Philosophy for reading: to what extent can Philosophy for Children improve the way lower-attaining Year 3 readers answer inferential and evaluative questions?

Christopher Fletcher

A Research & Development Project

Submitted for the MSc Learning & Teaching 2020

DEPOSIT AND CONSULTATION OF THESIS

One copy of your dissertation will be deposited in the Department of Education Library via the WebLearn site where it is intended to be available for consultation by all Library users. In order to facilitate this, the following form should be completed which will be inserted in the library copy your dissertation.

Note that some graphs/tables may be removed in order to comply with copyright restrictions.

Surname	Fletcher
First Name	Christopher
Faculty Board	Education
Title of Dissertation	Philosophy for reading: to what extent can Philosophy for Children improve the way lower-attaining Year 3 readers answer inferential and evaluative questions?

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I understand that the Department requires that I shall deposit a copy of my dissertation in the Department of Education Library via the WebLearn site where it shall be available for consultation, and that reproductions of it may be made for other Libraries so that it can be available to those who to consult it elsewhere. I understand that the Library, before allowing my dissertation to be consulted either in the original or in reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of this thesis belongs to me. I permit limited copying of my dissertation by individuals (no more than 5% or one chapter) for personal research use. No quotation from it and no information derived from it may be published without my prior written consent and I undertake to supply a current address to the Library so this consent can be sought.

3. I agree that my dissertation shall be available for consultation in accordance with paragraph 2 above.

Abstract

My research aimed to assess whether the philosophy for children (P4C) programme could help to improve the ways in which a two lower-attaining Year 3 readers answered inferential and evaluative comprehension questions. Additionally, my review of the literature on P4C and reading comprehension revealed that most of the studies tended to focus on quantitative measures and were devoid of qualitive analysis which could have illuminated how P4C could have helped readers to develop their higher-level comprehension skills. In light of this, I wanted to address this qualitative deficit by focusing on a small number of pupils and analysing the potential impact of the intervention from multiple angles. Consequently, I present my finding using the case study approach, which allows me to shed some light on how the pedagogy could help to improve readers' higher-level reading comprehension.

To monitor the impact of P4C on the pupils' responses to higher-level questions, I conducted two semi-structured group interviews and constructed a repeated measures design proposed by Thomas (2016). This involved setting three assessments which were initially intended to have been taken in two 10-week intervals: during the first, no intervention would be run; in the second the P4C sessions would have taken place. However, school closures, as a result of the Covid-19 pandemic, reduced the intervention period by three weeks (which could have reduced its impact) and prevented me from collecting enough data on two pupils from the initial group of four: they had to self-isolate for two weeks prior to schools closing.

Despite this, the test results from the two remaining participants (to whom I allocated the pseudonyms Jade and Ben) indicated that the intervention appeared to help each of them to develop their ability to answer evaluative questions. In addition, the data suggested that the intervention might have had an impact on Ben's inferential answers but perhaps had little impact on the Jade's. In addition, my analyses of the pupils' responses to inferential and evaluative questions, in conjunction with interview data, suggested that the P4C intervention might have helped the pupils to respond to higher-level questions by cultivating key critical thinking skills such as being able to present a line of reasoning which states and justifies a point of view. Moreover, Ben stated that the intervention had inspired him to use questioning to monitor his understanding of what he read more effectively. Using Kintsch and Rawson (2005)' model of comprehension, I suggested that this could have given Ben the chance develop a more accurate text-base and situation model; thus making it easier for him to answer more challenging comprehension questions. Additionally, my discussion evaluates the likelihood of a number of alternative explanations before concluding.

Introduction

Basaraba et al. (2013) frames reading comprehension in three levels of difficulty. The least challenging is the literal level, which involves readers retrieving and making sense of information directly stated in the text. This is followed by the inferential level which places a higher cognitive demand on readers because they have to blend their textual knowledge with other sources of information derived from their general knowledge or past experiences, for example, in order to deduce information which is not directly stated in the text. Finally, the most complex level is the evaluative stage, which requires readers to build on their literal and inferential comprehension by making critical judgements (such as deciding whether a characters' actions are morally justified) in order to create new and personalised meanings.

The genesis of this research project was a Key Stage 2 staff-meeting in which one of my colleagues (the Year 3 teacher) noted that a group of readers in her class were consistently able to answer questions which required a literal level of understanding. However, under the school's assessment system which places children into three categories (higher, middle or lower), she classified them as 'lower-attaining' readers. This was because they found it difficult to answer more challenging comprehension questions which tested their inferential and evaluative skills. In response to this, I was aware a study (Gorard et al., 2015) which suggested that the philosophy for children (P4C) programme (pioneered by Lipman et al. (1980)) could help to improve pupils' reading comprehension. This gave rise to the first aim of my research: I wanted to know whether P4C could equip this group of readers with the skills to demonstrate higher-level comprehension.

Moreover, whilst reading a number of key studies on P4C and reading comprehension (Haas, 1980; Lipman and Bierman, 1980; Williams, 1993; Gorard et al., 2015) to prepare for this research, I realised that they had a tendency to focus on numerical assessment data (test-scores,

for example), and did not, therefore, explore how P4C could help to improve pupils' reading comprehension. This prompted my second research aim: to find out how P4C could help the Year 3 readers to become more skilled at answering higher-level comprehension questions.

In order to shed more light on the nature of inferential and evaluative questions and why I believed that a P4C intervention could help pupils to answer them, I now turn to my review of the scholarly literature.

Literature Review

Introduction

In this literature review, I endeavour to:

- Establish an appropriate theoretical model of reading comprehension with which I can explore the role of evaluative and inferential questions;
- Examine potential barriers preventing primary school readers from being able to answer these questions;
- Present my rationale for using a P4C intervention to improve the way lower-attaining Year 3 readers answer higher-level comprehension questions;
- Identify a gap in the research literature which my research is intended to address.

Having explored the relevant research on these areas, I used my conclusions to inform my research questions.

The traditional 'common sense' view of reading

During the 1950s, the 'traditional' or 'common sense' view was used to explain how burgeoning readers learned to read and comprehend texts (Alexander and Fox, 2013; Perdede, 2008; Goodman, 1967). This model was influenced by Skinnerian behaviourism and Flesch's (1955), work which argues that educators should prioritise the teaching of phonics over whole-word methods of learning to read maintains that readers obtain information from the text by simply employing the decoding skills they learn. In behaviourist fashion, constant reading practice and error-correction, reinforced through rewards, leads to the formation of positive reading habits (Alexander and Fox, 2013; Perdede 2008; Dole et al., 1991). According to this

model, then, comprehension is simply the product of accurate decoding and the passive retrieval of textual information. Meaning is therefore located in the text alone, which is why this model is commonly referred to as a 'bottom-up' view of reading (Perdede 2008; Nunan 1991).

A merit of this outlook is that it draws attention to the importance of being able to decode: a skill which facilitates word recognition, without which reading itself, and therefore reading comprehension, would not be possible (Castles, 2018; Rose, 2006; Gough and Tunmer 1986). Consequently, Pardede (2008) notes that the prevalence of modern synthetic phonics programmes in schools is a testament to this insight.

However, as Samuels and Kamil (1988) make clear, the traditional approach does not sufficiently account for the complex array of processes occurring within the reader's mind as he or she interacts with the text in order to comprehend it. Poor comprehension is simply regarded as a consequence of the reader lacking sufficient phonics skills to understand the text. Therefore, it is clear that the traditional view of reading does not fully accord with Snow's (2002) definition of reading comprehension, which requires readers to play an active part in constructing their own understanding by incorporating information obtained from reading the text. Additionally, since the acts of inferring and evaluating (a major focus of this study) rely on complex interactions between the reader's mind and the text (Cromley and Azevedo, 2007; Cain and Oakhill, 1999; Graesser et al.,1994), this view provides little utility.

A top-down approach

The psycholinguist Kenneth Goodman developed what has since become one of the most frequently acknowledged 'top-down' models of reading: so-called because it placed the

reader's knowledge and thought process at the core of reading comprehension (Davoudi and Moghadam, 2015).

For Goodman (1967), instead of passively extracting meaning by simply decoding the text (as the traditional approach holds), readers must actively engage in psycholinguistic conjecture: taking into account the words they decode on the page and employing their own background knowledge of encountering similar syntactical, semantic and phonological cues in order to predict which words and phrases may appear next. They must then confirm, modify or reject the anticipated perceptual image they have created whilst progressing through the text. According to Goodman, this cyclical guessing game repeats itself until the reader has encoded the text with meaning. Therefore, as Davoudi and Moghadam (2015) and Grabe (2009) note, in Goodman's model, readers are essentially projecting their own understanding onto the text. This contrasts with the bottom-up model in which readers are thought to perform the opposite function: comprehending what they read by extracting meaning from the text.

However, Goodman's (1967) research concentrates on what happens when readers erroneously anticipate upcoming words: that is, they miscue. His model thus focuses exclusively on the processes involved when developing a word-to-word and sentence-to-sentence understanding. It does not, therefore, illuminate the role inferencing and evaluating plays in textual comprehension. Moreover, in direct contrast to traditional bottom-up view of reading, Cambourne (1976) argues that Goodman conflates oral and written modes of language by de-emphasising the need to convert the printed word into oral language; thus denigrating the role of decoding in the comprehension process.

In contrast, the next model I examine blends elements of bottom-up and top-down theories by recognising that both the reader and the text play important roles in the comprehension process.

The interactive schematic approach

Like Goodman, Rumelhart (1980) also recognises the importance of the reader's background and general knowledge in textual comprehension. However, unlike Goodman, Rumelhart (1977) argues that reading comprehension cannot be sufficiently understood using models which present it as a step-by-step linear act of cognition. Instead, he regards it as an even more complex process in which the reader's mind interacts with the text. Moreover, Rumelhart (1980) attempts to highlight the way in which readers retrieve knowledge from their long-term memories and shift it to their working-memories so that it can interact with the information presented in a text in order to facilitate their understanding.

Underpinning this approach is the Kantian concept of 'schemata': mental representations of concepts, objects, places and patterns of behaviour, for example, constructed through interactions between individuals and the world around them. Consequently, a reader's ability to comprehend a text is determined by how successfully he or she can accommodate the information retrieved from the text with existing schemata.

Rumelhart thus presents a more complex view of reading comprehension than Goodman (1967), which Sadoski et al. (1991) praise for being heuristically valuable and providing a counterweight to overly simplistic bottom-up models of reading. Despite this, however, the interactive model has faced criticism on a number of fronts.

For instance, Sadoski et al. (1991) argue that the concept of the schema has been reified in research literature: researchers have too often treated schemata as actual components of cognition rather than theoretical abstractions. In addition, others have regarded the notion of schemata as being insufficient for identifying the complex processes at play during reading or when accessing memory (Davoudi and Moghadam, 2015; Sadoski et al., 1991; Grabe, 1988;

Alba and Hasher, 1983). In light of this, I evaluate how well alternative views of reading could enable me to explain the higher-levels of comprehension in which I am interested.

The Simple View of Reading

Despite Rumelhart's interactive approach, the debate between proponents of the bottom-up approach such as Fries (1962), Gough (1972) and Rozin and Gleitman (1977) (who maintained that becoming a proficient decoder is equates to becoming a proficient reader) and those like Goodman (1973) and Smith (1982), who regarded decoding ability as an epiphenomenon, is known as the 'Reading Wars' (Castles et al., 2018; Kim, 2008). Out of this context, Gough and Tunmer (1986) constructed the Simple View of Reading (SVR).

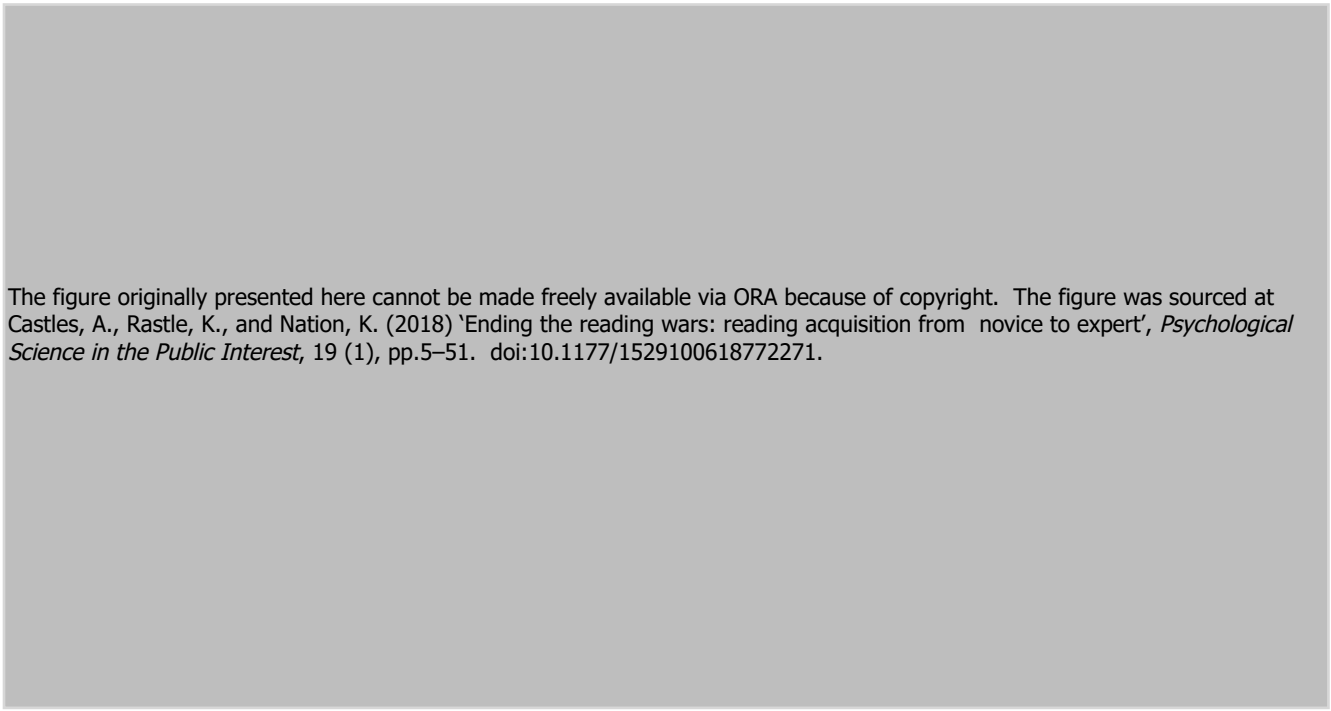


Figure 1: the Simple View of Reading Matrix adapted from Castles et al. (2018, p.27)

Intersecting both bottom-up and top-down ideas, the SVR presents reading comprehension as the product of decoding and linguistic comprehension: the latter being a reader's ability to interpret what they have decoded. If either skill is found lacking, then, the reader's ability to comprehend a text will suffer: if you cannot sufficiently decode the words (their definition of dyslexia) or are very poor at understanding what you read (hyperlexia), then, reading comprehension cannot be achieved.

Many studies (Lervåg, Hulme and Melby-Lervåg, 2017; Catts et al., 2005**;** Hoover and Gough, 1990) highlight the usefulness of the SVR as a formula for predicting reading ability and for identifying in which of the two broad areas a reader may be struggling: decoding or linguistic comprehension.

In light of its utility, Rose (2006) recommends a modified version of it as a framework to be used by teachers to inform their practice in his review of teaching early reading in England. In this instance, he switches the term 'decoding' to 'word recognition'. Although this suggests a broader set of skills, other than decoding alone, may be at work when recognising words (identifying familiar words by sight, for instance), Rose's definition of word recognition is consistent with Gough and Tunmer's (1986) view of decoding because, for him, it is achieved through the development of the reader's core phonics skills.

Although the SVR challenges the basic bottom-up and top-down views by acknowledging the importance of decoding and linguistic comprehension, it has many limitations. For example, Catts (2018) contends that the simplistic nature of the SVR has led to overly simplistic conceptions of reading comprehension; deprived of many crucial nuances. Instead, he argues, comprehension is multidimensional cognitive activity and that a person may demonstrate many levels of comprehension depending on what and why they are reading.

Moreover, the SVR conflates listening and silent reading comprehension because its authors (Gough and Tunmer, 1986; Hoover and Gough 1990; Gough, Hoover and Peterson, 1996) repeatedly make the claim that both process involve very similar skills. Uppstad and Solheim (2011) take issue with this; arguing that the acts of listening and reading place different cognitive demands on individuals, and, consequently, the individual's ability to comprehend.

With this in mind, it should be stressed that Gough and Tunmer (1986) never present the SVR as detailed model of reading, but rather as a formula incorporating the two most significant overarching components of reading to be used to predict reading performance. I therefore agree with the sentiments of Castles et al. (2018) who note that the SVR should be thought of as a useful framework, rather than a model, because it does not provide an account of how both areas of reading develop and function.

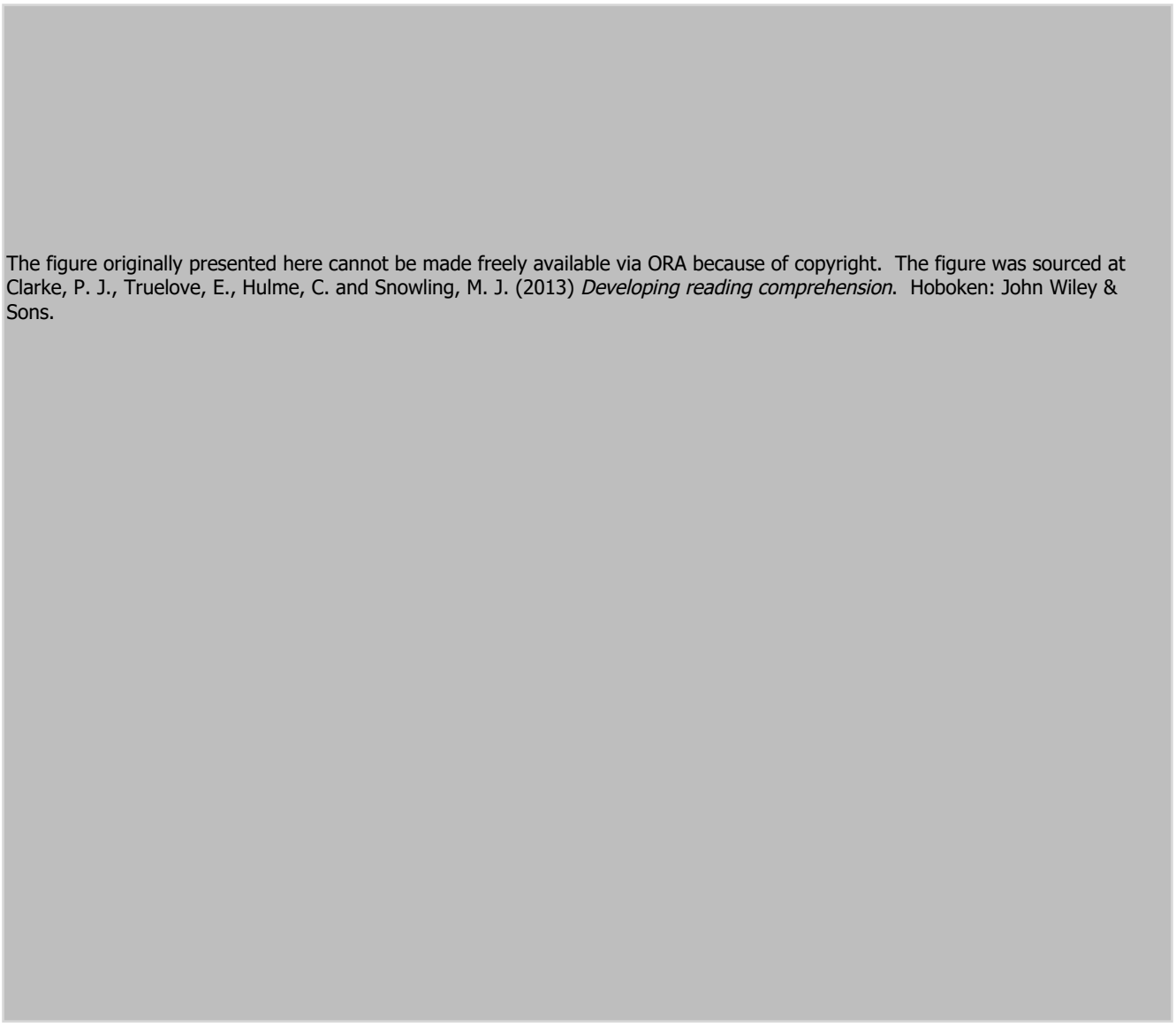
The text-base and situation model

Kintsch and Rawson (2005) extract key elements from the theories of reading presented in Kintsch and Dijk (1983) and Kintsch (1988;1998) and present them in an updated form. This paper is significant because it allows me to place the acts of inference and evaluation into a useful model of reading-comprehension and detail how they are integrated into the comprehension process without using an overly-complex theory which would require my analysing factors beyond the scope of this study.

The text-base

According to Kintsch and Rawson (2005), the initial bottom-up phase of text comprehension (the 'linguistic' stage) entails the reader decoding and recognising words and phrases from the

text. This triggers a semantic analysis whereby word meanings are activated and combined to form a network of interrelated ideas (or propositions) which are collectively known as the 'microstructure'. In order to form a 'macrostructure', a reader must organise the propositions of the microstructure into topics and themes which permeate the text. Together, the microstructure and macrostructure are referred to as the 'text-base' and consist of information explicitly expressed in the text. Figure 2 illustrates this.



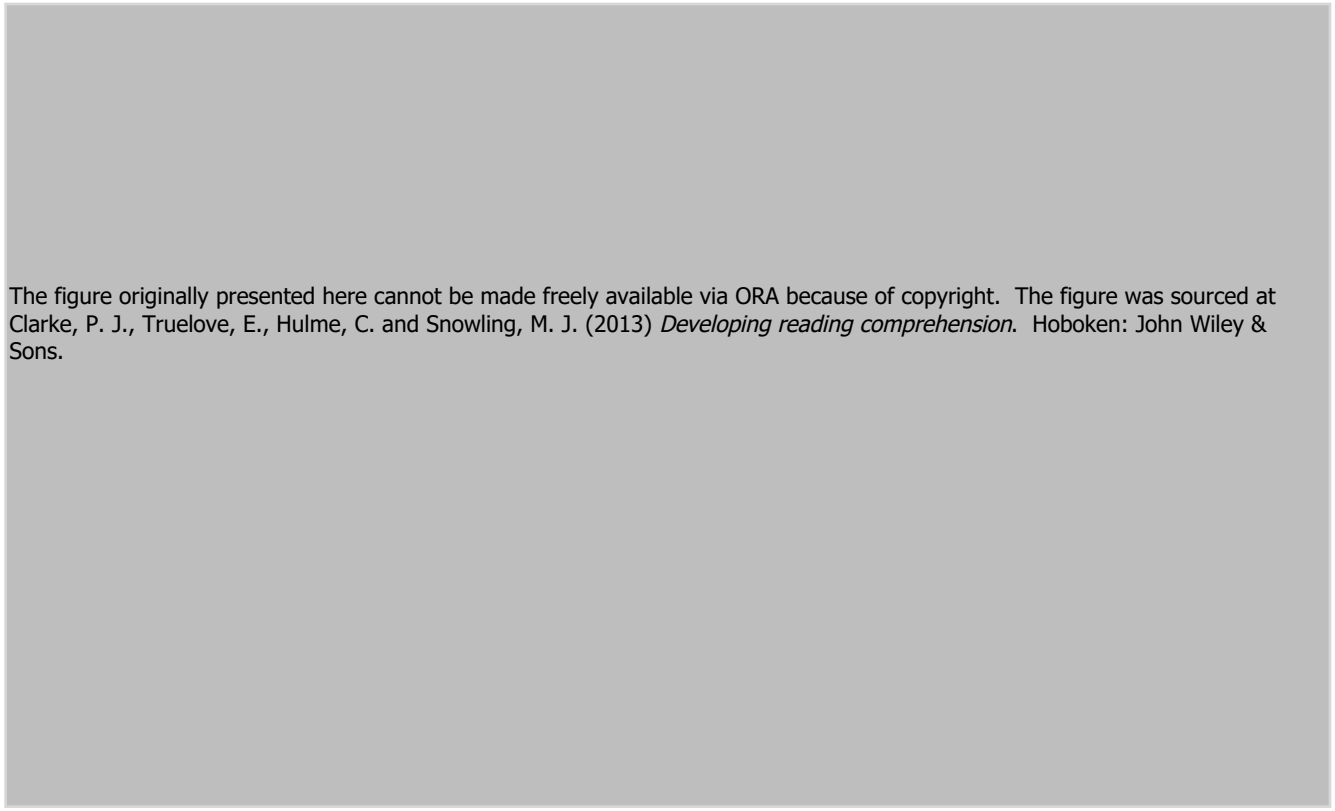


Figure 2: components of Kintsch and Rawson's text-base adapted from Clarke et al. (2013, p.6)

The situation model

Comprehension on a deeper top-down level involves the construction of a 'situation model'. This is a mental representation of the text which the reader becomes actively involved in creating and constantly refining by integrating the information retrieved from the text-base with his or her own general knowledge and reading goals in order to fill in textual gaps, make predictions and draw conclusions, for example. The reader must formulate both a functioning text-base and situation model in order to comprehend the text (Kintsch and Rawson, 2005). Figure 3 illustrates this.

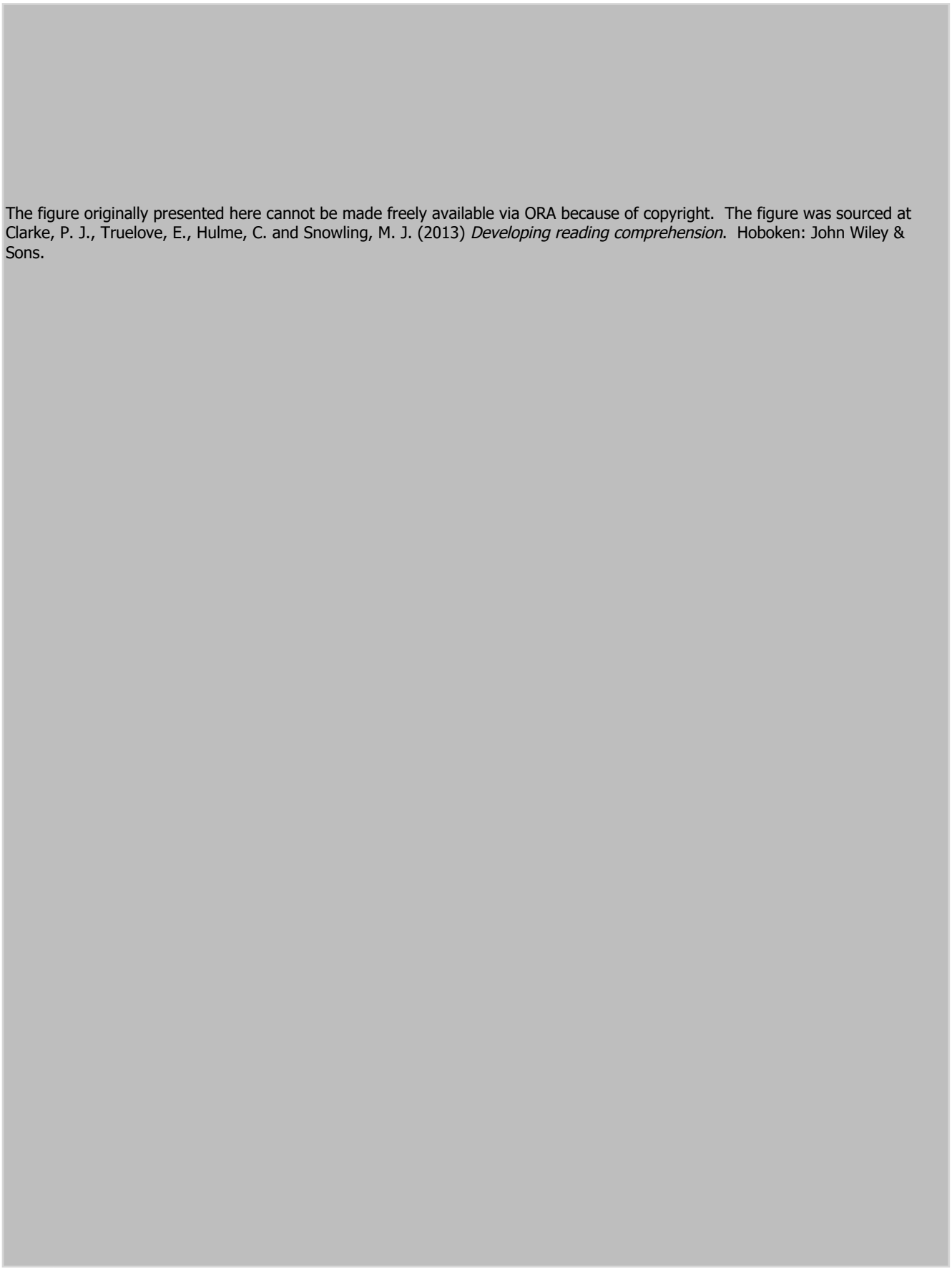


Figure 3: a graphic of Kintsch and Rawson's model adapted from Clarke et al. (2013, p.7)

Justification for incorporating Kintsch and Rawson's model into my theoretical outlook

As demonstrated, Gough and Tunmer's (1986) SVR and Scarborough's (2001) reading rope acknowledge the importance of both top-down and bottom-up processes: areas which were neglected by the traditional phonics-centred approach advocated by Flesch (1955) and the top-down view developed by theorists such as Goodman (1967). Yet, since these are frameworks, they do not provide conceptual tools to help explain how these dimensions of comprehension function. Moreover, whilst Rumelhart's (1980; 1977) interactive approach presents schematic tools for detailing how readers need to accommodate their existing understanding of various concepts with those they encounter on a textual level, Kintsch and Rawson's (2005) model is more useful to my study because it outlines different degrees of reading comprehension; thus helping me to situate theoretically the roles played by inferential and evaluative questions more precisely as I explore them.

Literal questions and text-based inferencing

According to Basaraba et al. (2013), reading comprehension can be viewed as a hierarchy comprising the least cognitively demanding literal level and the inferential and evaluative stages which are respectively more challenging.

Comprehension at the literal level requires readers to retrieve information explicitly stated in the text (Basaraba et al., 2013). In the context of Kintsch and Rawson's (2005) model, when readers answer questions testing their literal understanding (literal questions), they would need to operate on the level of the microstructure within the text-bases they have constructed to demonstrate their understanding or words, sentences and passages (Basaraba et al., 2013; Carnine et al., 2010; Perfetti et al., 2005).

The following passage provides an example of a literal question, since the answer to the question (what colour is Tom's hair?) is explicitly stated in the text following it, inviting the reader to locate and decode it.

> Question 1: In the sentence below, what colour is Tom's hair?
>
> As Tom looked in the mirror, he could see that his brown, wavy hair was styled to perfection.

Even though my study will not focus on literal questions, it is worth considering how readers develop a literal understanding of texts because the skills employed to achieve this provide the foundation on which the more cognitively challenging inferential and evaluative skills are built (Basaraba et al., 2013; Bowyer-Crane and Snowling, 2005).

For example, a number of theorists (such as Cromley and Azevedo, 2007; Bowyer-Crane and Snowling, 2005; Cain and Oakhill, 1999; Graesser et al.,1994) point out that in order to maintain an elementary level of coherence between the meanings of words, phrases and sentences, readers must actively exhibit anaphoric awareness by inferring inter-sentence connections in their minds. Given that such inferences are critical to establishing an intelligible text-base, Kintsch and Rawson (2005) refer to these as 'text-based inferencing'. When answering question 1, for instance, the reader would need to recognise that the pronouns 'he' and 'his' are linking the subject of the sentence (Tom) with an object (his hair) in order to answer the literal question. It is important to note, then, that even the process of answering literal questions is deeply rooted in the reader's ability to make simple inferences: as Dole et al. (1991) stresses, inference skills are at the heart of the comprehension process.

Knowledge-based inferential questions

As Kispal (2008) highlights, the number of scholars who have studied the process of inferring has led to the field becoming somewhat convoluted with competing and overlapping concepts and labels used to distinguish between various forms of inference. For example, Pressley and Afferback (1995) have identified nine types of inference as varied as being able to draw conclusions, confirming or refuting previous inferences and filling in omitted information, whilst Graesser et al. (1994) outline a total of thirteen: such as inferring authorial intent, characters' emotions, the causal consequences of actions and events and whether the inferences are made 'online' (during reading) or 'offline' (afterwards). Moreover, in thorough reviews of the literature on inference, Kispal (2008) and Graesser et al. (1994) note that there is no scholarly consensus regarding how many inference types there are or any agreement on how they should be labelled.

In light of this, a detailed analysis of the nuances between the numerous conceptions of inference categories would be beyond the scope of this literature review. Therefore, in order to navigate this complex terrain, Kispal's (2008) review of inference scholarship is invaluable. She notes that theorists such as Graesser et al. (1994), Barnes et al. (1996) and Cain and Oakhill (1999) tend to maintain a dividing line between two broad overarching inferences which encompass the inferential categories delineated by Grasser et al. (1994) and Presley and Afferback (1995): namely the aforementioned text-based inferences and 'knowledge-based' inferences.

In contrast to text-based inferences, 'knowledge-based' inferences (sometimes called 'gap-filling' or 'elaborative' inferences) require the reader to integrate their own general knowledge and real-life experiences with the text-base they have constructed in order to connect events and fill in information not directly stated but alluded to by the text (Kispal, 2008; Kintsch and

Rawson, 2005; Cain and Oakhill, 1999). The following Knowledge-based inferential question would test the reader's ability to make this type of inference.

> Question 2: in the following text, how do you think Katie was feeling?
>
> As Katie's face became increasingly red, she yelled at the top of her voice and threw her plate on the floor.

Since the sentence does not directly state how Katie was feeling, the reader would need to mentally construct the image of Katie presented in the text and make use of his or her own life experiences and general knowledge in order to recognise that the clues the text provides (Katie's increasingly red face, her yelling and the fact that she threw her plate) tend to point towards a person feeling angry. In other words, the reader's construction of a situation model enables them to make the knowledge-based inference necessary to answer the question. Importantly, it is knowledge-based inferential questions which my study is interested in: not the questions which test text-based inferencing skills.

Evaluative Questions

The most cognitively challenging stage of comprehension, according to Basaraba et al. (2013, p.356), is the evaluative level which is sometimes referred to as 'critical or applied understanding' since it requires 'readers to analyze and critically interpret the text based on their prior knowledge and experiences.' The elevation in cognitive challenge becomes evident when considering that this involves readers having to utilise multiple thinking skills like being able to generate a number of plausible answers to a question (divergent thinking), and rationally examining information in order to reach and substantiate their own judgements about an issue (critical thinking), whilst drawing on their memories of personal experiences and general knowledge in addition to the textual knowledge they have previously developed (Basaraba et

al., 2013; Vacca et al., 2009; McCormick, 1992; Rupley and Blair, 1983). In the eyes of Herber (1970), such higher-level thinking creates new personalised meanings for the reader that move beyond the realm of the text. This suggests that the evaluative level of comprehension enables readers to construct their own unique interpretations of texts in a way that is not possible with the literal and inferential levels, since they do not possess the same potential for generating multiple answers.

Before I demonstrate this, it is important to recognise that lower levels of reading comprehension provide the foundational knowledge on which a reader can proceed to build an evaluative understanding (Basaraba et al., 2013): a reader's literal understanding of what is directly stated in the text in conjunction with his or her knowledge-based inferential understanding of the relationships between certain characters and events, for example, will determine what information they have available to synthesise and analyse critically in order to construct an evaluative understanding of the text.

By offering and analysing various answers to the following evaluative question, I demonstrate how many of these elements can come into play.

> Question 4: In CS Lewis's *The Lion the Witch and the Wardrobe*, is Edmund a good or bad person?

This evaluative question is asking the reader to make a critical judgement on Edmund's morality. Viewing this through the lens of Kintsch and Rawson's (2005) model, I would argue that the richness of the situation model the reader has constructed is crucial to the way in which he or she could answer the question.

In the first half of the novel, for example, Edmund is shown mercilessly tormenting his little sister, Lucy; betraying her when she depends on him; having a gluttonous craving for Turkish Delight, which enables the White Queen to recruit him to her cause and betray his own family. Therefore, a reader whose situation model is limited to a basic image of this half of the tale may conclude that Edmund is an immoral character with no regard for others, including his own family (whom he betrays for the prospect of receiving his desired confectionary). However, a reader with a more finely-tuned situation model may be able to offer a more sophisticated answer.

There are important textual details which could give the reader a more critically nuanced perspective. For example, when initially encountering the White Queen, Edmund is ignorant of who she really is and unaware of her nefarious intentions. Given his ignorance, and the fact that she is offering him gifts, he may even consider her to be quite benevolent and while unable to appreciate the true danger his siblings would be in if he leads them to her. In addition, the text makes it clear that Edmund's obsessive craving for Turkish Delight is the result of an enchantment he has been subjected to; indicating that he is not responsible for the greed which affects his judgement. Moreover, a situation model more familiar with the latter half of the story will acknowledge that Edmund eventually fights alongside his brother, Peter, in the great battle against the White Queen; thus demonstrating courage and loyalty which is recognised when Edmund is subsequently knighted.

A reader with such a situation model of the text may be inclined to view Edmund perhaps as initially more ignorant and impressionable than bad and as a character who eventually proves himself to be more good than bad by atoning for his prior behaviour when risking his life to fight alongside his brother in a bid to defeat the menacing tyrant of Narnia.

As Kintsch and Rawson (2005) note a reader's experiences are key to the formation of the situation mode. With this in mind, readers who can call to mind their own experience of making a mistake which deeply upset someone whom they cared about may come to a more sympathetic judgement of Edmund's moral character, whereas those who recall a time when they were deeply hurt when feeling betrayed by someone's actions, may arrive at a more damning verdict of Edmund.

Although evaluative questions can offer more room for multiple responses, there are many factors to consider which can hinder a reader's ability to answer both evaluative and knowledge-based inferential questions.

Barriers to both inferencing and evaluating

A reader's ability to comprehend a text can be affected by many variables: for example, difficulties with cognitively processing textual information to possessing a restricted vocabulary which limits the number of words a reader is able to understand (Clarke et al., 2013; Kispal, 2008; Kintsch and Rawson, 2005; Scarborough 2001). Therefore, I have decided to focus only on the potential barriers hindering the development of a reader's knowledge-based inferential and evaluative abilities which I intended to address with my intervention. I therefore refrain from discussing areas such as decoding skills, for example, although they are considered to be crucial to comprehension (Kintsch and Rawson, 2005; Scarborough 2001; Gough and Tunmer, 1986). After examining these barriers, I introduce my intervention and justify why I believe it may be able to help readers struggling with these barriers.

Poor written and spoken language skills

A study by Nation et al. (2004) found that poor comprehenders tended to exhibit notable weaknesses in their spoken language skills, which were present in their early years of primary school. Clarke et al. (2013), referring to the same study, suggest that such a finding indicates that weaknesses in spoken language skills are present prior to the development of reading skills, and, therefore, a child's skill with spoken language is a good indicator of reading comprehension ability. In line with this thinking, studies by Carretti et al. (2016) and Munro (2011) have reported improvements in the reading comprehension abilities of children with poor language skills following interventions focusing on developing their spoken vocabulary. In addition, Wolf et al. (2019) found vocabulary to be a major indicator of both reading and listening comprehension and Tok and Mazl (2015) and Carretti et al. (2016) point out that poor comprehenders with weak oral and written language skills may experience difficulties expressing their own ideas in both modalities. This would thus hinder their ability to answer comprehension questions both orally and in written-form.

Poor comprehension monitoring

Clarke et al (2013, p.16) note that good comprehenders actively monitor their comprehension, develop their sensitivity to important textual details, possess 'zero tolerance of inconsistency' and adapt their situation models accordingly. Therefore, readers not using this approach may produce less coherent situation models; potentially diminishing the quality of their inferences and evaluations. These sentiments are echoed by many others (Kispal, 2008; Cain and Oakhill, 1998; Graesser et al, 1994) who regard active comprehension monitoring as a primary skill typically exhibited by readers who consistently draw logical inferences and lacking in those who do not.

Struggling to access relevant domains of knowledge

Additionally, many researchers (Cain et al., 2004; Harrison, 2004; Pressley, 2000; Long et al., 1996) recognise the importance of being able to access relevant general, background or textual knowledge when making knowledge-based inferences. Importantly, these inferential skills provide the foundation on which a reader is able to build his or her evaluative understanding (Basaraba et al., 2013). In order to answer both question types, then, a reader needs to be able to determine when a piece of knowledge is useful and needs to be held in his or her working memory for long enough to make the necessary evaluation or knowledge-based inference. This could be a memory of a personal experience perhaps similar to that of a character's (background knowledge); an awareness of the broader cultural or historical context in which the text was published or is set, for example, (general knowledge); and the reader's ability to recollect significant episode in a story (textual knowledge) which may, in turn, prompt the reader to draw from his or her background or general knowledge (Clarke et al., 2013; Kispal, 2008).

Whilst Long et al. (1996) note that good comprehenders appear to have a tendency to perform this task automatically, a study by Cataldo and Oakhill (2000) found that good comprehenders consciously used reading goals to regulate their approach to the text, whereas this did not seem to affect the ways in which poorer comprehenders approached the text. A notable example of a reading goal could be searching for information useful for answering questions: this could help readers to focus on details they believe to be relevant to meeting this aim. It follows, then, that pupils must learn to be mindful of their reading goals in order to develop higher-level comprehension skills.

The importance of critical thinking

My understanding of what critical thinking means is informed by Leicester and Taylor (2010, p.2) who present it as 'a toolbox of skills which enable children to think more deeply and

clearly about what they believe (and what they read or are told in the media etc.), and about what they should do' rather than uncritically accepting someone else's point of view. According to Leicester and Taylor (2010), some of the key elements of rudimentary critical thinking at the primary school level include, but are not limited to, pupils:

- **Developing their own points of view:** using a logical line of reasoning supported by evidence to reach their own judgement on a subject rather than uncritically accepting someone else's point of view;
- **Asking questions:** this is to guide and challenge thinking: for example, do I have good reasons to believe this?
- **Analysing**: breaking down information into smaller details in order to examine them more closely;
- **Evaluating**: examining the merits of beliefs, decisions, values and the reasons for holding them;
- **Thinking rationally**: using reasoning based on evidence and logic to justify viewpoints. This entails pupils being able to detect strong and weak arguments and evidence whilst being mindful of contradictions in lines of reasoning.

Attaining an evaluative understanding of a text requires the reader to think critically by analysing and evaluating the importance of information in order to form and justify their own judgement regarding a particular subject (Basaraba et al., 2013; Vacca et al., 2009; McCormick, 1992; Rupley and Blair, 1983). In addition to facilitating pupils' evaluative understanding of texts, such thinking skills are regarded as being an important part of a primary pupil's broader educational development, as evidenced by the National Curriculum (DfE, 2013), which makes numerous references to pupils being taught to apply critical thinking skills throughout its programmes of study: from reader's providing justifications for their judgements

to analysing and evaluating works of art, designs they have built, and computer programs they have created.

In addition, there appears to be a link between knowledge-based inferences and critical thinking. For example, a number of academics (such as Dwyer, 2017; Swatridge, 2014; Elder and Paul, 2002) whose work focuses on critical thinking regard such inferences as being integral to the reasoning process since, in order to draw logical conclusions the thinker has to infer information. In order to illustrate this, Swatridge (2014) references Bogart's famous claim that a dog biting a man is not news since it occurs relatively frequently but the opposite occurrence is news; noting that in order to understand Bogart's argument we must infer that a man biting a dog is unusual and therefore news-worthy.

Moreover, it could be argued that the process of answering a knowledge-based inferential question also requires a degree of critical thinking because the reader must be able to evaluate which pieces of information from various knowledge domains are conducive to drawing the inference before making it: as Dwyer (2017, p.117) recognises, there is some overlap between inferring and evaluating because 'good inference ability depends on good evaluation.' In addition, Clarke et al. (2013) note that good comprehenders draw logical knowledge-based inferences partly by having an intolerance towards contradiction. This is a trait of thinking rationally (Leicester and Taylor, 2010).

In light of the relationship between critical thinking, evaluating and inferring, I now present my intervention programme and outline why I decided to implement it to support Year 3 readers who struggled to answer evaluative and knowledge-based inferential questions.

My intervention

Philosophy for Children

Pioneered by the work of Matthew Lipman (who established the Institute for the Advancement of Philosophy for children (IAPC) at Montclair State University in 1974) and collaborators such as Ann Sharp, Philosophy for Children (P4C) is an inquiry-based pedagogy which aims to develop the critical thinking skills of primary and secondary-aged children in order to help them to 'become more thoughtful, more reflective, more considerate and more reasonable individuals' (Lipman et al., 1980, p.15).

The community of Inquiry

Lipman et al (1980) envisages the community of inquiry (CoI) as a group of pupils asking questions, sharing their experiences of, and thoughts regarding, a limitless number of subjects, whilst considerately evaluating and responding to each other's points of view through the engagement in purposeful and cumulative dialogue facilitated (as impartially as possible) by the teacher in a social environment in which everyone feels comfortable enough to freely speak, agree or disagree with each other, and change their minds (Lipman, 2003; Haynes, 2002; Lipman et al., 1980). P4C is therefore also routed in the Vygotskian notion that speech is critical to formulation (or reformulation) of thought itself (Jenkins and Lyle, 2010).

What is more, a number of studies suggest that P4C could help to improve pupils' critical thinking and reading comprehension skills.

Studies monitoring P4C's impact on reading comprehension and critical thinking skills

In the United States, Lipman and Bierman, (1980) found that 20 children from the 5th grade (aged 10-11) participating in 18 40-minute P4C sessions over a nine week period made significant gains in their logical reasoning and reading comprehension scores in a post-intervention when compared with the results of the 20 pupils in the control group. The gap

between the reading comprehension scores of the experimental group and control (in favour of the former) was still regarded as statistically significant in an assessment 2 ½ years after the intervention period.

On first glance, this early study suggests that regular interventions in P4C could help to boost the comprehension abilities of participants. However, the study does not go into detail about the types of comprehension questions asked. Additionally, Trickey and Topping (2004, p.371) note that since the sample sizes were very small and Lipman himself (a highly-trained philosophy professor and founder of the P4C pedagogy) conducted the interventions, the results 'might not be typical.' Lipman also acknowledges his own involvement as raising issues with the typicality of the findings and points towards a study conducted by Haas (1980), where classroom teachers trained in the pedagogy delivered the intervention, as being the next important step in P4C research (Lipman et al.,1980).

Haas (1980) used a larger sample size (n= 2 x 200) of 5th and 6th grade children (ages 10-12) with control group and P4C group sessions being led by teachers: the latter were trained to use the P4C approach. Statistically significant gains were reported amongst the P4C group in critical thinking and reading comprehension skills, which tested areas such as pupils' abilities to draw inferences. Although these results look promising, the study does not specify the types of inferences assessed in the reading test or detail the exact gains made in this area of reading comprehension.

Likewise, in England, Williams (1993) monitored the effects of 15 pupils (aged 11-12) who had 27 P4C sessions (each one hour long and delivered by teachers) over a 2 year period. When their pre and post-assessment scores were compared with the control groups (consisting of 17 pupils), the experimental group made significant gains in reading comprehension when

completing the London Reading Test and, judging from the questionnaire data, an improvement in critical reasoning skills: neither trend was found in the control group.

Despite the fact that the London Reading Test assessed literal, inferential and evaluate levels of comprehension, the study reports the raw-scores of the assessments and does not break-down the results to monitor the potential effects of the intervention on the various levels of comprehension. It is, therefore, hard to determine the effects of the intervention on each level of comprehension, even though the findings indicate that the programme could have a beneficial impact on participants' reading comprehension.

Gorard et al. (2015) monitored the impact of P4C sessions delivered by SAPERE-trained teachers (the main organisation promoting P4C in the UK) on attaining in reading, writing, maths and cognitive ability of Year 4 and 5 pupils from 48 schools across England. A major finding relevant to my study is that, judging from their analysis of pre and post-test scores using standardised Key Stage 2 assessments, they found that one academic years' worth of 1 hour P4C sessions per week resulted in a significant improvement in reading comprehension scores: the 1,550 pupils in the intervention group made approximately two additional months progress when compared with the 1,609 in the control group, and the interventions had the biggest impact on the test results of disadvantaged pupils (those eligible for free school meals). This is latter finding is particularly important since pupils from lower socioeconomic backgrounds tend to struggle to develop language and reading skills at the same rate as their more affluent peers (Save the Children, 2015; Gilkerson et al., 2017; Hart and Risley, 1995).

In contrast to these positive outcomes of P4C on reading comprehension, Fields (1995), working in Yorkshire, examined the effects of P4C using a population of 123 pupils (aged 7-8) selected from a state school and a private school and randomly allocated to experimental or control groups. The study took place over a two year period. During which time, Fields found

no significant difference in the reading scores of the control group and the experimental group. However, as Trickey and Topping (2004) point out, the study does not note the frequency or length of the P4C sessions or who delivered them. Additionally, it does not go into detail about the particular areas of reading comprehension assessed.

It should be noted, however, that Fields (1995) did report statistically significant differences in favour of the experimental group in her evaluation of logical reasoning skills using numerous standardised measures such as the New Jersey Reasoning Test and the Wechsler Intelligence Scale for Children.

Jenkins and Lyle (2010) analysed seven video recordings of seven P4C sessions conducted once a week with a Year 5 class in Wales. Focusing on the verbal interactions of four lower-attaining readers, they found that, as the intervention progressed, all four pupils began to demonstrate significant advances in their critical thinking skills, as evidenced by several key actions which began to emerge as the intervention period progress: pupils began to justify their arguments with evidence without being prompted; became more evaluative in order to consider the merits of different viewpoints; and increasingly referenced their own personal experiences in order to make connections with, or to contrast with, the opinions of others. These findings are important because being able to access and reference personal experiences and justify points of view with evidence is crucial to making both knowledge-based inferences and evaluating, whilst being able to critically evaluate different perspectives is key to reaching an evaluative understanding (Clarke et al., 2013; Basaraba et al., 2013; Kispal, 2008; Cain and Oakhill, 1998).

Despite the positive outcomes the intervention was reported to have on critical thinking, since the sample size is so small (4 pupils in a single primary class in Wales), the results of such a study are not generalisable. Additionally, although the study notes that recordings from seven

sessions were analysed and that the teacher delivering the sessions was trained in P4C, it does not specify whether this was the exact number of intervention sessions the focus pupils had, the approximate length of each sessions or the exact training the facilitator had.

The rationale for using P4C as an intervention to support lower-attaining Year 3 readers

As I have shown above, thinking critically is a key factor in being able to develop higher levels of reading comprehension. I therefore wanted to initiate an intervention which aimed to develop critical thinking skills with the intention of improving the ways in which lower-attaining Year 3 readers answer evaluative and knowledge-based inferential questions. Moreover, given that the research literature suggests that P4C can have a beneficial impact on both critical thinking and reading comprehension, it appeared to be an appropriate intervention to achieve this aim.

Furthermore, I believed that P4C might help to address the other potential barriers not studied in the literature. As noted above, pupils with poor comprehension skills can struggle to:

- Access relevant domains of knowledge;
- Monitor their comprehension;
- Use language to effectively present answers.

Significantly, Haynes (2002) notes that in seeking answers for philosophical questions, children draw on various domains of knowledge: especially their personal experiences. Therefore, engaging in P4C inquiries, which are driven by philosophical questions prompting pupils to access what they know (Haynes, 2002), might help pupils to improve the way they

access different areas of knowledge when answering higher-level comprehension questions. In addition, Carretti et al. (2016) suggest that some poor comprehenders may struggle to express themselves verbally or in written form. It seems to me, then, that P4C might have the potential to improve the way participants express themselves when answering comprehension questions because one of its goals is to encourage children to use language to express themselves clearly and precisely (Lipman et al., 1980). Moreover, I believe that the programme might also be able to help pupils to improve their comprehension monitoring because evaluating different viewpoints involves thinking rationally and, therefore, being mindful of contradictions (Leicester and Taylor, 2010). Similarly, effective comprehension monitoring involves having an intolerance towards contradiction (Clarke et al., 2013).

A gap in the literature on P4C and reading comprehension

It is also important to note that, although most of the aforementioned P4C studies (Haas, 1980; Lipman and Bierman, 1980; Williams, 1993; Gorard et al., 2015) indicate that P4C interventions can have a positive impact on pupils' reading comprehension, they do not illustrate the qualitative processes involved in any significant transformation in reading comprehension ability influenced by P4C.

In light of this gap in the literature, my study uses a case study approach and harnesses the literature I have explored to interpret the findings; thus enabling me to analyse and evaluate multiple perspectives from my participants in order to suggest possible answers to following research questions.

Research questions

1) Could a P4C intervention improve the way lower-attaining Year 3 readers answer knowledge-based inferential questions and evaluative questions?
2) How could a P4C intervention improve the way lower-attaining Year 3 readers answer knowledge-based inferential questions and evaluative questions?

Methodology

Introduction and note on the impact of Covid-19

Owing to the Covid-19 epidemic, schools across England closed on 18th March 2020 (BBC, 2020). I was therefore unable to complete my research: participants missed out on three planned intervention sessions; and two pupils missed a further two sessions, the final group-interview and assessment because they had to self-isolate for two-weeks I therefore acknowledge how far this affected certain parts of my study as I explore the following areas:

- My research approach;
- The context of the study and my position as a teacher-researcher;
- The participants and how they were selected;
- Setting-up the intervention in collaboration with other teachers;
- My chosen methods of collecting data;
- Ethical considerations.

My research approach

I adopted an eclectic approach which recognises the potential usefulness of mixing data-collection methods traditionally assigned to either positivist or interpretivist paradigms; employing whichever combination of qualitative and quantitative methods are felt to best serve the purpose of the study (Creswell, 2003; Cohen et al., 2011).

However, it is important to note that this study uses basic numerical data in a limited capacity: reading comprehension test scores were used when selecting participants (see selecting participants below) and are used in subsequent sections of the paper, when comparing results before and after the intervention period. This is because, as I have shown in the previous chapter, I believe that the dearth of qualitative research on P4C and its impact on reading comprehension marks a significant gap in the scholarly literature. Therefore, with the intention of helping to plug this gap, I present my research as an interpretive case study since it enables me to employ various research methods in order to examine the case in-depth from multiple angles; exploring how or why the desired effect of my intervention might or might not occur (Thomas, 2016; Cohen et al., 2011; Simons, 2009; Yin, 2009).

There are, however, some potential shortcomings of this approach which I now address.

Addressing the limitations of the interpretive case study

Although a lack of generalisability has been a key criticism of the case study (Thomas, 2016; Cohen et al., 2011), the limited scale of my research project, which involved a maximum of five participants, led me to rule out generalising. Moreover, by triangulating various sources of data the case study approach can construct a much more detailed picture than say quantitative methods alone can capture; offering particularisation rather than generalisation (Thomas, 2016). Moreover, by providing sufficient detail, my findings may be relatable and, therefore, useful to other teachers in similar situations: Bassey (1981, p.85) regards this as an 'important criterion for judging the merit of a case study.'

Another chief criticism of the case study is the charge that its reliance on the researcher's interpretations of events and data inevitably produces a biased and therefore unreliable study (Thomas, 2016; Cohen et al., 2011; Simons, 2009). Although my study will mainly rely on

interpretivist methods (interviews and observations, for example), I endeavour to strengthen the credibility of my critical analysis by considering alternative interpretations of data where possible, and taking into account multiple viewpoints derived from various sources of data: including quantitative test-results which can be less susceptible to biased interpretations. (Thomas, 2016; Cohen et al., 2011; Adelman et al., 1980).

The context of the study and my professional position within it

My research took place in a one-form entry, primary academy judged to be outstanding twice by Ofsted (2008; 2015). The school has 210 pupils on its role and is situated in a relatively affluent village in Leicestershire: according to internal records, the school receives pupil premium funding for 8% of its pupils; fewer than 1 % have free school meals; 10% of pupils are classed as having special educational need and/or a disability (SEND); and 4% speak English as an additional language (EAL). Although the study focuses on Year 3 pupils, I am also the school's Year 5 teacher but had no experience working with my target-pupils prior to conducting my research.

Selecting participants, constructing a repeated measures design and school closures

My focus was Year 3 pupils who exhibited good literal comprehension but consistently struggled to answer evaluative and knowledge-based inferential understanding questions. To identify these pupils, their teacher and I constructed a quasi-experimental approach which involved taking repeated measures of a single group of participants on several occasions before and after the intervention period (Thomas, 2016). Conducting this research in a one-form entry school in conjunction with the fact that so few pupils (five) eventually met criteria ruled out the possibility of assembling a group of similar participants to compare results and control for any progress which might have occurred regardless of the intervention (Cohen et al., 2011).

However, Thomas (2016) notes that by taking repeated measures of the same group's progress over a period of time before and after the intervention, the single group can serve as its own control. Figure 4 helps to illustrate how I intended to use this method.

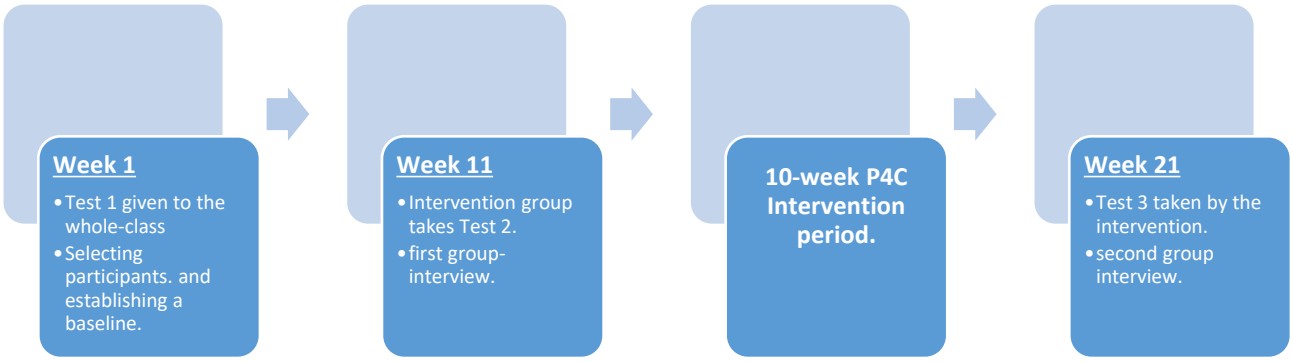


Figure 4: my repeated measures design

Using this design, I could develop a sense of the intervention's possible effectiveness by comparing the progress made between the first and second test (a ten-week period before the intervention period) with that made between the second and third assessments immediately after the intervention had finished: an equal period of time. a significant difference between the first comparison and the second would suggest that the intervention could have had an effect (Thomas, 2016).

As a consequence of school-closures, however, the intervention period was reduced from ten-weeks to seven, and I had to conduct the final assessment and group-interview three-weeks early. This might have curtailed some of the benefits the P4C programme could have had on pupils' higher-level comprehension skills.

Notwithstanding school-closures, I recognised that, on its own, this is still a very limited design with a number of limitations which prevent me from being able to write confidently about the effects of the intervention.

Addressing the limitations of the repeated measures design

There are several extraneous variables which could have influenced any effect on reading comprehension and would have been impossible to control for (Thomas, 2016; Cohen, 2011). For example, a participant might have read more at home between one of the periods; had a better sleep the night before one of the test; or for whatever reason it may have been that they would have made significant progress regardless of the intervention (Thomas, 2016; Thomas, 2009; Cohen et al., 2011). Although this could be remedied to some degree by applying measures such as counterbalancing, which would have involved splitting the participants into two groups and changing the orders in which they complete the conditions, this would have required more time and a larger selection of participants than I had (Allen, 2017).

Before we initiated this design, the Year 3 class-teacher and I endeavoured to reduce any influence that fewer or extra reading activities might have had on pupils' comprehension results by deciding that the participants should have the same degree of reading comprehension practice in the classroom as they did between Test 1 and 2: in addition to reading independently for 15-minutes per school-day, pupils would have one 15 minute, teacher-led guided-reading session each week in which they would practise answering literal and inferential questions; they would also read on a one-to-one basis to an adult (alternating between the teacher and LSA) for approximately 10 minutes twice a week, and. These sessions involved adults listening to pupils read a page or two and asking three literal and three inferential questions. Nevertheless, we recognised from the onset that achieving this each week would not always possible owing to circumstances beyond our control: participants being absent, for example.

Yet, despite such drawbacks, Thomas (2016, p158) states that examining the results of this sort of design in combination with data retrieved from multiple sources 'may well add an invaluable dimension'; thus enriching my analysis and discussion by presenting things from another angle.

Mode of analysis

To extract more value from these data, I used the constant comparative mode of analysis (recommended by Thomas, 2016;2009) which involved repeatedly analysing and comparing the participants' answer in order to elicit themes which I used to inform my analysis. Appendix D shows how I typed up and arranged a selection of pupils' responses to a number of inferential and evaluative questions from each paper. Doing this enabled me to conduct an in-depth qualitative analysis of the responses; thus adding another dimension to my interpretation of the results.

Designing reading comprehension tests and addressing their limitations

The Year 3 teacher and I jointly constructed three reading comprehension test (see Appendices A, B and C) specifically designed to assess how well pupils could answer evaluative and knowledge-based inferential questions. Three papers we produced with each comprising two short, age-appropriate extracts (one non-fiction and one fiction) and a total of 18 questions (three of each question-type following each text). The first paper served the joint purpose of helping us to select potential participants whilst establishing baseline reading comprehension data.

Using the standardised reading tests (NFER, 2019) we had access to as an alternative would have saved us a lot of time. Additionally, we could have made like-for-like comparisons of our pupils' scores with those from other schools in order to determine how well they had faired (Cohen et al., 2011). However, these assessments contained multiple-choice questions, which,

as Xu et al. (2016) note, can be susceptible to guess-work; thus potentially masking the respondent's true understanding. Additionally, the tests focused mainly on literal and knowledge-based inferential questions at the expense of the evaluative be ones. The evaluative questions carried more potential marks, but we wanted an even distribution of the three questions-types following each extract so that pupils had an equal opportunity to demonstrate their literal, inferential and evaluative skills.

We also constructed a simpler mark-scheme by splitting the results into three sections (corresponding to each question type) and awarding one mark per question for each accurate or reasonable answer regardless of question-type (as Table 1 shows): we decided it would be easier to compare how pupils performed at each level of comprehension if each section contained the same number of questions carrying an equal number of marks. Moreover, as I have noted above, coding the data yielded a qualitative insight into the significant variations in pupils' answers between assessments.

Question Type	**Criteria for receiving a mark**
Literal	Accurately relaying explicitly stated information requested by the question.
Knowledge-based inferential	A reasonable inference supported by at least one appropriate piece of textual information

Evaluative	Offering an answer supported by at least one relevant piece of textual knowledge and/or knowledge from other domains e.g. background knowledge.

Table 1: marking criteria

In a bid to attain some degree of quality assurance, we discussed the three papers with the Head Teacher and the English Subject Leader, who have both had approximately two decade's worth of experience using various mark-schemes and interpreting test-data. They felt that the tests would serve our intended purpose well but identified some proof-reading errors; and helped us to construct a classification table by dividing the 6 possible marks for each question type into 3 tiers; helping us to analyse the scores and determine who qualified as a lower-attaining comprehender who met out criteria.

Classification	**Score out of 6 for each section**
High	5-6
Medium	3-4
Low	0-2

Table 2: score classification

The assessments were conducted during the first day of each whole-school assessment week in their quiet classroom setting with no adult support. Pupils sat apart from each another in order to discourage copying. By taking these measures, we hoped that the class would treat this as another test; not regard it as a special assessment, which merited extra effort; or find it more daunting: we wanted the results to be as representative of the pupils' comprehension abilities as a written-test could be. However, we decided not to set a strict time-limit because we wanted to minimise time-restriction as a variable affecting performance and enable pupils to answer the questions as well as they could.

After marking the papers, we found five pupils who met our criteria by attaining a high score (5-6) in literal comprehension but a 'low' evaluative and inferential comprehension score (0-2).

The participants and the impact of Covid-19

Owing to restrictions triggered by the Covid-19 epidemic, two of my participants had to stay in isolation with their families three-weeks prior to the closure of schools. As a result, they missed out on the three intervention sessions, final group-interview and reading comprehension assessment. Because I was unable to gather as much data on them as I was on the other two members of the group, I decided to focus on the two remaining members of the group. To protect the participants' identities, I have given them pseudonyms. In addition, I address the sources used to gather this data later on in the chapter.

Target Pupils

- **Ben**: regarded as a higher attainer in maths and science but a lower-attainer in English, Ben comes across as quite a shy pupil who lacks confidence speaking in class: even in maths and science lessons (in which he frequently exhibits a good understand of the focus topics), he seldom puts his hand up to answer questions or to contribute to class discussions. He says this is because he thinks he is not as clever as some of his peers and feels he will give the wrong answers which will make him look silly. He is interested in books fiction and non-fiction books which involve sport.

- **Jade**: regarded as a lower-attainer in the core subjects (English, maths and science) She likes reading but especially loves to read books about music, adventures, myths or superheroes. Additionally, she is a budding musician who loves to play the drums and talk about rock music. As a result, she can regularly be seen tapping her table in class instead of concentrating on his class-work. In English lessons, Jade has developed a good, reciprocal working relationship with Ben (with whom she sits): they regularly help each other out with various tasks.

The intervention

In setting up the intervention, I collaborated with the Year 1 teacher who, like myself, has had experience conducting P4C inquiries with a range of Key Stage 1 and 2 classes over the past six years. These sessions, which are known as 'inquires' (Buckley, 2011; Haynes, 2002), were scheduled to take place for one-hour per week over the course of ten weeks during the Spring term. They took place on Monday afternoons in the ICT room (a comfortable and familiar setting) and included the following stages proposed by Buckley (2011):

The P4C

- **Getting started:** pupils participated in warm-up thinking games such as discussing frivolous questions like: would you rather be a birthday cake or a chocolate bar?
- **Stimulus**: this was either a story, video or picture intended to ignite philosophical thinking.
- **Thinking time**: individually, then as a group, pupils were given time to reflect on the stimulus and the effect it had on them.
- **Question making**: pupils then attempted to formulate interesting and open questions which went beyond the stimulus by being understandable to someone unfamiliar with the stimulus: for example, why do we eat some animals and keep others as pets?
- **Question airing**: participants shared their questions with each other.
- **Question choosing:** the pupils voted for the questions that they wanted to explore. We decided to use a secret-voting system where the facilitator read out the candidate questions whilst pupils closed their eyes and raised their hands to vote for the one they wanted to discuss.
- **First thoughts:** participants shared their initial thoughts pertaining to the chosen question.
- **Building:** regarded as the core of the inquiry. During the first two sessions, the teacher had to encourage the children to take charge of their own inquiry by ensuring they listened carefully and critically to each person's thoughts, allowing each other time to speak and justify their viewpoints.
- **Last thoughts:** pupils critically reflected on the answers which emerged during the inquiry. This stage also helped them to reflect on where they stood at the end of the inquiry.

- **Review:** finally, the group would appraise their inquiry by noting how successful the chosen question was at eliciting dialogue and collectively addressing any issues (such as participants forgetting to offer evidence to substantiate their opinions) and considering how the dialogue could be more fruitful in future.

Intended intervention schedule

The intervention sessions were scheduled to take place over a ten-week period during the spring term: between 1-2pm on Monday afternoons. This was for practical reasons, since the Year 1 teacher (the facilitator) had been granted this time out of his classroom to conduct the intervention. The sessions were held in a newly refurbished intervention room. Here, the pupils were able to sit around a circular table; allowing them to face each other when speaking, which we hoped would help them to concentrate more on what each other was saying, than if facing away from each other. Additionally, the padded chairs and freshly painted environment offered a comfortable environment for pupils to establish their CoI.

Addressing the limitations of the proposed intervention

Unfortunately, with the school's budget becoming increasingly tight, causing recent staffing cuts, the Head felt he could not justify paying for either of us to be officially trained to use the pedagogy. This represents a potential limitation of the study since Lipman (1980) argues that practitioners should be trained in P4C in order to get the most out of the pedagogy. Consequently, the standard of our P4C intervention could have been higher if it were conducted by a practitioner by an officially recognised P4C training provider such as SAPERE (Anderson, 2017).

Importantly, since beginning our teaching careers, the Year 1 teacher and I have regarded encouraging pupils to think for themselves as one of the most important roles of a teacher. So,

despite our lack of official training, we have been developing and refining our P4C practice over the past six years by scheduling regular P4C sessions with approximately 12 classes; studying the training and research literature (SAPERE, 2020; Anderson, 2017; Buckley, 2011; Haynes, 2002; Lipman, 1980); informally observing each other's sessions; and frequently discussing ways of improving our practice. What is more, during the intervention period, we scheduled informal 15-minute conversation each week in which to discuss the progress of the intervention and various ways in which we could improve the pupils' engagement in philosophical inquiry.

Although I accept that our lack of formal training constitutes a limitation, I believe the measures we took to address this enabled my colleague to facilitate a high-quality P4C intervention for the participants.

Gathering background data on the participants

I produced background profiles of each participant by gathering data from a mixture of sources:

- Pupil records on the school's database;
- An unstructured interview with their class-teacher;
- Summative assessment records;
- Pupils' reading diaries;
- The whole-class reading folder;
- A semi-structured group interview with the participants.

For example, I examined participants' records on the school's database to find out whether any of them were SEND or EAL pupils or whether any were from an economically disadvantaged background: as indicated by free school meal or pupil premium status.

I also examined the Year 3 whole-class reading record to ascertain how often participants read to adults in the classroom; inspected the reading diaries pupils had been taking home to get an idea how often participants read at home. However, in my experience, parents and carers do not always fill these in accurately. I therefore ensured that I asked participants how often they read to their parents and independently whilst at home during our interviews. Although the latter approach relied on pupils' honesty. Collectively these methods helped me to determine how often participants read.

Additionally, summative assessment data allowed me to identify pupils' levels in subjects across the curriculum; with interview data I learned about the participants' hobbies, interests and if they enjoyed reading. The summative data also showed that all participants were graded as having strong phonics knowledge: a key component of reading comprehension (Clarke et al., 2013; Kintsch and Rawson, 2005; Gough and Tunmer, 1986). Collecting such data was vital to my study because it enabled me to offer a more nuanced interpretation of my findings than I could have without it.

Semi-structured group interviews

I conducted two semi-structured group interviews with the participants: one the day after the second comprehension assessment (just before the intervention began) and another the day after the final assessment. This gave the interviewees a chance to discuss what they found difficult; whether they felt their ability to answer higher-level comprehension questions had improved between tests; and, crucially, whether they felt P4C had had an impact on this.

I opted for the semi-structured group interview format because it offered enough structure for me to prepare a schedule of questions (see Appendices E and F) which I felt were key to my

research whilst giving me the freedom to ask questions prompted by the interviewees' responses: I could not achieve this with either a strictly structured or an unstructured interview (Drever, 1995; Cohen et al., 2011). By using interviews I could ask many more open questions than I could have with a questionnaire; although I still had to ensure my questions were not leading pupils into giving certain response (Cohen et al. 2011).

Moreover, the group format was much more time-efficient than conducting separate interviews; helping me to manage the workload of being a full-time teacher, a researcher and a new father. However, I was aware of the risk that some pupils might dominate the discussion; thus preventing others from voicing their opinions as much (Cohen et al., 2011). To address this, I directed my questions at individuals in turn to provide each interviewee with a fair chance to offer their views, then asked if anyone wanted to add anything before moving on in case an interviewee wanted to change his or her mind. In addition, I was mindful that interviewees might hide their true opinions in a group-setting owing to such factors as peer pressure and wanting to save face amongst their peers or by offering answers they thought I wanted to hear in order to please me as a member as a teacher (Cohen et al. 2011). I thus informed pupils that they could notify me anytime if they had changed their minds about an answer.

The way in which an interview is conducted and captured can heavily influence the quality of data retrieved (Drever, 1995). I thus decided to audio-record both interviews using a school iPad placed on a table adjacent to ours in a bid to minimise the distraction it might cause, whilst still being close enough to capture the audio effectively. This also enabled me to play the recordings back so pupils could add to, or amend, their answers. The interviews took place in the Year 3 classroom: a familiar environment in which I thought the interviewees would feel comfortable. Additionally, to thank the children for giving up part of their lunchtimes, I put

some biscuits on the table. By creating a relaxed atmosphere, I hoped pupils would feel comfortable enough to provide full and truthful responses for me to analyse (Cohen. 2011; Thomas 2016).

The constant comparative method

Pupils' interview responses were also analysed using the constant comparison method, which involved comparing answers given by the respondents in both interviews several times in order to detect common themes(Thomas, 2016). Appendices G and H contain transcripts of the parts of both interviews which I refer to in my analysis and discussion section.

Ethical considerations

After gaining ethical approval from the University of Oxford's ethics committee, I discussed the intentions, possible benefits and potential risks of my research with the Head Teacher in order to gain his formal approval. I assured him that I would comply with the school's own data-protection policy at all times by preserving the anonymity of all those involved in my research. This involved not identifying the setting or staff-members by name and by providing all pupil-participants with a pseudonym. Additionally, all data was saved as password-protected files on my encrypted laptop and will be deleted following the completion of my MSc.

When seeking informed, verbal consent from pupil-participants, I notified them, in child-friendly terms, of my intentions, which included gathering as much useful information as I could from them to determine how an intervention might help to improve their reading

comprehension, whilst also making it clear that participating would involve my taking up some of their free-time and that they had the right to withdraw from the project at any stage.

When doing this, I was mindful of the uneven power-dynamics at work between the pupils and myself as a teacher: pupils might have felt obliged to comply with my wishes as an authority figure (Kellet, 2005; Alderson 2004; Lewis and Lindsay, 2000). I therefore asked each pupil for their informed consent on two separate occasions; made it clear that they would not be in trouble if they wanted to opt out; and that they could inform any member of staff or their parents of this, if this made the process easier for them. I asked pupils for their consent privately in order to reduce the potential influence of peer-pressure: they could have felt pressured into participating if their peers were (Thomas, 2009).

In addition, I spoke to the pupils' parents privately to communicate: the aims of the research; what it required from their children; and how their data would be used and stored. I made it clear that if any of them had reservations or objections about their child's role, then, I was happy to discuss the matter further with them or, if they so wished, their child would be removed from the study at any time.

Although my research was dependent on my collaborating with colleagues, I was also appreciative of the fact that they have busy lives and might not want to sacrifice their time to aid my research. Additionally, I understood that it is possible for colleagues to feel pressured into participating or not by other colleagues or their senior leadership team (Thomas, 2009). I thus privately sought consent from the Year 3 and 1 teachers and made sure they were also aware of my aims, the nature of their roles, the amount of time they might need and right to opt-out if ever they wished.

By taking these measures, I believe that my study adheres to the high standards of the University's ethics committee and the British Educational Research Association (2011).

Findings, analysis and discussion

Introduction

This chapter focuses on providing answers to the following research questions:

1) Could a P4C intervention improve the way lower-attaining Year 3 readers answer knowledge-based inferential questions and evaluative questions?

2) How could a P4C intervention improve the way lower-attaining Year 3 readers answer knowledge-based inferential questions and evaluative questions?

To answer them, I present and analyse my findings from my main sources of data-collection (three reading comprehension assessments and two group-interviews) in the form of two separate cases before engaging in a discussion which interprets my findings in relation to the literature I have reviewed.

Description of reading comprehension assessments

In order to answer my first research question, the Year 3 teacher and I set-up a repeated measures design (see Table 1 above). This involved the participants taking three separate reading comprehension tests (each of which included six literal, evaluative and knowledge-based inferential questions) which were originally intended to have been taken over two 10-week intervals: in the first interval, pupils received no P4C intervention, whereas in the second they participated in P4C sessions. A greater difference between the second and third test results

than the first and second test results could suggest that P4C had had an impact: positive or negative. However, the second interval was reduced to seven weeks owing to school closures. This could have limited the potential impact the intervention might have had on the participants' reading comprehension abilities.

N.B. Although this study focuses on evaluative (Eval.) and knowledge-based inferential questions (Inf.), we incorporated literal questions into each test paper to make each test as similar as possible. I have, therefore, kept the results in Table 3.

Ben

	Test 1			Test 2			Test 3		
	Lit.	Inf.	Eval.	Lit.	Inf.	Eval.	Lit.	Inf.	Eval.
Ben	5	2	1	6	1	2	6	3	5

Table 3: Ben's reading comprehension test-results

Analysis

There were no significant changes to Ben's scores when comparing Tests 1 and 2. In his second paper, he received one mark less for inferential questions and one mark more for evaluative questions. According to our classification table (Table 2), Ben's results placed him in the

lowest tier for inferential and evaluative comprehension. However, despite not receiving marks for the majority of these questions, Ben scored 2 and 1, for inferential and evaluative questions respectively on Test 1, and 1 and 2 on the second assessment. This suggests that he was cable of providing appropriate answers to both question-types prior to participating in the P4C intervention.

Reflecting on these results, Ben stated that he knew he was adept at answering literal questions because he enjoyed 'searching the words for the answers and enjoy it when they pop out at me.' However, he found the other questions 'confusing and hard to answer' and also noted that he was 'not good at finding answers to them.' This implied that Ben was looking for answers explicitly stated in the text (which is a strategy for answering literal questions) rather than using knowledge from multiple domains to infer and evaluate, for example. Ben also thought he struggled with the higher-level comprehension questions because he 'kept forgetting what [the text] was on about.'

However, a comparison between the second and third test-papers reveals a significant change in his rate of progress. He gained three marks for evaluative comprehension in Test 3 which moved him up to the highest tier of the classification table. He also increased his inferential score by two marks. This elevated him to the middle tier of inferential comprehension. By comparing a selection of Ben's inferential and evaluative answers from the three assessments, I identify significant qualitative differences between the ways Ben approached these questions before and after the intervention period.

A comparison of Ben's typical inferential answers from Test 1 and 2

Question 1 is about Fisher's (1995) *The Story of Gelert*, which involves a prince mistakenly killing a dog called Gelert who was faithfully protecting the prince's baby-son from a wolf.

Question 2 is refers to *The Hare and the Tortois*e (BBC, 2017): a traditional fable in which a complacent hare loses a race to a determined tortoise.

N.B. Mistakes in spelling and grammar are the result of the answers being quoted verbatim.

Test 1

Question 1: Why did the prince kill Gelert?

Answer: Because he looked really scary and he had no choice.

Test 2

Question 2: Why did the crowd cheer when the tortoise approached the finishing line?

Answer: They were really happy.

Analysis

These answers demonstrate weaknesses in Ben's inferential comprehension, which he consistently displayed in the first two papers, and reveal why he received few marks in this area. The marking criterion (Table 1) stipulates that in order to receive a mark for an inferential answer, it should consist of a 'reasonable inference supported by at least one appropriate piece of textual information.' However, these answers do not meet this condition.

When looking at Question 1, it is important to note that although Gelert might have looked scary (the text states that Gelert was covered in blood), this answer indicates that Ben had not made a necessary inference: that the prince erroneously believed that Gelert had murdered his

son and thus executes him in retribution. Likewise, although his response to Question 2 ('they were really happy') might also have been true, the text makes it clear that the tortoise is about to beat the hare when the crowd cheer. This makes it reasonable to infer that this is the cause of the cheering. Despite attaining a high literal comprehension score in both assessments (5 and 6 respectively), these answers indicate that Ben lacked a sufficient understanding of both texts to enable him to make reasonable inferences. This is consistent with his remark about finding it difficult to remember what he reads.

What is more, Ben displayed a similar lack of textual awareness even in the inferential answers for which he received marks. An analysis of his answer for Question 3 shows this.

Question 3:	Why did the hare think he could beat the tortoise easily?
Answer:	He thought he could beat him easily becos a hare is much faster than a tortos.

Although Ben's response accords with information presented in the story (the hare is depicted as being faster than the tortoise) it is hard to determine whether Ben was applying his general knowledge of the two animals or referring to information stated in the text: especially since the answer uses the indefinite article (which indicates *a* random hare and tortoise) rather than the definite article (which would indicate *the* creatures in the text).

In the first interview, Ben acknowledged that he was employing his general knowledge here. This is important because it shows that, in some cases, pupils like Ben could still score marks for inferential questions (even if they had poor textual awareness) by using their general

knowledge provided it was consistent with textual information. However, as shown above, poor textual knowledge (demonstrated, for example, by his unawareness of the prince's motivation for killing Gelert) cost Ben marks.

In contrast, his inferential answers in Test 3 suggest that he attained a higher degree of textual knowledge.

Question 4: Can plastic bags be dangerous to sea creatures?

Answer: Plastic bags are definly dangerous to sea creatures because they look like food and might eat the plastic and it could harm their insides because their bodies are not meant to eat plastic.

During the second interview, Ben noted that he was unaware that plastic could resemble food to sea-life before reading the text. Therefore, he had used this textual knowledge and combined it with his general knowledge (food is eaten) to develop a line of reasoning that led to a reasonable inference: the plastic could be eaten by sea-creatures which could make them ill. This application of textual and general knowledge to create a clear line of reasoning marks a significant difference between the quality of Ben's inferential answers before and after the intervention. Though, the most significant qualitative differences between Test 2 and 3 can be detected when comparing answers to evaluative questions.

A comparison of Ben's typical evaluative answers from Test 1 and 2

The first question is from Test 1 about *The Story of Gelert* (Fisher, 1995). This second is from Test 2 about *The Hare and the Tortoise* (BBC, 2017).

Question 5: Did the prince have a right to kill Gelert?

Answer: Yes because he had blood on him.

Question 6: Was the hare right to be proud of himself for being able to run fast?

Answer: The hare shud be proud because I think it is a good thing to run fast.

Analysis

The marking-criterion for evaluative questions in Table 3 stipulates that to be awarded a mark the answer must be 'supported by at least one relevant piece of textual knowledge and/or knowledge from other domains e.g. background knowledge.' Ben was, therefore not awarded marks for these answers because they lack supporting evidence. He also missed opportunities to elaborate points. For instance, Ben's answer argues that the prince does have a right to kill Gelert because 'he had blood on him', but does not explain that this gave the prince reason to believe that Gelert had murdered his son. Moreover, he does not use textual knowledge to support his reasoning and show a stronger evaluative understanding of the text. In addition, he tried to justify the first part of his response to Question 6 (the hare should be proud of running fast) with another opinion (it's good to run fast). Again, Ben did not use textual or general knowledge to substantiate his viewpoint. These responses are typical of the majority of his answers to evaluative questions from the first two papers and indicate why Ben was situated as a low-attainer in inferential and evaluative comprehension on our classification table.

However, a comparison of a sample of evaluative answers which earned marks from Tests 2 and 3 reveal important qualitative differences.

Evaluative answers which received marks from Tests 2 and 3

The first evaluative question is relates to a text on Anglo-Saxon kings.

Question 7: Was it right for kings to fight against other kingdoms?

Ben: Yes if they did not fight they could be killed by invading enemies.

Question 8, from Test 3, is about plastic pollution.

Question 8: Why do you think people still use plastic bags so much?

Answer: I think because if you are in the shop like my mum is a lot and she needs to carry things the text says plastic is durable and strong so it is useful for this. I also think that people like my mum and dad forget to bring their own bags because they have busy lives and bills to pay and other worries.

Analysis

Ben's response to Question 7 presents a line of logical reasoning which incorporates his general knowledge (kings invaded other kingdoms which led to conflict) to infer what might have happened to kings who did not fight (they might have been killed), which supports his evaluation: it was right for kings to fight.

It is, however, important to recognise that although the text mentions kings invaded other kingdoms, Ben stated that he was using his general, and not textual, knowledge when constructing this answer. Therefore, Ben was able to satisfy the marking criteria (Table 1) by integrating evidence from his general knowledge to corroborate his answer. However, as with his inferential answers for Tests 1 and 2, his evaluative answers suggest the he lacked important textual knowledge.

This was not the case with Question 8 where his answer includes a reference to textual information (plastics are strong and durable) to justify why people might still use plastic bags. Ben explicitly highlights that this is textual knowledge by declaring 'the text says'. Additionally, like his inferential responses in the final assessment, he blended this textual knowledge with his personal experience to support his line of reasoning. Ben's awareness of the fact that people like his parents often have busy lives and financial obligations on their minds was used as a potential reason for their forgetting to bring their own bags and resorting to the use of plastic ones. Importantly, Ben stated that the questions discussed in the P4C sessions encouraged him to reflect on his own life-experiences more when answering questions. In the final assessment, then, Ben had improved the quality of his evaluative answers by harnessing knowledge from multiple domains in order to substantiate his points of view.

Jade

	Test 1			Test 2			Test 3		
	Lit.	Inf.	Eval.	Lit.	Inf.	Eval.	Lit.	Inf.	Eval.
Jade	6	1	1	6	2	1	6	3	5

Table 4: Jade's reading comprehension test-results

Analysis

Like Ben, Jade's results from Tests 1 and 2 show no major changes in her ability to answer inferential or evaluative questions: she scored the same for evaluative questions and 1 mark more on inferencing in her second paper, which placed her in the lowest tier for inferential and evaluative comprehension on the classification table (Table 2). Like Ben, her results for Tests 1 and 2 suggest that she was able to provide appropriate answers to both question-types prior to participating in the P4C intervention but not consistently enough to score more marks: she received 1 mark for both question-types on Test 1, and 2 for inferential questions and 1 for an evaluative question on the second test.

In Test 3, though, Jade scored 3 marks for inferential comprehension, which moved her from the lowest tier of the classification table to the middle one. Yet, this was only an increase of 1 mark: she had also increased her first assessment score by 1 mark on Test 2, without participating in the intervention. However, on the final assessment, her evaluative comprehension score increased by 4 marks: this moved her from the lowest tier of comprehension to the highest and marked the biggest improvement in evaluative comprehension out of the two participants. What is more, this resulted in a radical change in

her perspective: ‘I still find inferential questions a bit hard but I can see I am getting better … I now enjoy the evaluative ones the most ‘cause I can do them.’

In light of these findings, I now explore a selection of Jade’s responses from the assessments to identify important changes to the way she answered inferential and evaluative questions over the duration of the study. To make a comparison with Ben’s responses easier, I examine Jade’s answers to the same questions.

Inferential answers which did not earn marks from Tests 1 and 2

Question 1 is from Test 1 and about the *The Story of Gelert* (Fisher, 1995) whilst Question 2 from Test 2 is about *The Hare and the Tortois*e (BBC, 2017).

Test 1

Question 1: Why did the prince kill Gelert?

Answer: Because the dog would of killed him if he didn’t.

Test 2

Question 2: Why did the crowd cheer when the tortoise approached the finishing line?

Answer: Becase [sic] they liked the race very much.

Analysis

When examining Ben's responses to the same questions suggested that he lacked significant textual knowledge. The same could be said for Jade's answers. Her claim that the dog would have killed the prince, in response to Question 1, indicates that she had not understood details which were essential to comprehending the story: Gelert is portrayed as a trustworthy dog who, after risking his life to protect the prince's son, is left incapacitated on the floor. It would, therefore, be unreasonable to infer, as Jade did, that the dog posed a threat to his master, and since the marking criteria (Table 1) states that an inferential answer should be 'reasonable', it was not given a mark.

Importantly, when asked how well she thought she understood both texts, Jade replied: 'I think my understanding was good.' Although her response to Question 1 indicates a lack of textual understanding, her answer to Question 2 could be interpreted differently. Although her answer ('they liked the race very much') may be accurate, it is too vague: if she would have specified that it was this aspect of the race which prompted the crowd to cheer, then, she would have made a reasonable inference. In its current form, thought, the answer does not show that Jade had understood the significance of the tale's main event: the hare arrogantly believed he would win easily, which made the humble tortoise's impending victory all the more desirable for the crowd; resulting in their cheering.

What is more, Jade stated: 'I am just rubbish at questions … I never know how I should say things.' Whereas Ben said he struggled to remember the events of the story, it is possible that Jade understood the story better than her answer suggested but was unaware of how to answer the question more precisely.

In contrast, although she only increased her inferential score by 1 mark in Test 3, her answers tended to be much clearer, as a comparison between her responses to Questions 3 and 4 shows.

A selection of inferential answers awarded marks from Tests 2 and 3

While Question 3 is about the same text from Test 2, Question 4 is from Test 3 and relates to a non-fiction text about plastic pollution.

Test 2

Question 3: Why did the hare think he could beat the tortoise easily?

Answer: Because the tortus was realy slow and he was realy fast.

Test 3

Question 4: Can plastic bags be dangerous to sea creatures?

Answer: Plastic bags are dangeros because sea creatures could eat the plastic bags because they look like their food and then they could get sick.

Both answers use supportive evidence presented in the text. Jade's response to Question 3 uses information conveyed in the text (the hare is faster than the tortoise) Whilst Ben's answer to this question includes the indefinite article ('*a* hare is much faster than *a* tortos) to refer to his general knowledge of hares and tortoises, Jade's use of the definite article ('*the* tortus was realy slow') suggests the answer is referring to the tortoise in the text, which, in turn, implies that Jade was using her textual knowledge. She also noted that this was the case in the first

interview. Therefore, Jade's use of textual knowledge in her responses to questions from the first two tests marks a notable point of difference between her inferential answers and Ben's.

However, her use of the pronoun 'he' makes the sentence's meaning ambiguous: it could be interpreted as meaning that the tortoise in the story was sometimes fast and sometimes slow or it could refer to the tortoise being slow and the hare being fast. Although, in the interview, Jade noted that she meant the latter statement, this answer reveals a lack of clarity which was not found in her inferential response to Question 4, from the final test.

This answer is written in a clearer manner. In contrast to her response to Question 3, which is a standalone subordinate clause, her answer to Question 4 is presented as a full sentence which articulates her main point using a dominant clause ('Plastic bags are dangeros'), then connects her reason for this view with the subordinating conjunction 'because' and features fewer pronouns and more nouns, which reduces ambiguity by making it clear when the answer is referring to 'plastic bags' and 'sea creatures'. Moreover, Jade attributed this notable difference in the structure of her inferential answers to the intervention: 'it showed me how to actually answer them … you know by saying your point and giving reasons.'

Furthermore, Jade also acknowledged that she did not know plastic bags could resemble food for sea creatures until she had read the text. Like Ben, then, she blended this textual knowledge with own general knowledge (food is eaten) to infer that the plastic could be eaten by sea-creatures and make them ill: thus producing a more detailed line of reasoning than her answer to Question 3 which relays only textual information: the tortoise is slow and the hare is fast.

Despite the fact that Jade's Test 3 inferential comprehension score (3 out of 6) only marked an increase of 1 mark (when compared with her Test 2 result), there were salient differences in

the quality of her inferential answers. However, by examining a selection of her answers to evaluative questions from all three assessments, I reveal a greater qualitative difference.

Evaluative answers which did not receive marks from Tests 1 and 2

Question 5 is from Test 1, about *The Story of Gelert* (Fisher, 1995). Question 6, from Test 2, is about *The Hare and the Tortoise* (BBC, 2017).

Test 1

Question 5: Did the prince have a right to kill Gelert?

Answer: No he did not becos he was a good dog

Test 2

Question 6: Was the hare right to be proud of himself for being able to run fast?

Answer: I think he shouldn't be that prod.

Analysis

Jade's attempt at justifying her view that the prince did not have the right to kill Gelert (by stating that he was a good dog) lacked sufficient evidence to satisfy the marking conditions outlined in Table 3 for evaluative questions: to receive a mark an evaluative answer must be 'supported by at least one relevant piece of textual knowledge and/or knowledge from other

domains e.g. background knowledge.' Despite this, the text provides a number of reasons for regarding Gelert as a 'good dog', which Jade did not refer to here: his loyal and selfless actions, for example.

However, in the first interview, I asked Jade if she could remember anything about the tale before we reviewed it. To my surprise, she was able to provide an accurate summary of the text and described Gelert as being a good dog who was loyal to the prince and had protected the prince's baby son. Even though she had not read the text for 10-weeks, Jade retained textual knowledge which she could have used to support her response to Question 5. This suggests, then, that although she possessed this knowledge, she was unsure how to incorporate it into her answer.

Whilst Jade attempted to provide a reason for her answer to Question 5, she did not for her response to Question 6, from Test 2. Instead, she gave the opinion that the hare should not have been so proud with offering any evidence to support it. Both of these answers, then, accord with Jade's view (expressed in the first interview) that she did not know how to answer inferential and evaluative questions.

Despite this, Jade did receive a mark for an answer to an evaluative question in Test 2. By comparing it with a response to an evaluative question from Test 3, I establish why it received a mark and identify significant differences in the quality of Jade's evaluative answers before and after the intervention.

Evaluative answers awarded marks from Tests 2 and 3

Test 2

Question 7: Was it right for kings to fight against other kingdoms?

Answer: I think it would only be rite if they had been ataked first because if someone has done something to hert you then you have a good reson to defend yourself if not you shud leave them alone.

Test 3

Question 8: Why do you think people still use plastic bags so much?

Answer: Peple might still use plastic bags because the text says plastic is really strong and durable. This is good for carrying things and pepel might not be aware of the dangers of using plastic and it getting into the sea and killing marin life also it is easier to get plastic bags from the shop because they are like 9p and even I can pay for them with my change so for an adult with a job this is nothing. If they were much higher in price then peple would not want to buy as many.

Analysis

Although Jade's answer to Question 7 is devoid of any reference to textual knowledge, it was still awarded a mark because she reached her conclusion by reasoning from her own moral standpoint: a person should only fight if someone attacks them because they then have a reason

to defend themselves. This contrasts with unsubstantiated opinions she offered for Questions 5 and 6. However, her response to Question 8 marks a substantial change in the calibre of her evaluative answers.

As with Ben's evaluative answers in the final assessment, Jade was able to harness different forms of knowledge to construct a more detailed answer. The opening sentence signposts her textual knowledge by using the phrase 'the text says' before relaying the fact that 'plastic is strong and durable.' The corollary of this is then inferred ('This is good for carrying things') and an additional reason for continued plastic bag use is provided (people's ignorance of its potential dangers to sea-life). Jade subsequently applied her own general knowledge (by referencing the ease at which plastic bags can be purchased cheaply) and life-experience: 'even I can pay for them with my change so for an adult with a job this is nothing.' Finally, Jade's line of reasoning ends by positing a logical solution: 'If they were much higher in price then peple would not want to buy as many.' Although the last part of the answer deviates from the question (which asks why people still use plastic bags, rather than how usage could be reduced), it is nonetheless important because it indicates that, after the intervention, Jade was more able to articulate and justify her opinions when answering evaluative questions. As a result of this, she increased her evaluative comprehension score by 4 marks.

I now discuss the implications of my findings and analyses in relation to the literature I reviewed in order to answer my research questions.

Discussion

Could a P4C intervention improve the way lower-attaining Year 3 readers answer knowledge-based inferential questions and evaluative questions?

A comparison between the test results in Tables 3 and 4 shows that Ben made greater progress at answering both question-types after participating in the P4C intervention than he had done in the period before it. It also shows that, whilst Jade made the same rate of progress at answering inferential questions after the intervention as she had done in the 10-week period prior to it, her inferential comprehension score did still improve. However, she made more progress than Ben at answering evaluative questions between the Test 2 (before the intervention) and Test 3 (after it). This suggests that P4C could have helped Ben to become more proficient at answering both forms of comprehension questions. It also suggests that it might have had less impact on Jade's ability to answer inferential questions but could have helped her to become more skilled at answering evaluative questions. However, it is important for me to acknowledge other factors that could have influenced these results.

Throughout the duration of the study, the pupils were engaged in weekly 15-minute guided-reading sessions with their teacher and two 10-minute one-to-one reading sessions with an adult which focused on answering literal and inferential questions. In addition, Ben noted that he read more often to his parents during the period in which the intervention took place. He also acknowledged that this involved answering inferential questions. This finding is consistent with Ben's reading record which was signed by his parents an additional twenty times during the intervention period, when compared to the ten-weeks between Test 1 and 2. Additionally, evaluative comprehension is dependent on inferential comprehension because it requires readers to use the textual understandings they have acquired, partly through knowledge-based inferencing, in order to make their own critical judgements about the text and reach an

evaluative understanding (Rupley and Blair, 1983; McCormick, 1992; Vacca et al., 2009; Basaraba et al., 2013). It seems possible, then, that although these sessions did not focus on evaluative comprehension, if they helped to improve the pupils' inferential skills, then they could have also helped to improve their evaluative comprehension, and, by extension, the ways in which they approached evaluative questions.

Having said that, the participants' progress in inferential comprehension was not as great as their progress in evaluative comprehension. Moreover, since the reading sessions did not focus on evaluative questions, it seems unlikely to me that they would have had a significant impact on the pupils' improved evaluative scores. What is more, P4C focuses on nurturing critical thinking skills such as being able to analyse information and evaluate evidence from a range of sources in order to form a critical judgement (Lipman et al., 1980; Haynes, 2002; Buckley, 2011). I argued in my literature review that these skills are also essential to evaluative comprehension. Therefore, it seems more likely to me that the P4C intervention played a larger part in facilitating the pupils' improved evaluative question scores in particular.

Furthermore, Jade said that she had begun to read to her little brother on most days for a month prior to the final assessment and believed this might have helped her to get better at comprehending what she read because she 'practised reading more … even though it is easy stuff.' Although reading to her little brother would have given her the chance to practise comprehending what she read, I think that the texts of choice (Peppa Pig and nursery rhyme books), which were predominantly aimed at pre-school and Key Stage 1 readers, would not have provided a sufficient challenge to develop Jade's higher-level comprehension abilities. As Snow (2002) notes that a text should be difficult enough to challenge the readers word recognition and comprehension skills in order to help them make progress.

Although my study, with so few participants , cannot assign causation (Thomas 2009, 2016), the improvements in the participants' reading comprehension results following the P4C intervention are consistent with the findings of several studies which also measured P4C's impact on reading comprehension (Gorard et al., 2015; Williams, 1993; Haas, 1980; Lipman and Bierman, 1980). However, my study differs from these in important ways. For example, instead of monitoring changes in general reading comprehension scores, my research utilises the levels of comprehension theory presented by Basaraba et al. (2013) to break down the participants' reading comprehension into three categories: from literal understanding (the least challenging) to the more cognitively demanding inferential and evaluative forms of comprehension. By presenting comprehension in this way, I was able to focus on questions which assessed the inferential and evaluative levels of comprehension. Using this theory my study suggests that P4C might have had more of an impact on the participants' evaluative, than their inferential, comprehension. This raises the question of why this might have been the case.

In the final interview, Jade was unsure why this might have been the case. However, Ben suggested that P4C might have had more of an impact on his ability to answer evaluative questions because 'there's not just one right answer. It's the same in our philosophy group. But with inference there is.' In light of this, I believe that Ben offered a rational potential explanation. The P4C inquiries concentrated on open questions which enabled the interlocutors to hold various opinions (Lipman et al., 1980; Haynes, 2002; Buckley, 2011). Similarly, evaluative comprehension also has an open nature which, according to Herber (1970), makes it possible for readers to possess personalised interpretations of the text. Therefore, since inferential comprehension does not tend to allow for the same level of diverse interpretation (Basaraba et al., 2013), this could have resulted in the intervention having more of an impact on the way participants' answered evaluative comprehension questions.

Furthermore, as noted in my literature review, the studies which monitor P4C's impact on reading comprehension (Gorard et al., 2015; Williams, 1993; Haas, 1980; Lipman and Bierman, 1980) tended to focus on quantitative data and were, therefore, unable to explore how P4C might have influenced pupils' reading comprehension. My study attempts to address this qualitative deficit by analysing a selection of the participants' answers to evaluative and inferential questions from all three assessments and by considering their interview responses. The results of this enable me to answer my next research question.

How could a P4C intervention improve the way lower-attaining Year 3 readers answer knowledge-based inferential questions and evaluative questions?

In Ben's final assessment paper, he appeared to display a higher degree of textual knowledge when answering both types of questions. What is more, he believed that P4C had helped to improve his textual knowledge. In light of this, it is important to note that one of the ways P4C attempts to develop pupils' critical thinking skills is by encouraging them to ask questions which challenge and clarify their thinking (Lipman et al., 1980; Haynes, 2002; Buckley, 2011). Ben thought that this aspect of the intervention motivated him to ask himself questions whilst reading in order to check his comprehension. If he did not understand what he had read or had forgotten parts of the text, he would re-read appropriate sections. Crucially Ben stated that 'doing this helps me to answer questions because I can see what's going on better in my head.'

Active comprehension monitoring such as this is regarded as a hallmark of proficient comprehenders (Graesser et al., 1994; Cain and Oakhill, 1998; Kintsch and Rawson, 2005; Clarke et al., 2013). Furthermore, by using questions to check his understanding could have resulted in Ben producing a more accurate text-base: his understanding of the information

conveyed in the text (Kintsch and Rawson, 2005). Moreover, the accuracy of Ben's text-base would have been critical to his ability to answer both evaluative and knowledge-based inferential questions. This is because the text-base, combined with his general knowledge and memories of personal experience, for example, might have enabled him to make knowledge-based inferences, construct a mental image of the text (a situation model) and, therefore, reach a deeper inferential and evaluative understanding of the text (Kintsch and Rawson, 2005; Kispal, 2008; Basaraba et al., 2013). This perhaps explains why Ben thought that questioning his comprehension of what he read helped him to 'see what's going on better in [his] head', which, in turn, helped him to improve the calibre of his responses to both types of questions.

In contrast to Ben, I found that Jade seemed to be able to construct text-bases which allowed her to accurately recall important textual details before the intervention period. This was evidenced by her ability to accurately summarise a text from the first assessment nearly 10-weeks after reading it, and by the fact that she was consistently able to recall texts during guided and one-to-one reading sessions according to her teacher. Moreover, Jade noted in the last interview, that she had always tried to ask herself whether she understood what she was reading and that 'this hasn't really changed much.' And, regarding her ability to create situation models, she also stated: 'I can't remember not being able to see it in my head.' This suggests that P4C did not have the same influence on Jade's ability to monitor her comprehension, or to construct text-bases and situation models. This was perhaps because the biggest barrier to her effectively answering inferential and evaluative questions was the fact that she was unsure how to construct her answers, as Jade noted herself in the first interview.

Whilst analysing her answers, I found a likely consequence of this: they lacked clarity and would also refer to textual information without explicating its significance. It is worth bearing in mind, though, that I also found that Ben's responses from the first two assessments implied

that he did not know how to answer inferential and evaluative questions either: he often presented an opinion without offering any justification, for example. These findings are consistent with the work of Tok and Mazi (2015) and Carretti et al. (2016) who point out that poor comprehenders often struggle to express what they want to say either verbally or in written form.

However, I also noted that in Test 3 the quality of both participants' answers had changed. Ben would state his opinion, then attempt to justify it and Jade's answers became much clearer because she would explain her reasoning more precisely and elaborate on certain points. What is more, both attributed these changes to answers to their P4C sessions. Jade believed that participating in the intervention showed her how to structure answers because someone (usually the facilitator) would ask a follow up question if she was not being clear enough. This, in turn, got her into the habit of expressing her ideas as clearly as she could. Ben thought that the intervention affected him in a similar way but also noted that it gave him the chance to practise verbalising his answers which also helped him to get better at writing them down. Both also noted that doing this helped them to think more clearly. This accords with Jenkins and Lyle's (2010) point about P4C being routed in the Vygotskian idea that speech is fundamental to the formulation of thought, which is why it embraces collective dialogue. Yet, given that speaking was prohibited during all three assessments, the pupils would probably not have benefited from being able to express and refine their thinking verbally before committing their responses to paper. Although it is possible that they still benefited from the practice P4C gave them at constructing answers: that is, presenting opinions before substantiating them with their reasoning and evidence from various sources.

However, these findings caused me to wonder why the pupils had attributed these improvements to their participation in P4C and not to say the one-to-one or guided-reading

sessions which, as I recognised above, could have also had a significant impact. Afterall, they provided opportunities for Ben and Jade to practise answering inferential, if not evaluative, questions. It may be the case, though, that the intervention had more of an impact because it had significant advantages over the other reading sessions. For example, the guided and one-to-one reading sessions only lasted between 10-15 minutes whereas the P4C inquires last for approximately an hour. Moreover, the Year 3 teacher noted that her guided-reading groups would discuss approximately four literal and four inferential questions and when reading on a one-to-one basis the adult would ask approximately three literal and three inferential questions. In contrast, the Year 1 teacher, who had facilitated the P4C inquires, informed me that sometimes they would spend the entire hour discussing one or two questions. This provided the pupils with more time to practise core elements of P4C such articulating, adapting and refining their own answers to questions after considering their peers' points of view and weighing up different sources of evidence (Lipman et al. 1980; Haynes, 2002; Buckley, 2011). In contrast, it is unlikely that the other sessions, with more questions and less time, would have been able to explore both questions and response in the same level of detail. This perhaps explains why the participants regarded the P4C intervention as being responsible for the differences in the ways they responded to inferential and evaluative questions.

Finally, my analyses of the reading assessment papers revealed that, after the intervention, Ben and Jade's answers to both types of questions began to incorporate knowledge from multiple sources: personal experiences, textual and general knowledge, for example. This marked a major improvement in the quality of their answers and broader comprehension skills in general because knowing when to access different domains of knowledge is critical to being able to construct situation models and to attain a higher-level of comprehension (Pressley, 2000; Cain et al., 2004; Harrison, 2004; Kintsch and Rawson, 2005; Basaraba et al., 2013). This also resulted in more elaborate lines of reasoning which sometimes went beyond the remit of the

question: Jade's proposed solution to decreasing plastic-bag usage was an example of this. However, the participants did not comment on whether this was a potential result of participating in the intervention or not. It is important to recognise, though, that developing the ability to make use of various sources of knowledge such as past experiences when expressing thoughts and constructing arguments is a key part of the critically reflective mindset which P4C aims to cultivate (Lipman et al., 1980; Haynes 2002). In addition, the Year 1 teacher noted that when facilitating the inquires he would often ask the pupils to think about a question from different angles by trying to relate it to their own experience or to imagine that they were in the shoes of a character from a text. He believed that this might have prompted them to harness multiple domains of knowledge more often when answering their Test 3 questions.

With my discussion borne in mind, I now conclude my study.

Conclusion, implications and future research

In this chapter I reflect on:

- My research aims;
- My key findings;
- Collaborating with colleagues;
- The implications of conducting this research
- Future research questions.

Research aims

The aim of this research and development project was to find out whether a P4C intervention could help lower-attaining readers from my colleague's Year 3 class improve the ways in which they answered evaluative and knowledge-based inferential questions. My colleague provided part of the motivation for this aim since she had noticed that a small group of readers in her class exhibited strong literal comprehension skills (they were able to recall information directly stated by the text) but struggled to answer more challenging comprehension questions which required them to infer and evaluate. I suggested that a P4C intervention might help them to develop critical thinking skills such as being able to justify an opinion with evidence and asking questions to enhance understanding, for example: skills which I argued are integral to being able to answer higher-level comprehension questions.

In addition, my desire to assess the impact of a P4C intervention also stemmed from a professional interest I have had in using the pedagogy to improve pupils' critical thinking skills since qualifying as a teacher almost seven years ago. Ever since, I had been eager to conduct my own research on the programme. After reviewing the research literature for this study, I discovered that the key studies on P4C and reading comprehension (Haas, 1980; Lipman and

Bierman, 1980; Williams, 1993; Gorard et al., 2015) focused on numerical assessment data and, therefore, lacked detailed qualitative assessments of P4C's impact on reading comprehension. This inspired me to employ a case study approach, which enabled me to interpret quantitative data whilst providing the sort of detailed qualitative analyses absent from the research literature.

In addition, the aforementioned studies did not demarcate between different levels of reading comprehension. This prompted me to adopt a levels theory presented by Basaraba et al. (2013) as a way to frame reading comprehension. The theory divides reading comprehension into three levels (the literal, inferential and evaluative) which are progressively more challenging for the reader. This enabled me to pursue my second research objective: to shed some light on how P4C might be able to help pupils to develop their ability to answer higher-level comprehension questions.

Summary of key findings

Firstly, in response to my first question, the results of the three reading assessments suggested that P4C might have helped to improve Ben's ability to answer inferential questions since he made more progress in this area after the intervention had taken place than he had done in the period before it. Yet, the fact that Jade's rate of progress after the intervention was the same as her rate of progress before it indicated that P4C may have had little impact on her ability to answer inferential questions. However, both made greater progress at answering evaluative questions after participating in the intervention. This implies that P4C might have helped to improve their responses to evaluative questions.

Furthermore, by analysing the pupils' responses to inferential and evaluative questions from the three assessment papers and considering their comments from two interviews, I was able

to answer my second question. Looking through the lens of Kintsch and Rawson (2005), I found that the intervention could have helped Ben to produce more accurate test-bases and situation models. This, in turn, could have helped him to attain a higher-level of textual understanding, which, in turn, might have helped him to answer inferential and evaluative questions more effectively. In contrast, I found no indication that the intervention had helped to improve Jade's comprehension of what she read. Instead, the main difficulty Jade appeared to have when answering higher-level comprehension questions was the fact that she was unsure how to answer them. However, there was evidence to suggest that participating in P4C had given both pupils the chance to learn how to structure their answers and draw from different domains of knowledge (such as textual and general knowledge, for example) in order to produce much clearer and detailed lines of reasoning.

Conducting collaborative research

Collaborating with colleagues posed a number of challenges: not least, trying to collectively manage our busy schedules so that we could work together. For me, though, the greatest challenges lay in attempting to maintain my effectiveness in the classroom as a full-time Year 5 teacher whilst collaborating with colleagues and getting used to my role as a new father. However, it also conferred benefits too. For example, working with colleagues made conducting the study and setting up the intervention less daunting because I could share the work-load. This was crucial in helping me to balance the requirements of the research with my wider commitments. Additionally, as a result of his experience collaborating in this study, the Year 1 teacher now wants to apply for a master's degree.

Implications and future research

I hope my research will help to address the qualitative gap in the scholarly literature on P4C and reading comprehension (albeit in a small way) by suggesting ways in which P4C might be able help pupils to answer higher-level comprehension questions. Moreover, I shared the findings of this study in a staff-meeting. Given the positive impact the intervention seemed to have on evaluative comprehension in particular, the Head Teacher is now reconsidering investing in SAPERE-certified P4C training for the whole-school and more members of staff (teachers and LSAs) would like to learn how to implement P4C in their timetables.

Furthermore, conducting this project has caused me to consider a number of other questions which a more in-depth study could focus on:

- How might the intervention affect the reading comprehension of different ability groups across one or multiple classes of primary-aged children?
- How might the intervention affect the reading comprehension of secondary school children?
- What impact could P4C have on mathematical reasoning skills?
- How could the principles of P4C be incorporated into other subjects across the curriculum?

Finally, this research project and the MLT course more broadly has strengthened my belief in the importance of educational research and inspired me to pursue doctoral study.

References

Alba, J. and Hasher, L. (1983) 'Is memory schematic?', *Psychological Bulletin* 93 (2), pp.203-231. doi:10.1037//0033-2909.93.2.203.

Albergaria-Almeida, P., Teixeira-Dias, J. J. C., and Martinho, M. (2011) 'How can critical thinking be recognized in the classroom?', in *Critical Thinking: Education in a Competitive and Globalizing World*, New York: Nova Science Publishers, pp.178-188.

Albrecht, J. E. and O' Brian, E. J. (2003) 'Updating a mental model: maintaining both local and global coherence', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19 (5), pp.1061-1070. doi: 10.1037/0278-7393.19.5.106.

Alexander, A., and Fox, E. (2013) 'A historical perspective on reading research and practice', in Alvermann, D., Unrau, N. and Ruddell, R. (eds.) *Theoretical models and processes of reading*. 6th edn. Newark, DE: International Reading Association, pp.3-46.

Alexander, R. J. (2001) *Culture and Pedagogy: International Comparisons in Primary Education*. Oxford: Blackwell Publishers.

Barnes, M.A., Dennis, M. and 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 (3), pp. 216-241.

Barzilay, R and Lapata, M (2008) 'Modeling local coherence: an entity-based approach', *Computational Linguistics*, 34 (1), pp.1-34. 10.1162/coli.2008.34.1.1.

Basaraba, D., Yovanoff, P., Alonzo, J. and Tindal, G. (2013) 'Examining the structure of reading comprehension: do literal, inferential, and evaluative comprehension truly exist?', *Reading and Writing*, 26 (3), pp.349-379.

Bassey, M. (1981) 'Pedagogic research: on the relative merits of search for generalisation and study of single events', *Oxford Review of Education*, 7 (1), pp.73-94.

Bensley, A. (2011) 'Rules for Reasoning Revisited: Toward a Scientific Conception of Critical Thinking', in *Critical Thinking: Education in a Competitive and Globalizing World*, New York: Nova Science Publishers, pp. 1-9.

British Educational Research Association (BERA) (2018) *Ethical guidelines for educational research*. Available at: https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018 (Accessed: 30 April 2020).

Bowyer-Crane, C. and Snowling, M. J. (2005) 'Assessing children's inference generation: What do tests of reading comprehension measure?', *British Journal of Educational Psychology*, 75 (2), pp.189-201.doi: 10.1348/000709904X22674.

Buckley, J. (2011) *Pocket P4C: Getting Started with Philosophy for Children*, Chelmsford: One Slice Books.

Cain, K. and Oakhill, J. (1998). 'Comprehension skills and inference-making ability: issues of causality', in Hulme, C. and Joshi, R.M. (Eds) *Reading and Spelling: Development and Disorders*, New Jersey, NJ: Lawrence Erlbaum.

Cain, K. and Oakhill, J. (1999) 'Inference making ability and it's relation to comprehension failure in young children', *Reading and Writing: An Interdisciplinary Journal*, 11, (5), pp. 489-503.

Cain, K., and Oakhill, J. (2007) *Children's comprehension problems in oral and written language: a cognitive perspective*. New York, NY: Guilford Press.

Cain, K., Oakhill, J. and Bryant, P. (2004) 'Children's reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills', *Journal of Educational Psychology*, 96 (1), pp.31–42.

Cambourne, B. (1976) 'Getting to Goodman: an analysis of the Goodman model of reading with some suggestions for evaluation', *Reading Research Quarterly*. 12 (4), pp. 605-636.

Carnine, D. W., Silbert, J., Kame'enui, E. J. and Tarver, S. G. (2010) *Direct Instruction Reading*, Boston, MA: Merrill.

Carretti, B., Motta, E. and Re, A. M. (2016) 'Oral and Written Expression in Children With Reading Comprehension Difficulties', *Journal of Learning Disabilities*, 49 (1), pp.65-76. doi: 10.1177/0022219414528539.

Castles, A., Rastle, K., and Nation, K. (2018) 'Ending the reading wars: reading acquisition from novice to expert', *Psychological Science in the Public Interest*, 19 (1), pp.5–51. doi:10.1177/1529100618772271.

Cataldo, M.G. and Oakhill, J. (2000) 'Are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text', Journal of Educational *Psychology*, 92 (4), pp. 791- 799.

Catts, H. W. (2018) 'The simple view of reading: advancements and false Impressions', *Remedial and Special Education*. 39 (5), pp.317-323. doi:10.1177/0741932518767563.

Catts, H. W., Adlof, S. M., Hogan, T. P. and Weismer, E. S. (2005) 'Are specific language impairment and dyslexia distinct disorders?' *Journal of Speech, Language, and Hearing Research*, 48 (6), pp.1378-1396.

Charles, P. and Stafura, J. (2014) 'Word knowledge in a theory of reading comprehension', *Scientific Studies of Reading*, 18 (1), pp. 22-37. doi: 10.1080/10888438.2013.827687

Clarke, P. J., Truelove, E., Hulme, C. and Snowling, M. J. (2013) *Developing reading comprehension*. Hoboken: John Wiley & Sons.

Cohen, L., Manion, L. and Morrison, K. (2011) *Research Methods in Education*, New York: Routledge.

Cresswell, J. (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, London: SAGE.

Cromley, J. and Azevedo, R. (2007) 'Testing and refining the direct and inferential model of reading comprehension', *Journal of Educational Psychology*, 99 (2), pp.311-325. doi:10.1037/0022-0663.99.2.311.

Cromley, R.A. and Azevedo, R. (2007) 'Testing and refining the direct and inferential mediation model of reading comprehension', *Journal of Educational Psychology*, 99 (2), pp.311-325.

Czujko-Moszyk, E. (2014) 'John Dewey's Community of Inquiry', *Lingua ac Communitas*, 24, pp.129-141.

Davoudi, M. and Moghadam, H.R.H. (2015) 'Critical review of the models of reading comprehension with a focus on situation models', *International Journal of Linguistics*, 7 (5), pp.172-187. doi: https://doi.org/10.5296/ijl.v7i5.8357.

Department for Education (2013) *The National Curriculum in England: Key Stages 1 and 2 Framework Document*. Available at: https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum. (Accessed: 1st August 2020).

Dole, A., Duffy, G., Roehler, R. and Pearson, D. (1991) '*Moving from the old to the new: research on reading comprehension instruction*', *Review of Educational Research* 61 (2), pp. 239-264. doi: 10.3102/00346543061002239.

Dwyer, C. (2017) *Critical Thinking: Conceptual Perspectives and Practical Guidelines*, Cambridge: Cambridge University Press.

Elder, L. and Paul, R. (2002) 'Critical thinking: distinguishing between inferences and assumptions', *Journal of Developmental Education*, 25 (3), p.34.

Fields, J. (1995) 'Empirical data research into the claims for using philosophy techniques with young children', *Early Child Development and Care*, 107, pp.115-128.

Fisher, R. (1995) *Stories for Thinking*, Oxford: Nash Pollock.

Flesch, R. (1955) *Why Johnny can't read – and what you can do about it.* New York: Harper.

Fries, C. C. (1962) *Linguistics and reading*. New York: Holt, Rinehart & Winston.

Gelerstein, D., del Río, R., Nussbaum, M., Chiuminatto, P. and López, X. (2016) 'Designing and implementing a test for measuring critical thinking in primary school', *Thinking Skills and Creativity*, 20, pp.40-49. doi:10.1016/j.tsc.2016.02.002.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L. and Paul, T. D. (2017) 'Mapping the early language environment using all-day recordings and automated analysis', *American Journal of Speech-Language Pathology*, 26 (2), pp.248-318. doi: 10.1044/2016_AJSLP-15-0169.

Goodman, K. S. (1967) 'Reading: a psycholinguistic guessing game'. *Journal of the Reading Specialist* 6 (4), pp.126-135.

Goodman, K. S. (1973) 'The 13th easy way to make learning to read difficult: a reaction to

Gleitman and Rozin', *Reading Research Quarterly*, 8 (4) pp.484-493. doi: 10.2307/747170.

Gorard, S., Siddiqui, N. and See, B. H. (2015) *Philosophy for Children Evaluation Report and Executive Summary*, Millbank: Education Endowment Foundation.

Gough, P. B. (1972) 'One second of reading', in Kavanagh, J. F. and Mattingly, I. G. (eds.) *Language by Ear and by Eye: The Relationship between Speech and Reading*. Cambridge, MA: MIT Press, pp. 331-358.

Gough, P. B. and Tunmer, W. E. (1986) 'Decoding, reading, and reading disability', *Remedial and Special Education*, 7 (1), pp.6-10. doi:10.1177/074193258600700104.

Gough, P.B., Hoover, W.A. and Peterson, C.L. (1996) 'Some observations on a simple view on reading', in Cornoldi, C and Oakhill, J (eds.), *Reading Comprehension Difficulties*, New Jersey: Lawrence Erlbaum Associates, Inc. pp. 1–13.

Grabe, W. (1988) 'Reassessing the term interactive', in Carrel, P., Devine, J. and Eskey, D. (eds.), *Interactive approaches to second language reading*. Cambridge: Cambridge University Press, pp. 56-70.

Graesser, A.C., Singer, M. and Trabasso, T. (1994) 'Constructing inferences during narrative text comprehension', *Psychological Review*, 101 (3), pp.371-395.

Graesser, A.C., Singer, M. and Trabasso, T. (1994) 'Constructing inferences during narrative text comprehension', *Psychological Review*, 101 (3), pp.371-395.

Haas, H. (1980) Appendix B: experimental research in philosophy for children, in in Lipman,

M, Sharp, A. M. and Oscanyon, F. (Eds) *Philosophy in the Classroom*, Philadelphia, PA, Temple University Press.

Harrison, C. (2004) 'Developing reading comprehension - what we have learned from research,' in Harrison, C. (eds.) *Understanding Reading Development*. London: SAGE

Hart, B. and Risley, T. (1995) *Meaningful Differences in the Everyday Experiences of American Children*, Baltimore: Paul H. Brooks Publishing.

Haynes, J. (2002) *Children as Philosophers: Learning Through Enquiry and Dialogue in the Primary Classroom*, London: Routledge.

Herber, H. L. (1970) *Teaching Reading in the Content Areas*. Englewood Cliffs, NJ: Prentice Hall.

Hoover, W. A. and Gough, P. B. (1990) 'The Simple View of Reading', *Reading and Writing*. 2 (2), pp.127-160. doi:10.1007/BF00401799.

Jenkins, P. and Lyle, S. (2010) 'Enacting dialogue: the impact of promoting Philosophy for Children on the literate thinking of identified poor readers, aged 10', *Language and Education*, 24 (6), pp. 459-472. doi:10.1080/09500782.2010.495781.

Kennedy, D. (2012) 'Lipman, Dewey, and the Community of Philosophical Inquiry', *Education and Culture*, 28(2), pp.36-53. doi: 10.1353/eac.2012.0009.

Kim, J. S. (2008) 'Research and the Reading Wars', *Phi Delta Kappan*, 89 (5), pp. 372-375.

Kintsch, W. (1988) 'The role of knowledge in discourse comprehension: a construction-integration model', *Psychological Review*, 95 (2), pp.163-182.

Kintsch, W. (1998) *Comprehension: a Paradigm for Cognition*. Cambridge: Cambridge University Press.

Kintsch, W. and Rawson, K. A. (2005) 'Comprehension', in Snowling, M. J. and Hulme, C. (ed.) *The Science of Reading: a Handbook*. Malden, MA : Blackwell, pp.211-226.

Kintsch, W. and Van Dijk, T. A (1983) *Strategies of Discourse Comprehension*. New York: Academic Press.

Kispal, A. (2008) *Effective teaching of inference skills for reading: literature review*. Nottingham: National Foundation for Educational Research and Department for Children, Schools and Families.

Leicester, M. and Taylor, D. (2010) *Critical thinking across the curriculum: developing critical thinking skills, literacy and philosophy in the primary classroom* Berkshire, England: Open University Press.

Lervåg, A., Hulme, C., and Melby-Lervåg, M. (2017) 'Unpicking the developmental relationship between oral language skills and reading comprehension: it's simple, but complex', *Child Development*. 89 (5), pp.1821-1838. doi:10.1111/cdev.12861.

Lipman, M. (2003) *Thinking in Education*, Cambridge: Cambridge University Press.

Lipman, M. and Bierman, J. (1980) 'Appendix B: experimental research in philosophy for children', in Lipman, M, Sharp, A. M. and Oscanyon, F. (Eds) *Philosophy in the Classroom*, Philadelphia, PA, Temple University Press.

Lipman, M., Sharpe, A. M., and Scanyan, F. S. O (1980) *Philosophy in the Classroom*, Philadelphia: Temple University Press.

Long, D. L., Seely, M. R., Oppy, B. J. and Golding, J.M. (1996) *The Role of Inferential Processing in Reading Ability*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McCormick, S. (1992) 'Disabled readers' erroneous responses to inferential comprehension questions: description and analyses', *Reading Research Quarterly*, 27 (1), pp. 54–77.

Munro, J. (2011) *Teaching Oral Language: Building a Firm Foundation using ICPALER in the Early Primary Years*. Camberwell, Victoria: ACER Press.

Nation, K., Clarke, P., Marshall, C. M., Durand, M. (2004) 'Hidden language impairments in children: parallels between poor reading comprehension and specific language impairment?', *Journal of Speech, Language, and Hearing Research*, 47 (1), pp.199-211. doi:10.1044/1092-4388(2004/017).

Nunan, D. (1991) *Language Teaching Methodology*. Hertfordshire: Prentice Hall International.

Perdede, (2008) 'A Review on Reading Theories and its Implication to the Teaching of Reading Conference Paper', *Presented in English Department bi-monthly Forum,* Universitas Kristen, Indonesia 27th June. Available at: https://www.researchgate.net/publication/321228081_A_Review_on_Reading_Theories_and_its_Implication_to_the_Teaching_of_Reading (Accessed: 18th August 2020).

Perfetti, C. A., Landi, N., and Oakhill, J. (2005) 'The acquisition of reading comprehension skill', in Snowling, M. J. and Hulme, C. (ed.), *The Science of Reading: A Handbook*. Malden, MA: Blackwell, pp. 227–247.

Pressley, M. (2000) 'What should comprehension instruction be the instruction of?', in Kamil, M. L., Mosenthal, P. B., Pearson, P. D. and Barr, R. (Eds.) *Handbook of Reading Research: Volume III*. New York, NY: Lawrence Erlbaum.

Pressley, M. and Afflerbach, P. (1995) *Verbal Protocols of Reading: The Nature of Constructively Responsive Reading*. New Jersey, NJ: Lawrence Erlbaum Associates.

Rose, J. (2006) *Independent Review of the Teaching of Early Reading: Final Report*, Department for Education and Skills. Available at: https://dera.ioe.ac.uk/5551/2/report.pdf (Accessed: 20th August 2020).

Rozin, P., and Gleitman, L. R. (1977) 'The structure and acquisition of reading II: the reading process and the acquisition of alphabetic principle', in Reber, A. S. and Scarborough, D. L. (ed.) *Toward a Psychology of Reading*. Hillsdale, NJ: Erlbaum, pp. 55-142.

Rumelhart, D. E. (1977). '*Toward an interactive model of reading*', in Dornic, S. (ed.) *Attention and Performance IV*, New York, NY: Academic Press.

Rumelhart, D. E. (1980). 'Schemata: the building blocks of cognition', in Spiro, R. J., Bruce, B. C. and Brewer, W. F. (ed.) *Theoretical issues in reading comprehension*, Hillsdale: Lawrence Erlbaum Associates. pp. 33-49.

Rupley, W. H., & Blair, T. R. (1983) *Reading Diagnosis and Remediation: Classroom and Clinic*, Boston: Houghton Mifflin.

Sadoski, M., Paivio, A. and Goetz, E. (1991) 'A critique of schema theory in reading and a dual coding alternative', *Reading Research Quarterly*, 26 (4), pp. 463-484.

Samuels, S., & Kamil, M. (1988) 'Models of the reading process', in Carrell, P., Devine, J. and Eskey, D. (ed.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press, pp.22-36.

Save the Children (2015) *Ready to Read: Closing the Gap in Early Language Skills so that Every Child in England can Read Well*, London: The Save the Children Fund.

Scarborough, H. S. (2001) 'Connecting early language and literacy to later reading (dis)abilities: evidence, theory, and practice', in Neuman, S., Dickinson, D. (eds.), *Handbook for Research in Early Literacy*. New York: Guilford Press.

Sharp, A., M. and Reed, R. F. (1992) *Studies in Philosophy for Children: Harry Stottlemeier 's Discovery*, Philadelphia: Temple University Press.

Simons, H. (2009) *Case Study Research in Practice*, London: SAGE.

Smith, F. (1982). *Understanding Reading*. New York: Holt, Rinehart & Winston.

Snow, C. (2002) *Reading for Understanding: Toward an R & D Program in Reading Comprehension*. Santa Monica, CA: RAND Corporation.

Swatridge, C. (2014) *The Oxford Guide to Effective Argument and Critical Thinking* Oxford: Oxford University Press.

Thomas, G. (2009) *How to do Your Research Project*. London: SAGE.

Thomas, G. (2016) *How to do Your Case Study*. London: SAGE.

Tok, S. and Mazl, A. (2015) 'The effect of stories for thinking on reading and listening comprehension: a case study in Turkey', *Research in Education*, 93 (1), pp.1-18 doi:10.7227/RIE.0006.

Trickey, S. and Topping, K. J. (2004) 'Philosophy for Children: a systematic review', *Research Papers in Education*, 19 (3), pp.365-380, doi: 10.1080/0267152042000248016.

Uppstad, P. H. and Solheim, O. J. (2011) 'Code and comprehension in written language-considering limitations to the simple view of reading', *L1-Educational Studies in Language and Literature*, 11 pp. 159-174.

Vacca, J. L., Vacca, R. T., Gove, M. K., Burkey, L. C., Lenhart, L. A. and McKeon, C. A. (2009) *Reading and Learning to Read*, Boston, MA: Pearson.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Williams, S. (1993) *Evaluating the Effects of Philosophical Enquiry in a Secondary School*, Derbyshire: Derbyshire County Council.

Wolf, M. C., Muijselaar, M. M.L., Boonstra, 'A. M. and de Bree, E. H. (2019) 'The relationship between reading and listening comprehension: shared and modality-specific components', *Reading and Writing* 32 (7), pp. 1747-1767. doi: 10.1007/s11145-018-9924-8.

Yin, R. K. (2009) *Case Study Research: Design and Methods*, London: SAGE.

Zulkifli, H. and Hashim, R. (2020) Philosophy for Children (P4C) in Improving Critical Thinking in a Secondary Moral Education Class *International Journal of Learning, Teaching and Educational Research* 19 (2), pp. 29-45.

Appendix A: a sample of Test 1

'The Story of Gelert' was taken from Fisher's (1995) *Stories for Thinking*.

Literal Questions

1) What type of animal was Gelert?

2) Which animal was looking for food?

3) How did the dog know that something was wrong?

Inferential Questions

4) Why did the prince kill Gelert?

5) Why did the prince think that Gelert had killed his son?

6) How do you think Gelert is considered brave?

_____
_____

Evaluative Questions

7) Did the prince have a right to kill Gelert?

_____
_____
_____
_____
_____

8) Do you think the prince was a good or bad person?

_____
_____
_____
_____
_____

9) What do you think the moral of the story is and why?

_____
_____
_____
_____
_____

Appendix B: a sample of Test 2

This version of 'The Hare and the Tortoise' is taken from BBC (2017).

Literal Questions

1) Who are the main characters in the story?

2) Why did the hare decide to have a rest?

3) Which character asked whether the tortoise was running yet?

Inferential Questions

4) Why did the crowd cheer when the tortoise approached the finishing line?

5) Why did the hare think he could beat the tortoise easily?

6) Why is the ending of the story surprising?

_____
_____

Evaluative Questions

7) Did the tortoise really win the race?

_____
_____
_____
_____
_____

8) What lesson could the hare have learned from this experience?

_____
_____
_____
_____
_____

9) Was the hare right to be proud of himself for being able to run fast?

_____
_____
_____
_____
_____

Appendix C: a sample of Test 3

The children read a paper called ‘Plastic Pollution in the Ocean’ found on Twinkl (2020).

Literal Questions

1) What is plastic pollution?

_____
_____
_____
_____
_____

2) How long can it take plastic to break down in the water?

_____
_____
_____
_____
_____

3) What can plastics be mistaken for in the water?

_____
_____
_____
_____
_____

Inferential Questions

4) Can plastic bags be dangerous to sea creatures?

_____
_____
_____
_____
_____

5) Plastic can take a very long time to break down in the water? Why is this a problem?

_____
_____
_____
_____
_____

6) If plastics are consumed by fish how could this be a dangerous to humans?

_____
_____
_____
_____
_____

Evaluative Questions

7) Why do you think people still use plastic bags so much?

_____
_____
_____
_____
_____

8) What do you think would be the best way for you to help prevent plastic pollution and why?

_____
_____
_____
_____
_____

9) What is the most important piece of information in this article and why?

_____
_____
_____
_____
_____

<u>Appendix D: a sample of answers to inferential and evaluative questions for comparative analysis</u>

Answers to inferential questions

Test 1

Question 1:	Why did the prince kill Gelert?
Ben:	Because he looked really scary and he had no choice.
Jade:	Because the dog would of killed him if he didn't.

Test 2

Question 2:	Why did the crowd cheer when the tortoise approached the finishing line?
Ben:	They were really happy.
Jade:	Becase [sic] they liked the race very much.

Test 2

Question 3:	Why did the hare think he could beat the tortoise easily?
Ben:	He thought he could beat him easily becos a hare is much faster than a tortos.
Jade:	Because the tortus was realy slow and he was realy fast.

Test 3

Question 4:	Can plastic bags be dangerous to sea creatures?
Ben:	Plastic bags are definly dangerous to sea creatures because they look like food and might eat the plastic and it could harm their insides because their bodies are not meant to eat plastic.
Jade:	Plastic bags are dangeros because sea creatures could eat the plastic bags because they look like their food and then they could get sick.

Answers to evaluative questions

Test 1

Question 5:	Did the prince have a right to kill Gelert?
Ben:	Yes because he had blood on him.
Jade:	No he did not becos he was a good dog.

Test 2

Question 6:	Was the hare right to be proud of himself for being able to run fast?
Ben:	The hare shud be proud because I think it is a good thing to run fast.
Jade:	I think he shouldn't be that prod.

Test 2

Question 7:	Was it right for kings to fight against other kingdoms?
Ben:	Yes if they did not fight they could be killed by invading enemies.
Jade:	I think it would only be rite if they had been ataked first because if someone has done something to hert you then you have a good reson to defend yourself if not you shud leave them alone.

Test 3

Question 8:	Why do you think people still use plastic bags so much?
Ben:	I think because if you are in the shop like my mum is a lot and she needs to carry things the text says plastic is durable and strong so it is useful for this. I also think that people like my mum and dad forget to bring their own bags because they have busy lives and bills to pay and other worries.
Jade:	Peple might still use plastic bags because the text says plastic is really strong and durable. This is good for carrying things and pepel might not be aware of the dangers of using plastic and it getting into the sea and killing marin life also it is easier to get plastic bags from the shop because they are like 9p and even I can pay for them with my change so for an adult with a job this is nothing. If they were much higher in price then peple would not want to buy as many.

Appendix E: interview schedule 1

1) What do you think of reading?
   - Why
2) What do you most like to read?

- Do you have a favourite type of book or author? If so why?
- Do you have any hobbies or interests you like to read about?

3) What are inferential questions?
   - how do you answer them?
   - What do you think of them?
4) What are evaluative questions?
   - How would you answer them?
   - What do you think of them
5) Are you good/bad at answering inferential questions?
   - Why?
   - How do you think you could get better?
   - Is there anything you find difficult about answering them.
   - Why do you think this is?
6) Are you good/bad at answering evaluative questions?
   - Why do you think this?
   - How could you get better at this?
   - Is there anything you find difficult about them?
   - Why do you think this is?

Explain and give examples of each question type first if interviewees are unsure.

Review last two assessments with the interviewees.

7) What can you remember about these texts?
   - What key bits of information can you recall?
   - Like/not like?
   - Explain

Look at a selection of questions with the pupils.

8) What did you think of this question?
9) What was your thinking there?
10) Where did you get this information from?
    - Were you using your knowledge of what you had read or general knowledge?
11) How could you have improved your answer here?
    - Why might this have helped?

Appendix F: interview schedule 2

1) What do you think of reading?
   - Why?
2) Are you good/bad at answering inferential questions?
   - Why?

- Have you changed the way you answer them?
- Is there anything you find difficult about answering them.
- Why do you think this is?

3) Are you good/bad at answering evaluative questions?
   - Why do you think this?
   - Have you changed the way you answer them?
   - How could you get better at this?
   - Is there anything you find difficult about them?
   - Why do you think this is?

Review last two assessments with the interviewees.

4) What can you remember about these texts?
   - What key bits of information can you recall?
   - Like/not like?
   - Explain
5) Why do you think you got a higher score for your answers to inferential questions?
6) Why do you think you got a higher score for your answers to evaluative questions?

Look at a selection of questions with the pupils.

7) What did you think of this question?
8) What was your thinking there?
9) Where did you get this information from?
   - Were you using your knowledge of what you had read or general knowledge?
10) How could you have improved your answer here?
   - Why might this have helped?

At the end of the interview ask about the intervention.

11) What did you think of the P4C intervention?

- Did you like/not like it? Why?
- Has it been useful/not useful to you in anyway in or out of school?

12) Has P4C helped/not helped you to think about things?
13) Has P4C helped/not helped you to say or write down what you mean?
14) Has the P4C intervention affected/not affected the way you read things?

- If so, how?
- If not, why?

Appendix G: transcript of responses from interview 1

*** = a shift in time.

… = pause

Test 1

Interviewer: What do you think of answering reading comprehension questions?

Ben: I really like erm, you know, the ones where you gotta start searching the words for the answers and enjoy it when they pop out at me… I think I am quite good at those questions.

Interviewer: I see. What about the other questions where the answer doesn't pop out at you? The ones where you have to work out the answer based on what you know and what you read.

Ben: They're confusing and hard to answer. Erm I don't think I am very good at those 'cause, um well I'm not good at finding answers to them.'

Jade: Yeah …

Interviewer: Why do you think this is?

Ben: Mmmm, well … well with those ones in the test, I tried to remember what I read but I kept forgetting what it was on about.

***

Interviewer: [Jade] what about you? What do you think of those questions? The ones where you have to work out the answer based on what you know and what you read. Inferring and evaluating.

Jade: Well, I think that I am just rubbish at questions. Well, those types of questions anyway. I never know how I should say things erm, how to answer them.

***

Interviewer: In these Tests 1 and 2, how well do you think you understood what was going on?

Jade: I think my understanding was good. I remember the first one now too. You know about the kind dog … erm, G, Gil, Gel … Gelert … Yeah.

Interviewer: Can you remember that story?

Jade: Yes, Gelert, is a really good dog. A nice hunting dog whose master wants to go out hunting and asks Gelert to look after his baby boy. Then, when he is out, the wolf comes in and licks his lips but Gelert, erm, fights him off but then he is hurt bad and his master comes home and can't see the baby 'cause there cot has fallen over and there is blood all over Gelert so he kills him because he thinks that he has killed his son. But he hasn't.

Appendix H: transcript of responses from interview 2

*** = a shift in time.

… = pause

Interviewer: In that case, how do you think it helped then?

Jade: My little sister would always ask me to read … Peppa Pig books and the nursery rhyme books that she loves. I think that helped me to get better at reading. I practised reading more this way, even though it is easy stuff.

Interviewer: Why do you think?

Jade: 'Cause I would then practice reading a lot more.

Ben: I think I was helped because I was reading lots more with Mum and Dad. They started testing me to see if I could remember what I read and they asked me inference questions too. So I got lots of practise as well.

Ben: I really enjoyed getting into the habit of asking questions about anything I could think of. That's something that I thought I could use when reading … um … so I can ask myself if I know what's really happening.

Interviewer: What do you do if you don't?

Ben: Re-read it again 'till I do I guess. I mean doing this helps me to answer questions because I can see what's going on better in my head.

Jade: I use questions like this too but I think I've always seen what's going on in a book in my head like a movie is plating or something like that. Um, I can't remember not being able to see it in my head

***

Interviewer: Your answers were much more detailed in that last test. Why do you think that was?

Ben: I started to think about my own memories and stuff. Um, how er … and yeah. What I knew about different things. And used that stuff to write my replies.

Interviewer: Why do you think you started to do this?

Ben: Well, in philosophy you do it when talking about different opinions and stuff … so I thought I would give it a go with the reading comprehension test. You know, think about what I know.

***

Jade: I think philosophy helped with being able to say what I thought better. You know: opinion, evidence, reasons for thinking something. I did this in the last test to answer the questions. I think I am much better at it know… Um, I used to hate the evaluative questions but know I think I am getting good at them. Erm. I actually enjoy doing them. I still find

	inferential questions a bit hard but I can see I am getting better … I now enjoy the evaluative ones the most 'cause I can do them
Ben:	Also in philosophy you gotta be clear about what you think.
Jade:	Yeah, 'cause if you're not then someone asks you more questions so they can help you to think more clearly I guess.
Interviewer:	Did this help you when answering inferential or evaluative questions?
Jade:	Ahh… definitely.
Ben:	Yeah, I agree. It also gave me a chance to practise saying my answers out loud … this really helped me think things through better erm in my um head.
Jade:	It showed me how to actually answer them … you know by saying your point and giving reasons.
	***
Interviewer:	So, why do you think that you both did better on evaluative questions?
Jade:	Um, I am not really sure.
Ben:	Maybe it's because in philosophy we have to ask and answer questions that there could be lots of answers for. There's not just one right answer. It's the same in our philosophy group. But with inference there is.