huge nest. They put bear cub into the nest. Lily and Joe then rushed back to their spaceship.

Item 1: Version 2

Lily and Joe's spaceship landed at the bottom of a cliff. Suddenly the ground shook.

The weight of the spaceship had made the cliff fall down. Huge rocks were crashing down all around them. Eventually, the rocks stopped falling.

Lily and Joe climbed out of their spaceship. They looked around.

They spotted a small rock that was still shaking. There was a tiny animal hiding under the rock. It was a bear cub.

The tiny bear cub was quite odd. He was not covered in fur. He was covered in cheese. (c) *RED*
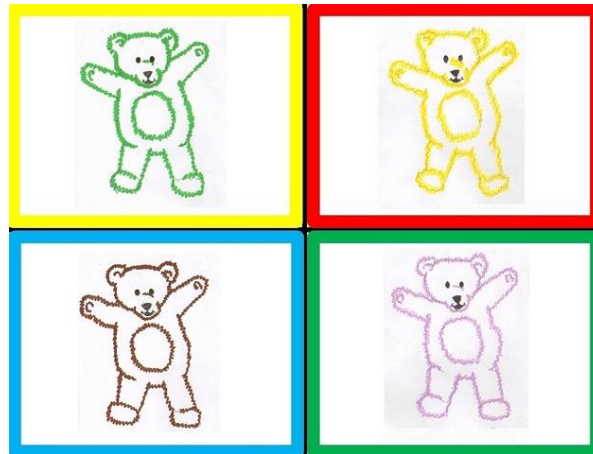
Lily ran over to the tiny bear cub. He was very upset.

"He must have fallen from the cliff. We have to take him back to his mum," said Joe. So Lily and Joe set off up the cliff.

The cliff was very steep. Climbing it was hard work.

After a few hours Lily and Joe came to a massive nest. There was a big green egg in the nest.

The giant egg began to crack loudly. Soon, instead of the egg there was a small bear cub. (e) *RED*
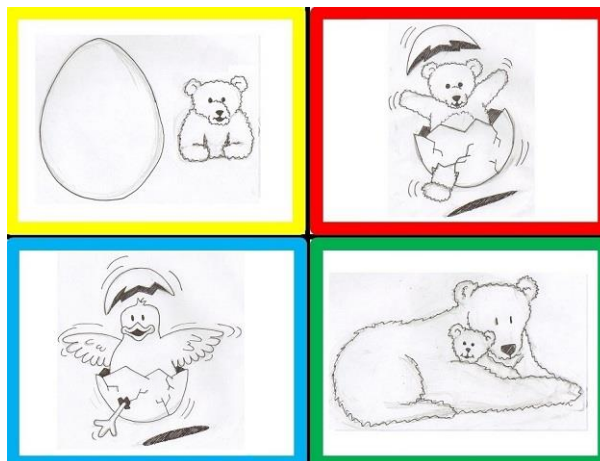
"This must be where our cub came from!" said Lily. They put their bear cub in to the big nest. Lily and Joe then rushed back to their spaceship.

Item 1: Comprehension Questions

The spaceship had caused the cliff to fall down **(ex)**

Lily and Joe were very tired after climbing the cliff **(inf)**

Item 2: Version 1

Lily and Joe landed on a planet covered in trees.

They began to explore the planet, but as they went deeper into the forest it began to get dark and scary. "Let's go back," begged Lily. "We can't, we're lost!" said Joe sadly.

Lily saw a stream. They could swim down it and out of the forest.

"This way!" Lily shouted. Lily was very thirsty. She bent down to get a quick drink.

The fluid in the stream was brown and steaming. "Watch out Lily, its coffee!" said Joe. (e) RED

The large wet stream was very hot
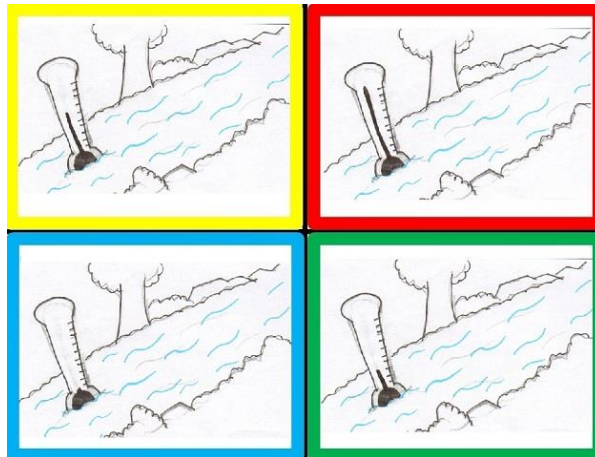
Lily and Joe would have to keep walking through the forests. However, as they went on they began to feel like they were being followed.

They heard a thud. They turned back to look. All they could see was trees.

THUD!!! They looked back again. There was still no one in sight, just trees. "This is very weird," said Lily.

Then they heard a huge clunk. They turned round and saw a large tree.

The tree was ten feet away. They turned around again. The same tree was now two feet away. (c) YELLOW

The large tree was moving closer to them
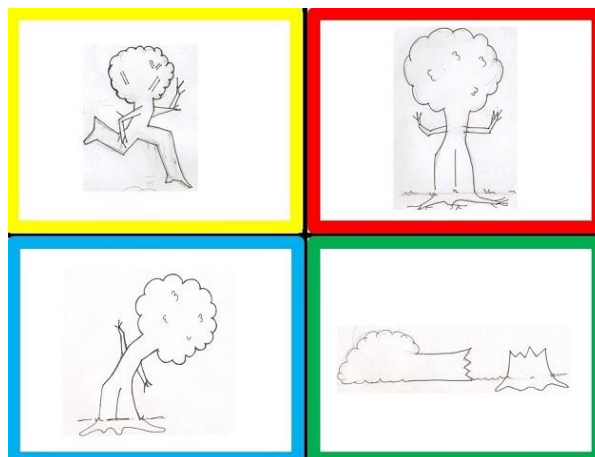
Lily and Joe were lost and now the trees were chasing them. Suddenly they heard a whizzing sound. Crashing through the trees came their spaceship! They were saved!

337

Item 2: Version 2

Lily and Joe landed on a planet covered in trees. They began to explore. They went deeper and deeper into the forest.

However, Lily and Joe soon began to feel like they were being followed.

They heard a thud. They span back to look. All they could see was trees. THUD!!! THUD!!! They looked back again. There was still no one in sight, just trees.

CLUNK!!! They turned around and saw a large tree.

The tree was ten feet away. They turned around again. The same tree was now two feet away. (c) YELLOW
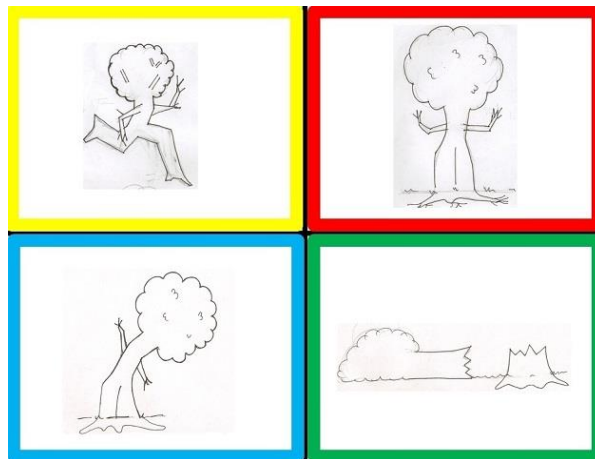
The tree was following them!!! In fact the tree was chasing them!

"Let's go back," begged Lily. "I wish we could, but we're lost!" said Joe sadly.

Lily spotted a large stream. They were saved. They could swim down it and out of the woods.

"Quick, this way!" she shouted. Lily ran over to the water as fast as she could. Lily was very thirsty. She bent down to get a quick drink.

The fluid in the stream was brown and steaming. "Watch out Lily, its coffee!" said Joe. (e) RED
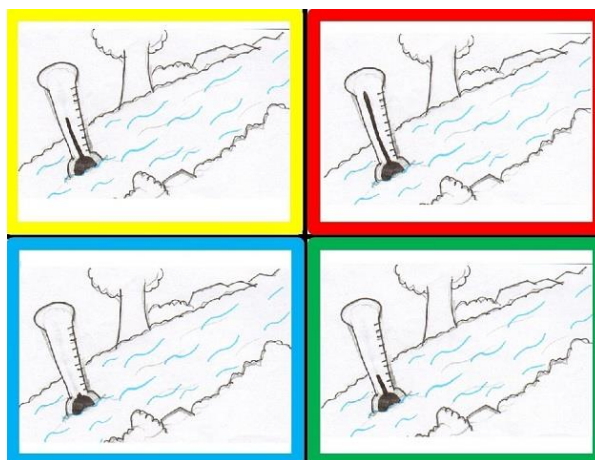
Lily and Joe were stuck. There was no way out. Suddenly they heard a whizzing sound. Crashing through the trees came their spaceship! They were saved!

Item 2: Comprehension Questions

Lily and Joe were lost **(ex)**

The woods scared Lily and Joe **(inf)**

Item 3: Version 1

Joe and Lily had just landed. They had just climbed out of their spaceship when Kaprunk! Kaprang! BANG! The spaceship vanished!!!

A few seconds later it reappeared miles away. Luckily a kind man was passing. He agreed to guide Lily and Joe to their spaceship.

They reached the spaceship quickly. But there was a problem.

The ship was on the other side of a deep hole. The hole was too wide to jump over. The man was not worried.

Arms spread wide he *jumped off the ledge*. A few *seconds* later he was on the other side. (c) *BLUE*

The kind man had flown across the large hole
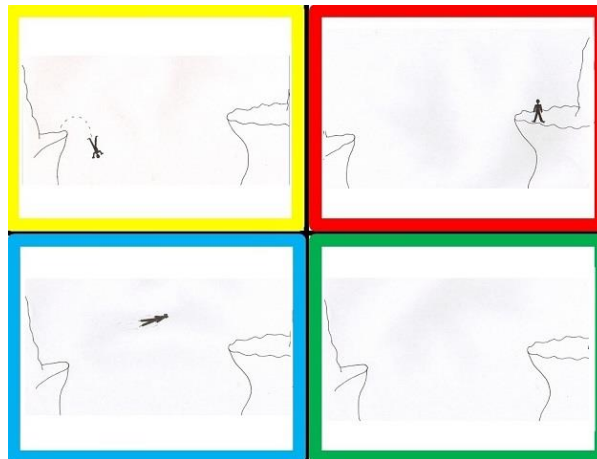
"Hurry up you two. Come on over" called the man.

"Could you just fly the ship back over to us?" asked Lily. He did.

They spent the rest of the day with the man. He told them fantastic story after fantastic story. He even told them what it is like to fly!

The day sped by. When it was time to leave Lily and Joe shook the man's hand. The man's hands were very odd.

They were shaped like human hands. But, they were like a large bird's wing in every other way. (e) *RED*

The man's hand was covered in feathers.

340
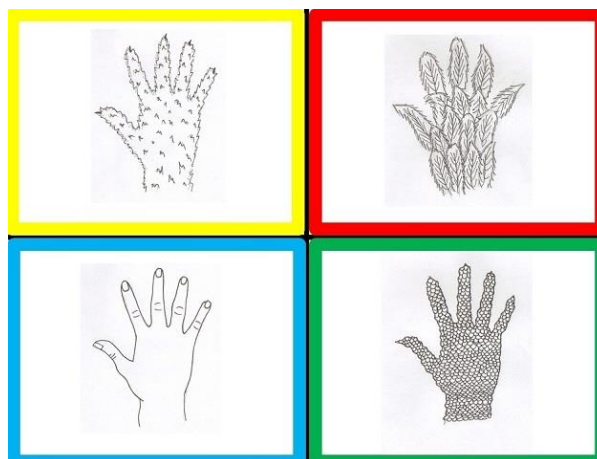
Lily and Joe thanked the kind man for his help. They then flew off on a new adventure.

Item 3: Version 2

Joe and Lily had just landed on an extremely rocky planet.

They had just climbed out of their spaceship when Kaprunk! Kaprang! BANG! The spaceship vanished!!! A few seconds later it reappeared many miles away.

The path to the ship didn't look easy though. Luckily a kind man was passing by. He agreed to guide Lily and Joe to their ship.

The man was just about the same as the men on Earth. However, his hands were very different.

They were shaped like human hands. But, they were like a large bird's wing in every other way. (e)
RED
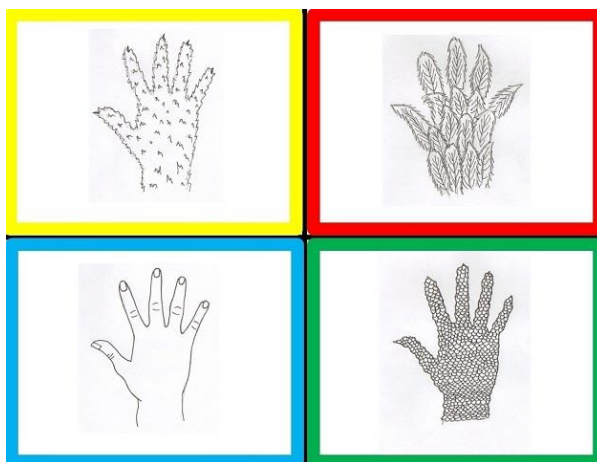
The kind man was very funny. He told Lily and Joe fantastic story after story. He had led a very good life.

The long journey to the ship flew by. In no time they were very close to the ship. But there was a big problem. A very big problem.

The ship was on the other side of a deep hole. The hole was far too wide to jump over.

The man was not worried.

Arms spread wide he jumped off the ledge. A few seconds later he was on the other side. **(c)** *BLUE*
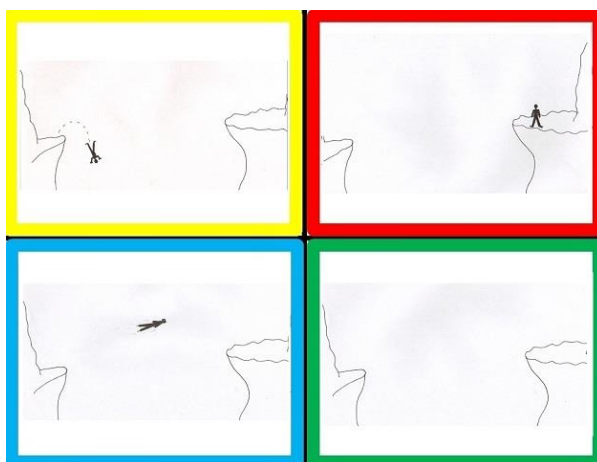
Lily and Joe thanked the kind man for his help. They then flew off on a new adventure.

Item 3: Comprehension Questions

The man told Lily and Joe stories **(ex)**

Lily and Joe enjoyed their time on the planet **(inf)**

Item 4: Version 1

Lily and Joe landed on a really sandy planet. The strong wind kept blowing sand into their faces.

Out of the blue an old piece of paper was blown into Joe's face.

Eventually the wind dropped. Joe looked at the piece of paper that had hit him. "It's a treasure map!" he shouted out.

First, the map took them into a forest. The forest was unlike any forest they had ever seen.

The trees were odd. The leaves were covered in tiny rubies that glinted in the sunlight. (e) BLUE

The leaves on the trees were red in colour
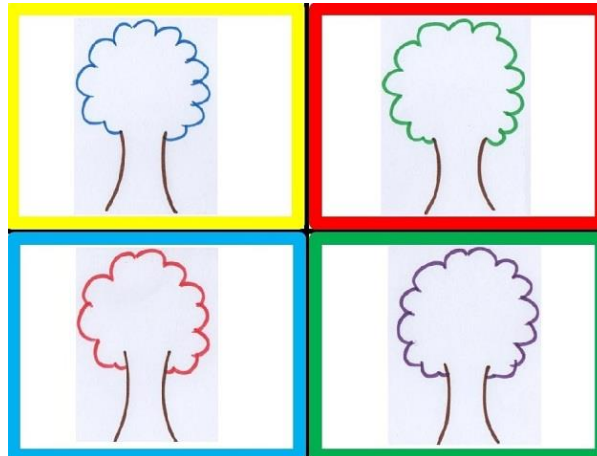
Once out of the woods Lily and Joe came to a lake. On the other side of the lake was a chest.

The lake was huge. The chest was too far away to swim to. Plus, there was no boat. There was no way for Lily and Joe to reach the chest.

They were just about to give up. But then Joe spotted a note on the map. 'Cross the lake like a jesus lizard.'

"Come with me," said Joe.

Lily and Joe slowly stepped onto the murky lake. They crossed the lake within no time. *(c) GREEN*

Lily and Joe had walked on the water
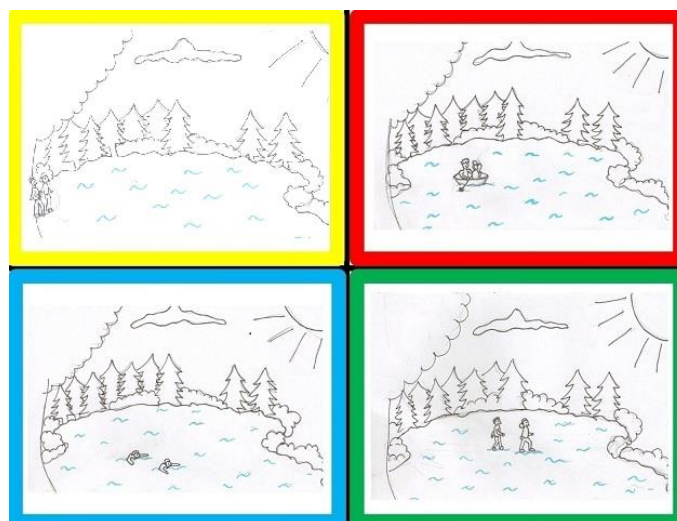
They reached the treasure chest. Shockingly, it was full of eggs. Lily and Joe laughed all the way back to the spaceship.

Item 4: Version 2

Lily and Joe landed on a windy planet. An old piece of paper blew into Joe's face. "It's a treasure map!" Joe cried out.

First the map took them to a lake. However, the lake was too large to swim across. Plus, there was no boat.

They were about to give up. But then Joe saw a note on the map. 'Cross the lake like a jesus lizard.'

"Come with me," said Joe.

Lily and Joe slowly stepped onto the murky lake. They crossed the lake within no time. (c) *GREEN*
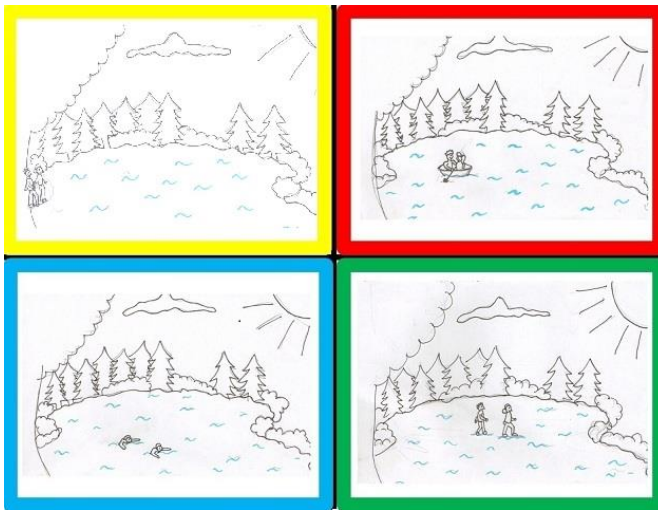
The old map then took Lily and Joe all over. They went over huge cliffs and harsh dry sands.

It was very hard work. They felt like they had been walking for days and days.

However, Lily and Joe knew it would be worth it. The chest would be full of huge gems and money.

Eventually, the map took them into a huge forest. It was not like any they had ever seen.

***The trees were odd. The leaves were covered in tiny rubies that glinted in the sunlight*. (e) BLUE**
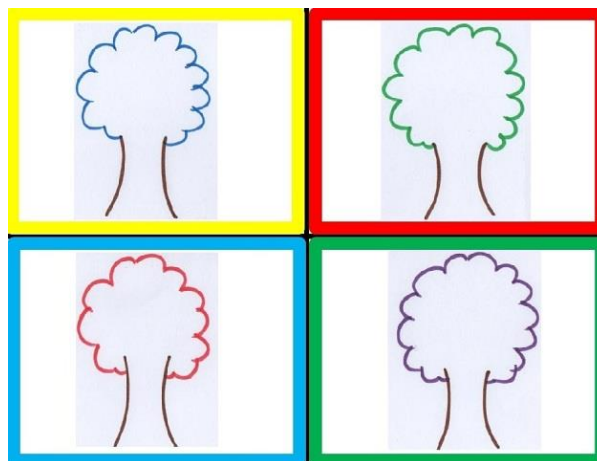
In the middle of the forest was the treasure chest. It was full of eggs! Lily and Joe laughed all the way back to the spaceship.

Item 4: Comprehension Questions

An old newspaper blew into Joe's face **(ex)**

Lily and Joe followed the map for hours **(inf)**

Item 5: Version 1

Lily and Joe landed on a planet full of huge mountains and big fluffy dogs.

One of the dogs ran up to them. Wag wag wag, went his big fluffy tail.

Lily and Joe spent most of the morning playing with the dog. They called him Chip. But, their fun soon came to a halt.

THUMP! THUMP! THUMP! The floor began to shake. Even the tall trees shook.

Lily and Joe looked up. The giant grey *leaves* above their heads were made of solid stone. (e) *YELLOW*

The giant grey leaves were very heavy
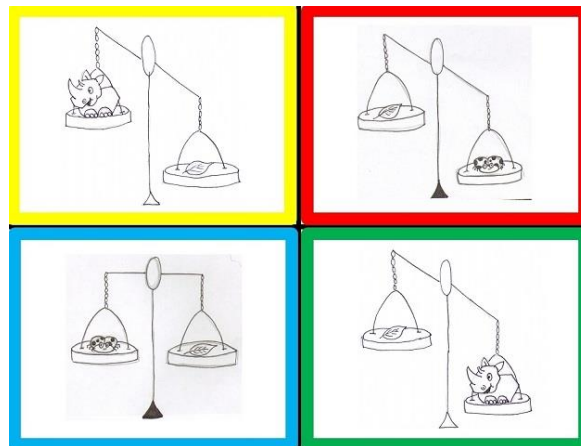
THUMP! THUMP! The noise began to get closer. There was a massive roar. Suddenly a tiger jumped out of the trees.

The tiger was huge. It had giant white teeth and huge claws. Lily, Joe and Chip ran as fast as they could.

Sadly, they had no idea where they were going. They ran straight into a dead end.

There was no way out. They were stuck!

The hungry tiger pounced. Almost instantly Joe's face was covered in warm sticky spit. (c) *GREEN*

The large hungry tiger had licked Joe's face
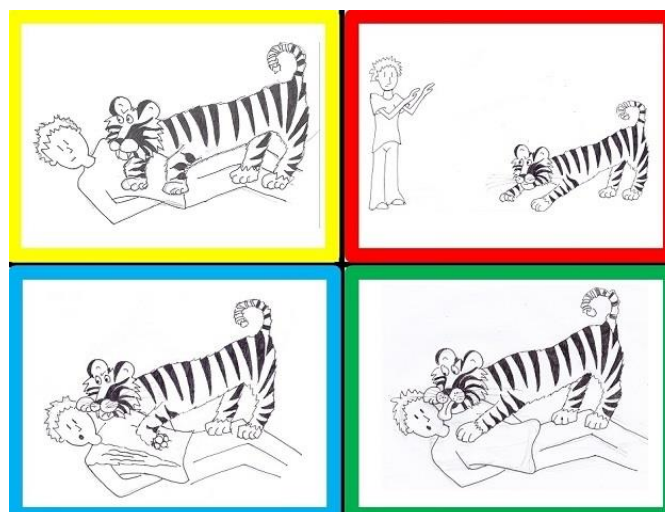
The tiger was not dangerous. He just wanted to play! Lily, Joe and Chip played with the lonely tiger all day. Lily and Joe were sad to leave the planet and their new friends.

Item 5: Version 2

Lilly and Joe landed on a planet full of huge mountains and big fluffy dogs.

One of the dogs ran up to them. Wag, wag, wag, went his big fluffy tail.

Lilly and Joe spent most of the morning playing with the dog. They called him Chip. However, their fun soon came to a halt.

There was a giant roar. Suddenly, a tiger leapt out of the trees.

The hungry tiger pounced. Almost instantly Joe's face was covered in warm sticky spit. (c) *GREEN*
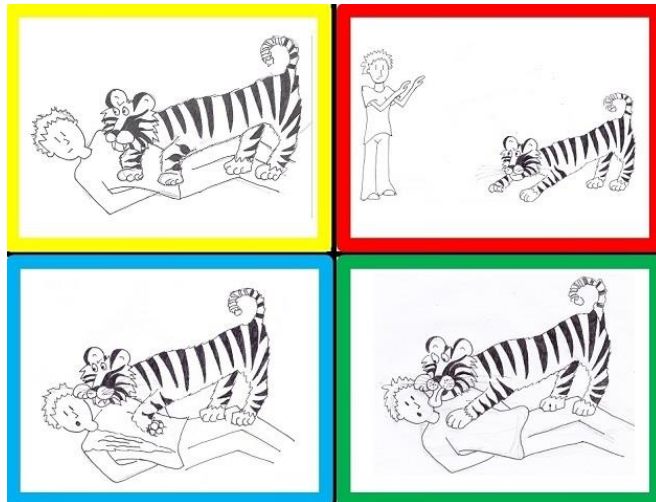
The giant tiger was not scary after all. Lilly, Joe and Chip played with the tiger all day. At first they chose to play hide and seek.

Sadly, the tiger was awful at hide and seek. He was far too big and heavy.

Every time he made a move the whole place shook.

THUMP! THUMP! THUMP! THUMP! The floor began to shake. Even the tall trees shook.

Lily and Joe looked up. The giant grey *leaves* above their heads were made of solid stone. (e) *YELLOW*
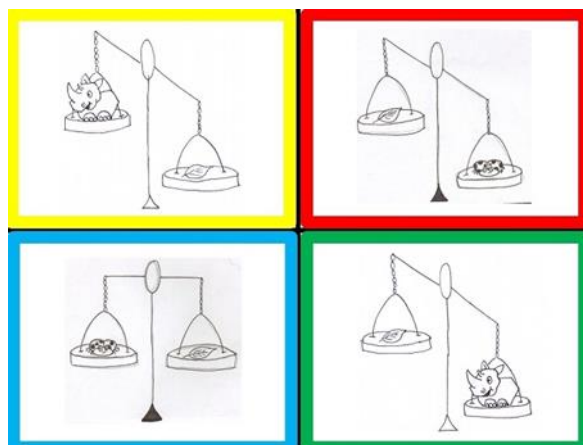
Lilly, Joe and Chip decided that hide and seek was too dangerous! They played chess with the clever tiger instead. Lily and Joe were sad to leave the fun planet and their new friends.

347

Item 5: Comprehension Questions

Chip was a small cat **(ex)**

The tiger wanted to eat Lily and Joe **(inf)**

Item 6: Version 1

Lily and Joe landed on a planet covered in fish, milk, and grass. "Wow! This planet would be perfect for cats," said Lily.

After a while Lily and Joe came to a huge brick wall with lots of guards in front of it.

 "Would you like to see 'Cat-World'?" asked a guard. "Sure," said Lily.

'Cat-World' was not what they had expected. It was full of lots of very sad cats.

"Hello," came a voice from below. They looked down to see a cat looking at them. (c) *BLUE*

The cat had spoken to Lily and Joe.
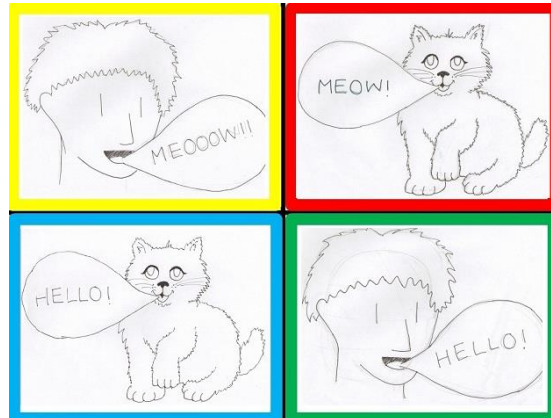
"Hi, are you OK?" asked Lily.

The sad cat shook his head. "Years ago only cats lived here. One day nasty men came. They caged all of us and put us in here."

"But there are so many trees," said Joe. "You could use them to get out."

"The trees are too soft for that," said that cat. Joe was quite puzzled. He reached out and touched one of the trees.

The trees were odd. The tree trunks were not made of wood. They were made out of huge pipe cleaners. (e) *GREEN*

The tree trunks in CatWorld were quite bendy
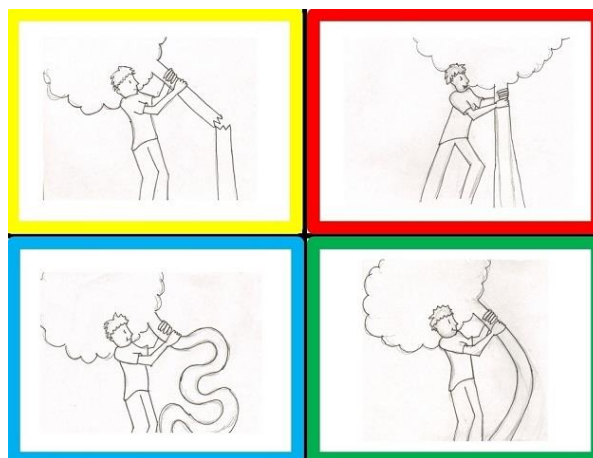
Joe had an idea. They could use the trees to make a catapult to fire the cats over the wall. One by one Joe and Lily set the cats free.

Item 6: Version 2

Lily and Joe landed on a planet covered in fish, milk, and grass. "This planet would be perfect for cats," said Lily.

After a while Lily and Joe came to a huge brick wall with lots of guards in front of it.

 "Would you like to see 'Cat-World'?" asked a guard. "Sure," said Lily.

'Cat-World' was dark and bare. There were just a few trees. Joe touched one of the trees.

The trees were odd. The tree trunks were not made of wood. They were made out of huge pipe cleaners. (e) *GREEN*
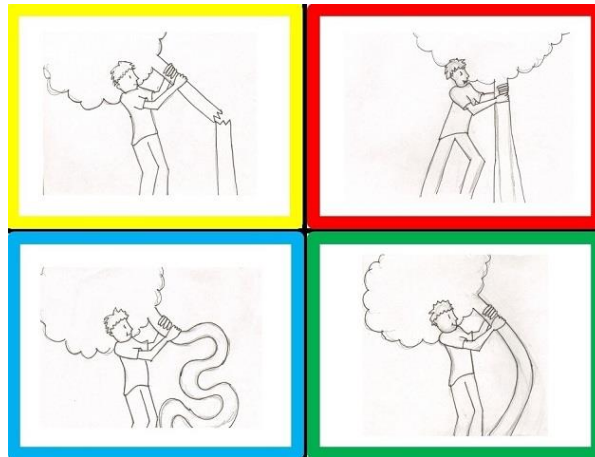
'Cat-World' was also full of sad cats.

Joe and Lily saw a sign… 'Many years ago only cats lived here. Then men came to this place. We tried to find a use for the cats. But, cats are of no use to us. So we put them in this giant cage. Then we took over.'

Lily and Joe were very upset by this.

"We have to help the cats," sobbed Lily.

"Hello," came a voice from below. They looked down to see a cat looking at them. (c) *BLUE*
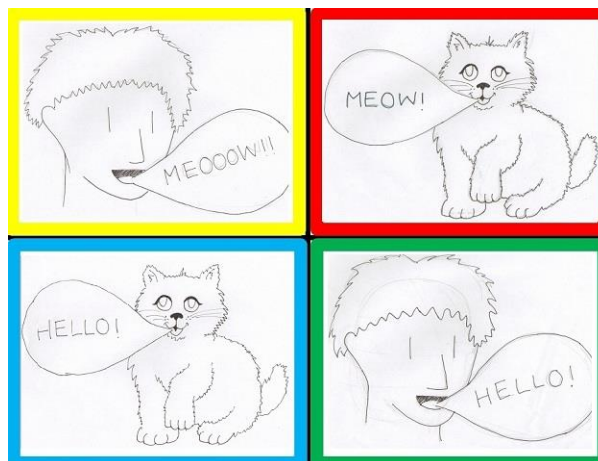
Joe had an idea. They could use the trees to make a catapult to fire the cats over the wall. One by one Joe and Lily set the cats free.

Item 6: Comprehension Questions

The cats in 'Cat-World' were happy **(ex)**

The cats wanted to escape **(inf)**

There are six real-world stories, these stories contain one real-world causal inference evoking critical sentence and one real-world elaborative inference evoking critical sentence. Causal evoking inference items can be identified by a '(c)' at the end of the sentence and elaborative evoking inference items can be identified be an '(e).'

Item 7: Version 1

Lily and Joe landed on a planet covered in big green fields.

The planet was home to lots of animals. There were ducks and pigs. There were sheep and geese. The animals did not look happy though. The animals looked scared.

Lily and Joe tried to get close to the animals but the animals ran away from them. Lily and Joe looked around. They found an eye patch on the floor.

Joe heard a long deep sound. Joe turned around to see a big cow walking towards him. (c) *BLUE*

The large cow had made the mooing sound.

The cow walked past Joe, then he looked back and mooed. "I think he wants us to follow him," Lily said.

Lily and Joe followed the cow to a large lake. In the middle of the lake was a pirate ship!

The pirates were throwing stones at any animal trying to drink from the lake. "I know what to do!" said Joe, "but I need a bag."

Luckily a snake skin floated past Joe. He took it and made a bag out of it. (e) *YELLOW*

The bag Joe made was covered in scales
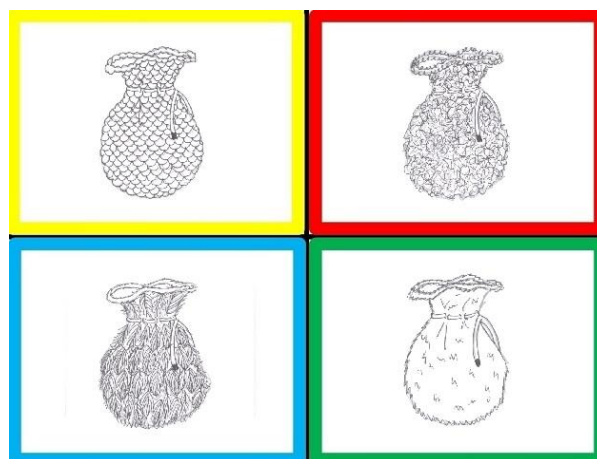
Joe set his transporter to 'pirate planet,' popped it into the bag and threw it at the pirate ship.

The pirates disappeared and the animals could drink from the lake again.

Lily and Joe landed on a planet covered in deep blue lakes. There were only a few small islands dotted here and there.

The ocean was crystal clear. It was also full of amazing fish, dolphins, and crabs.

The sea was also full of beautiful rocks and shells. "Let's take a rock home with us," said Joe.

"We can't," said Lily sadly. "Our swimming costumes don't have any pockets!"

Luckily a snake skin floated past Joe. He took it and made a bag out of it. (e) *YELLOW*
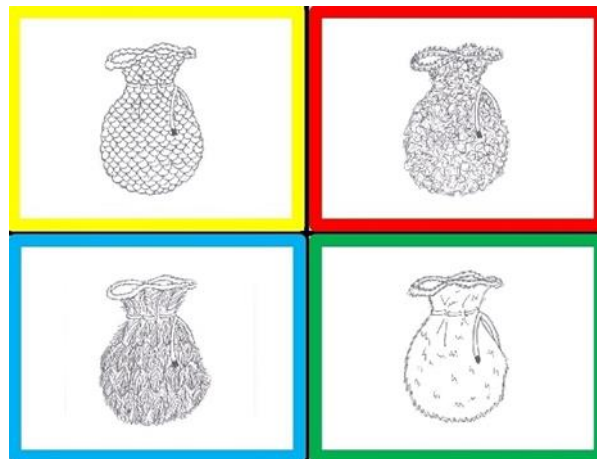
Lily put her favourite stone into the bag.

Lily and Joe hopped onto an island. The animals on the island looked scared. It did not take Lily long to work out why. In the middle of the island there was a huge pirate ship.

"We want the girl's gem!" shouted the pirates. Lily wanted to give the gem to them but she was too scared to go near them.

Joe heard a long deep sound. Joe turned around to see a cow walking towards in him. (c) *BLUE*
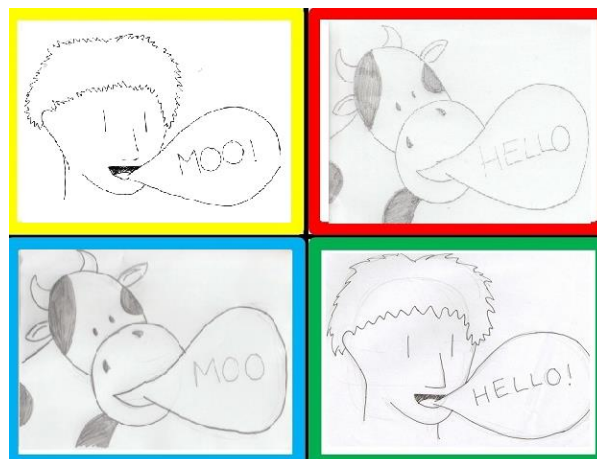
The brave cow took the gem from Lily to the pirates. There was a big bang and a whiz. Then the huge pirate ship shot into the air and disappeared.

Item 7: Comprehension Questions

The tiny black spot was actually a pirate ship **(ex)**

The whale liked Lily and Joe **(inf)**

Item 8: Version 1

Lily and Joe had flown straight into an asteroid storm. Giant rocks were crashing into the spaceship. It was going to get damaged.

Lily and Joe had to make a quick landing. They were able to land on a sandy beach next to a huge ocean.

The ocean was crystal clear. The best thing was the beach was empty.

But, there was a reason for that. The ocean was full of sharks! Lily dived into the sea.

However, a shark was approaching. It attacked. Blood began to spill out of Lily's arm.* (c) *YELLOW

The evil shark had bitten Lily's arm
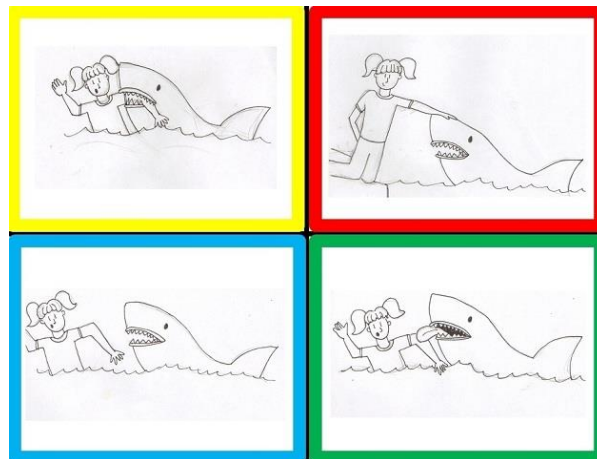
Lily cried out! Joe swam out to save her. But many more sharks began to pop up. Lily and Joe were stuck!

They began to panic. Out of the blue came a small boat.

There was a pink man in the boat. "Climb in," said the pink man. Lily and Joe were saved!

The man took Lily and Joe back to his town. In the middle of the town was a large statue.

The gigantic clay statue shone in the sun light. It must have been at least ten feet high.* (e) *BLUE

The statue was a burnt orange colour
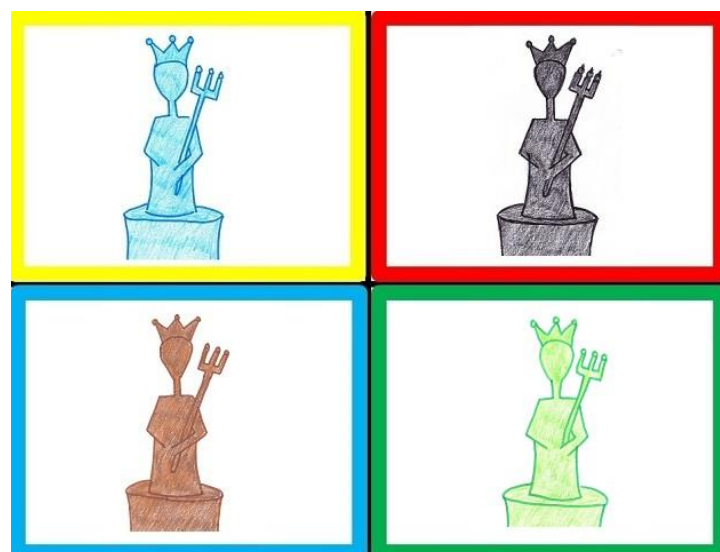
After a day in the town Lily and Joe went back to their spaceship. Joe fixed up Lily's arm and then they left the planet.

Item 8: Version 2

Lily and Joe had flown into an asteroid storm. Giant rocks were crashing into the spaceship. It was going to get damaged.

CRASH!! WHACK!!! A rock hit the wing. The wing fell off and landed in the ocean below.

The ship was out of control. Lily and Joe had to make an emergency landing.

They were able to land on a beach next to the huge ocean. There was a boat tied up next to a statue.

***The gigantic clay statue shone in the sun light. It must have been at least ten feet high.* (e)** *BLUE*
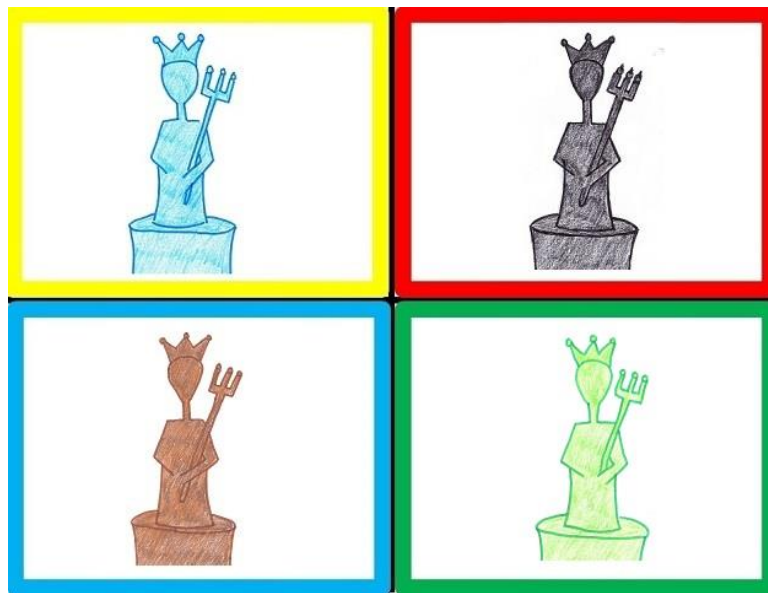
Lily and Joe could see the wing in the ocean. However, they could also see some very evil sharks.

Lily and Joe climbed into the boat. They made it to the wing in no time.

Sadly, the heavy wing had started to sink. Lily and Joe could not reach the wing from the boat.

One of them would have to jump in for it. Lily dove into the sea. She grabbed the wing.

***However, a shark was approaching. It attacked. Blood began to spill out of Lily's arm.* (c)** *YELLOW*

357
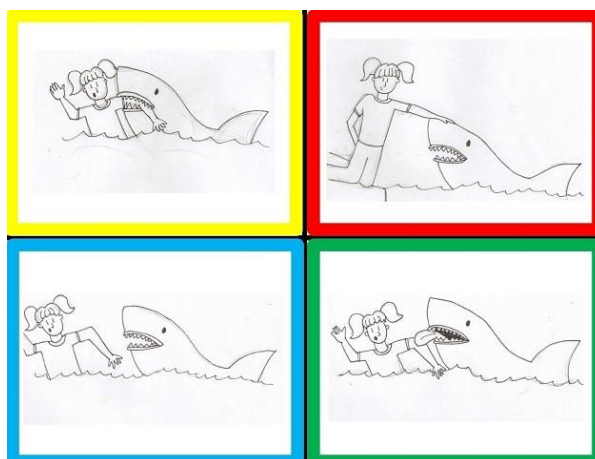
Joe pulled Lily and the wing into the boat. Joe put the wing back on the spaceship. They then flew away from the awful planet.

Item 8: Comprehension Questions

The statue was tiny **(ex)**

The sharks were hungry **(inf)**

Item 9: Version 1

Lily and Joe landed on a planet covered in ice.

The only animals around were large hippos. However, legend said a lost tribe of humans also lived on the planet.

A photo of this tribe would be worth millions. Lily and Joe decided to look for the tribe.

They came to a very thin piece of ice. Did they dare to cross it? On the other side of this thin ice was a hairy hippo.

The hippo looked at Lily and Joe. As the hippo walked its giant feet cracked the ground. (e) RED

The hippo was very, very heavy
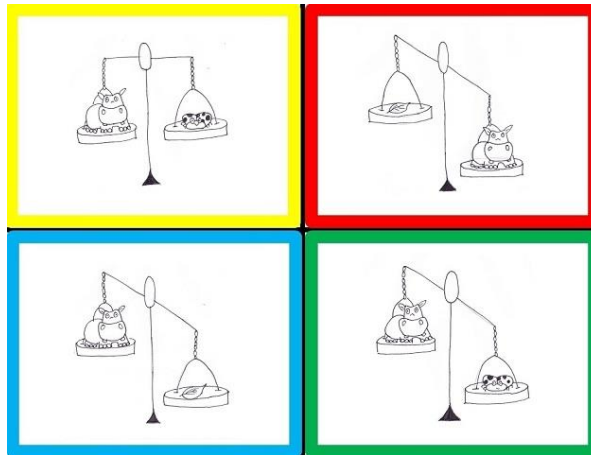
Joe and Lily saw something where the hippo had been stood. What was it? It was pink and blue. Was it a teddy bear?

"Do you think it is from the lost tribe?" asked Joe. "It must be!" cried Lily.

"We could go and take a look," said Joe. "But the ice is so thin," said Lily.

"I think we'll be able to make it," said Joe.

They cautiously crept on to the thin ice. Crack! Crack! Ice cold water swept all over them. (c) BLUE

Lily and Joe had fallen through the ice
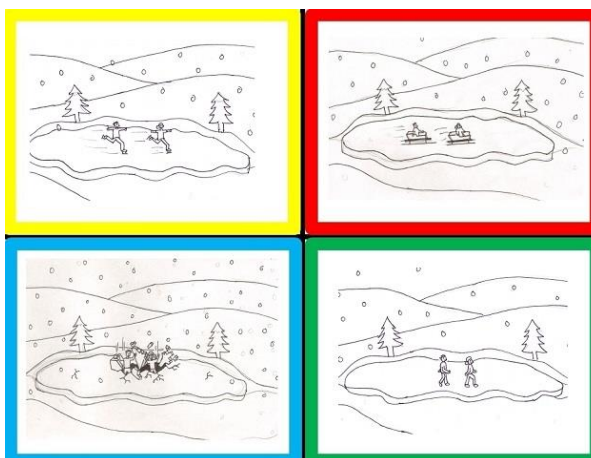
Lily and Joe began to drown. Suddenly they were pulled out of the water. They looked around. However, they never found out who pulled them out of the water.

Item 9: Version 2

Lily and Joe landed on a planet covered in ice.

The only animals around were hairy hippos. However, legend said a lost tribe of humans also lived on the planet.

A photo of this tribe would be worth millions. Lily and Joe decided to look for the tribe.

Joe and Lily saw a small pink teddy bear. It must have been from the tribe. It was on the other side of some very thin ice.

They cautiously crept on to the thin ice. Crack! Crack! Ice cold water swept all over them. (c) *BLUE*
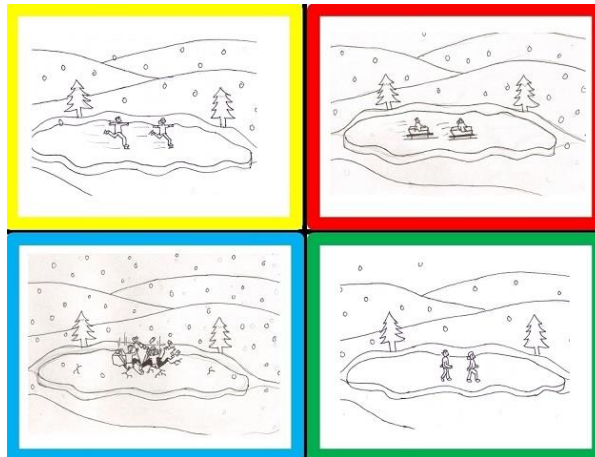
Lily and Joe began to drown. Suddenly, they were pulled out of the water. They quickly looked round. However, there was no one about.

Lily and Joe tried to find the lost tribe all day. After hours of searching they gave up.

They headed back to the ship. They went by the patch of ice they had fallen into.

 A hairy hippo was stood on the other side of the ice.

The hippo looked at Lily and Joe. As the hippo walked its giant feet cracked the ground. **(e)** *RED*
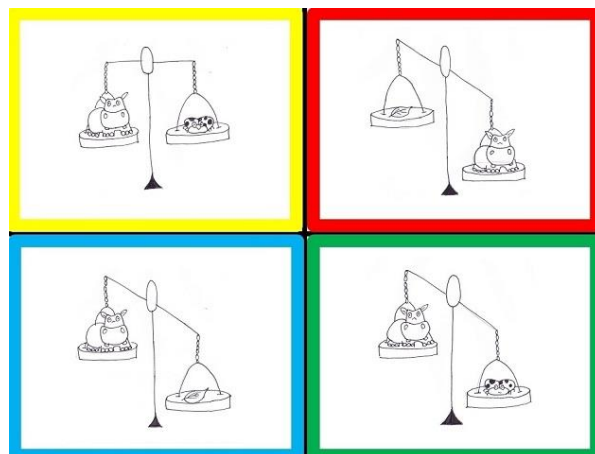
Could the hippo have saved them? Lily and Joe would never find out who saved them. They would always be very, very thankful though.

Item 9: Comprehension Questions

The hippo fell through the ice **(ex)**

Lily and Joe were glad that they had been saved from drowning **(inf)**

Item 10: Version 1

Joe and Lily landed on a lovely planet. Some friendly aliens called Blibs lived on the planet.

The Blibs were about as tall as a desk chair. They were like big shaky blobs of green jelly. They didn't have arms or legs.

The Blibs loved to grow stuff. The Blibs used their plants to make magical medicines.

The Blibs noticed that Joe looked ill.

The caring Blibs gave Joe a special medicine. It was made from mashed up bananas. (e) *GREEN*

The medicine was yellow in colour
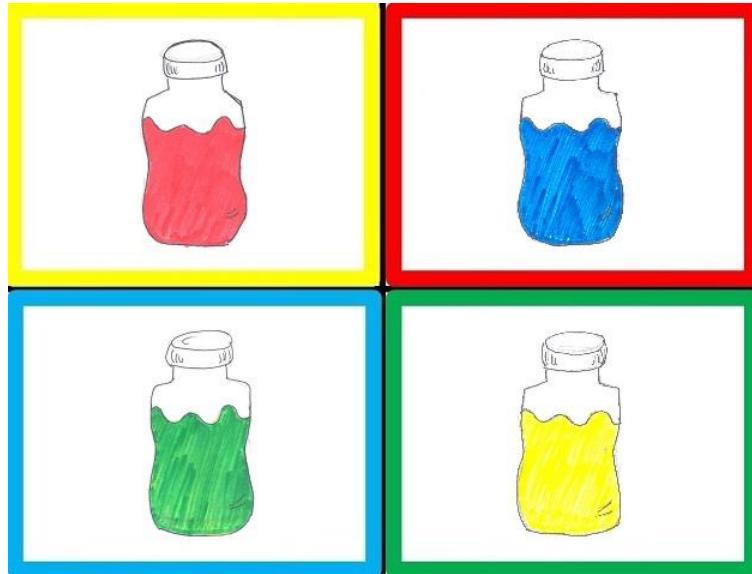
Joe was starting to feel better when whole place went pitch black. The sun was a giant light bulb and it had just blown out.

Sadly, it was too high for the Blibs to reach. An engineer put a new light bulb in every year. Sadly, he was not due for weeks.

"The plants will die!" cried the Blibs. "Calm down!" said Joe. "I will put a new bulb in for you."

Joe put in the new bulb. But the bright light dazzled him and he slipped.

Joe fell to the ground. He landed on his leg. There was a crack and Joe cried out in pain. (c) *BLUE*

One of the bones in Joe's leg had broken
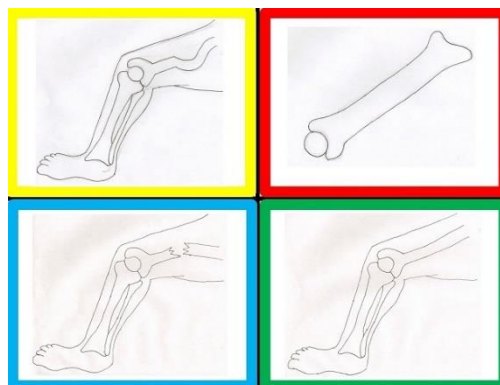
Luckily the Blibs had a special medicine to mend broken bones. So within a week or two Joe's leg was all better.

Item 10: Version 2

Joe and Lily landed on a planet in darkness. Friendly aliens called Blibs lived on the planet. Their sun was a giant light bulb and it had just blown out.

An engineer changed the bulb every year. Sadly he was not due for weeks.

The Blibs were all very upset. "Calm down!" said Joe. "I will change the bulb for you."

Joe put in the new bulb. But the light dazzled him and he fell.

Joe fell to the ground. He landed on his leg. There was a crack and Joe cried out in pain. (c) *BLUE*
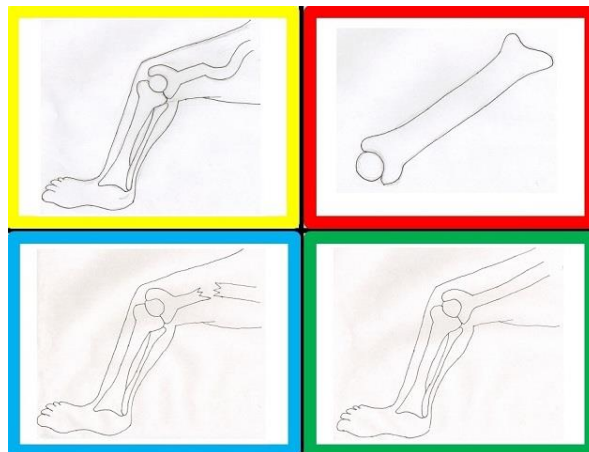
"OUCH!!! OUCH!!! OUCH!!! OUCH!!!" Joe cried out in pain. The fall had broken Joe's leg.

He was in an awful lot of pain! The pain was so bad Joe felt like he was going to pass out.

Luckily, the smart Blibs very, very good at growing things. The whole planet was covered in weird yet wonderful plants.

The Blibs used their plants to make magical medicines.

***The caring Blibs gave Joe a special medicine. It was made from mashed up bananas.* (e)** *GREEN*
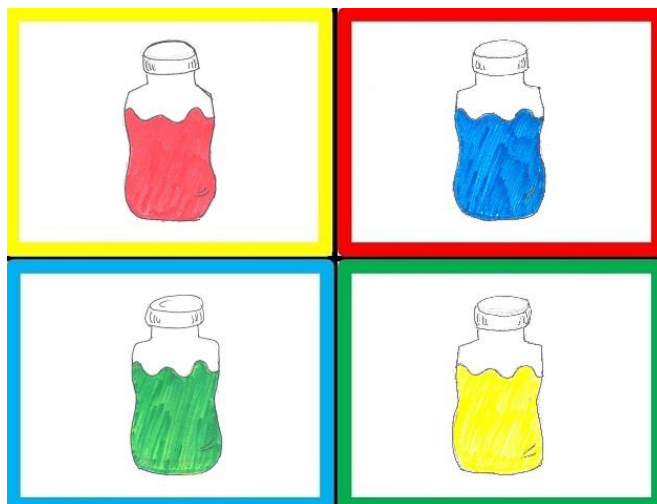
Joe took the medicine. It was really yummy. The best thing was it was a special medicine that could heal broken bones. Within no time at all Joe's leg was all better.

Item 10: Comprehension Questions

The Blibs were not good at growing plants **(ex)**

The Blibs do not like the dark **(inf)**

Item 11: Version 1

Lily and Joe landed on a very, very snowy planet. First Lily and Joe wrapped up warm.

They put on five coats and seven pairs of socks. Then they got out of their nice and warm spaceship. The planet was ice cold.

"Wow it's freezing!!" said Lily. "How could anyone live here? It's far too cold!"

It seemed as though someone did live on the ice cold planet though.

Lily and Joe spotted a little white house. The tiny round house was made out of snow. (e) *YELLOW*

The tiny white round house was very cold.
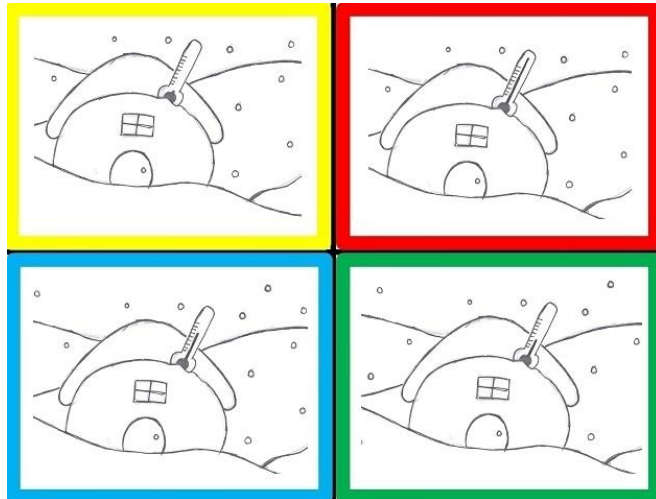
Lily and Joe ran up to the tiny house. They knocked on the snowy door.

A gruff old man answered. As soon as he saw Lily and Joe he smiled. The man told Lily and Joe that he was the only human left on the planet.

"Aren't you lonely?" asked Joe. "No, I have many friends," said the man. "Let me show you."

They all went outside. The old man whistled at a group of penguins.

One penguin jumped into the cold ocean. A few seconds later it was at their feet. (c) *YELLOW*

The penguin had swum to Lily and Joe.
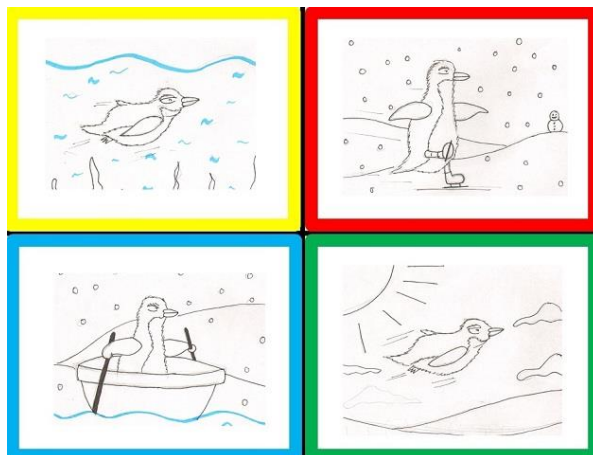
"You see the penguins are my friends," said the man happily. Lily and Joe spent the afternoon with the man and his penguins. They then flew off on another adventure.

Item 11: Version 2

Lily and Joe landed on a snowy planet. First Lily and Joe wrapped up warm. Then they got out of their spaceship.

The planet was ice cold. The first thing they saw was a family of penguins.

The penguins were on an iceberg in the sea. The penguins looked at Lily and Joe.

There were two large penguins and two baby ones. They all squawked at Lily and Joe.

One penguin jumped into the cold ocean. A few seconds later it was at their feet. (c) *YELLOW*
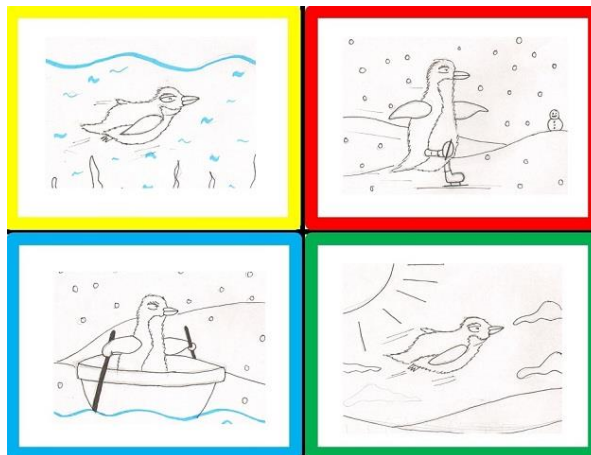
The penguins let Lily and Joe stroke them for a while. They then jumped back into the ocean and swam away.

Lily and Joe decided to explore the rest of the planet. There wasn't much to see.

Every now and then Lily and Joe came across a penguin. That was all though.

Lily and Joe were just about to head back when they saw something.

Lily and Joe spotted a little white house. The tiny round house was made out of snow. (e) *YELLOW*
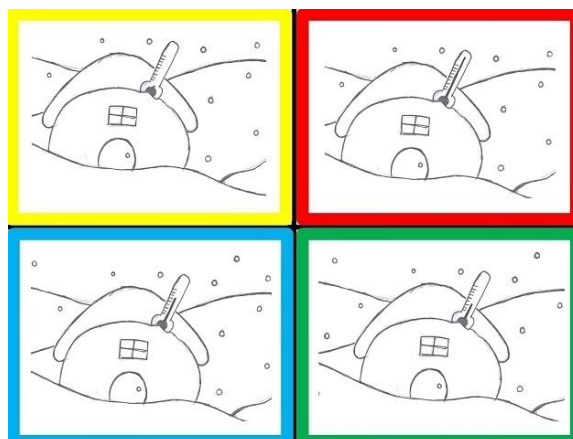
An old man was in the house. He had been stuck on the planet all alone for years! He missed his family. He asked Lily and Joe to take him back to his own planet. Lily and Joe did.

Item 11: Comprehension Questions

The planet was freezing cold **(ex)**

Lily and Joe liked the planet **(inf)**

Joe and Lily had just landed on a planet covered in concrete.

They came to a large sign which said: 'Welcome to Concretia. We are working hard. We want to cover every inch of this place in concrete.'

Lily and Joe did find some very dry soil though. "At last," said Joe. "Soil!!!"

"We need to plant some seeds" said Lily. "This place needs to come back to life."

Lily planted a tiny seed. A century later an oak tree stood in its place. (c) *RED*

The tiny seed had grown into an oak tree
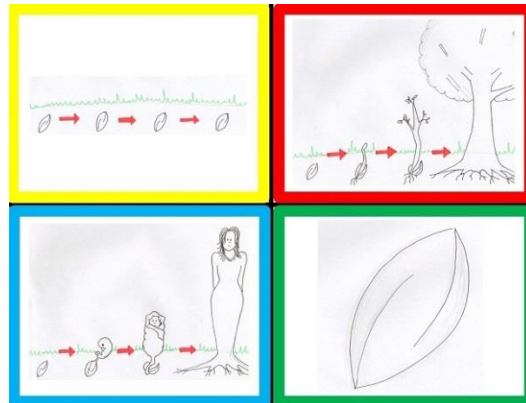
The evil robots that lived on Concretia were not happy about this though. In fact they were very, very mad at Lily and Joe.

Within seconds a large army of robots were stood in front of Lily and Joe.

"KILL THEM BOTH!!!" ordered one of the huge robots.

All of the huge metal robots began to throw big blocks of concrete at Lily and Joe. Luckily Joe spotted a large shield.

Joe grabbed the steel shield. He used it to protect him and Lily from the huge rocks. (e) *GREEN*

The shield was hard and inflexible

Lily and Joe ran back to the spaceship as fast as they could. As they flew away Lily threw some seeds out of the window. She hoped they would find a way to grow.

Item 12: Version 2

Joe and Lily had just landed on a planet covered in concrete.

"This place is awful," said Joe. "Where are all of the plants?"

Suddenly Lily and Joe were attacked by an army of huge robots.

The robots were throwing big blocks of concrete at Lily and Joe. Luckily Joe spotted a shield.

Joe grabbed the steel shield. He used it to protect him and Lily from the huge rocks. (e) GREEN
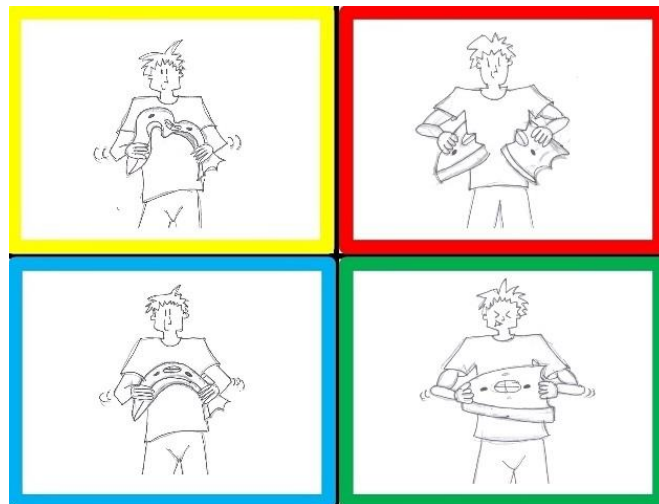
Lily and Joe ran away from the robots as fast as they could. They out ran the robots. However, Lily and Joe realised that they were lost.

Lily began to cry. "Don't cry," said a tiny mouse. "I'll take you back to your ship." Lily and Joe followed the mouse back to their ship.

"Is there anything we can do for you," asked Joe. "Some grass and plants would be nice," said the mouse. "But nothing grows in concrete."

Lily smiled. She gave the mouse a bag of seeds. "These seeds will grow, *even in concrete*."

Lily planted a tiny seed. A century later an oak tree stood in its place. (c) *RED*
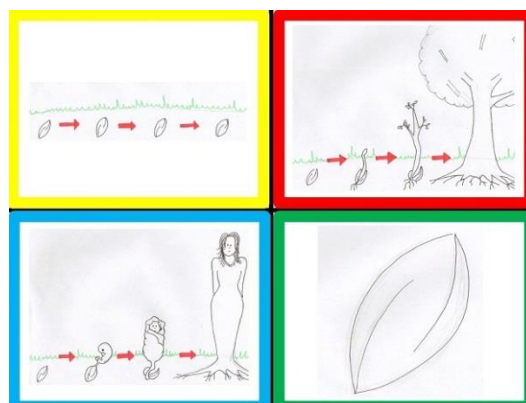
The mouse promised them that he would plant all of the seeds and bring the planet back to life. Lily and Joe then left the awful planet.

Item 12: Comprehension Questions

The planet was covered in grass **(ex)**

The robots were glad that Lily and Joe had come to their planet **(inf)**

There are six literal stories, these stories contain one real-world critical sentence and one counterfactual-world critical sentence. Real-world critical sentence items can be identified by a '(rw)' at the end of the sentence and counterfactual-world sentence items can be identified be a '(cfw).'

Item 13: Version 1

Lily and Joe landed on a planet full of caves. They decided to go caving.

Caving was fun, but really hard work. Soon, Lily and Joe were hungry and tired. They decided to head back.

However, there was a problem. One of the holes they'd carefully crossed earlier was now full of water.

To make it worse a fish was in the hole! A fish with razor sharp teeth and huge claws.

Surprisingly the fish was massive. The fish was also covered in pink and green fur. (cfw) RED

The fish was covered in pink and green fur
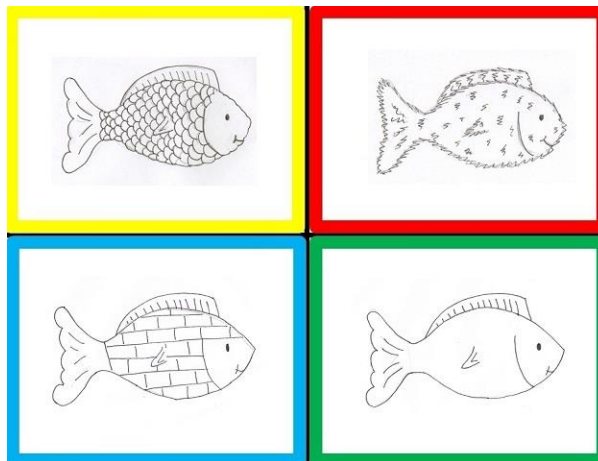
Lily and Joe froze. The fish was scary.

"We can wait here until the fish swims back out of the hole," said Joe.

But the cave was filling with water too. "We can't! If we stay in the cave we'll drown," said Lily.

Suddenly, the fish made a sound. The sound was a low deep growl. Lily gazed at it.

Lily did not want to wake the huge fish up. So when Lily spoke to Joe she whispered. (rw) *BLUE*

When Lily spoke to Joe she had whispered
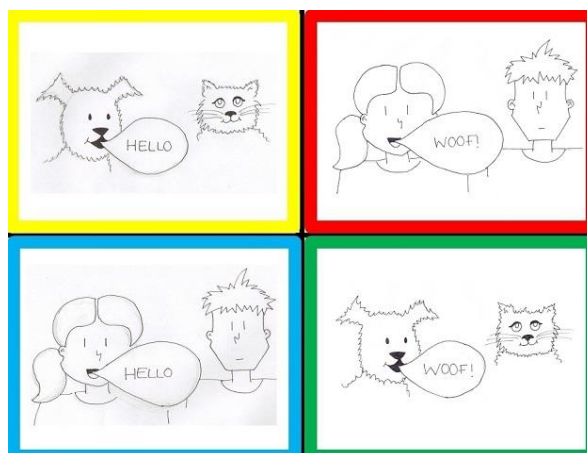
"The fish is snoring!" said Lily. "It's asleep." There was a thin ledge next to the hole the fish was in. Lily and Joe tiptoed around the ledge. They ran out of the cave and back to the spaceship.

Item 13: Version 2

Lily and Joe landed on a planet full of caves. They decided to go caving.

Soon, Lily and Joe were hungry and tired. They decided to turn back.

However, there was a problem. One of the holes they had crossed earlier was now full of water. To make it worse a huge fish was in the hole!

Suddenly, the fish made a sound. It was a low deep growl. Lily gazed at it.

Lily did not want to wake the huge fish up. So when Lily spoke to Joe she whispered. (rw) *BLUE*
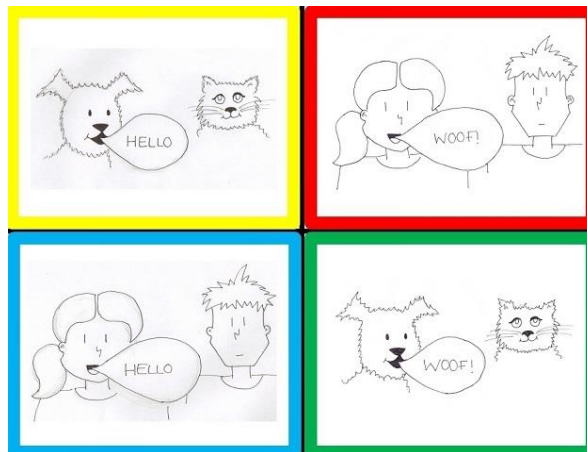
"I think the fish is asleep," Lily said. "It's snoring!"

Luckily, there was a thin ledge next to the hole. Lily and Joe were almost out of the cave when the ledge behind them broke.

A rock fell into the water. The fish moved. But it didn't wake up.

As Lily and Joe passed the fish they realised that it was a very odd fish.

Surprisingly the fish was massive. The fish was also covered in pink and green fur. (cfw) *RED*
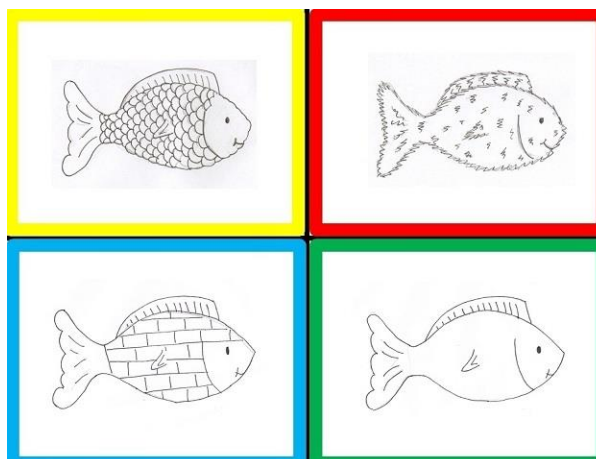
Lily and Joe ran out of the caves as fast as they could. They were safe at last. "That was a close one!" said Lily. "I know!" said Joe. They both began to laugh at the sleeping fish.

Item 13: Comprehension Questions

The fish was asleep **(ex)**

A fish was blocking Lily and Joe's exit **(inf)**

Item 14: Version 1

Lily and Joe landed on a dusty planet. Everyone was holding plant pots.

The people were all walking through a tunnel. Lily and Joe followed them. On the other side of the tunnel was a lake.

CRASH!!! Suddenly rocks began to fall down all over the place. Lily and Joe hid under a tree.

Eventually, the rocks stopped falling. But huge rocks were now in front of the tunnel. They were all stuck!!!

The people in this place wore tops made from elastic. Elastic is quite stretchy. (rw) _GREEN_

The elastic was quite stretchy
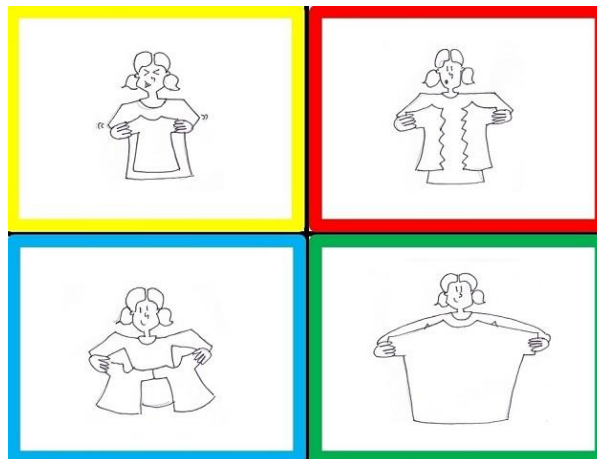
This gave Lily an idea. They could use the tops to make a trampoline. Everyone could then jump over the tunnel.

Lily told the men and women her idea. They all began to help her. They all gave Lily their tops.

Since there was now a way out of the tunnel they began to relax. They put their plant pots down next to the lake.

Every plant then did quite a strange thing.

**These plants did not suck up water from their roots. The plants sucked up water from a huge straw. (cfw) YELLOW**

The plants sucked up water from a huge straw
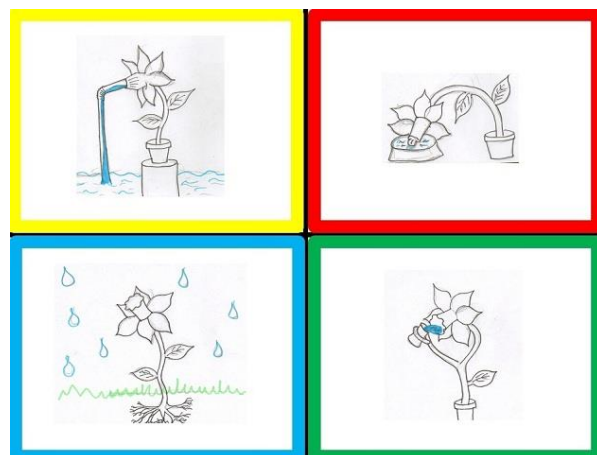
Lily and Joe watched the plants drink for the rest of the day. It was an amazing sight. They then bounced over the tunnel and back to their spaceship.

Item 14: Version 2

Lily and Joe landed on a dusty planet. Everyone was holding plant pots. The people were walking through a tunnel.

"Where are you going?" asked Lilly. "It hasn't rained for weeks. So we're taking the plants to the lake." said a man.

Lily and Joe followed him. On the other side of the tunnel was a lake.

All of the plants were put down next to the lake. The plants then did quite an odd thing.

These plants did not suck up water from their roots. The plants sucked up water from a huge straw. (cfw) *YELLOW*
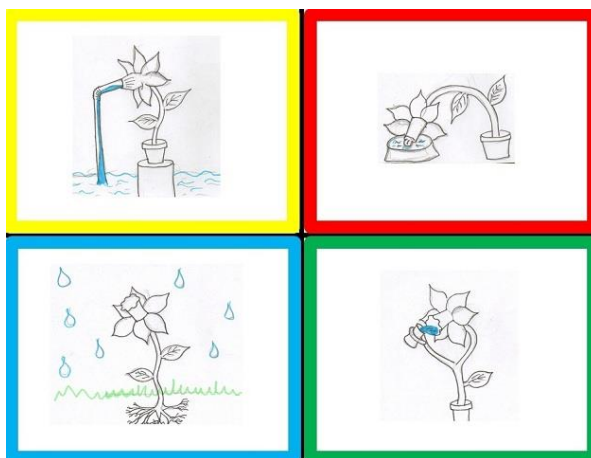
Lily and Joe watched the plants drink. It was a very cool sight. Each plant drank for about an hour.

Lily and Joe were just about to head back to their spaceship when… CRASH!!! Huge rocks were falling from the sky.

Lily and Joe hid under a large tree. Eventually, the falling came to a stop.

Sadly, two huge rocks were now in front the tunnel. They were all stuck!!

The people in this place wore tops made from elastic. Elastic is quite stretchy. (rw) *GREEN*
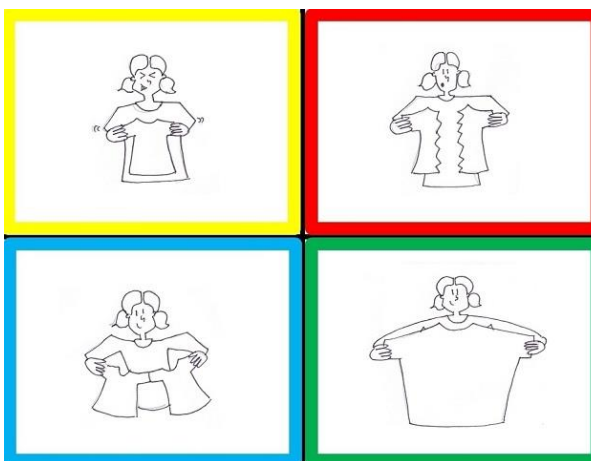
This gave Lily an idea. They could use the tops to make a trampoline. They could then jump over the tunnel. Everyone was free in no time at all.

Item 14: Comprehension Questions

The people gave Lily their trousers **(ex)**

The people were happy that the tunnel was blocked **(inf)**

Item 15: Version 1

Lily and Joe landed on a planet in chaos. There had been a murder.

The kind old bear that lived on a small island just outside of the city had been found dead.

 The island was surrounded by a fast flowing river. There was no bridge and no boats. The only way to reach the island was to swim.

A slimy fish had been seen at the scene. But he had not been seen since. Lily saw the fish.

The large fish ran across the green grass. He left a trail of gunky slime as he went. (cfw) *YELLOW*

The large fish had run across the green grass
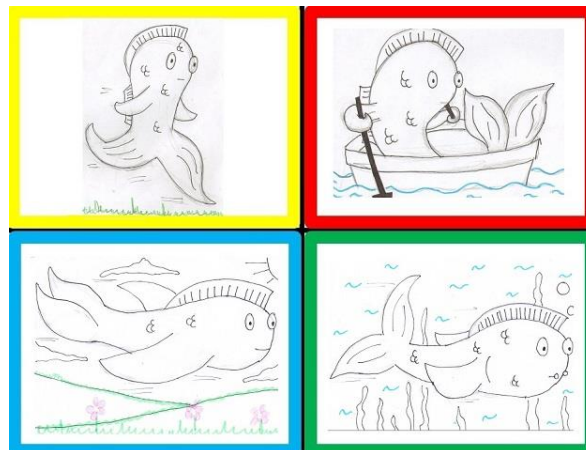
"I didn't do it!" cried the fish, "I can't swim." So the fish couldn't have killed the bear. But if he didn't do it who did?

Joe had an idea. Lily, Joe and the fish went to the bear's house.

Joe shook the bear. It began to move. "He's not dead!" said Joe, "he was hibernating."

"Well I could have told you that," said a large rhino that lived with the bear.

The rhino was huge. It had huge feet. The rhino was also very, very heavy. (rw) *GREEN*

The rhino was very, very heavy
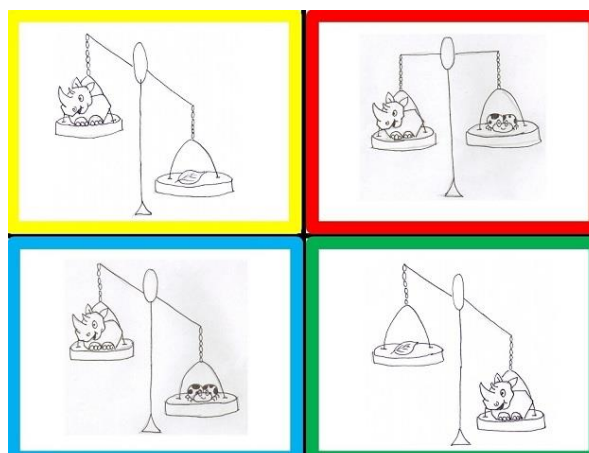
Lily, Joe, the fish, and the bear all returned to the city. Once there they told the police what had really happened.

378

Item 15: Version 2

Lily and Joe landed on a planet in chaos. There had been a murder. The old bear that lived on a small island just outside of the city had been found dead.

The island was surrounded by a fast flowing river. There was no bridge and no boats. The only way to reach the island was to swim.

A slimy fish had been seen at the scene. But he had not been seen since.

Suddenly, a rhino spotted the fish.

The rhino was huge. It had huge feet. The rhino was also very, very heavy. (rw) *GREEN*
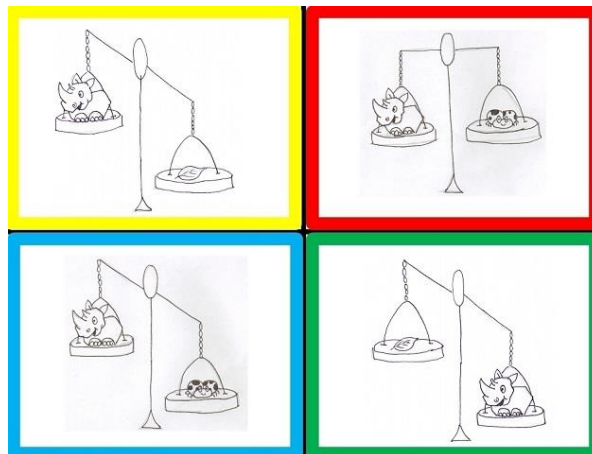
"Over there!" shouted the Rhino. "I didn't do it!" cried the fish, "I can't swim." So who had killed the bear?

Joe had an idea. Lily, Joe and the fish went to the bear's house.

Joe shook the bear. It began to move. "He's not dead!" said Joe, "he was hibernating!"

Lily, Joe and the fish went back to the city. But the police still thought the fish was guilty. They began to chase after him.

The large fish ran across the green grass. He left a trail of gunky slime as he went. (cfw) *YELLOW*
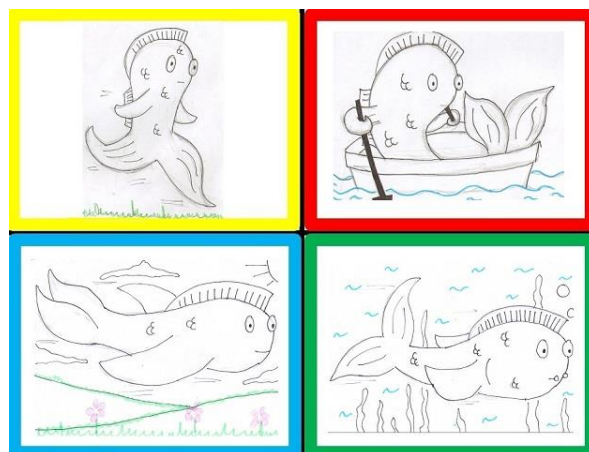
Lily and Joe told the police that the fish didn't kill the bear. The fish was then freed.

Item 15: comprehension Questions

The fish could swim **(ex)**

The fish was scared of the police **(inf)**

Lily and Joe landed on a cold frosty planet. They were greeted by a man. "Welcome," he beamed. "Here, every city is made of ice-cream."

"Do you have a strawberry ice-lolly city?" asked Lily. "Yep, I'll take you there" smiled the man.

The strawberry ice-lolly city was surrounded by thick walls of rock. "The walls keep the ice-cream cold," said the man.

"Mmm they look tasty," said Lily. She had to have one.

Lily licked the ice-lolly. She screamed. The ice-lolly was so hot it had burnt her tongue. (cfw) *BLUE*

The ice lolly was very, very hot
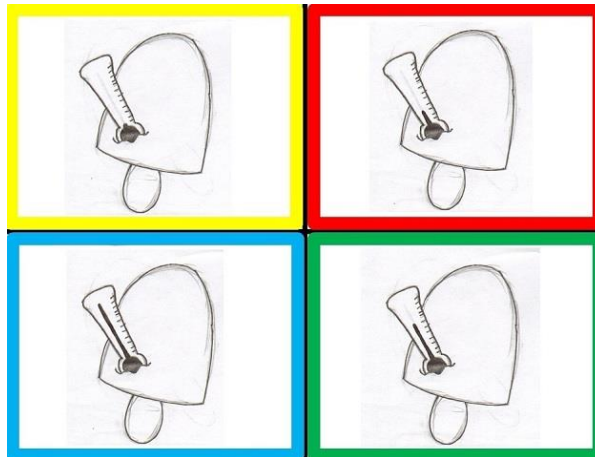
The man began to laugh. "I am Evil-tron!!" he said. "I tried to take over the universe. So the space police trapped me here. But, now I can steal your spaceship and escape."
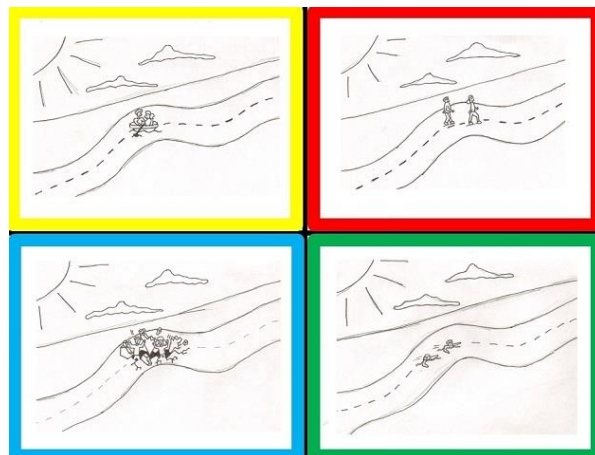
The man laughed then ran off. He shut the door on his way out. Lily and Joe were stuck!

Joe had an idea. They could use an ice lolly to melt the walls.

Lily and Joe were free.

Lily and Joe ran back to the spaceship along the hard road. They quickly made it back. (rw) *RED*

Lily and Joe ran along the hard road

The man was trying to open the spaceship but it wouldn't open. As soon as Lily and Joe appeared it flew to them. They jumped in and flew off.

Item 16: Version 2

Lily and Joe landed on a frosty planet. They were greeted by a man. "Welcome," he beamed. "Every city is made of ice-cream."

"Do you have a strawberry ice-lolly city?" asked Lily. "Yep, let's go" smiled the man.

Once there the man began to laugh. "I am Evil-tron!!" he said. "I tried to take over the universe. So the space police trapped me here. But, now I can steal your spaceship."

"Quick!" cried Joe.

Lily and Joe ran back to the spaceship along the hard road. They quickly made it back. (rw) *RED*
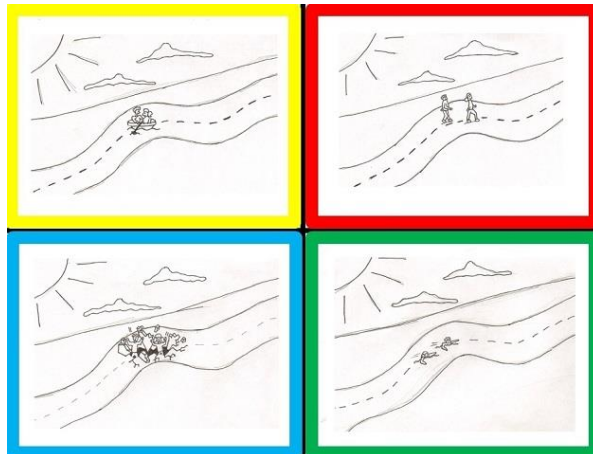

Evil-tron had got back to the ship before them. He had laid a trap. Giant walls rose out of the road around Lily and Joe. They were trapped!

"That evil man is going escape and it's all our fault," cried Lily.

"He won't, the spaceship only opens for us," said Joe.

This made Lily feel much better. "Since we're stuck I might as well have an ice-lolly," said Lily.

Lily licked the ice-lolly. She screamed. The ice-lolly was so hot it had burnt her tongue. (cfw) *BLUE*
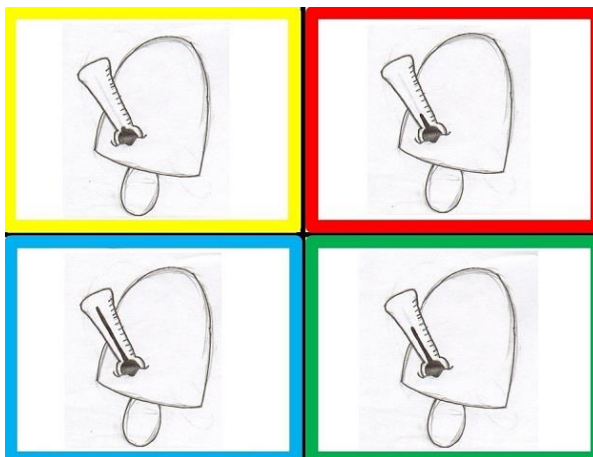
Joe had an idea. They could use an ice lolly to melt the walls. Lily and Joe were free. The spaceship flew over to them and they flew off.

Item 16: Comprehension Questions

The planet was made of ice-cream **(ex)**

Evil-tron was a nice guy **(inf)**

Item 17: Version 1

Lily and Joe landed in a small town. The town was a little bit spooky.

"This place is odd," said Joe. Every house was the exact same shade of blue.

Every lawn had the same tree in it. Every tree was the exact same shape. And every leaf was the same size.

It got worse. Lily and Joe were very surprised. The whole town was perfect.

Every blade of grass was exactly the same. Every blade of grass was quite short and green. (rw) BLUE

Every blade of grass was quite short and green
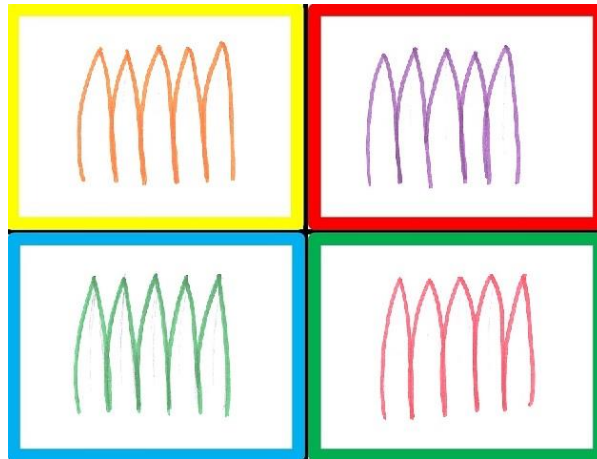
Lily and Joe went into a café. A man came over. He looked just like the man sat in front of them. More men came into the café.
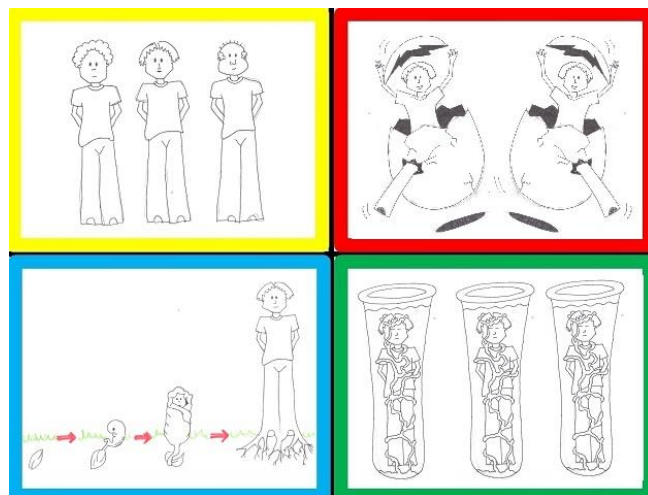
These men were all very alike too. They all made their way over to Lily and Joe. Lily and Joe were stuck.

"Must be the same. Must be the same," the men chanted.

Lily and Joe realised the truth…

The men were clones. They all looked the same because the men had been grown in giant test tubes. (cfw) GREEN

The men had been grown in giant test tubes

Lily had an idea. "We are the same. You are different," she chanted. The clones looked at each other and then began to fight. Luckily, this meant Lily and Joe were able to sneak out and back to their spaceship.

Item 17: Version 2

Lily and Joe landed in a small town. They went into a café.

A man came over. He looked just like the man sat ahead of them.

More very alike men came into the café. They made their way over to Lily and Joe. Lily and Joe were trapped.

"Must be the same. Must be the same," the men chanted. Lily and Joe realised the truth…

The men were clones. They all looked the same because the men had been grown in giant test tubes. (cfw) *GREEN*
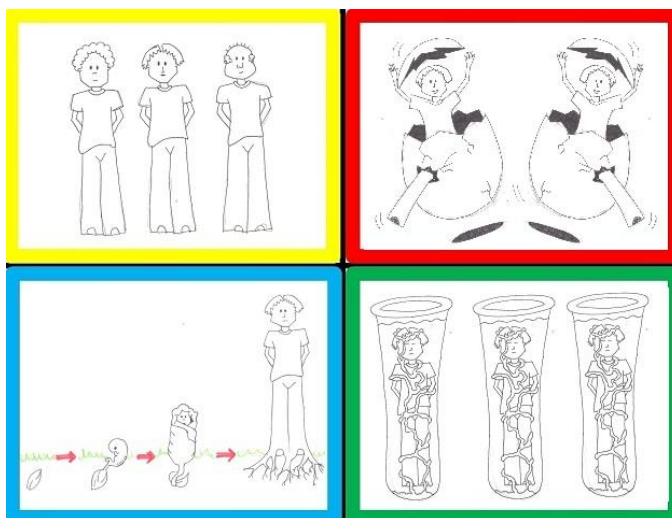

Lily had an idea. "We are the same. You are different," she chanted.

The odd men faced each other. Then they began to fight. This meant Lily and Joe were able to sneak away.

They ran back to their ship. They saw that the whole town was just as weird as the men.

The whole town was the same. Every house was the exact same shade of pale blue.

Every blade of grass was exactly the same. Every blade of grass was quite short and green. (rw) *BLUE*
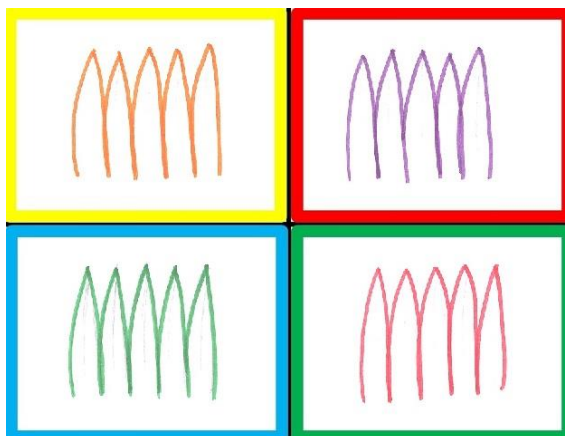
They got back to the spaceship in no time. "What a lucky escape," said Joe. "Must be the same, must be the same," chanted Lily. Joe froze. "Only joking! You should have seen your face though" laughed Lily.

Item 17: Comprehension Questions

Everything on the planet was the same **(ex)**

The clones did not want to hurt Lily and Joe **(inf)**

Item 18: Version 1

Lily and Joe landed on a planet covered in deserts. There were also lots of cool purple rocks.

Lily and Joe decided to look for a rock as a gift for their friend Newt. Eventually they spotted a rock shaped like an N.

However, it was on the other side of a large crack in the ground. The crack was full of snakes.

Lily leapt, but the gap was too wide. She fell right into the crack.

"ARGH!!! ARGH!!! ARGH!!!" Lily cried out in pain. The large evil snake had bitten Lily's ankle. (rw) *YELLOW*

The large snake had bitten Lily's ankle.
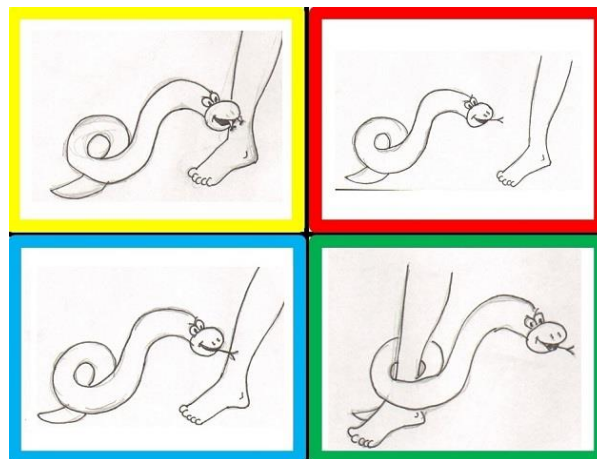
Bravely Joe went into the pit and got Lily out. Lily was as pale as a ghost and very shaky.
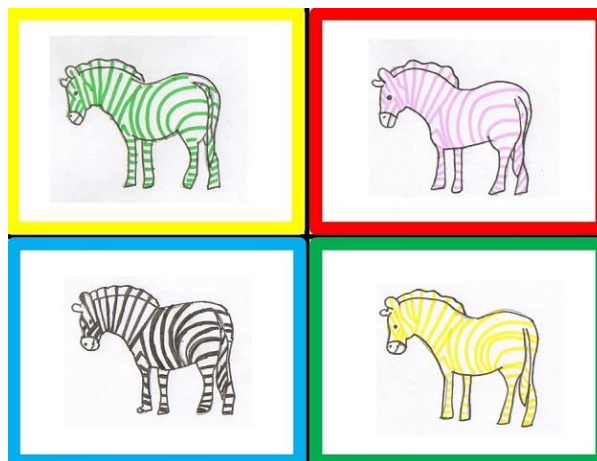
Joe had some pills that would save Lily. Sadly, they were back in the spaceship.

It would take him hours to carry Lily back. Lily didn't have a few hours.

Joe would have to find a faster way back. Luckily, the answer right in front of him.

There was a large herd of very odd zebras. The zebras were covered in huge pink stripes. (cfw) *RED*

The zebras were covered in huge pink stripes.

Joe approached the zebras. One of them bowed down so Joe and Lily could ride on its back. Within no time they were back at the spaceship. Lily got better very fast.

Lily and Joe landed on a planet covered in deserts and beautiful purple rocks.

Lily and Joe decided to look for the perfect rock as a present for their friend Newt. Newt collected weird rocks.

Lily and Joe searched all day. They saw many wonderful things. But they just couldn't find the right rock.

They saw giant pink trees, huge birds, and tiny cats. In front of them was a sight just as weird.

There was a large herd of very odd zebras. The zebras were covered in huge pink stripes. (cfw)
*RED*
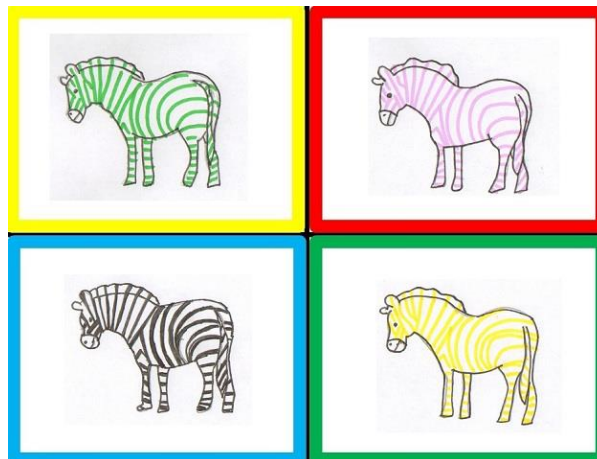
After a while they found the right rock.

It was on the other side of a giant hole. The hole was full of snake after snake.

"Can you jump that far?" asked Joe. "If you fall in the snake pills are in the ship. It will take hours to get back to them."

Lily leapt, but the gap was too wide. She fell straight into the crack.

"ARGH!!! ARGH!!! ARGH!!!" Lily cried out in pain. The large evil snake had bitten Lily's ankle. (rw)
YELLOW
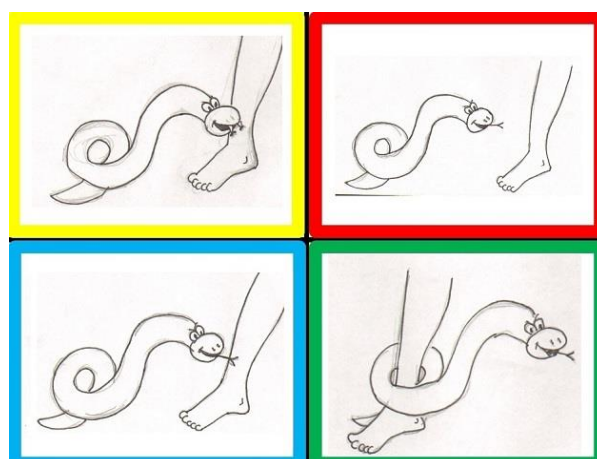
Joe pulled Lily out of the pit. He approached the zebras. One of them bowed down so they could ride on its back. Within no time they were back at the spaceship.

Item 18: Comprehension Questions

The planet was covered in deserts **(ex)**

The zebras would fit in on Earth **(inf)**

Appendix 4.5 – Reading Grades

	Version 1	**Version 2**
Story 1	1.66381	1.546566
Story 2	1.389934	1.034305
Story 3	1.319643	1.882885
Story 4	1.833403	1.605384
Story 5	1.263378	1.757986
Story 6	1.383646	1.315577
Story 7	2.558578	2.357026
Story 8	1.82406	2.113874
Story 9	2.047927	2.82226
Story 10	1.724031	1.693808
Story 11	2.595006	2.684251
Story 12	2.743687	2.823654
Story 13	2.22301	2.050983
Story 14	2.504829	2.050983
Story 15	1.705825	1.599674
Story 16	2.015049	2.072054
Story 17	1.580346	1.446664
Story 18	2.937035	1.890408

Appendix 4.6 – Practice Story

Lily and Joe landed on a new planet. They climbed out of their rocket to look around. Lily felt unsure. She turned around to head back, but it was too late. Their rocket was rising into the sky. "Someone has stolen our rocket!" cried Joe. The thieves did not get far though. Suddenly, there was a loud bang. The rocket crashed. The bang was so loud it knocked Lily's hat off her head.

**Lily bent down and picked up her big hat. Lily's large hat was yellow and made from straw. (rw) GREEN**

Lily's hat was made from straw.
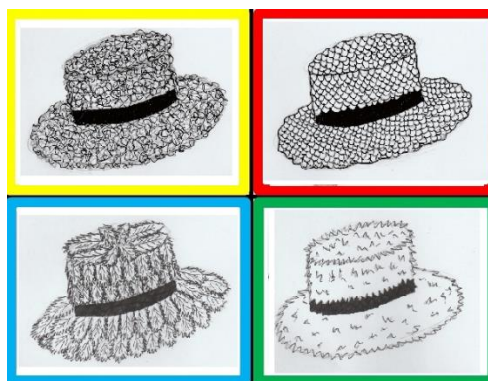
Lily climbed a tree to see if she could see where the rocket had landed. "It's that way!" she shouted. "We just need to find the river and follow it." Lily and Joe jumped into the river and swam down it. They saw green fish, red ducks, and purple eels. They Finally, Lily and Joe could see their rocket.

The rocket was surrounded by hundreds of sheep. The sheep were quite odd. The sheep were green. (CFW) YELLOW

The sheep were green.
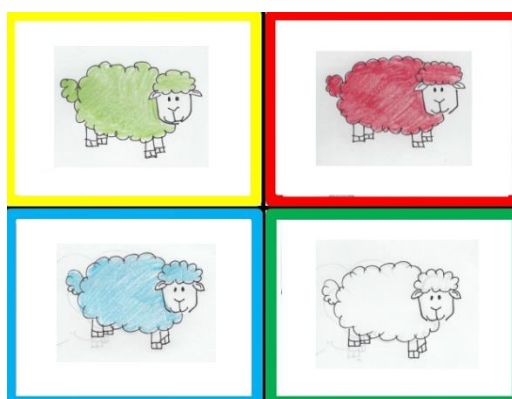
Lily and Joe jumped out of the river. They ran to their rocket. It was empty. The thieves had run away. Lily and Joe jumped into their rocket and flew away.

Comprehension Questions

Lily was feeling happy at the start of the story? (inf)

Lily climbed a tree to find the rocket? (exp)

Appendix 4.7 – Remaining Interactions for SVT Response Time (Stage 2)

Plausibility * Age: $F(1,28) = 1.20$, $p = .28$, $\eta p^2 = .04$

Inference * Age: $F(2,56) = 2.26$, $p = .11$, $\eta p^2 = .01$

Plausibility * Inference Type: $F(2, 56) = .35$, $p = .71$, $\eta p^2 = .01$

Plausibility * Inference Type * Age: $F(2,56) = 1.21$, $p = .31$, $\eta p^2 = .04$

Appendix 4.8 – Remaining Interactions for IST Reading Speeds (Stage 2)

Plausibility * Age: $F(1, 28) = 0.95$, $p = .34$, $\eta p^2 = .03$

Inference Type * Age: $F(2, 56) = 0.76$, $p = .47$, $\eta p^2 = .03$

Plausibility * Inference Type: $F(2, 56) = 0.80$, $p = .46$, $\eta p^2 = .03$

Plausibility * Inference Type * Age: $F(2, 56) = 2.42$, $p = .10$, $\eta p^2 = .08$

Appendix 4.9 – Remaining Accuracy Interactions (Stage 2)

Plausibility * Age: $F(1, 56) = 0.19$, $p = .66$, $\eta p^2 < .01$

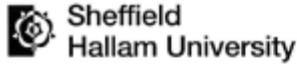Plausibility * Task: $F(1, 56) = 1.06$, $p = .31$, $\eta p^2 = .02$

Plausibility * Age * Task: $F(1, 56) = 0.19$, $p = .66$, $\eta p^2 < .01$

Plausibility * Inference Type * Age: $F(2, 112) = 1.35$, $p = .27$, $\eta p^2 = .05$

Plausibility * Inference Type * Task: $F(2, 112) = 1.69$, $p = .19$, $\eta p^2 = .06$

Plausibility * Inference Type * Age * Task: $F(2, 112) = 1.64$, $p = .20$, $\eta p^2 = .06$

Appendix 5.1 - Ethical Approval

Sheffield Hallam University

SHARPENS YOUR THINKING

Our Ref AM/SW/3-2012

Rebecca Hancock
46 Fircroft Avenue
Nottingham
NG8 3AL

14th February 2012
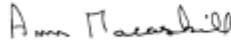
Dear Rebecca

Request for Ethical Approval of Research Project

Your research project entitled "**The Development of Counterfactual-world Inference Generation Abilities during Adolescence**" has been submitted for ethical review to the Faculty's rapporteurs and I am pleased to confirm that they have approved your project.

I wish you every success with your research project.

Yours sincerely

Professor A Macaskill
Chair
Faculty Research Ethics Committee

Appendix 5.2 – Information Sheets, Consent Forms, and Debriefs

Email to Schools/Sports Groups

Dear *Head Teacher*,

I'm Rebecca Hancock. I'm a PhD student at Sheffield Hallam University. My PhD project will be exploring the development of reading comprehension and text production skills during adolescence.

I was wondering if you would be willing to let me work with around 5 students from *year*.

What Would the School Receive for Taking Part?

Once the study is completed I would provide you with feedback. This feedback would offer suggestions to improve reading comprehension and text production skills.

I would also be happy to run a class/assembly on 'what it is like to go to university' or 'Psychology as a subject.'

Your school would also receive a certificate from Sheffield Hallam University stating that you've been involved in a project that aims to promote literacy skills in young people.

What Would Your Students Receive for Taking Part?

All students that take part will receive a small goody bag, typically filled with small items of stationary.

We also find that young people generally enjoy the types of tasks I would like them to complete.

Each student would also receive a certificate.

What Would Your Students Have to Do?

I have developed a variety of computer and paper based tasks to assess reading comprehension, text production, and related skills.

Students would complete the task in two sessions. The first session is a one-to-one session which should take no longer than an hour. The second session is a group session, where all 5 students complete the tasks at the same time. Again this session should take no longer than an hour.

Of course, when these sessions take place would be decided through discussion with the students' class teachers.

If you think your school would be interested in my project or if you would like any further details please do not hesitate to contact me.

With Best Wishes,

Rebecca

Tier 1 – Organisation Consent

Dear *Head teacher,*

Thank you for taking the time to find out more about my research project. This letter will provide you with more information about the project.

The project is looking at how the ability to generate inferences from fantasy and fiction texts may change with age. Therefore, this project is looking for changes across time, and so the young people will need to complete the tasks both now and then again 18months later. I would be looking to work with children from year 5, year 7, and year 9. You can chose which age group(s) you would like me to work with and within these age groups the groups of children (i.e., which tutor group) you would like me to work with.

The tasks and activities will be assessing:

- intelligence
- reading comprehension abilities
- text production skill
- counterfactual-world and real-world inference generation abilities
- ability to access real-world knowledge
- inhibition skill
- working memory capacity
- strength of belief-bias

All tasks are suitable for children and should not differ too much from the tasks they complete in class. Some tasks will be completed on a laptop and some tasks will be paper based. The laptop will be provided by the researcher.

Parental consent will be gained from the parents and both verbal and written consent will be gained from the children before they take part in the study.

The young people would be seen individually during school time in a quiet area of your school. Several sessions may be needed but each session will last no longer than 40minutes and all sessions will be arranged through discussion with the class teacher. This should minimise any disruption to learning. It is thought approximately 2-3 sessions will be needed, however, this may vary depending on each child's age and abilities.

All data collected will be treated with the strictest confidence. All data will be anonymised, by giving each child a unique but random code. This, instead of their name, will be recorded on all tasks. Any data stored on a computer will be password protected and any paper based data will be stored in a locked cupboard.

Of course all young people would be able to withdraw from the study at any time during data collection. The young people will be told if at any point they want to stop then they can. Furthermore, children will be told that the tasks are voluntary and so if they do not wish to complete a task they do not have to.

Children's data can also be withdrawn by yourself or parents/guardians up until 31st December 2012 by contacting me directly by e-mail (dsrh9@exchange.shu.ac.uk). Data collected in

18months will also be able to be withdrawn up to 2 weeks after data collection. A more specific time will be given closer to the time.

If you have any further queries about the study please do not hesitate to contact me by e-mail. You can also contact my supervisor Dr Claudine Bowyer-Crane by e-mail or telephone (claudine.bowyer-crane@york.ac.uk; 01904 434398). If you are happy for the pupils in your school to take part in this research please fill in the form below.

Kind regards

Rebecca Hancock

Project exploring the development of inference generation abilities: OPT-IN form

I have read and understood the information given to me about this study. I give permission for the pupils in my school to take part in the research project looking at the development of inference generation abilities. I understand that participation in this study involves:

- Assessment at time 1 and then again 18months later at time 2

- Distribution and collection of parental consent forms on behalf of the researcher

- Children completing several individual test sessions which will last no longer than 40minutes a session

- Children completing a variety of computer and paper based activities

I also understand that:

- All data will be anonymised

- All data will be securely stored, either in a locked cupboard or when on a computer, password protected

- The children can withdraw from their data from the study up until 31st December 2012

- The children do not have to take part if they do not wish

- The children can refuse to complete any tasks

- Data can also be withdrawn by either the school or the child's parents/guardians

- I can contact the researcher or her supervisor at any point during this study if I have any questions of queries

Name of School:…………………………………………………………

Head teacher:………………………………………………………..

Signed:…………………….(Head Teacher) **Date:**………………………

Sheffield Hallam University

SHARPENS YOUR THINKING

I am writing to ask whether you would be happy for your child to be involved in a research project organised by Sheffield Hallam University, which your child's school has kindly agreed to support.

The project is looking at young people's ability to understand fantasy and fiction texts . The young people taking part would be seen on a one-to-one basis by myself to complete activities based around:

- understanding text
- Writing stories
- reading between the lines
- accessing knowledge stored in memory
- memory
- biases for everyday information

Your child would complete these tasks individually during school time over several sessions, approximately 2-3 sessions. Each session will take place in the school. Sessions will last no longer than 40minutes and will be arranged through discussion with the class teacher. This means your child shouldn't miss out on any important classroom activities. We also find children generally enjoy taking part in this type of research, especially since some of the tasks are completed on a laptop.

All data collected will be treated with the strictest confidence. Instead of their name, your child will be given a unique code and this will be put on any work they do. All data will be stored in locked cupboards, and any data stored on the computer will be password protected.

Your child would be able to withdraw from the study at anytime during data collection. These tasks are not tests. Therefore, if your child doesn't want to take part they don't have to and if they don't want to complete a task then they don't have to.

 You would also be able to withdraw your child's data from the project up until the 20th December by contacting me or my supervisor directly either by telephone or by e-mail.

If you have any queries about the research project, please do not hesitate to contact us directly. My email address is dsrh9@exchange.shu.ac.uk. My project supervisor's (Dr Lisa Reidy) email address is ssllr@exchange.shu.ac.uk and her telephone number is 01904 434398.

If you are happy for your child to take part in this project please return the form on the next page.

With best wishes

Rebecca Hancock
Psychology postgraduate student
Sheffield Hallam University

Sheffield Hallam University

SHARPENS YOUR THINKING

I have read and understood the information given to me about this study. I give permission for my child to take part in the project looking at the ability to read between the lines. I understand that, if my child wants to, they will:

- Complete tasks now and then again 18 months later

- Complete tasks over several sessions, with each session lasting no longer than 40minutes

- Complete some tasks on a laptop

- Complete some tasks on paper

I also understand that:

- My child's data will be made anonymous

- My child's data will be stored safely, in either a locked cupboard, or when on a computer, password protected

- My child can withdraw their data up until the 31st December 2012 (and 2 weeks after data collection when they complete the tasks again in 18months)

- My child does not have to take part if they do not want to

- My child does not have to complete any tasks that they do not want to

- I can withdraw my child's data if I wish

- I can contact the researcher or her supervisor if I have any questions or queries.

....

Project on inference generation abilities: OPT-IN form

I give permission for my child to take part in the research project looking at the ability to read between the lines.

Name of Child:...

Name of Parent/Guardian ....

Signed...

Class teacher:...

Can you read between the lines?

I am looking at how young people like yourself understand what they read. I am inviting you to take part in this study, however, if you don't want to take part then you do not have to. If you do take part you will help me to understand how people read between the lines.

If you decide to take part you will complete a variety of tasks which will look at how you read between the lines, understand what you have read, how you use your memory, how you make decisions, and how easily you can remember things you've learnt.

We will complete these tasks away from your classmates in a quiet area of your school. It may take a couple of sessions (maybe 2 or 3) to complete all of the tasks. Each session will last no longer than 40minutes.

None of the answers you give will be shared with your class teacher or classmates. Once I take your answers home I will keep them in a locked cupboard. All of the answers you produce on a computer will need a password to be opened. Only I will have this password.

If you do decide to take part you can ask me to remove your answers from the study up until the 20th December 2013. And remember, these tasks are not tests and if you want to stop at any time then you can.

Do you have any questions? I'm more than happy to answer them for you. If you decide to take part and think of any questions then just ask them.

If you would like to take part in this study please write your name and the date on the lines below

Name...................

Date ....

Appendix 10—debrief

Sheffield
Hallam University

SHARPENS YOUR THINKING

Your child has now completed the study looking at young people's ability to understand fantasy and fiction texts. Thank you very very much for allowing your child to take part. I hope that they enjoyed the experience.

This study was conducted because up to 10% of young people struggle to understand text. Inference generation – the ability to read between the lines - is thought to be very important when it comes to understanding text. This means, if we understand how young people read between lines specially designed tasks can be created for those children who struggle to understand text. By taking part in this project your child has helped us to understand inference generation that little bit more. Thank you very much

Remember if you would like to withdraw your child's data please contact myself or my supervisor before 17th July 2012. Our contact details are below.

Researcher: Rebecca Hancock
email: dsrh9@exchange.shu.ac.uk
Project Supervisor: Dr Claudine Bowyer-Crane
email: dscbc1@exchange.shu.ac.uk
Tel:0114 225 5964

With best wishes

Rebecca Hancock
Psychology postgraduate student
Sheffield Hallam University

THANK YOU

Undergraduate Consent

Thank you for taking the time to find out more about my research project. This sheet will provide you with more information about the project.

The project is looking at how the ability to generate inferences from fantasy and fiction texts may change with age and genre. The tasks and activities you will complete will be assessing:

- intelligence
- reading comprehension abilities
- writing skills
- counterfactual-world and real-world inference generation abilities
- ability to access real-world knowledge
- inhibition skill
- working memory capacity
- strength of belief-bias

Some tasks will be completed on a laptop and some tasks will be paper based.

You will be seen individually at times convenient for yourself. Several sessions may be needed but each session will last no longer than 40minutes. You will receive psycredits for your participation. These will be awarded after all sessions have been completed.

All data collected will be treated with the strictest confidence. All data will be anonymised, by giving you a unique but random code. This, instead of your name, will be recorded on all tasks. Any data stored on a computer will be password protected and any paper based data will be stored in a locked cupboard.

Of course you have the right to withdraw from the study at any time during data collection up until 2 weeks after taking part by contacting me directly by e-mail (dsrh9@exchange.shu.ac.uk). If at any point you wish to stop then you can. Furthermore, all tasks are voluntary and so if you do not wish to complete a task you do not have to.

If you have any further queries about the study please do not hesitate to contact me by e-mail. You can also contact my supervisor Dr Lisa Reidy by e-mail or telephone (ssllr@exchange.shu.ac.uk |0114 2255813). If you are happy to take part in this research please fill in the form on the next page.

Kind regards

Rebecca Hancock

Project exploring the development of inference generation abilities: OPT-IN form

I have read and understood the information given to me about this study. I understand that participation in this study involves:

- Completing a variety of computer and paper based activities over several sessions

I also understand that:

- All data will be anonymised

- All data will be securely stored, either in a locked cupboard or when on a computer, password protected

- I can withdraw my data from the study up until 2 weeks after taking part

- I do not have to take part if I do not wish to do so

- I can refuse to complete any tasks

- I can contact the researcher or her supervisor at any point during this study if I have any questions of queries

Name:……………………………………………………………..

Signed:……………….................................….   **Date:**………………………

Sheffield Hallam University

SHARPENS YOUR THINKING

Thank you for completing the study looking at young people's ability to understand fantasy and fiction texts. Thank you very very much for taking part. I hope that you enjoyed the experience.

This study was conducted because up to 10% of young people struggle to understand text. Inference generation – the ability to read between the lines - is thought to be very important when it comes to understanding text. This means, if we understand how young people read between lines specially designed tasks can be created for those children who struggle to understand text. By taking part in this project you have helped us to understand inference generation that little bit more. Thank you very much

Remember if you would like to withdraw your data please contact myself or my supervisor before 30th January 2013. Our contact details are below.

Researcher: Rebecca Hancock
email: dsrh9@exchange.shu.ac.uk
Project Supervisor: Dr Lisa Reidy
email: ssllr@exchange.shu.ac.uk
Tel:0114 2255813

With best wishes

Rebecca Hancock
Psychology postgraduate student
Sheffield Hallam University

Psycred code:MED8NG-ZES5W2-CTYK0N

Appendix 5.3 - Order of Tasks

Session 1	Session 2
forced-choice picture-selection task	written language task
vocabulary task	Knowledge task
inhibition task	belief bias task
reading comprehension task	
working memory task	
matrix reasoning task	

Appendix 6.1. – Individual Analysis for Vocabulary and Matrix Reasoning Raw Scores

There was found to be a significant main effect of age for vocabulary score ($\chi2$ (3, N = 200) = 116.31, p < .001). Six non-parametric independent t-tests were conducted to explore further. Year 5 were found to have weaker vocabulary skills than Year 7 *(U =* 360.00, p < .001, N = 100), Year 9 (U = 79.50, p < .001, N = 100), and adults (U = 43.00, p < .001, N = 100). Year 7 were found to have weaker vocabulary skills than Year 9 (U = 324.00, p < .001, N = 100) and adults (U = 262.50, p < .001, N = 100). Year 9 were also found to have weaker vocabulary skills than adults (U = 551.00, p = .002, N = 100).

A main effect of age was found for matrix reasoning scores ($\chi2$ (3, N = 200) = 46.01, p < .001). Six non-parametric independent t-tests were conducted to explore further. Year 5 were found to have weaker matrix reasoning skills than Year 7 (U = 605.50, p < .001, N = 100), Year 9 (U = 542.00, p = .006, N = 100), and adults (U = 426.00, p < .001, N = 100). There was found to be no difference between the matrix reasoning abilities of Year 7 and Year 9 (U = 687.50, p = .576, N = 100). However, Year 7 (U = 757.00, p = .001, N = 100) and Year 9 (U = 538.00, p = .001, N = 100) were found to have weaker matrix reasoning skills than adults.

Appendix 6.2 – Individual Contributions of Controls

Beta values for Text Quality Regression Models

	Year 5	Year 7	Year 9	Adults
Working Memory Capacity	.359*	.159	.019	.014
IQ Raw Score	.312	.233	.347	.070
Single Word Reading Score	.267	.297	.112	.174

p < .05

Beta values for Reading Comprehension Regression Models

	Year 5	Year 7	Year 9	Adults
Working Memory Capacity	.024	.096	.007	.093
IQ Raw Score	.238	.009	.087	.235
Single Word Reading Score	.246	.523*	.381	.073

p < .05

Appendix 7.1 – Comparisons for Main Effect of Age (Accuracy)

Year 5 vs Year 7: t (99) = 2.44, p = .02

Year 5 vs Year 9: t (99) = 2.95, p = .004

Year 5 vs Adults: t (105) = 4.93, p < .001

Year 7 vs Year 9: t (98) = 0.72, p = .48

Year 7 vs Adults: t (104) = 2.03, p = .04

Year 9 vs Adults: t (104) = 1.03, p = .30

Appendix 7.2 – Remaining Interactions for Accuracy Data
Plausibility * Age: F (1, 203) = .12, p = .95, ηp^2 < .01

Plausibility * Inference Type * Age: F (1.85, 383.70) = .63, p = .70, ηp^2 = .01

Appendix 7.3 – Additional Analysis Comparisons for Main Effect of Reading Comprehension Age (Accuracy)

7-9;12 years vs 10-13;12 years: t (98) = 2.82, p = .006

7-9;12 years vs 14-15;12 years: t (95) = 2.72, p = .008

7-9;12 years vs 16 years and over: t (106) = 2.78, p = .006

10-13;12 years vs 14-15;12 years: t (97) = 0.37, p = .71

10-13;12 years vs 16 years and over: t (108) 0.15, p = .88

14-15;12 years vs 16 years and over: t (105) = .022, p = .82

Appendix 7.4 – Remaining Interactions for Additional Analysis Accuracy Data
Plausibility * Reading Comprehension Age: F (1, 206) = .99, p = .40, ηp^2 = .03

Inference Type * Reading Comprehension Age: F (1.65, 352.56) = 1.01, p = .41, ηp^2 = .03

Plausibility * Inference Type * Reading Comprehension Age: F (1.65, 352.56) = 1.46, p = .20, ηp^2 = .04

Appendix 7.5 – Comparisons for Main Effect of Age (Reading Speed)

Year 5 vs Year 7: t (99) = 4.28, $p < .001$

Year 5 vs Year 9: t (99) = 8.81, $p < .001$

Year 5 vs Adults: t (105) = 10.58, $p < .001$

Year 7 vs Year 9: t (98) = 4.47, $p < .001$

Year 7 vs Adults: t (104) = 6.01, $p < .001$

Year 9 vs Adults: t (104) = 1.31, $p = .19$

Appendix 7.6 – Remaining Interactions for Reading Speed Data

Plausibility * Age: F (1, 180) = 1.15, $p = .33$, $\eta p^2 = .02$

Inference Type * Age: F (1.92, 346.25) = .014, $p = .99$, $\eta p^2 = .003$

Plausibility * Inference Type: F (1.92, 346.25) = .03, $p = .97$, $\eta p^2 < .001$

Plausibility * Inference Type * Age: F (1.92, 346.25) = .09, $p = .99$, $\eta p^2 = .001$

Appendix 7.7 – Additional Analysis

Additional analysis was conducted when exploring accuracy patterns as Year 9 displayed a qualitatively different pattern to the other three groups. This was not the case for reading speed, such a main effect of age was not observed. The Year 9 sample comprises a high number of young people with reading comprehension levels below age-related expectations. Therefore, it is possible that reading speed patterns were skewed by the mix of reading abilities in each age group. Therefore, reading speeds were explored using the four new reading comprehension age groups created previously: 7-9;11 years (n = 49), 10-13;11 years (n = 51), 14-15;11 years (n = 48), and 16 years and over (n=59). See Table 7.8 for a summary of reading speed means for each condition. A mixed factorial ANOVA, with reading speed of critical text as the dependent variable was conducted.

Data was inspected prior to analysis. Since two versions of each story were created, analysis was conducted to determine if one version of the stories was more difficult than the other. There was found to be no difference between the number of errors made on Version 1 and Version 2 for those with a reading age of 7-9;11 years (t (47) = 1.81, p = .08), 10-13;11years (t (49) = 1.04, p = .31), 14-15;11 years (t (46) = 1.39, p = .17) or 16 years and over (t (57) = 0.83, p = .42). Data was therefore collapsed across both versions. Inspection of the skewness and kurtosis statistics revealed that for all conditions the data was normally distributed (z-scores for skewness and kurtosis were less than 1.96, the level suggested by Field (2005) for sample sizes less than 200). Homogeneity of variance was violated such that the largest standard deviation was more than four times larger than the smallest standard deviation (Howell, 1987). However, Field suggests ANOVA is robust enough to deal with this violation. Mauchly's Test of Sphericity was significant for inference type. Girden (1992) suggest that Huynh-Feldt, not the Greenhouse-Geisser correction, should be used when the estimates of sphericity are greater than 0.75, as in these cases the Greenhouse-Geisser correction is likely to be too conservative. Consequently, a parametric ANOVA was conducted to explore the effects of age, plausibility and inference type on reading speeds, with Huynh-Feldt statistics reported for those analyses including inference type.

Mean number of errors (and Standard Deviations) for real-world literal, real-world coherence-inference-evoking, real-world elaborative-inference evoking, counterfactual-world literal, counterfactual-world coherence-inference evoking, counterfactual-world elaborative-inference evoking conditions for all Reading-Age groups

	Real-World Information			Counterfactual-World Information		
	Literal	Coherence	Elaborative	Literal	Coherence	Elaborative
7-9	6633.72 (2126.51)	7075.66 (2507.11)	6510.65 (2051.05)	6899.19 (1865.28)	7312.65 (2485.18)	6640.50 (2179.13)
10-13	4776.77 (1254.54)	4975.08 (1592.89)	5062.18 (1319.17)	5227.32 (1626.09)	5341.75 (1679.28)	5316.26 (1593.20)
14-15	4687.07 (1043.21)	4792.88 (1616.85)	4584.23 (1633.77)	4917.31 (1246.55)	5047.95 (1399.56)	5037.26 (1746.59)
16	4678.28 (1549.50)	4947.85 (1891.43)	4662.15 (1620.71)	4677.04 (1360.02)	5187.80 (1774.62)	5468.35 (1773.56)

A 2 (Plausibility: real-world, counterfactual-world) * 2 (Inference Type: literal, coherence, elaborative) * 4 (Reading Comprehension Age: 7-9:11 years, 9-13:11 years, 14-15:11 years, 16 years) mixed factorial ANOVA was conducted. A main effect of plausibility was observed, with counterfactual-world critical sentences being read significantly slower than real-world critical sentences (F (1, 206) = 10.01, p = .002, ηp^2 = .08). A main effect of inference type was also observed (F (1.89, 351.69) = 4.19, p = .02, ηp^2 = .03). Three post hoc pairwise comparisons were conducted. There was a trend for coherence critical sentences to be read slower than literal critical sentences (t (206) = 2.51, p = .01) and elaborative critical sentences (t (206) = 2.16, p = .03), however, these differences did not reach significance due to the correction applied. There was no significant difference between reading speeds for literal critical sentences and elaborative critical sentences (t (206) = 0.37, p = .57). An interaction was observed between inference type and age (report). Six post hoc pairwise comparisons revealed that 7-9 year olds read critical texts significantly slower than 10 -13 year olds (t (98) = 4.81, p < .001), 14-15 year olds (t (95) = 5.88, p < .001), and 16 years and over (t (106) = 5.42, p < .001). None of the remaining comparisons were significant (p > .42 for all remaining comparisons). None of the interactions were significant (p > .39 for all interactions).

Appendix 8.1 – Example Stimuli – Knowledge

Experimental Items

1. mammals give birth to live young
2. most men look quite different
3. most trees grow from tiny seeds
4. without help humans can't fly
5. penguins can't fly but can swim
6. fish can't walk but they can swim
7. cats can't talk but can meow
8. (human speech)
9. (whale song)
10. trees are rooted to the ground
11. tigers could bite a human
12. some sharks could bite a human
13. some snakes could bite a human
14. the bones in humans can break
15. thin ice can crack if walked on
16. humans can't walk on water
17. solid rocks can be walked on
18. plants soak up water through roots
19. a lot of bears have brown fur
20. humans are covered in skin
21. snakes skin is covered in scales
22. fish are covered in scales
23. coffee is often quite hot
24. snow is usually cold
25. ice lollies are very cold
26. mostly, leaves on trees are green
27. usually grass is green
28. zebras have black and white stripes
29. Most bananas are yellow
30. most clay is brownish
31. hippos are very heavy
32. rhinos are very heavy
33. leaves are not very heavy
34. most tree trunks are very hard
35. solid steel doesn't bend
36. concrete walls are quite solid

Fillers (7syllables)

1. Birds are covered in scales
2. Nearly all pink pigs can fly
3. Dolphins can't walk but can fly

413

4. Lions can't talk but can woof
5. Hippos can't swim but can fly
6. Many dogs have bright green fur
7. Pigs are covered in feathers
8. Most snakes are covered in fur
9. Lava is often quite cold
10. Bumble bees have purple stripes
11. Mud is often bright yellow
12. Water is often bright pink
13. Milk is usually green
14. Feathers are very heavy
15. Elephants are very light
16. Many insects are heavy
17. Iron bars bend easily
18. The human skull is bendy
19. Humans can walk on cold milk
20. Plants use their leaves to make cakes
21. Almost all chickens can speak
22. Whales bark at each other
23. Trees jump around to find food
24. Worms often attack humans
25. Squirrels often attack men
26. Goldfish often attack men
27. Humans can walk on thin air
28. Most humans can't walk on grass
29. Rain is usually hot
30. Orange juice is very hot
31. Foxes are often purple
32. Thick wood is easy to bend
33. Rubber is quite hard to bend
34. A lot of dogs hatch from eggs
35. All girls are identical
36. Plants give birth to baby plants

Appendix 8.2 – Example Stimuli – Belief Biases

Premise	Conclusion
All dogs have feathers. A dalmatian is a dog	Dalmatians have feathers
All dogs are bigger than donkeys. All donkeys are bigger than elephants	Dogs are smaller than elephants
All wood sinks in water. Oak is a type of wood	Oak floats on water
Snow makes driving conditions safer. It always snows in the winter	It is safer to drive in the winter
All apples are blue or pink. Tim's apple is not pink	Tim's apple is blue
All fish live in trees. A trout is a fish	Fish live in water
All footballs are the same shape as tennis balls. All tennis balls are the same shape as rugby balls	A football is the same shape as a rugby ball
All plants make carbon dioxide. A sunflower is a plant	A sunflower makes carbon dioxide
All elephants are hayeaters. All hayeaters are light	Elephants are not light
All fruit is unhealthy. A banana is a fruit	Bananas are healthy
All birds have wings. A canary is a bird	A canary has wings
All snakes use their tongues to smell with. A cobra is a snake	Cobras use their tongues to smell with
All buses have more wheels than cars. All cars have more wheels than motorbikes	Motorbikes have more wheels than buses
Rain causes the water in rivers to rise. It rains a lot in Autumn	In Autumn the water levels in rivers drop
All squirrels are red or grey. Jack's squirrel is not red	Jack's squirrel is grey
All Summers are warmer than Autumns. All Autumns are warmer than Winters	Winters are warmer than Summers
All vegetables are healthy. A carrot is a vegetable	A carrot is unhealthy
All humans produce carbon dioxide. Sally is a human	Sally produces carbon dioxide
All igneous rocks are formed by magma. Granite is an igneous rock	Granite is formed by magma
All quadrilaterals have four sides. A square is quadrilateral	A square has three sides

Appendix 8.3 – Contributions of Controls

Contribution of Controls to Real-World Coherence Inference Generation Skill

	Year 5	Year 7	Year 9	Adults
IQ Raw Score	-.23	-.20	-.31	-.38*
Working Memory Capacity	-.22	-.34	-.13	<-.01
Single Word Reading Score	-.41***	-.08	-.06	-.13

*p<.05 **p<.01 ***p<.001

Contribution of Controls to Real-World Elaborative Inference Generation Skill

	Year 5	Year 7	Year 9	Adults
IQ Raw Score	-.23	-.35	-.15	-.22
Working Memory Capacity	-.15	-.21	-.12	-.03
Single Word Reading Score	-.17	>-.01	-.06	-.18

*p<.05 **p<.01 ***p<.001

Contribution of Controls to Counterfactual-World Coherence Inference Generation Skill

	Year 5	Year 7	Year 9	Adults
IQ Raw Score	.04	.01	.17	.08
Working Memory Capacity	.02	.13	.10	.15
Single Word Reading Score	.24	.21	.30	.14

*p<.05 **p<.01 ***p<.001

Contribution of Controls to Counterfactual-World Elaborative Inference Generation Skill

	Year 5	Year 7	Year 9	Adults
IQ Raw Score	.20	.17	.25	.24
Working Memory Capacity	.21	.07	.14	.01
Single Word Reading Score	.11	.08	.08	.06

*p<.05 **p<.01 ***p<.001

Appendix 8.4 – Composite Analysis

*Beta Values for Step 2 Variables when Exploring Accessibility of Knowledge in Step 3
and Real-World Coherence Inference Generation Skill as Dependent Variable*

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.04	.20	.19	.11
Inhibitory Control	.02	.16	.17	.02
Amount of Knowledge	-.53***	-.56***	-.38*	-.36**

*p<.05 **p<.01 ***p<.001

*Beta Values for Step 2 Variables when Exploring Accessibility of Knowledge in Step 3
and Real-World Elaborative Inference Generation Skill as Dependent Variable*

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.11	.21	.12	>.01
Inhibitory Control	.13	.42*	-.10	-.54***
Amount of Knowledge	.01	-.23	-.70	-.19

*p<.05 **p<.01 ***p<.001

*Beta Values for Step 2 Variables when Exploring Accessibility of Knowledge in Step 3
and Counterfactual-World Coherence Inference Generation Skill as Dependent
Variable*

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.75***	.37*	.33*	.42**
Inhibitory Control	.05	>.01	.40*	.24
Amount of Knowledge	.13	.26	.05	.12

*p<.05 **p<.01 ***p<.001

*Beta Values for Step 2 Variables when Exploring Accessibility of Knowledge in Step 3
and Counterfactual-World Elaborative Inference Generation Skill as Dependent
Variable*

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.54***	.46***	.32	.42*
Inhibitory Control	.07	.77**	-.59**	-.02
Amount of Knowledge	.01	.10	.08	.09

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Amount of Knowledge in Step 3 and
Real-World Coherence Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Accessibility of Knowledge	.29	.43**	.51**	.63***
Strength of Belief Bias	.01	.02	.18	.17
Inhibitory Control	.15	.16	.01	.04

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Amount of Knowledge in Step 3 and
Real-World Elaborative Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Accessibility of Knowledge	.53*	.11	.45*	.40***
Strength of Belief Bias	.01	.15	.03	.04
Inhibitory Control	.12	.42*	-.13	-.50***

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Amount of Knowledge in Step 3 and
Counterfactual-World Coherence Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Accessibility of Knowledge	.21	.28	.18	.13
Strength of Belief Bias	.80***	.47**	.27	.44**
Inhibitory Control	.08	>.01	.41*	.25

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Amount of Knowledge in Step 3 and Counterfactual-World Elaborative Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Accessibility of Knowledge	.10	.09	.16	.09
Strength of Belief Bias	.57**	.43**	.34*	.41**
Inhibitory Control	.05	.77***	-.63***	-.05

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Inhibitory Control in Step 3 and Real-World Coherence Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.02	.10	.09	.14
Amount of Knowledge	-.48**	-.47**	-.29	-.23*
Accessibility of Knowledge	.18	.19	.48**	.57***

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Inhibitory Control in Step 3 and Real-World Elaborative Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	>.01	.11	.20	.05
Amount of Knowledge	-.11	-.22	-.64***	-.15
Accessibility of Knowledge	.56***	.01	.37**	.44**

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Inhibitory Control in Step 3 and Counterfactual-World Coherence Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.79***	.43**	.32	.46**
Amount of Knowledge	.08	.17	.03	.08

	Year 5	Year 7	Year 9	Adults
Accessibility of Knowledge	.19	.19	.13	.08

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Inhibitory Control in Step 3 and Counterfactual-World Elaborative Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.57**	.27	.32	.43**
Amount of Knowledge	>.01	.06	.22	.12
Accessibility of Knowledge	.11	.04	.06	.11

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Strength of Belief Bias in Step 3 and Real-World Coherence Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Amount of Knowledge	-.48**	-.43**	-.33*	-.26*
Accessibility of Knowledge	.17	.24	.48**	.57***
Inhibitory Control	.01	.13	.09	.08

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Strength of Belief Bias in Step 3 and Real-World Elaborative Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Amount of Knowledge	-.17	-.16	-.60**	-.11
Accessibility of Knowledge	.57***	.07	.33*	.38**
Inhibitory Control	.02	.12	-.45*	-.33

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Strength of Belief Bias in Step 3 and Counterfactual-World Coherence Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Amount of Knowledge	.08	.30	.01	.17
Accessibility of Knowledge	.02	.05	.24	.09
Inhibitory Control	.02	.12	.45*	.33

*p<.05 **p<.01 ***p<.001

Beta Values for Step 2 Variables when Exploring Strength of Belief Bias in Step 3 and Counterfactual-World Elaborative Inference Generation Skill as Dependent Variable

	Year 5	Year 7	Year 9	Adults
Amount of Knowledge	.03	.06	.02	.05
Accessibility of Knowledge	.08	.20	.09	.09
Inhibitory Control	.02	.64***	-.59**	-.11

Appendix 8.5 – Beta Values for Regression Analysis

Beta Values for Real-World Coherence Regression Models

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.02	.14	.10	.14
Inhibitory Control	.01	.17	.09	.06
Accessibility of Knowledge	.18	.19	.47***	.58**
Amount of Knowledge	-.48**	-.47**	-.31	-.24*

*p<.05 **p<.01 ***p<.001

Beta Values for Real-World Elaborative Regression Models

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.01	.21	.20	.02
Inhibitory Control	.17	.42*	-.04	-.49***
Accessibility of Knowledge	.57***	>.01	.36*	.38**
Amount of Knowledge	-.17	-.23	-.65***	-.11

*p<.05 **p<.01 ***p<.001

Beta Values for Counterfactual-World Coherence Regression Models

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.79**	.42**	.29	.42**
Inhibitory Control	.06	>.01	.43	.26
Accessibility of Knowledge	.20	.19	.19	.11
Amount of Knowledge	.06	.17	.08	.10

*p<.05 **p<.01 ***p<.001

Beta Values for Counterfactual-World Elaborative Regression Models

	Year 5	Year 7	Year 9	Adults
Strength of Belief Bias	.57**	.45**	.35*	.43**
Inhibitory Control	.06	.77***	-.61**	-.04
Accessibility of Knowledge	<.01	.05	.15	.11
Amount of Knowledge	.02	.07	.06	.12

*p<.05 **p<.01 ***p<.001