Situation Models and Children's Reading Comprehension: What Role Does Visual Imagery Play?

by

Maroulia Katsipis, B.A. (Honours)

This thesis is presented for the degree of Doctor of Philosophy at Murdoch University.

2016

Declaration

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary institution.

....

Maroulia (Lia) Katsipis

Acknowledgements

I would like to extend my sincerest gratitude to my thesis supervisors, Dr Bethanie Gouldthorp and Dr Helen Davis. Bethanie, your continued support of this research from the very beginning has been invaluable. Helen, thank you for becoming involved as a supervisor with such interest and enthusiasm for this project. The knowledge and guidance you have both provided has been of the greatest value.

I would also like to thank Dr Jeffrey Coney and Dr Suzanne Dziurawiec for their contributions in the earliest stages of planning this research project.

A big thank you goes to all the school principals, teachers and students who agreed to participate in this research. Also, thank you to the team at Project Kids for providing me with a great space for writing and to share ideas in.

Finally, special thanks go to my partner, Reade, and my family and friends. I will forever be grateful for your endless support, encouragement and belief that I would see this thesis through to the end.

Abstract

Individual differences in children's reading comprehension have been attributed to the level at which a reader is able to construct a coherent meaning-based mental representation of the situation described in a text (i.e., a "situation model"). However, although there is evidence that situation models contain perceptual information such as visual imagery, it is yet to be established whether visual imagery contributes to children's reading comprehension via its role in situation model construction. To investigate this, three studies were conducted with children in Grades 4 and 5 (age range: 8.08-11.17 years) as part of the current thesis.

Study 1 explored the utility of several measures of visual imagery and examined whether this construct is best captured by the differentiation of separate visual imagery processes in this younger population. Fifty-nine children completed five measures of visual imagery, each designed to capture a distinct subcomponent of the visual imagery system, including image generation, image maintenance, image scanning, image transformation, and image strength/vividness. It was found that the visual imagery measures were not highly related to one another and thus each represented a unique construct. However, not all of the included measures proved to be valid and reliable.

Utilising the measures of visual imagery that were found to have adequate psychometric properties in Study 1, Study 2 then examined the influence of different subtypes of visual imagery (image maintenance, image scanning and image transformation) on individual differences in reading comprehension. In addition, this study further investigated existing criticisms that traditional measures of reading comprehension do not capture all of the skills involved in situation model construction, by including two separate measures of reading comprehension: a traditional standardised measure (the Neale Analysis of Reading Ability), and a newer measure designed from cognitive

theory, which measures higher-level comprehension processes separate to the effects of lower-level reading ability (the Diagnostic Assessment of Reading Comprehension; DARC). It was found that each subtype of visual imagery differentially predicted reading comprehension. In addition, each measure of reading comprehension was differentially influenced by variations in word reading ability and verbal working memory, with evidence that the Neale was more influenced by lower-level reading skills and simple verbal working memory, whereas the DARC was more influenced by non-verbal reasoning and complex verbal working memory. However, visual imagery was not found to be a reliable predictor of reading comprehension; although, this may have been due to an incongruity between the type of imagery that occurs during objective tasks of visual imagery and the visual simulation of narrative events.

Thus, Study 3 was designed to disrupt good and poor comprehenders' visual imagery *during* reading in order to determine whether good comprehenders show more reliance on visual imagery during comprehension than poor comprehenders. Unexpectedly, however, good comprehenders showed limited evidence of engaging in higher-level comprehension processes (i.e., predictive inferencing) even when imagery was not impaired. Despite this, important implications regarding the use of both textbase and imagery-based representations were revealed, as poor comprehenders displayed increased difficulty maintaining a verbal load during reading compared to a visuospatial load. This suggests that in comparison to good comprehenders, poor comprehenders may have a greater reliance on textbase over imagery-based representations during reading.

Overall, this thesis adds to the literature that suggests not all reading comprehension measures are interchangeable in regards to the underlying skills that they measure. Further, visual imagery may be relevant to reading comprehension; yet, it is likely that this relationship will be further established through careful conceptualisation and

measurement of visual imagery versus visual simulation. These findings have

implications regarding the use of existing comprehension measures in research and

practice, and may also aid future research that investigates the role of visual imagery in

higher-level comprehension processes.

Table of Contents

Chapter 1. Introduction

1.1 Overview

Despite efforts in identifying and remediating reading difficulties, a lack of proficiency in reading comprehension is a problem that continues to affect many children both in earlier years and once they have reached high school, across a range of countries (OECD, 2014). Understanding what we read is central to literacy, and has wider implications for an individual's educational, social and economic outcomes, including being a predictor of subsequent vocational and academic training (OECD, 2001). Thus, more research is clearly needed to advance our understanding of the skills and cognitive processes that support reading comprehension, in order to identify specific areas for remediation and prevent literacy failure.

Reading comprehension is defined as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (Snow, 2002, p. 11). It is now a common notion that comprehension goes beyond lower level text-processing skills and depends on a reader's ability to construct a coherent meaning-based mental representation of the situation described in a text, often referred to as a "situation model" (van Dijk & Kintsch, 1983). Situation models are theorised to contain extensive information about the persons, events, actions and objects described in a text (van Dijk & Kintsch, 1983), and are distinguished from two lower levels of text representation. The lowest level is the surface form representation, which is a representation of specific words and syntax, and the second level is the propositional textbase representation, which is an abstract representation of the ideas present in the text (van Dijk & Kintsch, 1983). Thus, together, the surface form and textbase are merely a mental representation of the text itself; a result of lower word-level processing

that utilises lexical information and basic relations between individual words in sentences. Conversely, the situation model is a coherent representation of the *meaning* of a text and is created through higher message-level processing; that is, the combination of syntactic, semantic and pragmatic information embodied in discourses (Kintsch, 1988; van Dijk & Kintsch, 1983). Importantly, readers go beyond mere linguistic processes when constructing a situation model as they combine previously acquired knowledge that is stored in long-term memory with information explicitly mentioned in the text (Kintsch, 1988; van Dijk & Kintsch, 1983).

It has now become a central tenet of most models of reading comprehension that maintaining global coherence via the construction of a situation models is the ultimate goal of comprehension. These models explain coherence as being achieved by mapping currently incoming discourse on to the preceding discourse content (McNamara & Magliano, 2009). Thus, the construction of a situation model enables the continual monitoring and integration of information in order for this mapping process to take place. As information that is monitored and integrated with the current situation model representation may not only include that which is explicitly stated in the textbase, but also implicit information (such as background knowledge activated from long-term memory), situation models are also the vehicle for the generation of knowledge-based inferences. These inferences have been identified as being particularly important for reading comprehension (Cain & Oakhill, 1999; Cain, Oakhill, & Bryant, 2004a; Graesser, Singer, & Trabasso, 1994; Kendeou, Bohn-Gettler, White, & van den Broek, 2008; Lynch & van den Broek, 2007). Additionally, as situation models are meaning-based representations stored in long-term memory, they may aid off-line recall of the information contained within the text following reading (Radvansky & Dijkstra, 2007).

Situation models are often described as developing in a manner similar to a physical scene, and as such are often conceptualised as being a mental simulation of what is described in a written text (Zwaan, 1999b). Accordingly, several researchers have found evidence that situation models contain perceptual symbols such as visual imagery (Bergen, Lindsay, Matlock, & Narayanan, 2007; Dijkstra, Yaxley, Madden, & Zwaan, 2004; Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012; Zwaan, Madden, Yaxley, & Aveyard, 2004; Zwaan, Stanfield, & Yaxley, 2002), and this imagery has even been identified as a *vital* component for the construction of a coherent situation model (Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004). From an embodied cognition perspective, it is possible that this visual and motor simulation may aid in a deeper experience and understanding of the situation described in a text (Fischer & Zwaan, 2008). Indeed, non-linguistic mechanisms such as imagery have long been recognised as an important part of discourse comprehension. For example, Paivio's (1986) dual coding theory suggests that two separate, but interconnected, subsystems are involved in coding mental representations: a verbal subsystem specialised for dealing with language; and a nonverbal subsystem specialised for dealing with nonverbal objects and events such as imagery. Comprehension, especially of concrete language, is proposed to be dependent on both these subsystems (Paivio, 1986). In addition, visual imagery during reading is proposed to lead to higher reading engagement (Green & Brock, 2002), and reading engagement has been found to be a significant predictor of reading comprehension (Guthrie & Wigfield, 2000; Wigfield et al., 2008).

However, although there is evidence that constructing a situation model involves multiple skills, a focus on a single-component approach to reading comprehension has limited our understanding of the unique contribution that each of these different skills

and processes make to one's ability to comprehend written language (see Hannon & Daneman, 2001, for an extended discussion). It has been noted that, in both research and education, comprehension is often measured globally: a score is based on readers' answers to questions following short text passages, rather than the differentiation of different components or levels of comprehension. Thus, this approach only taps a single dimension of reading skills, often being those operating at the lower-level (Hannon & Daneman, 2001).

Yet, it has been found that higher-level cognitive abilities, such as working memory capacity, generating inferences, and monitoring coherence, are dissociated from lower-level skills, such as phonological awareness and reading fluency (Kendeou, Savage, & van den Broek, 2009a; Kendeou, van den Broek, White, & Lynch, 2009b; Oakhill, Cain, & Bryant, 2003), and skills other than lower-level processes uniquely predict comprehension level (Kendeou, van den Broek, White, & Lynch, 2009b; Landi, 2010). Therefore, it is possible that a number of children may struggle with reading comprehension due to higher-level processing difficulties, but go unidentified because they do well on traditional tests of reading ability and comprehension which emphasise their accurate and fluent word reading skills. As poor comprehension skills may compromise learning in other areas, it is important to investigate these higher-level processing difficulties in order to develop tools that identify these children and provide targeted interventions. Additionally, while visual imagery has been demonstrated to be a vital component of situation modelling, and maintaining online coherence of a narrative (Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004), few studies have isolated the role of visual imagery in situation model construction specifically in relation to higher-level reading comprehension, or explored what type of imagery skill may be the

most important in constructing and updating a situation model, in order to aid comprehension.

Thus, the aim of the current thesis is to investigate the role of different subtypes of visual imagery in reading comprehension separately and via situation model construction. Additionally, the current thesis seeks to provide additional evidence that currently utilised measures of reading comprehension do not tap into all of the skills necessary for comprehension of written texts, and newer measures based on cognitive theory may be more useful for identifying specific skill deficits that lead to comprehension difficulties. The practical implications of this research may be far-reaching, as knowledge of all of the skills involved in reading comprehension is vital for improving the measurement of reading comprehension in research and practice. Specifically, this research may produce further evidence that current methods of assessing reading comprehension are not sufficient for identifying all types of reading difficulties, and may aid in the development of more effective ways to identify individuals with comprehension difficulties by identifying possible deficits in situation modelling and visual imagery ability. Consequently, this could aid in the future development of literacy interventions aimed at imagery production and/or situation model construction.

Accordingly, three studies will be conducted meet this aim. The first study will investigate the psychometric properties of several imagery measures, to identify which would be the most informative when used with children, and to further determine whether imagery would be best measured as a single skill, or as several subskills. The second study will then examine the influence of several different subtypes of visual imagery on individual differences in reading comprehension, along with other

constructs known to be relevant to reading comprehension, such as word-reading ability and verbal working memory, using two measures of comprehension: a traditional standardised measure, and a newer measure developed from cognitive theory. Finally, the third study will determine whether good and poor comprehenders differ in their generation of predictive inferences (a higher-level skill central to the construction of a coherent situation model), and examine the role of visuospatial imagery in this inferencing process. The following subsections provide a review of the literature that outlines the potential relationship between visual imagery ability and reading comprehension, theories and evidence of situation models and their relationship to comprehension, and criticisms of previous methods of measuring comprehension, in order to provide a context and rationale for the current studies.

1.2 Visual Imagery in Language and Reading

1.2.1 Dual Coding Theory

Imagery has long been theorised to play a central role in cognition, stemming as far back as the writings of Aristotle, who asserted that "thought is impossible without an image" (Aristotle, 350BC/1961). In relation to reading comprehension, the role of imagery in understanding language has its most predominant theoretical roots in dual coding theory, which was proposed by Paivio in 1971. Dual coding theory proposes that two separate but interconnected subsystems, referred to as a dual coding system, are involved in coding mental representations: a verbal subsystem specialised for dealing with verbal representations such as language; and a nonverbal subsystem specialised for dealing with nonverbal objects and events such as imagery (Paivio, 1971; 1986).

Dual coding theory highlights the importance of non-linguistic mechanisms such as imagery in discourse comprehension, by proposing that all mental representations retain some of the qualities of the sensory experiences from which they are derived. As these experiences can be non-linguistic, reading may therefore activate mental images that retain the visual properties of their referents, including features such as size, colour and shape (Paivio, 1971; 1986). Subsequently, it has been argued that comprehension, especially of concrete language, is dependent on both verbal and non-verbal representations (Paivio, 1986). This is likely because information that is coded in two forms is expanded on, thus strengthening memory of it, while also deepening comprehension and consequently facilitating understanding and recollection of a text (Sadoski, Goetz, & Fritz, 1993). Consequently, researchers have found that non-linguistic mechanisms such as imagery, including not only visual imagery, but also motor, auditory, gustatory, haptic, olfactory and affectual information, are an important part of written discourse comprehension and recollection (Paivio, 2007; Sadoski & Pavio, 2004). However, the visual modality of these images has received the most attention in the literature, and is the focus of the current study, so will be addressed here in more detail.

1.2.2 Perceptual Symbols and Embodied Cognition

Although historically, theorists emphasised the importance of mental imagery in human thought and cognition (e.g., Galton, 1883), the twentieth century also bought with it theories of knowledge that centred around objectivism and symbol manipulation, inspired by developments in logic, statistics, programming language, and computer science (see Barsalou, 1999). In a comprehensive review, Barsalou (1999) describes how these theories projected the view that all cognitive representations are inherently non-perceptual, or what he describes as "amodal". Thus, these theories suggested that

knowledge is represented by symbols that do not resemble the external referent that produced them (see also van Dijk & Kintsch, 1983). From this viewpoint, since symbols are amodal, they are proposed to activate cognitive and neural systems separate from those used during perception itself. They are also essentially arbitrarily linked to the perceptual states that produce them, in the same way that words have arbitrary relations to the objects they represent (Barsalou, 1999).

However, the past few decades has seen a renewed interest in the role of visualisation during reading as, following on from dual coding theory, more recently developed theories have asserted that sensory modalities such as the visual system are involved in the representation of language and memory (Glenberg, 1997). One such prominent theory is perceptual symbols theory (Barsalou, 1999), which argues that cognition involves modal systems that utilise the same neural regions involved in actual perceptual experience to construct perceptual symbols that represent knowledge. Thus, unlike amodal systems, Barsalou (1999) proposed that cognitive representations are derived directly from perceptual experience and bear a strong relationship to their external referents.

Theories such as perceptual symbols theory are embedded in an embodied cognition framework, which suggests that cognitive processes including language are grounded in the same systems that govern direct perception and bodily action (Barsalou, 1999; Glenberg, 1997; M. Johnson, 1987; Lakoff, 1987; Lakoff & Johnson, 1980). Several studies support this view of embodiment in language comprehension by demonstrating that individuals construct representations that contain simulations across a range of perceptual modalities (i.e., visual, tactile, auditory, gustatory, olfactory, kinesthetic, and somatic) during narrative reading (Olivetti Belardinelli et al., 2009; Palmiero et al.,

2009). For example, participants are faster to make motor responses that are consistent with the direction of action described in a text (e.g., moving their hand to a button further away from their body, when reading a sentence that implies motion away from the body [e.g., "close the drawer"]; Glenberg & Kaschak, 2002), and activation of neural substrates within the motor and pre-motor cortex involved in performing an action (e.g.., "kick") have been found in response to processing the meaning of words denoting that action (Pulvermuller, 2005). Thus, when using an embodied, or "grounded cognition" (Barsalou, 2008) framework to investigate reading comprehension, it has been suggested that readers construct a perceptual and motor simulation of the situation described in a text in order to represent the text's meaning (Barsalou, 2008; Glenberg, 1997; Johnson-Laird, 1983), which develops in a manner similar to a real physical scene. These representations are often referred to in the literature as "mental models" (Johnson-Laird, 1983), or "situation models", a term coined by Van Dijk and Kintsch (1983) and which will be used in the current review of the literature.

1.3 Situation Models

1.3.1 Definition and Historical Overview

Research on situation models was initiated in the late 1970s and early 1980s with the first major model of comprehension that focused on higher-level cognitive processes. This model was proposed by Kintsch and van Dijk (1978), and has provided a foundation for most subsequent models of reading comprehension. These include the construction–integration model (Kintsch, 1988), the event-indexing model (Zwaan, Langston, & Graesser, 1995a), the resonance model (Albrecht & Myers, 1995; Albrecht & O'Brien, 1993; Myers & O'Brien, 1998; O'Brien & Myers, 1999), the causal network model (Trabasso & Sperry, 1985; Trabasso, van den Broek, & Suh, 1989), the structure

building model (Gernsbacher, 1990), the constructionist model (Graesser et al., 1994), and the landscape model (Tzeng, van den Broek, Kendeou, & Lee, 2005; van den Broek, Young, Tzeng, & Linderholm, 1999). These models are primarily focused on understanding the characteristics of the mental representation that results from processing and understanding discourse. Thus, although these models acknowledge the role of lower-level processes in reading, such as fluency, decoding, phonological processing and vocabulary, they are concerned foremost with higher-level comprehension processes (see also McNamara & Magliano, 2009, for a complete review of each of the models mentioned here).

The most comprehensive of these models of text comprehension is arguably the construction-integration model (Kintsch, 1988), which was an extension of the original comprehension model proposed by Kintsch and van Dijk (1978) and one of the first to move away from memory and schema-based accounts of discourse understanding, towards investigating the processes and strategies that actually take place during comprehension. The fundamental assumption of this model is that two phases occur during comprehension: construction, which refers to the activation of information contained in the text and additional information from the reader's knowledge base (both relevant and irrelevant), and integration, which refers to spreading activation that results in stronger activation for concepts linked to those that are activated and less activation for unrelated constructs (Kintsch, 1988; 1998). The more influential aspect of this model was, however, that it built upon the novel view that discourse comprehension comprises three different levels of representation: the surface form, the propositional textbase, and the situation model (Kintsch, 1988).

The term "situation model" was introduced based on the proposal that readers not only construct a representation of the text, but also generate a mental model of the situation described by the text (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Thus, situation models are used to describe what is essentially a coherent representation of a text's underlying meaning (van Dijk & Kintsch, 1983). Situation models are considered to be the highest level of text comprehension, and are distinguished from lower levels of text representations, including the lowest level - the surface form representation - which is a representation of specific words and syntax (van Dijk & Kintsch, 1983); and the second level - the propositional textbase representation - which is an abstract representation of the ideas present explicitly in the text (van Dijk & Kintsch, 1983). In comparison, situation models contain extensive information about the persons, events, actions and objects described in a text (van Dijk & Kintsch, 1983; Zwaan, Radvansky, Hilliard, & Curiel, 1998). Thus, together, the lower-levels are merely a mental representation of the text itself, whereas the situation model representation is a result of deeper processing of the meaning embedded within the discourse (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). A further critical distinguishing feature of situation model representations is that they include not only explicit knowledge found in the textbase, but also implicit information, such as a reader's background knowledge (van Dijk & Kintsch, 1983).

Advancement of situation model theory took place in the 1990s with the introduction of the event-indexing model (Zwaan, Langston, & Graesser, 1995a). This model extended the construction-integration model and previous research by clearly defining the different dimensions contained within the situation model. Previous research had focused mostly on a single dimension of situation model construction, namely, either temporal, spatial or causal information; whereas the event-indexing model proposed that situation models contain extensive information along multiple dimensions, including

space, time, protagonist, causation, and intentionality, which are simultaneously monitored and updated by a reader (Zwaan, Langston, & Graesser, 1995a; Zwaan, Magliano, & Graesser, 1995b).

Central assumptions of the event-indexing model also include that event relations and causal sequences drive situation model construction, and that situation models are dynamic, and thus convey events that take place in space and time with readers connecting events along these dimensions (Zwaan et al., 1998; Zwaan, Langston, & Graesser, 1995a; Zwaan, Magliano, & Graesser, 1995b). The event-indexing model also clearly outlines situation model construction in three stages, proposing that situation model construction begins upon reading the first clause of a text, as the meaning of the clause becomes activated in working memory, resulting in "the current model" which is up for construction (Zwaan & Radvansky, 1998; Zwaan, Langston, & Graesser, 1995a). Secondly, "the integrated model" is that which is being constructed, via integration processes that combine incoming information (either from the textbase or referential background knowledge) into the current situation model, a process often referred to as "updating" (Zwaan & Radvansky, 1998; Zwaan, Langston, & Graesser, 1995a). The "complete model" is then that which is stored in long-term memory after the entire text has been processed (Zwaan & Radvansky, 1998).

It has been proposed that, as the event-indexing model accounts for how texts with event-sequences are processed, it is particularly applicable to narrative comprehension (McNamara & Magliano, 2009). Consequently, this model has served as the major framework for explaining comprehension of this type of discourse. However, as will be highlighted throughout the following sections, there is substantial overlap between each of the other models. For example, they all explain how comprehension is achieved

through situation model construction by positing that a continual monitoring and integration of information must take place in order to achieve coherence of a text (McNamara & Magliano, 2009). Thus they are not mutually exclusive when explaining comprehension.

1.3.2 Situation Models as the Vehicle for Reading Comprehension

It is often claimed that comprehension necessarily relies on the construction of a detailed situation model. Two main components of situation modelling have been investigated in regards to their effects on individual differences in comprehension: coherence monitoring and inference generation.

1.3.2.1 Coherence

It is acknowledged in nearly all models of reading comprehension that paramount to comprehension is the process of mapping incoming discourse information to the prior discourse context in order to achieve both local and global coherence (McNamara & Magliano, 2009). Global coherence refers to coherence that results from the integration of incoming discourse with information that is no longer available in working memory (i.e., information that was encountered much earlier in the text) and relevant background knowledge (Albrecht & O'Brien, 1993; D. L. Long & Chong, 2001). In contrast, local coherence involves integrating information from incoming sentences with information still currently accessible to working memory (i.e., information relating to one or two sentences immediately prior; Albrecht & O'Brien, 1993; D. L. Long & Chong, 2001). Thus, local coherence can largely be achieved through a textbase representation, whereas global coherence is the result of the construction of a situation model that relies heavily on long-term memory.

Studies investigating global coherence have often relied on the measurement of a "contradiction effect" using reading times or eye-tracking methodology. A contradiction effect occurs when a reader takes longer to read a critical sentence that is, although locally coherent at the grammatical level, not consistent with information presented earlier in the text (Albrecht & O'Brien, 1993; O'Brien & Albrecht, 1992). For example, a text may state that "Mary is a vegetarian", and then later that "Mary ordered a hamburger". The second statement is thus inconsistent with information provided earlier, although it still makes sense at the textbase level. The longer reading time of the second example sentence suggests that the reader is engaged in maintaining global coherence of the situation described in the text rather than just creating local coherence, thus indicating a situation model has been constructed (O'Brien & Albrecht, 1992).

Earlier research on coherence resulted in the proposal of *the minimalist hypothesis*, which suggested that readers do not automatically establish or maintain global coherence, and will only do so when there is a break in local coherence and background knowledge is necessary to interpret the passage (McKoon & Ratcliff, 1992). However, subsequent studies have found that global inconsistencies affect online comprehension even when local coherence is maintained (Albrecht & Myers, 1995; Albrecht & O'Brien, 1993; Huitema, Dopkins, Klin, & Myers, 1993; Myers, O'Brien, Albrecht, & Mason, 1994; O'Brien & Albrecht, 1992; O'Brien, Rizzella, Albrecht, & Halleran, 1998). Thus, it appears that strategies relying foremost on textbase processing may not provide complete comprehension. This is also reflected in behavioural studies using the contradiction effect, which have found that both adults (D. L. Long & Chong, 2001) and children (Oakhill, Hartt, & Samols, 2005b; van der Schoot, Reijntjes, & van Lieshout, 2011) classified as good comprehenders maintain both local and global coherence, whereas poor comprehenders predominantly maintain local coherence only. These

studies demonstrate that individual differences in reading comprehension may be a reflection of situation modelling ability. Specifically, whereas good comprehenders build and update situation models, poor comprehenders construct predominantly textbase representations only. However, beyond this, few studies have examined whether individual differences in reading comprehension are related to global coherence. Further, the measurement of comprehension differs greatly between these studies, from being based on answers to comprehension questions about the experimental passage read during the task (i.e., D. L. Long & Chong, 2001), to being measured with a standardised test of reading comprehension (i.e., Oakhill, Hartt, & Samols, 2005b; van der Schoot et al., 2011).

1.3.2.2 Inference

Paramount to constructing a situation model that is both integrated and coherent is the generation of knowledge-based inferences. Knowledge-based inferences are those that incorporate information extraneous to the text (i.e., background knowledge such as world knowledge and episodic knowledge of past events including previously encountered textbases), with information provided explicitly in the text, to fill in missing details (Cain & Oakhill, 1999). Knowledge-based inferences are often required to maintain global coherence of the situation described in a text (Albrecht & O'Brien, 1993; Graesser et al., 1994; Kintsch, 1988). Conversely, coherence or "bridging" inferences are used to maintain local coherence at the level of the textbase (i.e., by linking together premises stated explicitly in a text; Bowyer-Crane & Snowling, 2005; Graesser et al., 1994; McNamara & Magliano, 2009). Based on this, it is generally assumed that if a reader's activation of knowledge is confined to what is explicitly stated in the text, their situation model will be less globally coherent, resulting in a level

of understanding that does not exceed the textbase level (McNamara & Magliano, 2009).

A variety of knowledge-based inferences may occur via situation model construction, and can be the result of either passive inferential processes (which take place automatically) or strategic processes (which require readers' working memory and attentional resources; van den Broek, Rapp, & Kendeou, 2005). However, it has been suggested that during narrative comprehension these processes are generally more automatic (due to the familiarity of topics and ease of reading that accompanies these types of texts) as compared to other more demanding texts, such as expository texts used for learning (McNamara & Magliano, 2009). In relation to the event-indexing model, it is proposed that inferences are drawn for each of the five dimensions (i.e., space, time, protagonist, causation, and intentionality). Evidence for this integration of implicit knowledge has come from several studies that have shown that during reading comprehension inferences are made in relation to (i) spatial relations (Rinck, Williams, Bower, & Becker, 1996; Tversky, 1993), (ii) future events (Fincher-Kiefer, 1993), (iii) the characteristics of objects and protagonists, including their emotions (Gernsbacher, Goldsmith, & Robertson, 1992) and gender (Oakhill, Garnham, & Reynolds, 2005a), (iv) causal antecedents and consequences (Kuperberg, Paczynski, & Ditman, 2011), and (v) protagonists' goals (D. L. Long & Golding, 1993).

In contrast to studies on coherence, there is a large amount of evidence that suggests knowledge-based inference generation is related to narrative comprehension in both adults (Perfetti, Landi, & Oakhill, 2005), and children (Cain & Oakhill, 1999; Cain, Oakhill, Barnes, & Bryant, 2001; Elbro & Buch-Iversen, 2013; Oakhill, 1984) even as young as 4 years old (Kendeou et al., 2008; Tompkins, Guo, & Justice, 2013). Further,

this relationship exists whether a story is to be comprehended via oral, picture, or text presentation (Kendeou et al., 2008), indicating the involvement of a situation model rather than a textbase representation. This is not surprising, as knowledge-based inferences provide extended information about several narrative features (Graesser et al., 1994).

As the event-indexing model highlights the importance of making causal connections during reading, inferences about causal antecedents and consequences of events are hypothesised to be made routinely during situation model construction and updating. This is a view shared by the causal-network model (Trabasso et al., 1989; Trabasso & Sperry, 1985), which suggests that the primary basis for constructing a coherent situation model of a narrative is the generation of causal inferences. Indeed, an extensive amount of research has supported the notion that making causal connections is an important part of the reading comprehension process (Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Fletcher & Bloom, 1988; Lynch & van den Broek, 2007; Lynch et al., 2008), and the ability to draw causal inferences has also been demonstrated to be directly related to level of reading comprehension (Cain & Oakhill, 1999; 2006; Kendeou et al., 2008; Tompkins et al., 2013; van Kleeck, 2008). It has been proposed that children obtain a greater understanding of the "how and why" of the events described in a text if they can understand this causal structure (Kendeou et al., 2005).

Of course, as these causal connections are not often explicitly stated in a text, this understanding will likely be deeper if the child can go beyond what is mentioned in a text and infer these causal connections (Kuperberg et al., 2011). In addition, causal inferences can relate to several story dimensions, including initiating an event, action or

problem; potential solutions to problems; and consequences of events and actions, including the emotional responses of a character (Graesser et al., 1994; van Kleeck, 2008). As such, children with adequate low-level reading skills, but poor comprehension, have been found to draw fewer causal inferences than good comprehenders (Cain & Oakhill, 1999; 2006), and the number of causal inferences made during reading has been shown to be a predictor of comprehension ability (Kendeou et al., 2008). Causal inferencing has even been found to contribute to reading comprehension over and above other inference types (such as inferences about places, and character dialogue; Kendeou et al., 2008; Tompkins et al., 2013).

Another consistent finding in the reading comprehension literature is that both younger and older children's inference skills predict variance in comprehension that goes over and above that contributed by lower level reading skills (Cain et al., 2004a; Kendeou et al., 2008; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012; Oakhill & Cain, 2012; Tompkins et al., 2013). In addition, cross-sectional evidence demonstrates a significant contribution of inference to listening comprehension in pre-schoolers, even after controlling for age, verbal memory, receptive vocabulary, and verbal IQ (Florit, Roch, & Levorato, 2011). Inference generation has been found to be a greater predictor of narrative comprehension than lower-level skills such as vocabulary knowledge and grammar in 4- to 6-year-olds (Kendeou et al., 2008; Lepola et al., 2012; Tompkins et al., 2013). In addition, longitudinal studies have found that these earlier contributions of inferencing to reading comprehension remain at a later age (Lepola et al., 2012; Silva & Cain, 2015). Specifically, Silva and Cain (2015) found they remained after one year, and were independent of grammar and literal comprehension (Silva & Cain, 2015), and Lepola et al. (2012) found that inference making skills at ages 4 and 5 uniquely contributed to narrative comprehension at age 6. The predictive power of inference

generation on reading comprehension also appears to increase with age, as the percentage of variance predicted by inference generation has been demonstrated to increase from age 4 to 6, and again from age 6 to 8 (Kendeou et al., 2008). Thus, it appears that inference skills are central to reading comprehension in even the preliminary stages of reading development (Silva & Cain, 2015).

In addition, the evidence that inference making skills in earlier years uniquely contribute to narrative comprehension at later ages suggests a causal effect of inferencing on comprehension (Lepola et al., 2012; Oakhill & Cain, 2012). The causal relationship of inference generation to reading comprehension is also supported by intervention studies: instruction aimed at increasing inference generation was found to improve the listening comprehension of first-grade children in comparison to those who did not receive training (A. H. Paris & Paris, 2007), and also the reading comprehension of 7- to 8-year-old children who received inference training, in comparison to those who received decoding training (Yuill & Oakhill, 1988). Additionally, Cain and Oakhill (1999) matched skilled and less-skilled comprehenders for reading accuracy, sight vocabulary, and chronological age, and included a comprehension-age match group of younger normally developing children, whose comprehension ability was equivalent to that of the less skilled comprehenders. It was found that the comprehension-age match group performed better at text-connecting inference generation than the poor comprehenders. This suggests that inference making skills do not occur as a consequence of proficient reading comprehension, but rather that poor inferencing ability leads to comprehension deficiencies.

Thus, it appears that when an individual has more knowledge about a particular topic or domain, their situation model will be more coherent, resulting in deeper comprehension.

However, it has also been suggested that the strategy used to select information for activation and integration affects the ease and success of inference generation and consequently, situation model updating (McNamara & Magliano, 2009). While earlier theories proposed information is selected on the basis of recency of mention (Kintsch & van Dijk, 1978), such an approach does not take into account selection and integration of background knowledge (O'Brien & Albrecht, 1992; van Dijk & Kintsch, 1983). Thus, several strategies for how this mapping process takes place have been proposed, each based on the models of comprehension outlined earlier (see pp. 16-17). However, it is unlikely that comprehension is a result of any one of these strategies, but rather varies depending on the demands of the reading task, reader's goals, and textual constraints (McNamara & Magliano, 2009). As such, there currently appears to be no consensus as to which is the most commonly used during situation model construction, and there remains clear overlap between the proposals put forth by these models. Specifically, a central tenet of all of these models is that the presence of situational cohesion (connections related to actions and events) is a key component of successful reading comprehension.

For example, the event-indexing model proposes that events (including actions of protagonists) are the main focal point for basis of situation model monitoring and updating (Zwaan, Langston, & Graesser, 1995a), and that events are indexed along five dimensions (see pp.18-19) based on how many features they share with the current model (Zwaan & Radvansky, 1998; Zwaan, Langston, & Graesser, 1995a). In contrast, strategies based solely on causal reasoning have also been proposed. For example, theories based on the causal-network model (Trabasso et al., 1989; Trabasso & Sperry, 1985) suggest that readers keep active the most recent causal antecedent without a consequence (Fletcher & Bloom, 1988). Although some authors claimed that causal

monitoring only occurs at the textbase level (McKoon & Ratcliff, 1992), in cases where the consequence of an event is not explicitly stated in a text, readers may have to use background knowledge to infer the likely consequence, in order to maintain coherence. Thus, it has since been recognised that some causal connections occur at the level of the situation model (Fincher-Kiefer & D'Agostino, 2004; Kuperberg et al., 2011).

Lastly, proponents of memory-based models such as the resonance model (Albrecht & Myers, 1995; Albrecht & O'Brien, 1993) have proposed that the events, actions, thoughts and objects that are foregrounded and kept active in working memory, are those which are relevant to the visual perspective of the protagonist (O'Brien & Albrecht, 1992). This theory is supported by various studies that demonstrate that both adult and child readers adopt the point of view of the protagonist during narrative comprehension, and maintain information relevant to their actions, events, thoughts, and objects in their possession (J. B. Black, Turner, & Bower, 1979; Bower & Morrow, 1990; O'Brien & Albrecht, 1992; O'Neill & Shultis, 2007; Rall & Harris, 2000; Ziegler, Mitchell, & Currie, 2005), even when information about the protagonists perspective is implied rather than explicitly stated (Morrow, Bower, & Greenspan, 1989; Rall & Harris, 2000). This model aligns with the view that situation models are perceptual simulations of what is described in a text, with readers becoming "embodied" in the narrative experience. Similar to the event-indexing model, these models add to existing frameworks of comprehension by explaining how visuospatial and perceptual information is important in situation model construction and updating.

1.3.3 Perceptual Information in Situation Models

As outlined, drawing from theories of embodied cognition (Lakoff, 1987; Lakoff & Johnson, 1980) and perceptual symbols (Barsalou, 1999), it is often acknowledged that

situation models resemble a perceptual simulation of the scene described in text, which is supported by the same neural areas that produce actual perception and bodily movement (Barsalou, 2008; Speer, Reynolds, Swallow, & Zacks, 2009). Several studies have supported the notion that simulations of motor movement (Kaschak et al., 2005; Zwaan et al., 2004) and perceptual information including visual and auditory imagery (Bergen et al., 2007; Brunyé, Ditman, Mahoney, Walters, & Taylor, 2010; Klin & Drumm, 2010; Stanfield & Zwaan, 2001; Zwaan et al., 2002) are activated as part of the situation model, although simulations of visual information have received the most attention in the literature.

As some proponents of memory-based models of comprehension (i.e., the resonance model; Albrecht & Myers, 1995; Albrecht & O'Brien, 1993) proposed that readers adopt the perspective of the protagonist as a strategy to maintain relevant information for integration (O'Brien & Albrecht, 1992), empirical investigation of evidence for this proposition provided initial support for the notion that visual imagery is activated during situation model construction. For example, it was found that adults read a deictic verb of motion (i.e., come/go) more quickly if it is spatially consistent with the point of view of the main protagonist (J. B. Black et al., 1979), and these findings have since been replicated and extended to children (Rall & Harris, 2000; Ziegler et al., 2005).

Furthermore, as actions and motion are played out in space, several studies have also examined whether spatial information about the environment is represented in situation models. Findings of these studies suggested that this is the case, as readers are faster to recognise target objects described as being located closer to (i.e., in the same room), rather than further away from, the reader's focus of attention (i.e., the location of the protagonist; Haenggi, Kintsch, & Gernsbacher, 1995; Morrow et al., 1989; Morrow,

Greenspan, & Bower, 1987; Rinck et al., 1996; Rinck, Bower, & Wolf, 1998). Also, items physically associated with a protagonist (e.g., an item they were carrying, or wearing) are recognised faster than disassociated items (e.g., an item the protagonist had just set down or removed; Glenberg, Meyer, & Lindem, 1987; Radvansky & Copeland, 2006). This facilitation effect occurs even if an object further away was mentioned more recently (Morrow et al., 1987) or when the name of the target rooms are not explicitly mentioned (Haenggi et al., 1995; Rinck et al., 1998), thus indicating that these spatial-separation effects are not simply due to name-based lexical priming.

However, although these findings provide supporting evidence that situation models contain spatial information, they cannot completely conclude that *visual imagery* is involved in these spatial representations. For example, Rall and Harris (2000) note that the results from their study could not make the distinction between whether participants adopted an internal perspective of the character, or rather an external view of the scene described (i.e., as an observer), treating the location of the character as a landmark or "anchor", from which they code any movement that is described in the narrative (i.e., towards, or away, from the anchor). Should a reader be constructing a model from this external view, it is possible that they are constructing a propositional representation of the locations of the characters and objects in a story.

Thus, what is perhaps the most compelling evidence that these representations are not merely propositional comes from the findings of perceptual mismatch studies. These studies utilised sentences that included a manipulation of a target object's implied orientation (e.g., "The man hammered a nail into the floor" versus "The man hammered a nail into the wall"; Stanfield & Zwaan, 2001), or shape (e.g., an egg in a carton versus an egg in a frying pan; Zwaan et al., 2002). After reading these sentences, participants

viewed images of the target object, and decided whether the object had been mentioned in the previous sentence. Significantly faster response times were found when the image matched the orientation or shape implied by the sentence than when it did not match (Stanfield & Zwaan, 2001; Zwaan et al., 2002). These studies were among the first to indicate that readers not only construct perceptual simulations of objects, but that this simulation occurred at the situation model level, as the correct orientation or shape was only *implied* by the text, but not explicitly mentioned in the textbase (Stanfield & Zwaan, 2001). These findings have since been replicated, in populations of both younger and older adults (Dijkstra et al., 2004; Zwaan & Pecher, 2012), and with children aged 7 to 13 years old, whose responses to the picture verification task showed evidence of the mismatched picture effect when both listening to texts presented aurally, and when reading written sentences out loud (Engelen et al., 2011).

More recently, the perceptual mismatch effect has been utilised to determine that these representations are also dynamic (i.e., include simulations of motion; Zwaan et al., 2004). This was achieved by presenting participants with sentences that described the motion of a ball either toward or away from an observer (e.g., "The pitcher hurled the softball to you"), followed by a pair of images of the ball that represented towards or away movement (i.e., by presenting the second image as either slightly larger or smaller than the first image). Similar to previous studies, participants were faster to judge whether the two objects were the same when the implied movement of the balls in the images matched the movement described in the sentence (Zwaan et al., 2004).

However, studies investigating colour have been inconsistent. When presenting participants with sentences such as ''John looked at the steak in the butcher's window'' followed by a picture of a red (match) or brown (mismatch) steak, contrary to studies on

shape, orientation and motion, Connell (2005) found significantly faster response times to the mismatching than to the matching items. In contrast, however, when attempting to replicate the findings of Connell (Connell, 2005; 2007), Zwaan et al. (2012) found the opposite pattern of results: a mismatch effect did occur between the colour implied by the text, and that of the presented object. Although the findings of Zwaan et al. (2012) appear to be more theoretically logical, both these sets of findings have been interpreted as support for perceptual simulation. Specifically, Connell argues that colour, as opposed to orientation and shape, is not as salient as other object properties, or important for object recognition, and is therefore encoded with less stability in mental representations (Connell, 2007). Consequently, there is minimal interference when perceptual input mismatches perceptual simulation on an unstable property, thus this unimportant unimodal mismatch can easily be ignored. Yet, given the mixed and limited research in this area, further research on colour simulations in reading comprehension is required to resolve these discrepancies.

Regardless, the studies reviewed here suggest that linguistic input is not represented merely as propositions but rather perceptual symbols that bear a resemblance to their referents, including their shape, orientation and motion. Importantly, these object features were activated even when they were not explicitly mentioned in the text. Thus, participants were likely constructing situation model representations by activating implicit knowledge of these structures from long-term memory.

Furthermore, Horton and Rapp (2003) found evidence that readers do mentally simulate what appears to be the *visual perspective* of a protagonist's point of view. As situation models reflect a reader's knowledge of an on-going situation, Horton and Rapp (2003) hypothesised that if situation models utilised perceptual information, the availability of

such information would change as a function of the narrative. Further, this effect would occur when the perceptual availability was implied, rather than explicitly indicated in the text. To investigate this, Horton and Rapp (2003) presented participants with narratives of situations that either resulted in part of the protagonist's view being blocked (e.g., a large truck in front of a mailbox) or did not describe any occlusion of vision (e.g., a bicycle in front of a mailbox). It was found that participants were slower to respond to verification questions about objects when they had been blocked from the vision of the protagonist than when the object had not been blocked by the critical event. Further, in a second experiment it was found that this effect did not generalise to other objects in the narrative that had been mentioned prior to the critical event, but had not been blocked from view. Thus, these results provide evidence that it was not just a shift in event that caused all prior story information to become less accessible, but rather reduced availability of information about objects only occurred for those objects that were no longer part of the protagonist's perceptual perspective. Therefore, it appears that readers do in fact represent story information in a manner that is somewhat analogous to actual visual perception (Horton & Rapp, 2003).

In addition, Bergen et al. (2007) provided stronger evidence for the use of visual imagery during reading by using a dual-task paradigm. In this study, Bergen et al. (2007) found that listening to literal sentences about real space (e.g., "the ant climbed/dropped") interfered with performance of a visual task (deciding whether an object is a circle or a square), when the object was in the same location on the screen as denoted by the verb in the sentence (i.e., top of the screen for "climbed"). However, this effect did not occur when listening to metaphorical sentences that contained motion verbs, and therefore did not denote literally perceivable action (e.g., "stock prices climbed/dropped"), or abstract verbs (e.g., "wane"). This study provides further

evidence that visual imagery is activated to create a mental simulation of the *meaning* of a sentence but not in response to simple lexical associations. More recently, this finding has been supported by neuroimaging studies investigating motor simulation in reading comprehension. Specifically, these studies have provided evidence towards a weak version of the embodiment hypothesis (that is, activation of motor information is dependent on context; Raposo, Moss, Stamatakis, & Tyler, 2009; Schuil, Smits, & Zwaan, 2013). For example, by finding activation of the motor cortex in response to action verbs that are embedded in literal sentences (e.g., "kick the ball") but not non-literal sentences (e.g., "kick the habit"; Schuil et al., 2013).

However, it is also recognised that evidence for the "strong" version of this embodiment hypothesis has been found, which suggests that activation of sensory-motor regions of the brain can occur during comprehension of action verbs regardless of whether they are presented in a literal (e.g., "he grasped the cup") or non-literal (e.g., "he grasped the concept") context (Boulenger, Hauk, & Pulvermuller, 2009; Jirak, Menz, Buccino, Borghi, & Binkofski, 2010; see also Gallese & Lakoff, 2005), although this hypothesis has been largely investigated in relation to motor, rather than visual, imagery. Thus, more information regarding the role of visual and motor simulation in both the comprehension of literal and non-literal language is needed.

Regardless, several neuroimaging studies have revealed neural activity that is consistent with the activation of visual imagery during language comprehension, and the neural substrates involved in actual bodily movement have also been found to overlap with those that are activated while reading words, or extended passages, that denote the perceptual input or movement (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; R. F. Goldberg, Perfetti, & Schneider, 2006a; 2006b; Hauk & Pulvermuller, 2004; Hauk,

Johnsrude, & Pulvermuller, 2004; Just, Newman, Keller, McEleney, & Carpenter, 2004;

Pulvermuller, 2005; Speer et al., 2009). Similarly, significant activation of brain regions

involved in visual, tactile, auditory, gustatory, olfactory, kinesthetic, and somatic

perception has been found in response to phrases describing experiences via these

senses (Olivetti Belardinelli et al., 2009; Palmiero et al., 2009). In addition, Just et al.

(2004) examined neural activation while participants read or listened to high-imagery

sentences (e.g., "the number eight when rotated 90 degrees looks like a pair of

spectacles") or low-imagery sentences (e.g., "although now a sport, marathons started

with Greek messengers bringing news"), and made judgments about their accuracy. For

high imagery sentences, more activation was found in regions that are activated in other

mental imagery tasks, such as mental rotation (particularly, the intraparietal sulcus), for

both auditory and visual presentation of the sentences, compared to low-imagery

sentences, thus indicating a neural mechanism for language-evoked imagery is not

dependent on the mode of presentation.

Further, Speer et al. (2009) found activation in brain areas involved in the manual

manipulation of objects (i.e., pre-central and parietal regions associated with grasping

hand movements), the navigation of spatial environments (i.e., right and left

parahippocampal cortex areas), and the processing of goal-directed human activity

(Brodmann's Area and the pre-frontal cortex) when those aspects of the narrated

situation changed during reading. This suggests that the brain regions involved in the

actual performance of these activities are also involved in story comprehension, and

readers use and dynamically update perceptual and motor representations in the process

of narrative comprehension (Speer et al., 2009).

Moreover, by comparing the activation found in their study with previous work, Speer et al. (2009) suggested that the collection of neural regions associated with situational changes overlapped with those that are also activated during imagination or observation of these activities. They proposed that this largely resembled a pattern of activation that had been found to correspond to the act of "projecting one's self into a remembered, anticipated, or imagined situation" (p. 997). Thus, it appears that readers' situation models are constructed from dynamic sensory and motor representations, which may reflect a more general neural mechanism that enables cognition to be grounded in real-world experiences in order for individuals to communicate ideas and experiences more efficiently and vividly. However, while it can be concluded from these studies that visual simulation during language processing may largely be unconscious and automatic, it is yet to be fully ascertained whether such visual and motor simulation is indeed *necessary* for comprehension.

Some studies have, however, provided evidence that alludes to this possibility. For example, Fincher-Kiefer and colleagues found that situation models not only *include* perceptual information, but may *require* it for construction (Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004). Using a contradiction effect (see p. 21) to measure situation model construction, Fincher-Kiefer (2001) found that readers had more difficulty maintaining global coherence when required to hold high-imagery sentences in memory, than when holding low-imagery sentences (which were equated on word length, comprehensibility and truth agreement). Specifically, under the high-imagery load condition, participants did not show evidence of identifying critical sentences that were, although locally coherent at the grammatical level, not consistent with information presented earlier in the text. Based on dual-logic theory (Baddeley, 1992), which proposes that a disruption of a cognitive process will occur if the resources it

requires are utilised simultaneously by a separate task, it was concluded that the failure to maintain global coherence was due to the perceptual resources required for situation model construction being utilised in the visual memory task.

Similar effects were found in a later study, when a visual memory task interrupted readers' ability to draw predictive inferences from a text (Fincher-Kiefer & D'Agostino, 2004). In this study, Fincher-Kiefer and D'Agostino (2004) presented a group of adult participants with short texts designed to elicit either a predictive inference (experimental condition) or no inference (control condition), under one of two between-group conditions: either while holding a visuospatial memory load (an array of five dots within a 4 x 4 grid) or a verbal memory load (a string of six letters), and measured reaction time to subsequently presented target words related to the inference. It was found that participants given a verbal memory load showed the typical facilitation effect to predicted inference targets but participants given a visuospatial memory load showed a reduced facilitation effect. It thus appears that perceptual resources are also vital in order to construct predictive inferences, likely because predictive inferences are knowledge-based and thus occur at the level of the situation model.

In support of this interpretation, Fincher-Kiefer and D'Agostino (2004) also found that an additional visuospatial load did not disrupt inferencing when the experimental passages were designed to elicit *bridging* inferences, rather than predictive inferences. Bridging, or textbase inferences, differ from knowledge-based inferences as they are used to maintain local coherence of a narrative at the textbase level (i.e., to make links between premises in a text), rather than being elaborative and requiring the integration of background knowledge (Fincher-Kiefer & D'Agostino, 2004; see also van Kleeck, 2008). Thus, it was inferred that bridging inferences do not require perceptual resources,

as unlike predictive inferences, they are not involved in situation model construction. Thus, it appears that visuospatial resources may possibly aid predictive inference generation because they allow readers to envision the interaction of objects and events described in a text, in order to draw conclusions about the likely consequences of these events, which can then be incorporated into the overall situation model. This is an interpretation shared by other researchers who describe how the simulation of described situations activates supplementary information, such as affordances, emotional responses, and typical situational constrains, which can be incorporated in to the situation model to aid in the comprehension process (Marmolejo-Ramos, de Juan, Gyax, Madden, & Roa, 2009).

It is therefore possible that a visually rich, more dynamic situation model may advance reading comprehension, as it enables a reader to become embodied in the story experience, facilitating meaning generation. However, while the aforementioned studies provide evidence that imagery is an important process in building a coherent situation model, they have only assessed how this affected online comprehension during a controlled task. More research is needed on whether individual differences in overall reading comprehension level are related to this use of visual and spatial information during situation model construction, in order to determine whether this may be an area for more targeted reading interventions in a developmental context.

1.3.4 Developmental Studies

Research on the development of situation models in children has been lacking up until recent years. However, there is now increasing evidence that children also use situation models to represent the meaning of a text (Barnes, Raghubar, Faulkner, & Denton, 2014; O'Neill & Shultis, 2007; Pyykkönen & Järvikivi, 2012; Rall & Harris, 2000;

Uttal, Fisher, & Taylor, 2006; van der Schoot et al., 2011; Ziegler et al., 2005; Ziegler & Acquah, 2013), and similarly to those of adults, these representations are dynamic (Fecica & O'Neill, 2010) and include perceptual symbols such as visual imagery (Engelen et al., 2011).

The spatial properties of children's situation models are also evident from a young age (Barnes et al., 2014; Nyhout & O'Neill, 2013; Rall & Harris, 2000; Ziegler et al., 2005; Ziegler & Acquah, 2013). Rall and Harris (2000) aurally presented 3- and 4-year-olds with stories and measured accuracy of recall of the story as the dependent variable (rather than reading time, which is often explored in adult studies), and found that children recall a narrative more accurately when verbs denoting motion are consistent with the protagonist's perspective, but make substitution errors (e.g., replace *come* with *go*) on verbs that are inconsistent with that perspective. Replications of this study by Ziegler et al. (2005) also found that this effect remains with unfamiliar stories, and regardless of whether the protagonist is depicted as being *good* or *bad*, thus indicating that perspectives are adopted to assist with maintaining understanding of the story, not simply to empathise with the character.

Additionally, children show variability in their situation model constructions, which may be a predictor of reading comprehension ability. For example, in a series of studies using an eye-fixation methodology, van der Schoot and colleagues (van der Schoot et al., 2011; van der Schoot, Horsley, & van Lieshout, 2010; van der Schoot, Vasbinder, Horsley, Reijntjes, & van Lieshout, 2009) found that from ages 10 through to 12 years, poor comprehenders do not build as rich and elaborate situation models as good comprehenders, although they do build adequate textbase representations to maintain local coherence. For example, poor comprehenders tend to exclude situation-relevant

information from their situation models that could be used to identify later contradictions in text information (van der Schoot et al., 2011). Specifically, they fail to maintain global coherence (van der Schoot et al., 2011); spend as much time processing information that is less relevant to the goal of the text as information that is relevant (whereas good comprehenders spend more time processing relevant information only; van der Schoot et al., 2011); allocate more of their processing capacities to textbase variables (e.g., number of syllables, word frequency, and number of new concepts) than situation model variables (e.g., gaps in temporal or spatial story information; van der Schoot et al., 2010); and take longer to resolve ambiguous word references (i.e., refer back to information presented earlier in a text to generate an inference about the meaning of the currently encountered expression; van der Schoot et al., 2009). Additionally, Pyykkönen and Järvikivi (2012) found that 8-year-olds display difficulties in comprehending sequential temporal events, which may originate from their inability to revise their situation model representation of the events when required by the text (e.g., when the events described are not presented in chronological order), although these authors did not examine this in relation to the children's overall reading comprehension level.

Further, it appears that visual imagery may play an important role in children's situation model construction that has effects on comprehension. For example, van der Schoot et al. (2010) found that situation model instruction aimed at encouraging children to enhance their imagery abilities resulted in more correct answers to comprehension questions, and enabled poor comprehenders to redistribute more resources from textbase processing to situation modelling (as indicated by slower reading times and eye-fixation on situation model versus textbase variables). Although, it appeared their situation models were still not as extensive as good comprehenders, and instruction did not result

in poor comprehenders having better memory performance for situation model information. However, this may be a reflection of the type of instruction used, which the authors note was a direction to do something rather than any form of teaching or activity. Thus "instruction" here simply meant participants were asked to "imagine the events and developments described in the story" (van der Schoot et al., 2010, p. 824), in contrast to asking them to *understand* what the text is about. As with other reading and educational interventions, one can assume that more intensive activity-based intervention is likely required for further improvements in reading comprehension. Further, the authors could not conclude which aspect of situation model construction the imagery-based instruction specifically contributed to (i.e., inference making, updating or integration; van der Schoot et al., 2010).

More recently, Nyhout and O'Neill (2013) investigated how children's spatial situation models affect story recall, by measuring 7-year-olds' performance on reconstructing the layout of a described neighbourhood. It was found that performance on this task was better when the layout was presented as a narrative as compared to a description, although both included the same amount of spatial information (Nyhout & O'Neill, 2013). This was interpreted by the authors as being due to readers being able to build a situation model representation centred around the character's motivations and actions in the narrative condition, possibly by adopting the character's point of view, whereas in the description condition this was not possible. This view is supported by Ziegler et al. (2005), who found that although a shift of perspective can occur for stories that lack a principal protagonist, this shift is easier when there is a principal protagonist involved, thus indicating that imaginatively placing oneself into a story lends to ease of understanding (Ziegler et al., 2005).

More extensive evidence of the importance of visuospatial information in children's situation model construction and reading comprehension has recently been provided by Barnes et al. (2014) using a task in which children aged 9 to 16 memorised a physical model of a market place, and then read stories describing a protagonist traversing the same marketplace. During reading, children were periodically presented with the names of two objects from the market and had to indicate if these objects were from the same or different shops. Children were faster at identifying objects in areas traversed by the protagonist than objects in locations the protagonist had not passed through, indicating participants had adopted the protagonist's mental perspective (Barnes et al., 2014). Further, objects that were in locations that were not mentioned but relevant from the protagonist's perspective were responded to faster than those in explicitly mentioned but less relevant locations. Thus, the effects could not be interpreted simply as a result of lexical-priming of the objects due to reading the shop name, but rather, indicated that mental access to these objects was based on their spatial location in a situation model. This was also supported by the fact that the objects contained within the marketplace had weak semantic associations to the shops at which they were located (thus limiting the effects of pre-existing associations between these shops and objects; Barnes et al., 2014).

Furthermore, Barnes et al. (2014) found that faster access to this inferred spatial information in the situation model predicted reading comprehension but not decoding. Specifically, decision times to probes of objects in locations not explicitly mentioned, but traversed by the protagonist, uniquely predicted reading comprehension after accounting for word decoding. Thus, it was concluded that the ability to update a situation model of the text based on inferred information might be particularly important for reading comprehension, especially implicit spatial location and object

information from the protagonist's perspective. However, the authors did note that the texts used, in conjunction with the method of having children first memorise the marketplace, may have encouraged a strategy that resulted in the activation of a greater amount of visuospatial information during situation model construction than what might occur in other reading situations, or with other types of text (Barnes et al., 2014).

Lastly, embodied accounts of how reading becomes meaningful also support the notion that the development of a visually rich story representation may enhance reading comprehension in early years. Although not focusing explicitly on situation model construction, Glenberg and colleagues (Glenberg & Kaschak, 2002; Glenberg & Robertson, 1999; 2000) proposed the indexical hypothesis, which asserts that some children fail to obtain meaning from a text as they do not consistently map (i.e., "index") written words to the objects the words represent. Thus, even when the words are read and pronounced correctly, these children fail to derive any meaning from a text (i.e., reading is merely an exercise in word naming, which fails to engage the reader, or lead to comprehension; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004). Empirical studies have supported this view by demonstrating that interventions that aim to strengthen indexing, by encouraging simulation of the actions and events described in a text using physical objects, or images on a computer, result in better text memory and comprehension in young children (Glenberg et al., 2004; Glenberg, Goldberg, & Zhu, 2011; Marley, Levin, & Glenberg, 2007; 2010). Importantly, and in relation to mental models, these improvements in comprehension have also been shown to result not only from physical simulation but also the *imagined* manipulation of the story objects and events (Glenberg et al., 2004).

1.3.5 Working Memory and Situation Model Construction

As has been shown in the current literature view, two main components of situation modelling that have been identified as important for comprehension include coherence monitoring (in particular, global coherence) and knowledge-based inference generation. However, the critical role identified for these skills does not completely rule out the importance of other cognitive abilities in the construction of situation models and determination of reading comprehension. As highlighted, previous literature also indicates a role for visual imagery. Further, inference itself draws on other linguistic skills and cognitive resources. In particular, working memory has been implicated in both inference generation and maintaining coherence, and both verbal and visuospatial working memory components have been shown predict overall reading comprehension level.

For example, in light of evidence for the event-indexing model, which proposes that readers monitor the temporal and causal dimensions of situation models separately from the spatial dimension (Zwaan, Magliano, & Graesser, 1995b), Friedman and Miyake (2000) suggested that separate working memory subsystems are implicated in the construction and monitoring of different situation model dimensions. In support of this, it was found that participants responded faster and more accurately to questions that probed spatial information (i.e., whether readers had placed characters in the described locations) for spatially simple texts (e.g., a description of a one-storey building) compared to spatially complex texts (e.g., a description of a two-storey building); however, response times for spatial probes did not differ significantly from causally explicit texts (i.e., in which all causal connections are explicitly stated in the text) to more demanding causally implicit texts (i.e., the reader has to infer causal connections in order to maintain coherence). In contrast, this causal demand manipulation had the same effect on causal probe questions (which assessed whether readers had drawn the

correct causal inference), whereas the spatial demand manipulation did not (Friedman & Miyake, 2000). Furthermore, no interaction was found for reaction time or accuracy between the spatial and causal probes, and scores on a separate visuospatial working memory measure (spatial span) correlated with the spatial probe reaction times, but verbal working memory scores (sentence span) did not (Friedman & Miyake, 2000). Verbal working memory scores did however correlate with causal probe accuracy. It was therefore inferred that the spatial and causal aspects of situation models are maintained and elaborated separately, most likely in different subcomponents of working memory (Friedman & Miyake, 2000). However, other roles for these working memory components in situation modelling and comprehension have also been identified, which will be explored in the following section.

1.3.5.1 Verbal Working Memory

A vast amount of research has established that verbal working memory is related to many aspects of language, including vocabulary learning, sentence processing, and inference, as well as reading comprehension in general (Cain, Oakhill, & Lemmon, 2004b; Carretti, Borella, Cornoldi, & De Beni, 2009; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Seigneuric & Ehrlich, 2005). In addition, the relationship between verbal working memory and reading comprehension in children remains after controlling for additional skills such as word-reading and vocabulary (Cain et al., 2004a; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000; Sesma, Mahone, Levine, Eason, & Cutting, 2009), and individual differences in pre-schoolers' working memory capacity have been found to make an independent prediction to listening comprehension that goes over and above lower level skills (Florit, Roch, Altoè, & Levorato, 2009). The role of working memory is argued to be especially important in language comprehension as it supports situation model construction by enabling a reader to maintain relevant

information so that it can be integrated with incoming information into the meaning-based model, and more connections can be made between concepts in a text (Daneman & Carpenter, 1980; Just & Carpenter, 1992). As such, readers with limited working memory capacity may demonstrate inadequate inference making and comprehension monitoring due to constraints on how much information they can keep active as they read.

Several studies examining discourse comprehension have supported this proposition by finding that verbal working memory (e.g., measures of reading span or digit span) is related to the ability to draw the inferences required to build a coherent situation model representation, including making global inferences about a text (Masson & Miller, 1983) and generating coherence inferences at the level of the textbase (Singer & Ritchot, 1996; Singer, Andrusiak, Reisdorf, & Black, 1992). Additionally, verbal working memory appears to be crucial for drawing causal inferences (Friedman & Miyake, 2000), including predictive inferences (Pérez, Paolieri, Macizo, & Bajo, 2014). This is likely because the reader needs to maintain the content of a causal antecedent until they encounter the causal consequent in order to make this causal connection (Fletcher & Bloom, 1988). In addition, this relationship does not appear to be due to poor comprehenders having reduced memory for the text as a whole as, although they show difficulties with comprehension questions that require an inference, they do not show difficulties with answering questions about literal information found in the text (Cain & Oakhill, 1999).

In relation to coherence monitoring, adults with high working memory capacity show superior ability to integrate local and global information across the text, whereas those with low working memory capacity demonstrate difficulties concurrently maintaining

both global and local coherence (Whitney, Ritchie, & Clark, 1991). Similarly, Orrantia, Múñez and Tarín (2014) found that verbal working memory capacity was a determinant of whether 11-year-old children identified inconsistencies that were separated by text. Additionally, in children aged 9 to 10 years old, who were matched for vocabulary and word recognition skills but differed in comprehension ability, skilled comprehenders performed better than poorer comprehenders on measures of verbal working memory, and more accurately monitored sentence level anomalies (Oakhill, Hartt, & Samols, 2005b). Indeed, good comprehenders not only perform better than poor comprehenders at identifying contradictions in a text, but this difference between the groups appears to be even more pronounced when the contradictory sentences are separated by additional text than when they are adjacent (Oakhill, Hartt, & Samols, 2005b; Yuill, Oakhill, & Parkin, 1989). Thus, it appears that good comprehenders are better at maintaining global coherence than poor comprehenders because of their greater working memory capacity (Oakhill, Hartt, & Samols, 2005b; Yuill et al., 1989). Supporting these findings, Kim (2014) used structural equation modelling to demonstrate that the role of verbal working memory in the listening comprehension of kindergarten aged children is mediated by comprehension monitoring (measured via an inconsistency detection task).

Thus, it appears that working memory makes its contribution to comprehension through its effects on integration of information and coherence monitoring. Consistent with this, working memory tasks that require both storage and additional processing of information have more often been found to correlate with children's reading comprehension than tasks that assess passive storage capacity (Daneman & Merikle, 1996). However, Cain, Oakhill, and Bryant (2004a) found that after controlling for word reading ability and verbal IQ, the relationship between reading comprehension and both inference making and comprehension monitoring were not *entirely* mediated

by verbal working memory and each component provided its own unique variance. Thus, additional resources must play a role in these higher-level skills and text representations. For example, as evidence suggests that situation models contain perceptual and spatial information, visuospatial working memory or visual imagery may also play a role. However, few studies have investigated the role of visuospatial working memory specifically in relation to component skills such as inference generation and comprehension monitoring, and findings regarding its contribution to overall reading comprehension have been mixed.

1.3.5.2 Visuospatial Working Memory

In light of the evidence that situation models contain spatial and perceptual information, it seems intuitive that short-term visual storage may play a role in comprehension monitoring and inference generation. However, the role of visual working memory in reading comprehension has been less thoroughly investigated than that of verbal working memory. Initial correlational evidence supported a relationship between visuospatial working memory (VSWM) tasks and measures of reading comprehension (Bayliss, Jarrold, Baddeley, & Gunn, 2005; Bayliss, Jarrold, Gunn, & Baddeley, 2003; Haenggi et al., 1995). Further, spatial separation effects (i.e., faster response times to objects described as being in the same room as the protagonist, than those in a different room), were found to correlate with scores on the Card Rotation test (Haenggi et al., 1995). Additionally, Denis and Cocude (1997) found that measures of spatial situation model construction (i.e., faster response times to scan across locations described in texts as being spatially distant), correlated with scores on the Minnesota Paper Form Board, another measure of VSWM. Thus, this VSWM and comprehension relationship is possibly mediated by situation model construction.

However, when controlling for lower-level reading skills and verbal working memory, the findings have been mixed. While some studies have found that VSWM measures remain predictors of performance on standardised tests of reading comprehension after controlling for these additional variables (Goff, Pratt, & Ong, 2005; Pham & Hasson, 2014), most studies suggest this relationship does not exist. For example, Swanson and Berninger (1995) found that, in a sample of 91 children, differences in reading comprehension on the Peabody Individual Achievement Test (PIAT) were related to differences in verbal working memory measures, but not visuospatial measures such as remembering a visual sequence of dots within a matrix, or remembering the sequence of directions on an unlabelled map. Further, Seignuric, Ehlrich, Oakhill and Yuill (2000) tested children in the fifth grade on a spatial working memory task that required recall of the placement of coloured lines, and found scores on this task did not significantly correlate with performance on a French standardised reading comprehension measure, or emerge as a predictor of reading comprehension in a multiple regression analysis after controlling for vocabulary and decoding, whereas several tests of verbal working memory did.

In addition, Nation, Adams, Bowyer-Crane, and Snowling (1999) found no differences between groups of children classified as good or poor comprehenders who were matched for lower-level reading abilities, on either a test of spatial visualisation, or spatial working memory span. Similarly, Cataldo and Oakhill (2000) found no differences between fifth graders classified as either good or poor comprehenders on the Pelmanism card game, which requires participants to remember the spatial location of cards in order to find matching pairs. Lastly, Nyhout and O'Neill (2013) did not find a relationship between the Neale Analysis of Reading Ability and a task that assessed mental rotation and visual transference in a sample of 38 7-year-olds.

Thus, it appears that VSWM resources are not required for narrative comprehension, or perhaps, required at such a minimal level that even young children or individuals with low VSWM ability have the capacity to produce and utilise spatial information for this purpose. In line with this, some researchers have argued that complex spatial information is not *routinely* accessed during situation model construction, but only when necessary for comprehension (for example, to interpret maps or track character movements within a described environment; Hakala, 1999; W. Langston, Kramer, & Glenberg, 1998; Zwaan & van Oostendorp, 1993). Hence, as most previous studies investigating whether spatial information is activated during narrative comprehension used texts that emphasised layouts of buildings and spatial relationships, it may explain why they consistently found evidence of this spatial dimension.

Accordingly, in contrast to narrative studies, there is a large amount of evidence that suggests VSWM is required for the comprehension of explicitly spatial and expository texts (although, notably, the majority of these studies have been done with adults). Within the context of undergraduate students' learning from scientific texts, Sanchez and Wiley (2014) found that individuals with low multi-object dynamic spatial ability (MODSA) were poorer than individuals with high MODSA at comprehending geoscience expository texts that likely require dynamic mental imagery in order to be understood (i.e., formation and movement of tectonic plates). Further, low MODSA individuals developed less understanding in text conditions that were either non-illustrated or accompanied by static image conditions, than when the text was accompanied with dynamic images. This was interpreted as being due to low MODSA individuals being less able to generate their own dynamic internal imagery in order to build a spatial mental model of the text to support comprehension, which was required in the non-illustrated and static conditions. Although this study did not measure

narrative comprehension, it does demonstrate that visuospatial resources are needed to connect ideas and that individual differences in dynamic spatial ability may affect the formation of situation models and text comprehension, even when the text does not portray explicitly spatial information.

Furthermore, Kruley, Sciama and Glenberg (1994) also concluded that VSWM plays a role in the construction of situation models of expository texts that are accompanied by pictures (as images may evoke situation model construction), as comprehension of these texts interfered with memorisation of the spatial layout of dots on a grid, but comprehending stories without pictures did not disrupt this additional spatial task (Kruley et al., 1994). Also, using a dual-task paradigm, a vast number of studies have shown that a concurrent visuospatial task can interrupt comprehension of texts that explicitly convey spatial information, such as route descriptions or directions. For example, De Beni, Pazzaglia, Gyselinck, and Meneghetti, (2005) found that recall of a non-spatial text was interrupted by a concurrent verbal task (articulatory suppression), but not a visuospatial task (spatial tapping), whereas recall of a spatial text (a route description of a farm) was interrupted by both verbal and spatial tasks. This suggests that if visuopatial information is conveyed by a text, some of this information is translated into a visual representation. Other tasks using route descriptions of open environments as the spatial text stimuli have shown similar results (Gyselinck, De Beni, Pazzaglia, Meneghetti, & Mondoloni, 2007; Gyselinck, Jamet, & Dubois, 2008; Gyselinck, Meneghetti, De Beni, & Pazzaglia, 2009; Meneghetti, De Beni, Gyselinck, & Pazzaglia, 2013; Meneghetti, Gyselinck, Pazzaglia, & De Beni, 2009).

However, the findings are less clear when also considering individual differences in VSWM in dual-task studies. For example, Gyselinck et al. (2007) found spatial

interference impaired comprehension of spatial texts for individuals with high visuospatial working memory capacity only (as measured by the Corsi block task). This is possibly because individuals with high VSWM capacity rely more heavily on that component when processing texts, thus are more sensitive to the effects of interference; whereas low VSWM capacity individuals may rely more on verbal information for text processing, thus an additional VSWM load would have nothing to interfere with (Gyselinck et al., 2007). A verbal load also interfered with comprehension for both groups, suggesting that verbal working memory is immediately activated when linguistic material is to be processed (Gyselinck et al., 2007). These findings are interesting, as they suggest that the activation of spatial information is not *necessary* for comprehension of spatial texts. However, this finding has not been extended to narrative texts. Further, it was not assessed whether the low VSWM individuals, who were potentially relying on verbal information only, also had lower overall comprehension than high VSWM individuals (i.e., even when there is no additional load task).

Conversely, when taking multiple measures of individual differences in visuospatial working memory and visual imagery into account, Gyselinck et al. (2009) found that neither spatial or verbal interference impaired spatial text recall for individuals with high visuospatial imagery ability (as measured using the Vividness of Visual Imagery Questionnaire (VVIQ) and a mental rotation task), whereas recall for low visuospatial individuals was impaired by both types of concurrent tasks. Thus, it may be that high VSWM individuals have additional processing capacities that enable them to deal with additional spatial information. A possible explanation offered for this was that the high VSWM group had proficiencies in both the generation of vivid visual images (visuospatial capacity) *and* spatial manipulation. As these abilities have been shown to rely on different functions within the VSWM system, it its likely that participants were

able to manage additional spatial information as they were utilising both these components. This highlights the importance of also considering the multiple visual and spatial functions that may be required for text comprehension, however this distinction is not often made in the literature, and most studies investigating reading comprehension have only used a single measure of VSWM, most often one that simply assesses general visuospatial capacity and/or processing requirements.

Further, although some authors argue that VSWM and visual imagery are largely overlapping constructs (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Tong, 2013), in a review discussing the architecture of VSWM, Quinn (2008) identifies factors potentially distinguishing VSWM from visual imagery. For example, that visual imagery supports conscious depictive representations and receives direct input from internal visual sources, whereas VSWM maintains previously presented visual stimuli, but not necessarily in depictive format. To elaborate, information maintained in VSWM is a visual memory, but not a visual mental image per se. However, visual memories held within VSWM can be used to create conscious visual mental images within the visual imagery system. Additionally, theories of visual imagery provide more specific information about imagery subcomponents and how they are differentiated from one another. For example, in arguably the most detailed theory of visual imagery, Kosslyn's computational model of visual imagery (Kosslyn, 1980; 1994; Kosslyn, Brunn, Cave, & Wallach, 1984), it is argued that visual imagery is supported by several distinct types of processes, including image generation, maintenance, scanning and transformation.

Therefore, it is possible that a failure to find a relationship between VSWM and narrative comprehension may be due to less alignment of the type of visuospatial processing that occurs in many working memory tasks and the type of visual imagery

that occurs during narrative reading, which is likely to be more depictive. Previous studies focusing on VSWM may also have only been tapping into a fraction of an individual's visual and/or visuospatial abilities, which were unrelated to narrative comprehension, but more closely aligned with the processes that take place when reading texts that convey explicitly spatial information such as route descriptions.

This proposition is highlighted well by the findings of Nyhout and O'Neill (2013); that, verbal working memory was predictive of recall performance on a descriptive version but not a narrative version of a route description. This is likely because the narrative condition required processes beyond those that are purely verbal (e.g., the construction of a visually rich situation model by adopting the character's point of view). However, a measure of VSWM was not related to overall comprehension in this study, so perhaps it is other imagery processes that play a role in narrative representations. Yet, few studies have focused specifically on *visual imagery* processes, rather than VSWM, in relation to narrative comprehension.

1.3.5.3 Visual Imagery

The majority of research investigating the role of mental imagery in reading comprehension was conducted in the 1970s and 1980s, prior to the introduction of situation models. While some of these studies did find that measures of visual imagery were related to reading comprehension (S. A. Long, Winograd, & Bridge, 1989; Sadoski, 1983; 1985), findings were often mixed and other studies found no relationship (Cramer, 1980). Further, many earlier studies simply focused on visual imagery by investigating improvements in reading comprehension after intervention aimed at improving visualisation. Although several of these studies found improvements in comprehension or story recollection (F. L. Clark, Deshler, Schumaker, Alley, &

Warner, 1984; Gambrell & Jawitz, 1993; Oakhill & Patel, 1991; Pressley, 1976), even

when examining higher-level components of reading such as comprehension monitoring

(i.e., Gambrell & Bales, 1986; Giesen & Peeck, 1984), conflicting findings were also

abundant and these interventions were not always successful (Maher & Sullivan, 1982),

or showed limited improvements over other training strategies such as verbal instruction

(Moore & Kirby, 1988; Rose, Cundick, & Higbee, 1983).

The majority of these studies were, however, conducted prior to the advancement of

cognitive models of reading comprehension. Although they originated from dual coding

theory, these discrepancies in findings may have been due to not having a more detailed

reading comprehension framework, such as situation model theory, from which to build

upon. Consequently, several of these studies also relied on text recall as a measure of

comprehension. In addition, many of these studies utilised subjective ratings of imagery

vividness and thus often failed to differentiate between different types of visual imagery

abilities. For example, it is unlikely that static image generation and the vividness of

such images are the only processes involved in reading comprehension. As de Koning

and van der Schoot (2013) note in a review addressing visualisation as a strategy for

reading comprehension, the reading interventions that appear to be most successful are

those aimed at building a dynamic mental representation of a text, rather than creating

static "pictures" in the mind, thus highlighting the importance of investigating the

nature of the entire representation and the range of skills involved in constructing this.

Indeed, more recent and compelling evidence that visual imagery training and

encouraging internal generation of perceptual simulations can lead to better text recall

and comprehension has since been established (Center, Freeman, Robertson, &

Outhred, 1999; Glenberg et al., 2004), including findings that imagery training may aid

poor comprehenders' resolution of pronouns, thus may play a role in integrative

processes such as inferencing (Francey & Cain, 2015). In addition, it has since been

proposed that guiding story comprehension through the use of perceptual and motor

representations may be a language mechanism that has evolved in humans to allow

individuals to communicate experiences more effectively and vividly (i.e., one must

simulate another's behaviour in order to understand it; Fischer & Zwaan, 2008; Klin &

Drumm, 2010). Yet, beyond this, few studies have empirically investigated this

relationship by measuring individual differences in visual imagery ability using

objective measures, and the link between imagery-rich situation models and

comprehension has remained largely theoretical. Further, it is now known that

comprehension is a multi-dimensional process, yet the question of what specific parts of

this process are affected by visual imagery remains largely unanswered.

1.4 Multicomponent Views and "Levels" of Comprehension

Capitalising on situation model research and cognitive psychology, in recent years

reading research has started to adopt a multi-component view of reading, which

differentiates reading into two subsets: the "lower level" skills, which include basic

processing abilities such as reading fluency, phonological processing, word recognition

and decoding, and knowledge of grammar, vocabulary and syntax; and the "higher

level" skills, which include cognitive processes that contribute to building a situation

model of the meaning of a text. For example, the ability to integrate background

knowledge with information provided in the text to draw inferences, monitor one's own

comprehension and maintain coherence, and having sufficient knowledge of, and ability

to use, appropriate text structure are all examples of higher-level reading skills (Cain et

al., 2004a; Hannon & Daneman, 2001; Kendeou, van den Broek, Helder, & Karlsson,

2014).

Although cognitive theories of reading comprehension have differed in terms of the specific categorisation of these components, most emphasise this dissociation between word-reading ability and comprehension. For example, the simple view of reading (Gough & Tunmer, 1986) describes reading as being the product of both decoding and comprehension, therefore reading disability can result in three ways: (i) from a failure to decode, (ii) a failure to comprehend, or (iii) both. Several studies have supported the simple view of reading by demonstrating that skills that support comprehension are dissociated from those that support reading ability in both children (Kendeou, Savage, & van den Broek, 2009a; Kendeou, van den Broek, White, & Lynch, 2009b; Oakhill et al., 2003) and adults (Landi, 2010), and each level has been found to account for separate variance in overall reading comprehension (Kendeou, van den Broek, White, & Lynch, 2009b; Landi, 2010).

However, it should be noted that these levels are not completely discrete, and the contributions of each of these components to comprehension can depend on additional factors such as individual differences in topic knowledge, and text features such as domain or difficulty (McNamara & Magliano, 2009). Research also shows that both levels begin to develop at preschool age (4 years old) prior to the start of formal reading education, but become more independent as predictors of reading comprehension with age. For example, the two skill sets have been found to become less interrelated from age 4 to 6, and again between ages 6 and 8 (Kendeou, van den Broek, White, & Lynch, 2009b). It has been suggested that automisation and efficiency of lower level reading processes frees cognitive resources to allow for the development of these more taxing higher-level processes of reading comprehension (LaBerge & Samuels, 1974; Perfetti, 1985). Consistent with this, lower-level processes become automated earlier than higher-level processes, these are typically developed by early to mid-childhood

(Kendeou, Papadopoulos, & Spanoudis, 2012), whereas higher-level cognitive processes undergo vast developmental changes from early childhood (i.e., age 8) through to adolescence, and thus take more time to mature and become automated (Luna, Garver, Urban, Lazar, & Sweeney, 2004). Consistent with this, evidence shows that reading comprehension is compromised when decoding and skills that support this process, including phonological awareness and accurate word identification, are poor (Gottardo, Stanovich, & Siegel, 1996; Nation & Snowling, 1998; Perfetti, 1985; Perfetti & Hart, 2001; Shankweiler et al., 1999; Storch & Whitehurst, 2002), especially in the earlier years of reading (Rupley, Willson, & Nichols, 1998; Willson & Rupley, 1997).

Accordingly, while it is acknowledged that lower level skills such as decoding are a necessary component in the comprehension of written text (i.e., comprehension would be near impossible if one cannot identify, or retrieve the meaning of the words on a page) it is now clear that decoding is not the *sole* requirement of successful comprehension. Such a sentiment has been supported by studies that have identified groups of children who demonstrate comprehension difficulties despite competence in word reading and lower level reading skills (Oakhill, 1994; Perfetti et al., 2005; Yuill & Oakhill, 1991). These children have been described as being *unexpectedly* poor at comprehending (Cain, 2009; Cain & Oakhill, 2007) because their reading comprehension is below the level predicted by their word reading ability and their chronological age. Thus, it has been suggested that the comprehension difficulties of these children arise from impairments in higher-level cognitive skills (Cain, 2009; Cain et al., 2001; Kendeou et al., 2014; Nation, 2005).

However, although the complexity of reading comprehension has now been captured in cognitive models of reading comprehension that describe the interaction of multiple

processes and levels of meaning that occur during narrative comprehension, an earlier focus on a single-component approach to reading comprehension has limited our understanding of the unique contribution that each of these different skills and processes makes to one's ability to comprehend written language (Hannon & Daneman, 2001). In particular, many of these measures rely predominantly on lower-level word reading ability (Francis et al., 2006; Keenan, Betjemann, & Olson, 2008; Rowe, Ozuru, & McNamara, 2006; Spooner, Baddeley, & Gathercole, 2004). Additionally, as many standardised comprehension measures rely on offline questioning following reading, it has been argued that most of these measures only provide an indication of the *product* of reading comprehension, rather than the *processes* that take place to provide this outcome (S. E. Carlson, Seipel, & McMaster, 2014a; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007). Consequently, many standardised measures of reading comprehension have been criticised for their limited ability to identify poor comprehenders and their specific skill deficits. As the way in which comprehension is measured may impact on whether it is likely to show a relationship with other variables, these criticisms will be explored in more detail in the following section.

1.4.1 Criticisms of Traditional Standardised Comprehension Measures and Current Directions in Measurement

Although widely used in research and practice, many standardised tests of reading comprehension have been criticised on accounts of poor construct validity. Particularly, it has often been argued that many of these comprehension tests do not accurately capture the skills required to extract meaning from a text, but rather, comprehension scores on these measures largely reflect an individual's lower-level reading ability, or even additional constructs that are extraneous to comprehension.

For example, standardised measures of reading comprehension have been found to be heavily reliant on a reader's decoding ability (Francis et al., 2006; Keenan et al., 2008; Rowe et al., 2006; Spooner et al., 2004). As an example of this, when assessing the item difficulty of the Gates-MacGinitie Reading Test (GMRT) as administered to 7th and 9th graders, Rowe et al. (2006) found that item difficulty correlated with text passage features such as word frequency and sentence length, but not item characteristics such as whether an inference was required to answer the comprehension question. This finding was replicated in a later study, which found that comprehension scores on the GMRT were primarily influenced by vocabulary difficulty, and other text-level features (Ozuru, Rowe, O'Reilly, & McNamara, 2008).

Further, Spooner et al. (2004) argues that combined measurement of accuracy and comprehension on the Neale Analysis of Reading Ability can underestimate the comprehension ability of children with poor decoding skills. Specifically, Spooner et al. (2004) found that poorer decoders, as identified by low accuracy scores on the Neale, attained Neale comprehension scores that were lower than what would be predicted by their age and listening comprehension level. In contrast, skilled decoders achieved comprehension scores higher than would be predicted (Spooner et al., 2004). As children were matched for level of listening comprehension, rather than inferring that comprehension ability is generally dependent on decoding ability, it was concluded that the decoding and comprehension measures of the Neale cannot be separated, and that comprehension scores on the Neale largely reflect level of reading accuracy (Spooner et al., 2004).

Indeed, when comparing different comprehension measures, several studies have found evidence that suggests comprehension measures should not be used interchangeably, as

scores on these measures differ greatly with regards to the amount of variance that is accounted for by lower versus higher-level skills (Bowyer-Crane & Snowling, 2005; Cutting & Scarborough, 2006; Keenan et al., 2008; Nation & Snowling, 1997). Intercorrelations between comprehension measures have also been variable, and mostly low, suggesting these measures do not all tap the same component skills (Keenan et al., 2008).

Thus, in addition to demonstrating that comprehension measures are influenced by both word-level and higher-level reading skills, the studies outlined demonstrate how comprehension measures may identify different groups of children as having problems with reading comprehension, depending on where skill deficits lie. This has also highlighted the question of which reading and cognitive skills standardised tests actually measure. Specifically, some researchers have found these measures do not assess skills that are likely related to comprehension, such as verbal working memory (Cutting & Scarborough, 2006). Further, it has been found that individuals can score above chance even when they do not actually read the passages of an established reading measure: the Gray Oral Reading Test (GORT; Keenan & Betjemann, 2006). This suggests that many of the questions can be answered using prior knowledge alone (i.e., are "passage independent") and students are likely to perform above their actual comprehension ability (Keenan & Betjemann, 2006).

Thus, the limitations of standardised measures inevitably limit the conclusions that can be drawn regarding exactly what it is that makes good and poor comprehenders differ. Yet, few measures have been designed to overcome the limitations of existing comprehension measures, and provide a standardised assessment of higher-level processes, which would be viable for use in educational settings. Exceptions to this

include a few measures built from cognitive theory. For example, The Diagnostic Assessment of Reading Comprehension (DARC; August, Francis, Hsu, & Snow, 2006) was developed in recognition of the problem that the decoding and word recognition requirements of measurement tools can prevent accurate measurement of other cognitive processes necessary for comprehension (i.e., inferencing and accessing background knowledge), and aims to measure comprehension skills independently of decoding ability. Thus, the DARC controls for the level of decoding that is required while measuring comprehension by using simple and highly decodable words in texts that require inferences drawn at the text level and via knowledge integration, as well as text memory and knowledge access (August et al., 2006), to identify where specific skill deficits lie. Similar to the DARC, the Bridging Inferences Test, Picture Version (Bridge-IT, Picture Version; Pike, Barnes, & Barron, 2010) was also developed to assess children's ability to draw inferences (although, only at the textbase level) as well inhibit irrelevant text information during reading.

Additionally, Magliano and colleagues developed a computer-based assessment, which measures several comprehension processes found to lead to a coherent situation model: the Reading Strategy Assessment Tool (RSAT; Magliano, Millis, Levinstein, & Boonthum, 2011). The RSAT presents readers with texts one sentence at a time, and requires open-ended answers to indirect questions aimed at gauging readers' thoughts on how well they understood the text, or direct questions requiring elaboration which assess comprehension level, for example questions regarding why events occurred in the text (Magliano et al., 2011). Answers are then analysed for evidence of different types of comprehension processes (e.g., paraphrases, inferences, and elaborations). Extending on this, the pen-and-paper format Multiple-choice Online Cloze Comprehension Assessment (MOCCA) was recently developed by Carlson et al.

(2014a) in an attempt to capitalise on the strengths of measures such as the RSAT (e.g., identification of specific comprehension processes that occur *during* reading), but also overcome some of their limitations, such as requiring computer-administration and therefore being unfamiliar to readers and inefficient for educators to administer and score. It is considered a cloze task as readers are required to choose among four alternatives in order to complete a missing sentence (rather than a single word as in traditional cloze tasks). The four possible choices each reflect a specific reading comprehension process, including causal inferences, paraphrases, local bridging inferences, and lateral connections. Thus, the measure gives specific information about which process a reader is most often relying on during comprehension. The best response in this test is considered the one that requires the reader to make a causal inference, as causal inferences are considered to reflect a coherent situation model representation of the text (S. E. Carlson, Seipel, & McMaster, 2014a).

Although the development of these measures is an important step forward in reading comprehension measurement, as they have been guided by cognitive theory that accounts for the multi-dimensional and complex nature of reading comprehension, they have so far not been used extensively in research and each has its limitations (S. E. Carlson, Seipel, & McMaster, 2014a). Further information about how these tools compare to traditional measures would be valuable for developing these existing measures and constructing other informed measures that identify where specific skill deficits lie, in order to provided targeted interventions and give more explanatory power to research.

1.5 Rationale and Aims of the Current Thesis

From the literature reviewed thus far, situation models can be viewed as a perceptual simulation of the events described in a narrative situation, within which a reader situates him- or herself and vicariously experiences via the view of the protagonist. Further, drawing from embodied cognition, this perceptual simulation has been theorised to be central to reading comprehension in both adults and children. However, empirical evidence of this is lacking, with few studies explicitly examining how visual imagery via situation model construction relates to individual differences in overall reading comprehension ability.

In addition, in contrast to investigations of the role of visuospatial working memory, studies exploring the effects of individual differences in visual imagery processes are few. For example, although visual imagery has been found to be an important component of situation model construction, virtually no research exists that has investigated whether individual differences in visual imagery are related to situation model processes such as generating knowledge-based inferences, or coherence monitoring. The role of visual imagery in situation model constructions of texts that are not inherently spatial in nature also requires further clarification. Furthermore, it may be that some subtypes of visual imagery play a more important role in situation model construction than others, yet this proposition has not been explored.

Furthermore, although the recent development of measures based on cognitive theories of reading comprehension represent an important step forward, as they account for the multi-dimensional and complex nature of reading comprehension, more information is needed regarding how these measures compare to traditional standardised measures of reading comprehension. Information about the importance of other higher-level processes would also be valuable for strengthening the connection between theory and

practice, in order to aid the development of existing measures and construction of other informed measures that identify where specific skill deficits lie, and to provide targeted interventions and give more explanatory power to research.

Thus, the aim of the current thesis is to investigate the role of several different types of visuospatial working memory and visual imagery processes in children's reading comprehension, both via other skills involved in situation model construction, and in relation to other higher-level cognitive skills potentially involved in reading comprehension. The current thesis also seeks to provide additional evidence of whether currently utilised measures of reading comprehension assess all of the skills necessary for comprehension of written texts, and how these measures relate to newer measures that are based on cognitive theory and aim to determine where specific reading comprehension difficulties may exist within an individual.

Consequently, three studies were designed to meet these aims. The first study aims to provide more information about the components of the visual imagery system, by exploring the psychometric properties of several imagery measures when used with children, and examining whether imagery is best conceptualised as a single skill, or as several subskills. The second study then compares the influence of these potential subtypes of visual imagery, and additional higher-level cognitive functions such as verbal working memory on individual differences in reading comprehension, both when measured by a traditional standardised measure and by a newer measure that focuses on higher-level comprehension skills. Finally, the third study tests whether good and poor comprehenders differ in their generation of predictive inferences when reading narrative texts, and whether the use of visuospatial imagery is necessary for this inferencing process, thus accounting for group differences.

Chapter 2. Study 1

2.1 Visual Imagery as a Multi-Dimensional Construct: A Study on the Utility of Various Imagery Measures Used With Children.

Visual imagery has been examined in relation to several cognitive processes over the past four decades. However, our understanding of how this construct should be defined and measured has varied both in theory and practice. One of the earliest and most prominent imagery debates concerns the nature of the representation that underlies the experience of visual imagery (Kosslyn, 1994; 2005; Kosslyn, Ganis, & Thompson, 2003; Pylyshyn, 1973; 2003). Kosslyn, the major proponent of the pictorial representation view, presents a core premise that imagery is a *depictive representation.* For example, the internal representation, or "image" is picture-like and resembles visual perception, thus an image also retains the spatial properties of its external referent and can be generated, inspected and manipulated by the same processes used in visual perception (Kosslyn, 1975; 1980; 1994). Conversely, Pylyshyn has long argued that imagery consists of *propositional representations*, that is, the internal representation is inherently non-perceptual, but rather "sentence-like" or "descriptive", and bears no resemblance to its external reference (Pylyshyn, 1973; 2002).

While resolution of this debate is not of central focus to the current study, and, indeed, the existence of one type of internal representation does not preclude the other, consideration of the debate is important with regards to the current study's main purpose: to determine the utility of several measures of imagery for subsequent investigations of the role of visual imagery in children's reading comprehension. In light of this, research from an embodied cognition framework has provided convincing

evidence that textual input activates *perceptual information* about referents during reading and comprehension of written texts (Bergen et al., 2007; Engelen et al., 2011; Klin & Drumm, 2010; Stanfield & Zwaan, 2001; Zwaan et al., 2002; 2004; Zwaan & Pecher, 2012), including their shape (Engelen et al., 2011; Zwaan et al., 2002; Zwaan & Pecher, 2012), orientation (Engelen et al., 2011; Stanfield & Zwaan, 2001) and motion (Zwaan et al., 2004). Thus, imagery here is also conceptualised as being a pictorial representation of the external or described stimuli.

Historically, individual differences in imagery ability have commonly been assessed using subjective measures. Such measures ask participants to provide introspective reports about the vividness of their imagery, by rating and describing this internal experience. The first of these was Galton's Breakfast Table Questionnaire (Galton, 1883), which was further developed to construct the first standardised measure of visual imagery: the Questionnaire Upon Mental Imagery (QMI; Betts, 1909). Later, one of the scales of the QMI was revised and expanded upon to create the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and subsequently, an additional revised version: the Vividness of Visual Imagery Questionnaire-2 (VVIQ-2; Marks, 1995). Although further questionnaires have since been devised to tap into additional visual imagery constructs (i.e., the Test of Visual Imagery Control [TVIC]; Gordon, 1949; and the Verbalizer-Visualizer Questionnaire [VVQ]; Richardson, 1977) vividness appears to be the most frequently investigated aspect of imagery. Hence the VVIQ has remained one of the most relied upon measures of visual imagery throughout recent years (McAvinue & Robertson, 2006).

However, concerns exist regarding the use of subjective measures such as the VVIQ as an overall indication of visual imagery. One major reason for this is due to the lack of

correlation between these measures and more objective measures of visual imagery and spatial ability (Dean & Morris, 2003; Durndell & Wetherick, 1976; Ernest, 1977; Lequerica, Rapport, Axelrod, Telmet, & Whitman, 2002; Poltrock & Brown, 1984; Richardson, 1977). These studies have highlighted the possibilities that participants are unable to accurately introspect on visual imagery processes, or that the phenomenological experience of imagery is unrelated to the cognitive processes involved in tasks of spatial ability (Dean & Morris, 2003). However, it has also been proposed that a focus on a single construct such as vividness as the ultimate indicator of imagery has meant that these subjective measures do not reflect the multiple imagery processes that underlie this ability (Dean & Morris, 2003; Lequerica et al., 2002; McAvinue & Robertson, 2006). Indeed, although only providing unitary scores, both the VVIQ and the Test of Visual Imagery Control (TVIC) have been found to have multiple factors (Dean & Morris, 1991; K. D. White & Ashton, 1977). Thus, it appears that imagery ability goes beyond being a singular construct, yet a consistent approach to the study of visual imagery and its subcomponents has not been adopted in the literature; although, frameworks for this do exist.

Specifically, two predominant models have been used to study the internal experience and representation of visual information. Firstly, with the introduction of the multi-component working memory model (Baddeley & Hitch, 1974) visual imagery came to be commonly conceptualised as being supported by the visuospatial sketchpad (Baddeley, 1986; Baddeley & Andrade, 2000; Baddeley & Hitch, 1974), leading to discussion of imagery processes beyond vividness. Baddeley and Hitch (1974) first described the visuospatial sketchpad component as being responsible for the generation and integration of visual, spatial, and kinesthetic information, which may be temporarily stored and manipulated, implying the multi-process nature of VSWM. Converging

research evidence has since supported the notion of distinguishable subcomponents within VSWM, however, there appears to be no general consensus as to the number and nature of these components (see Mammarella, Pazzaglia, & Cornoldi, 2006, for a review).

At the broadest level, VSWM function can be separated into two distinct types of processing: *visual* processing (i.e., internally generating and maintaining image qualities such as shape, colour and size) and *spatial* processing (i.e., generating and maintaining image qualities such as the orientation, location and sequence of objects; Mammarella et al., 2006). Accordingly, behavioural studies using the dual-task paradigm show that a spatial task (such as tapping out a spatial pattern on a grid matrix) disrupts maintenance of a concurrent spatial load, but not a concurrent visual load (such as viewing irrelevant black and white drawings), and vice versa (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Hecker & Mapperson, 1997; Klauer & Zhao, 2004; D. G. Pearson, Logie, & Gilhooly, 1999). Corresponding neurological evidence has also revealed that two separate neural pathways are activated during different types of VSWM tasks: the parvocellular pathway during tasks that involve retaining and recognising the identity of visual objects, and the magnocellular pathway during tasks that involve retaining a sequence of movement or spatial information (Hecker & Mapperson, 1997). Subsequently, to provide a clearer distinction of these subcomponents in relation to Baddeley's model, Logie (1995) outlined a visual "cache" for the temporary storage of visual information, and an inner "scribe" for the rehearsal and thus maintenance of motor-spatial sequences within the VSWM system.

However, as theories of working memory developed, others argued that it is more appropriate to define visual imagery beyond the type of visual and spatial processing

that occurs within visual working memory (Hiscock, 1978; Kosslyn et al., 1984; Poltrock & Brown, 1984). Thus, a second model, which is arguably one of the most detailed theories of the imagery representation system, was developed. This second model is that presented by Stephen Kosslyn (Kosslyn, 1980; 1983; 1994; Kosslyn et al., 1984), and serves to further define the concept of visual imagery and categorise different processing subsystems within the visual imagery system.

The key structure in Kosslyn's model is the visual buffer, which is the medium for holding short-term visual information (i.e., the depictive representation; Kosslyn, 1980; Kosslyn et al., 1984) and is indicated by neural activity in the same areas of the visual cortex that supports visual perception (Kosslyn, 1994; 2005). According to this model, visual memories can also be stored in a long-term memory structure, which is known as the pattern activation system (PAS) or "associative memory", along with their visual, spatial, semantic, and other properties (Kosslyn, 1994). Yet, although images may be stored in this long-term memory system, they must be generated in the visual buffer in order to be accessible to cognitive awareness, thus the visual buffer can also receive input from higher-level cognition and long-term associative memory via efferent connections. The visual buffer has, therefore, been described as the gateway through which different parts of the cognitive system receive visual input and, thus, the primary mechanism for generating visual mental images (whether this be as visual memories of recently perceived scenes, or as visual images generated from prior knowledge or verbal descriptions; Kosslyn, Thompson, Sukel, & Alpert, 2005).

Under Kosslyn's model, once images are generated within the visual buffer, they are amenable to three types of processing. Thus, this model differentiates four major components of visual imagery ability: (i) image generation (formation of a visual image

in the visual buffer), (ii) image maintenance (retaining the visual image over a period of time), (iii) image inspection (interpreting object or spatial characteristics of the image), and (iv) image transformation (manipulating or reorganising the image in some way; Kosslyn, 1980; 1983; Kosslyn et al., 1984). Behavioural studies conducted with both adults and children support such a division of functions, as measures designed to capture the separate components set out in Kosslyn's model do not highly correlate with one another (Kosslyn et al., 1984; Kosslyn, Margolis, Barrett, Goldknopf, & Daly, 1990; Poltrock & Brown, 1984). In addition, neurological evidence of these divisions has also been provided by Kosslyn et al. (2004), as measurements of normalised regional cerebral blood flow (rCBF) during performance of these different imagery tasks shows that, while there is some overlap in the brain areas that predict performance in each component measure, in all cases, variations in rCBF of at least one brain area predict performance in only one of the four tasks. Thus, while some of the processes drawn upon by these imagery tasks are shared, most can be considered distinct, as each appears to draw upon an independent area of the brain (Kosslyn et al., 2004).

Therefore, to overcome the limitations of unidimensional imagery questionnaires, Dean and Morris (1991, 2003) designed a self-report measure based on the properties of the imagery system as proposed by Kosslyn, by including items which asked participants to rate their performance at generating, maintaining and transforming mental images of 2-dimensional (2D) and 3-dimensional (3D) shapes, similar to those used in objective tests of imagery. Other items were also included to assess the pictorial quality of participants' mental image such as ease of evocation, clarity, detail and vividness. A factor analysis of these ratings revealed four separate factors, all of which corresponded with the processes or properties of the imagery system identified by Kosslyn (1980, 1984): (i) ease of image formation (generation), (ii) pictorial stability (maintenance),

(iii) ease of rotation (transformation) and (iv) relative size of the image (structural properties of the visual buffer). It was also found that ratings on this new imagery measure did not correlate with scores on the VVIQ, but did correlate with two objective spatial ability tasks that require the use of imagery: the Comprehensive Ability Battery Space test (CAB-S) of 2D mental rotation and the Vandenberg and Kuse Test of 3D Mental Rotation. Thus, it appears that individuals can successfully introspect on a number of imagery processes, although these may be functionally distinct to the singular construct of vividness.

From the studies reviewed, it appears that imagery goes beyond the unitary concept of "vividness" and is made up of several distinct processes that may be measured both objectively and subjectively. Despite this, models and measures that serve to delineate some of the key subskills of visual imagery are not often adopted in practice. Although a distinction is often made between *visual* and *spatial* imagery, researchers continue to conceptualise and measure these as a unitary construct. This is concerning, as commonly used measures of these constructs may only tap into a fraction of an individual's imagery ability. It is proposed that an adoption of clearer definitions of specific imagery processes is needed to provide more valuable information regarding the role of visual imagery in cognition, and the neurological bases of these functions.

Furthermore, research regarding the development of children's visual imagery and component processes is lacking. With regards to the development of overall visual imagery ability, earlier studies suggested that the vividness of an individual's imagery increases with age (Galton, 1883). More recently, Coulbeau, Royer, Brouziyne, Dosseville, and Molinaro (2008) proposed that the visual complexity of imagery representations increases as a child gets older (i.e., from age 2 to 6 years old). However,

to establish this, the Draw-A-Man test was used to measure the complexity of participants' mental representations. This test is scored based on the quality and detail of a child's drawing of a human figure, thus is clearly subject to additional confounds, such as developing motor skills and perceptual awareness, hence the theoretical underpinnings of this measure and what it actually assesses have been questioned (Kamphaus & Pleiss, 1992).

With regards to specific visual imagery processes, developmental research has been mostly confined to only one subcomponent: image transformation (rotation). These studies have found that children as young as 4 years old can rotate "child friendly" 2D objects (e.g., monkeys; Estes, 1998; Marmor, 1975; 1977) and that by age 5 and 6 most children can rotate more complex 2D forms, although more slowly and less accurately than adults (Estes, 1998; Kosslyn et al., 1990). However, children have extensive difficulty with 3D mental rotation (Jansen, Schmelter, Quaiser-Pohl, Neuburger, & Heil, 2013), therefore, these stimuli are generally not utilised with younger age groups. Perhaps it is due to this focus on the development of mental rotation that these tasks appear to be the most commonly applied measure of visual and/or spatial imagery in studies with children, when trying to determine how imagery is relevant to other cognitive abilities.

However, some research does exist that has examined the development of multiple subcomponents of imagery skills. For example, Isaac and Marks (1994) found that children aged 7 to 16 years old were capable of both forming visual images, and forming visual images of movement, and that these two skills were not correlated. Although, this study only included subjective measures of imagery vividness (the VVIQ and the Vividness of Movement Imagery Questionnaire [VMIQ]; Isaac, Marks, &

Russell, 1986) which, as outlined, may not be good indicators of the cognitive processes underlying imagery performance. In addition, Kosslyn et al. (1990) found that each of the imagery processes proposed in Kosslyn's original model appear to be independent by 5 years of age. On the other hand, studies on motor imagery suggest that mental simulation of movement (i.e., rotation) becomes more distinct from other imagery processes as children mature (Frick, Daum, Walser, & Mast, 2009; Funk, Brugger, & Wilkening, 2005). Beyond this, however, research regarding the development of different forms of visual imagery appears to be scarce.

In summary, it appears that it is important to try to differentiate imagery constructs in a population of interest prior to correlating these with other abilities. Yet, although attempts have been made to examine the subcomponents of imagery skill in adults, it appears research with children is lacking. With the exception of Kosslyn et al. (1990) there is little evidence that confirms whether a differentiation of imagery skills exists within a younger population, or develops later in life. Thus, the first aim of this study was to determine whether different imagery skills appear as distinct within a population of children aged 8 to 11 years old. In addition, little is known about the psychometric properties of more objective measures of imagery; thus, a second aim of the current study was to investigate the psychometric properties of the measures used to assess these subcomponents.

To achieve this, several measures of imagery were chosen on the basis that they each fit one of the subcomponents outlined by Kosslyn's model of visual imagery (Kosslyn et al., 1984), but that they also had the potential to be developed for efficient administration in a school environment. This was important, as beyond subjective questionnaires, few measures exist that have psychometric data available and meet these

criteria. Furthermore, an additional measure of visual imagery strength was also included, due to the promise of this recently developed measure to provide a more objective assessment of the vividness of an individual's visual imagery. As this measure had not yet been used with children, its inclusion would potentially provide valuable information about its utility in future developmental research. The current research was thus exploratory, in that it aimed to determine the utility and psychometric properties of these measures; however, it also sought to determine whether each of the measures included gauged a separate component of visual imagery skill. In this regard, the current study investigated whether all the included imagery measures were strongly correlated, thus indicating that visual imagery is indeed a single undifferentiated construct. Alternatively, fitting with previous research (i.e., Kosslyn et al., 1990) if these measures were not found to be highly correlated with one another, it would be taken as an indication that imagery is best conceptualised and measured as a group of distinct sub-abilities.

2.2 Method

General Method

Participants

Fifty-nine children (32 female and 27 male) in Grades 4 and 5, from three primary schools in Perth, Western Australia participated in this study. The total sample had an age range of 8.08 to 11.17 years ($M = 10.16$, $SD = .64$). All participants had normal, or corrected-normal vision, were free from cognitive impairment and spoke English as their first language.

General Procedure

Participation in this study took place at the child's school, during school hours, in a quiet area separate from the classroom. Children completed all tasks over two sessions, and in the same order. In the first session, which was completed individually, the order of tests was: the binocular rivalry task, the image maintenance task, and the image scanning task. The second session consisted of two pen and paper tasks (the object imagery task and the mental rotation task), which are designed for administration in either individual or group format. Thus, in order to minimise testing time at each school, these tasks were completed as a small group session consisting of 3-4 children, Each session lasted 45 minutes to 1 hour, including a 10-minute break between tasks, and no longer than three weeks elapsed between testing sessions.

Measures

Task 1: Object Imagery Task (Image Generation)

To measure image generation, an object imagery task (OIT) was developed based on a measure devised by Mehta, Newcombe and De Haan (1992), which requires participants to generate visual imagery in order to determine which object out of three alternatives is the least similar on the basis of its shape.

Materials

Using stimuli selected from Snodgrass and Vanderwart's (1980) standardised set of pictures, 23 lists of the names of three objects were constructed. Twelve lists were of non-living objects (i.e., tools, kitchen utensils, items of furniture, and parts of a building) and 11 lists were of living objects (i.e., fruits, vegetables and animals). The three objects within each list all belonged to the same semantic category (as per the categories set out by Snodgrass and Vanderwart; 1980) and were selected on the basis of overall visual similarity: for each list, two of the three items were judged in pilot

trials completed by adults to be more similar visually (at least 90% agreement). Thus, the most dissimilar item could only be identified on the basis of visual appearance, not by semantic associations with the other objects. In addition, using the ratings provided by Snodgrass and Vanderwart (1980), care was taken to ensure all items within each list had similar ratings of image agreement (i.e., the degree to which an individual agrees that a picture of the object resembles their mental image of this object) and familiarity (i.e., how familiar an individual is with the object based on viewing an image of the object). Examples of the lists used in this task can be found in Appendix A.

Procedure

This measure was presented to participants as an "odd-one-out" task in pen and paper format. Participants were instructed to read each list individually, and imagine each of the objects in that list, paying attention only to the outline shape of the objects and not their size, colour or patterns. If the objects were animals, they were to pay attention only to the shape of the animal's head. Participants then circled the name of the object in each list that they thought was the odd one out based on its shape only. This task was completed without a time limit. Scoring consisted of 1 point for every list in which the correct item had been circled, and zero points for lists in which an incorrect item had been circled, resulting in a total possible score of 23. This task took approximately 10-15 minutes to complete.

Task 2: Image Maintenance

To assess the ability to maintain an imaged pattern, a computer task was adapted from Kosslyn et al. (1984). This task required participants to memorise a pattern contained within a grid, and then once the pattern was removed, decide whether two probes fell in grid cells that were previously filled by the pattern.

Materials

Two conditions of maintenance load were included in this task, each consisting of 20

unique patterns: a "light load" consisting of a 4 x 5 cell grid (117 x 140 pixels/55mm x

66mm), with lines 1 pixel wide, and patterns composed by filling 20% of the grid cells

black, with at least one two-cell block (two filled cells that were horizontally or

vertically adjacent); and a "heavy load", which consisted of a 5x7 cell grid (117 x 140

pixels/56mm x 67mm), with lines 1 pixel wide, and patterns formed by filling 20% of

the grid cells black, with at least one three-cell block (three horizontally or vertically

adjacent filled cells). In both conditions, all filled cells were connected (at either the

corners of cells, or by full cells), to ensure all patterns had similar spread over the grid,

keeping level of complexity controlled for within conditions. No patterns consisted

wholly of a simple straight line or recognisable shape or letter, in order to prevent

participants relying on word related placement (i.e., "the middle column", or "a

square") and to ensure the use of visual imagery. For an example of the stimuli used in
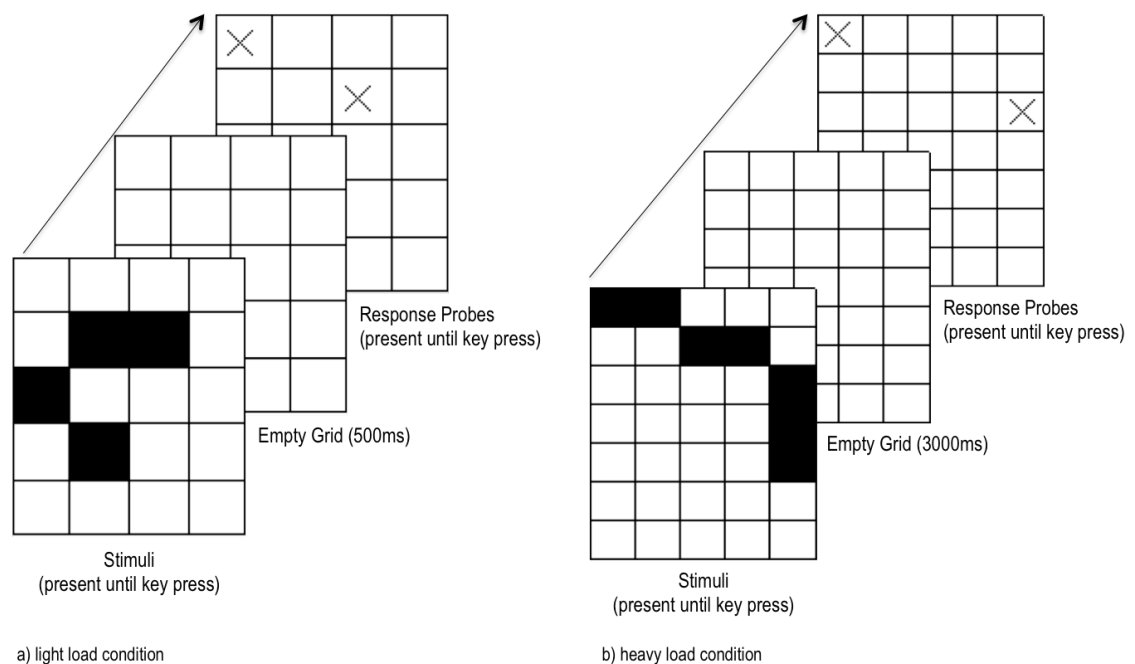
each condition of this task, refer to Figure 2.1.

Figure 2.1. Example of the stimuli used in the image maintenance task.

In both conditions, each pattern was followed by two X shaped probes placed in the centre of two of the grid cells. In each condition, 50% of the probe pairs consisted of probes that were *both* placed in cells previously filled by the pattern. For the other 50% of the pairs, *only one* of the probes fell in a cell previously filled by the pattern and the probe that did not fall in a previously filled cell was placed in a cell adjacent to one that was previously filled (to ensure all probes were placed a similar distance from the pattern, increasing the likelihood that any decisions about the placement of the probes would have to be made by having a clear image of the previous pattern, not simply the general distance of the probes from the pattern). In both conditions, all probes were separated by one full cell (either vertically or horizontally), and were either in the same row or column, or separated by no more than one full cell width or height in the other direction. For an example of probe placement, see also Figure 2.1.

Stimuli were presented on a Toshiba Satellite C660 notebook with the monitor set at 1280 x 720 screen resolution, 32-bit colour and 85 hertz refresher rate using DirectRT version 2010 software (Jarvis, 2006) run on an Intel Core i3 processor with a Windows XP operating system and 2 GB Ram. A DirectIN (Empirisoft Corporation) 305mm x 75mm response box was connected to the laptop via USB cable. The response box had nine buttons on it (corresponding to numbers 1-9 on the computer keypad), however only two were labelled and could be used to provide responses in this study: the far left (1) button was labelled "yes" and the far right button (9) was labelled "no".

Procedure

Prior to starting the task, verbal instructions were given to participants using an example pattern and probes printed on an A4 sheet of paper. To ensure they understood the instructions, participants were required to provide a verbal response to the example

pattern and probes. If the participant gave an incorrect answer, the instructions were clarified, until the participant gave a correct answer (up until a maximum of three incorrect responses, at which the task was discontinued). Participants were instructed to respond as quickly as they could, while still being accurate. Participants were then given the opportunity to ask any questions, and completed five practice trials using the response box and computer with stimuli that did not appear in the test trials.

During the test trials, all 20 light load trials preceded the 20 heavy load trials, but trials within each condition were presented in a random order. Participants were seated at a distance of 420mm from the computer screen (thus the grid stimuli subtended approximately 7.5° x 9.0° of visual angle) on which they viewed each grid pattern and pressed any button on the response box once they had memorised it. Upon the button press, the filled squares were removed by the computer, leaving the grid empty for 500ms in the light load condition, and 3000ms in the heavy load condition. After this delay, the two probes were presented. Participants pressed the button labelled "yes" on the response box if both of the probes fell in squares previously filled by the pattern, or pressed the button labelled "no" if only one of the probes fell in a square previously filled by the pattern. Refer to Figure 2.1 for an example of both a required "no" response (example a) and a required "yes" response (example b). sReaction time (RT) and accuracy of this decision were recorded by DirectRT software (Jarvis, 2006) as an indication of image maintenance.

Following responses to the probes, the screen remained blank for 500ms before the next trial began. No feedback (correct/incorrect) was given following any trials. Rest breaks were offered to participants after completing a block of seven trials via a message on the computer screen. Once participants had completed all light load trials, they were

informed via a message on the computer screen that they had completed "level one", and were about to proceed on to "level two", which consisted of patterns with "more filled squares" and a longer time gap between the pattern and the "X's". The task finished once the participant had completed all 40 trials. This task took approximately 10-15 minutes to complete.

Task 3: Image Scanning

To assess the ability to scan across a visual image, a computer task was adapted from Kosslyn et al. (1984). This task required participants to memorise a pattern contained within a grid, and then once the pattern was removed, decide whether an X shaped probe was placed in a cell that was previously filled, or, if the probe was O shaped, decide whether it was placed in a cell *opposite* to one previously filled.

Materials

This task included two conditions: a control (X) and a scan (O) condition. Stimuli for all conditions consisted of a 174mm x 174mm square grid (which was as large as possible to fill the entire computer screen) with lines 1 pixel wide, consisting of five cells on each side and a hole in the centre (i.e., a square ring). Twenty-eight unique patterns were formed by filling three cells of the grid, with the criteria that each of the three filled cells was on a different side of the grid, and all filled cells were separated by at least three empty cells. Twenty of the patterns were followed by an "O" shaped probe, and 20 were followed by an "X" shaped probe, placed in the centre of a single grid cell. Half of the 'X' probes were placed in a cell that was previously filled by the pattern and the other half were placed in a cell that was not filled but was adjacent to a previously filled cell. In the 'O' condition, half of the probes were placed in a cell opposite to one that was previously filled and the other half were placed in a cell opposite to one that

was not previously filled. No 'O' probes fell on a previously filled cell and no 'X'

probes fell opposite a previously filled cell. A sample of the stimuli used in the task can

be found in Figure 2.2. Stimuli were displayed centrally using the same software and

laptop computer used in the maintenance task, however, with the screen resolution set at

1280 x 729, so the grid filled the entire laptop screen. The DirectIN response box and

button labels used in the maintenance task were used for this task also.
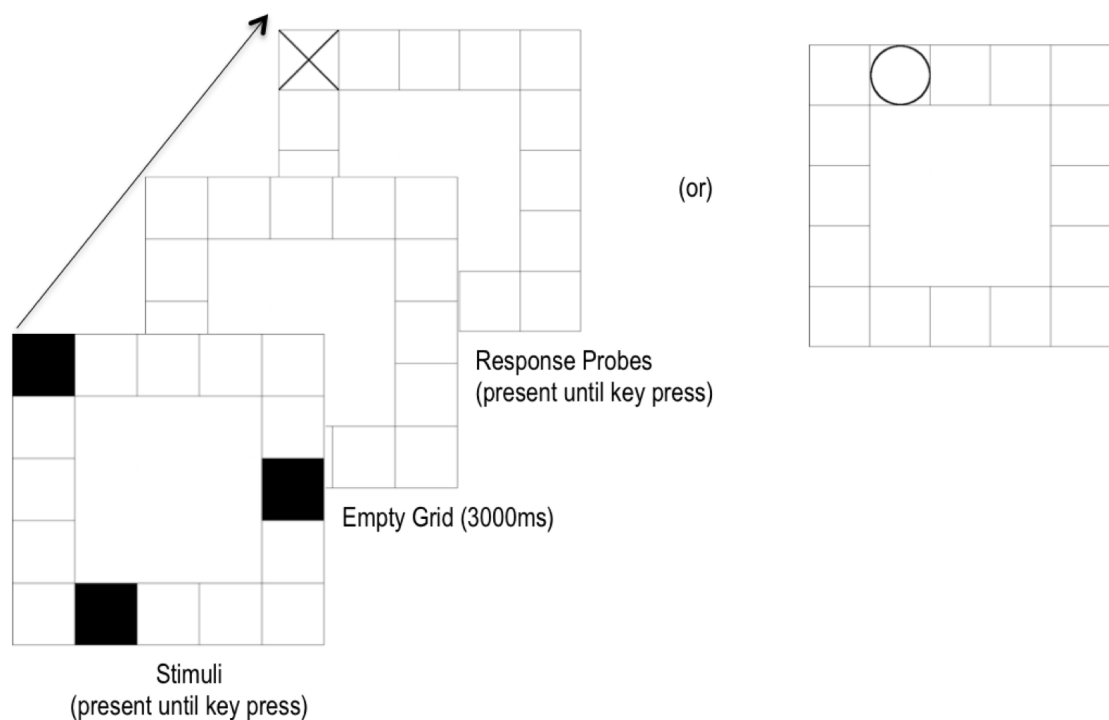
Figure 2.2. Example of the stimuli used in the image scanning task.

Procedure

Prior to starting the task, verbal instructions were given to participants using an example

pattern and probes and practice trials were administered using the same procedure as the

maintenance task. During the task, participants were seated at a distance of 285mm

from the computer screen (thus the grid stimuli subtended approximately 35° of the

visual angle). This manipulation of visual angle ensured participants could not attend to

the entire pattern at once, thus requiring them to shift their visual attention (i.e., scan)

over the entire image. Participants viewed each grid pattern on the computer screen, and pressed any button on the response box once they had memorised it. Upon the button press, the filled squares were removed by the computer for 20ms, following which the probe was presented. If an 'X' appeared, participants indicated whether or not the grid cell that contained the probe had previously been filled, by pressing the corresponding button (Yes/No) on the response box. If an 'O' appeared, participants indicated whether or not the grid cell directly *opposite* the cell containing the probe had previously been filled, by pressing the corresponding button (Yes/No) on the response box (note that the example depicted in Figure 2.2 demonstrates a required "yes" response for each condition). The RT and accuracy of these decisions were recorded by the experimental control software (DirectRT) as an indication of image scanning. Following responses to the probes, the screen remained blank for 500ms, before the next trial began. All trials were presented in a random order. No feedback (correct/incorrect) was given following any trials. Rest breaks were offered to participants after completing a block of seven trials via a message on the computer screen. The task concluded once the participant had completed all 40 trials (20 'X' and 20 'O' trials). This task took approximately 10-15 minutes to complete.

Task 4: Mental Rotation Task (MRT; Image Transformation)

As children of the current study's age group have demonstrated difficulty or inability to complete mental rotation with 3D objects (Jansen et al., 2013), a 2D mental rotation task was used in the current study. This task was the Primary Mental Abilities (PMA) Spatial Relations test (L. L. Thurstone & Thurstone, 1947).

Materials

The Primary Mental Abilities (PMA) Spatial Relations test (L. L. Thurstone & Thurstone, 1947) is a pen-and-paper task. All test items from the original test were used. This included three practice items and 20 test items. Each item consisted of a target stimulus, and a row of six rotated forms of the target item. Within each row, either two or three items were rotated forms of the target item, whereas all others were rotated mirror images of the target item.

Procedure

The standardised written instructions from the Spatial Relations test were presented to participants. Children were also given the opportunity to ask any questions or have the instructions clarified prior to starting the task. A departure from the standardised instructions, however, was that the measure was administered untimed, rather than with the 5-minute time limit normally applied. This was done to reduce test anxiety that may impede performance and to measure imagery ability without the effects of processing speed. Following the task instructions, participants were given the opportunity to ask any questions and completed all three practice items. Participants then continued on to the test items, marking each item in every row that they thought was *not* a mirror image of the target stimulus. Participants were instructed to work quickly but without making mistakes. The standard method of scoring recommended by Thurstone and Thurstone (1947) was applied to this test: participants scored 1 point for every correctly marked item in each row, and 1 point was deducted for every incorrectly marked item in each row, resulting in a total possible score of 54. On average, it took participants 10-15 minutes to complete the test items of this task.

Task 5: Binocular Rivalry (Imagery Vividness)

A binocular rivalry task developed by J. Pearson, Clifford, and Tong (2008) was used to measure imagery vividness.[1] Binocular rivalry involves presenting two different patterns to both eyes, resulting in one pattern reaching perceptual awareness while the other is suppressed. J. Pearson et al.'s (2008) measure was developed based on evidence that the strength of an individual's visual imagery can induce a bias effect with regards to which pattern subsequently reaches perceptual awareness. For example, it has been found that when asked to imagine a target image (i.e., a red horizontal grating), then that image will emerge as dominant during a subsequent binocular rivalry display of the imagined image and a rivalry image (i.e., a green vertical grating; J. Pearson et al., 2008; J. Pearson, Rademaker, & Tong, 2011). Moreover, this effect is strongest for individuals who report strong visual imagery ability, in contrast to when visual imagery is weak, in which cases the bias effect is not apparent (J. Pearson et al., 2011). Thus, individuals can be differentiated based on those whose imagery has a strong bias effect on perceptual rivalry (high imagery strength) from those whose imagery does not have a strong bias effect (low imagery strength). The convergent validity of this measure has also been established with positive correlations between scores of this measure and the VVIQ-2 ($r = .72$; J. Pearson et al., 2008) and discriminant validity has been found with a correlation between this task and a measure of visual working memory ($r = .52$; Keogh & Pearson, 2011). In addition, whereas self-ratings of imagery ability correlate with the perceptual bias found in this task, ratings of effort do not (J. Pearson et al., 2011).

Stimuli

A central bull's-eye fixation point (0.8° diameter) was used to aid binocular convergence. A plaid stimulus was presented to participants centrally (to both eyes), by

---

[1] The author gratefully acknowledges Dr Joel Pearson for providing the program files for the binocular rivalry task.

physically summing a green vertical grating and red horizontal grating into a single image presented in an annulus around the central fixation point (colour values for each grating were set as follows: green: CIE, x = .285, y = .610; red: x = .618, y = .342; see Appendix B for a copy of the stimulus used). Within the plaid stimulus, luminance of each colour component was set to 40% of the luminance of the original rivalry gratings, and was displayed on a black background. Participants wore red-green 3D glasses throughout all trials in order to present the red image to the participant's left eye and the green image to the participant's right eye.

Materials

Stimuli were presented using MATLAB (version R2010a) and the Psychophysics toolbox add on (Brainard, 1997) on the same laptop computer and operating system used in the maintenance and scanning tasks. A fixed viewing distance of 57cm for all experiments was obtained using a chinrest and participants were instructed to maintain fixation on the bull's-eye (a fixation point) at all times throughout the experiment. The DirectIN response box used in the maintenance and scanning tasks was used for this task also, however with the following labels above each key: the "1" button was labelled with a picture of a green circle, the "2" button was labelled with a picture of a half red and half green circle, and the "3" button was labelled with a picture of a red circle, and below each key the buttons corresponding from 1 through to 4 were labelled as such.

Procedure

This task was administered in a darkened room to increase testing efficiency, as previous research has demonstrated that background luminance can interfere with generation and storage of visual imagery (Sherwood & Pearson, 2010). To prevent

perceptual bias due to eye-dominance, the relative strength of the rivalry gratings was

matched across eyes for each participant prior to beginning the rivalry task. To achieve

this, participants underwent an eye dominance test, which involved adjusting the

relative contrast of the two gratings to determine the point at which perceptual

competition was most balanced and therefore most liable to disruption. Prior to

completing the eye dominance test, verbal instructions were given to participants using

examples of the rivalry gratings printed on A4 paper. Participants were instructed to

maintain central fixation throughout, and indicate, upon hearing a tone from the

computer, whether they saw the green or red stimulus. During the procedure,

participants viewed the rivalry display (every 10.75 seconds) accompanied by a tone to

indicate that they were required to respond by pressing one of the assigned keys.

Following their response, participants were shown the pattern that appeared dominant at

full contrast, as adaptation to a high-contrast stimulus has been shown to result in

weaker neural responses to that pattern when it is subsequently presented during rivalry,

thus increasing the probability of its perceptual suppression and a reversal in perceptual

dominance (J. Pearson & Clifford, 2005). Therefore, following the procedure set out by

Pearson et al. (2008), the contrast of the two rivalry gratings was adjusted until the

intervening stimulus caused a perceptual switch on 80% to 90% of rivalry presentations,

indicating balanced perceptual competition. For example, if an intervening stimulus

induced switches in dominance from the red grating to the green grating but not from

green to red, then the contrast of the red grating was increased until switches could be

effectively induced in either direction. The resulting contrast values were then used to

balance the relative strength of the rivalry patterns in the imagery task.

Prior to starting the imagery rivalry task, verbal instructions were given to participants

using examples of the rivalry gratings and probes printed on A4 paper. During the task,

participants were required to maintain central fixation throughout each block of trials. A central cue (either a ''G'' or ''R'') was presented at the beginning of each trial for 1000ms. If the cue was a "G" participants were to form a mental image of a green vertical grating; however, if the cue was an "R", they were to imagine a red horizontal grating (see Appendix B). This cue was randomised on each trial and appeared an equal number of times. Participants then rated the strength of their visual imagery using the keys labelled 1-4 (1 = almost no imagery, 2 = some weak imagery, 3 = moderate imagery, 4 = strong imagery almost like perception), and immediately following, viewed the rivalry display for 750ms and reported on the dominant pattern by pressing one of the three assigned keys (1) a green vertical grating (3) a red horizontal grating, or (2) an approximately equal mixture of the two patterns (due to binocular combination or piecemeal rivalry), under no time limit. To minimise response conflict, participants were required to use their left hand to complete the imagery rating task and their right hand for rivalry responses. Participants completed 40 rivalry trials. Ten catch trials were also included at random, to provide an indication of whether demand characteristics or response biases were affecting rivalry decisions: on these trials participants were shown an image consisting of a physical blend of the two gratings, mimicking the appearance of piecemeal rivalry, rather than an actual rivalry display. It was expected that on these trials participants would not show bias in favour of the imagined stimuli if no response was occurring.

Following completion of the experiment, the percentage of imagery trials in which perception of the binocular rivalry display was biased in favour of the imagined grating pattern was produced by MATLAB as an indication of the strength of the individual's imagery (a higher percentage of times that imagery biased subsequent perception indicated higher imagery strength). As an indication of test validity these percentages

were also binned according to the four levels of rated vividness to determine whether participants were more likely to see the imagined pattern during the subsequent rivalry display on trials in which they reported greater vividness.

2.3 Results

All statistical analyses were conducted at an alpha level of .05, except for where indicated.

Data Screening and Reduction

For each condition of the image maintenance and image scanning tasks, mean accuracy scores (percentage correct) and RT (ms) on correct trials were calculated for each participant. Trials in which a participant responded more than double their mean RT for that condition were considered to likely reflect a lapse in concentration, thus were coded as errors along with incorrect responses. In order to increase the reliability of the data set, any participants who made more than five errors in a single condition were excluded from further analyses. This resulted in data from one participant being excluded from the maintenance task, and three participants being excluded from the scanning task. One participant's RT data in the maintenance task were also removed due to a computer error that resulted in missing data.

All data were screened for outliers using the deletion criteria of +/- 3 standard deviations from the mean score for each variable, including for each condition of the maintenance and scanning tasks, and the difference in RT and accuracy between each conditions. This resulted in the following outliers being detected and subsequently removed prior to further analyses: two outliers in the maintenance task (one participant's mean difference in RT between conditions, and one participant's mean

accuracy score in the heavy load condition); six outliers in the scanning task (one participant's mean RT in the scan (O) condition, one participant's mean accuracy score in the scan condition, two participants' mean accuracy scores in the control (X) condition, one participant's mean difference in RT between conditions, and one participant's mean difference in accuracy between conditions); finally, one outlier was detected in the object imagery task, and one participant's data from this task were removed due to failure to comply with task instructions. No outliers were detected in the mental rotation or binocular rivalry tasks; however, five participants' MRT data were also excluded due to failure to complete this task.

Sampling Distribution

Prior to analysing the data from each test, and calculating correlations between variables, the assumptions of normality, linearity, and homoscedascity for all measures were assessed. Descriptive statistics for each of the variables can be found in Table 2.1. Measures of skewness indicated that all measures were within normal range (-1.0 - 1.0), with the exception of the RT score for the maintenance task (difference in RT between the two conditions; 1.07). However, visual inspection of histograms and Q-Q plots demonstrated no serious departures from normality for any variables, including the maintenance task data, therefore, all data were assumed to be normally distributed.

Table 2.1

Descriptive Statistics for all Visual Imagery Measures

Measure	*n*	Mean (*SD*)	Median	Minimum Score	Maximum Score	Skew
OIT (% correct)	57	80.77 (11.41)	82.61	43.48	100.00	-.90
Image Maintenance (RT; ms)	56	595.25 (412.22)	514.81	-193.16	1618.51	.63
Image Maintenance (% correct)	57	2.54 (2.28)	2.00	-3.00	8.00	-.06
Image Scanning (RT; ms)	55	584.92 (456.47)	521.13	-585.13	2021.23	1.07
Image Scanning (% correct)	55	-0.34 (11.31)	0.00	-37.50	18.75	-.54
MRT (% correct)	54	67.93 (30.79)	80.56	3.70	100.00	-.98

Note. OIT = object imagery task; MRT = mental rotation task. Image maintenance and scanning scores reflect the difference in unit of measurement between the two conditions in the denoted task.

Analysis of Experimental Manipulations and Psychometric Properties of Imagery Measures

Image Maintenance

Paired-samples *t*-tests were conducted separately on the RT and accuracy data to assess the difference between the conditions (light load and heavy load). The results indicated that participants responded significantly faster in the light load condition (*M* = 1322.16ms, *SD* = 371.50) than the heavy load condition (*M* = 1952.47ms, *SD* = 540.75; *t*(55) = 10.44, *p* < .001, *d* = 1.40). Participants also had a greater percentage of correct responses in the light load condition (*M* = 94.91, *SD* = 6.91) than the heavy load condition (*M* = 81.96, *SD* = 9.80; *t*(55) = 8.534, *p* < .001, *d* = 1.13).

Pearson's *r* correlations were also conducted between the accuracy and RT scores for each condition (light load and heavy load) to examine whether there were any speed-accuracy trade-off effects. No positive correlations were found, indicating there was no trade-off effect (i.e., slower responses were not correlated with greater accuracy, or vice versa).

Image Scanning

Paired-samples *t*-tests were conducted separately on the RT and accuracy data to assess the difference between the conditions (control trials and scan trials). The results indicated that participants responded significantly faster in the control condition (M = 1456.20ms, SD = 521.50) than the scan condition (M = 2041.12ms, SD = 606.84; $t(54)$ = 9.50, $p < .001$, d = 1.28), thus indicating scanning had occurred. There was, however, no difference in accuracy between the control and scan conditions (M = 90.33%, SD = 8.72, and M = 89.74%, SD = 8.68, respectively; $t(52)$ = .42, p = .673, d = 0.06).

Pearson's *r* correlations were also conducted between the accuracy and RT scores for each condition (control and scan) to examine whether there were any speed-accuracy trade-off effects. No positive correlations were found, indicating there was no trade-off effect.

Object Imagery Task (OIT)

To assess the reliability of the OIT, Cronbach's alpha was calculated. This measure appeared to have adequate internal consistency (α = .595), however, none of the items appeared to correlate with the total scale to a good degree (highest r = .48), and split-half reliability was low (r = .39).

Mental Rotation Task (MRT)

To assess the reliability of the MRT, Cronbach's alpha was calculated. This measure appeared to have high internal consistency ($\alpha = .941$) and all items correlated with the total scale to a good degree (lowest $r = .41$). Split half reliability was also high ($r = .85$).

Binocular Rivalry (Imagery Strength)

Data from the binocular rivalry task were collected from 23 participants. A further 12 participants also attempted the perceptual balance stage of the task, however a perceptual switch could not be induced on 80-90% of the trials with these participants, even after three attempts. Within the data from those children who completed the entire task, as shown in Figure 2.3, the expected linear relationship between strength of rated imagery and subsequent perceptual effect was not found, $F(1, 22) = .117$, $p = .74$, $\eta^2 = .01$, suggesting that stronger imagery did not induce a perceptual bias of the imagined stimuli and thus poor construct validity within this sample. For this reason, data from this measure were excluded from further analyses.
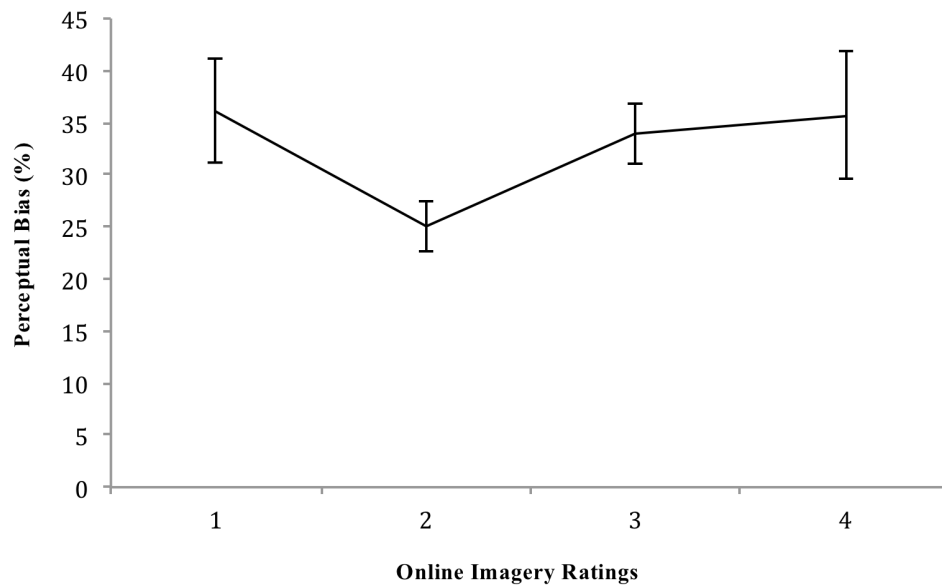
Figure 2.3. Relationship between online imagery ratings and percentage of trials in which the imagined grating was reported as dominant in the binocular rivalry task. Error bars represent +/-1 SE of the mean.

Correlations Between Imagery Measures

To assess whether each imagery measure assessed a different component of visual imagery ability, a series of bivariate Pearson's correlations was conducted between all four imagery measures. The scores from each measure used in this analysis included: the percentage correct on the OIT, the percentage correct on the MRT, the difference between conditions (light load and heavy load) in the maintenance task for both RT and accuracy (i.e., the relative accuracy and efficiency of maintaining images), and the difference between conditions (scan and no scan) in the scanning task for both RT and accuracy (i.e., the relative accuracy and efficiency of scanning across images). Missing data in this analysis were excluded listwise, thus, only participants with complete data sets were included ($n = 47$). Due to multiple comparisons, a Holm-Bonferroni adjustment was applied to these correlations to control for Type I error. As shown in

Table 2.2, after applying the Holm-Bonferroni correction, no significant correlations
were found between any of the imagery measures.

Table 2.2

Correlation Coefficients Between all Visual Imagery Measures (n = 47)

	Measure	1	2	3	4
1	OIT	-			
2	MRT	.32	-		
3	Image Maintenance	.08	-.07	-	
4	Image Scanning	-.07	-.09	.34	-

Note. OIT = object imagery task; MRT= mental rotation task.
* significant at adjusted alpha level following Holm-Bonferroni correction (original
alpha = .05).

Gender Differences

To determine if there were gender differences on any of the imagery measures, a series
of independent sample t-tests were conducted. No significant differences between male
and female participants were revealed on any of the imagery measures (highest $t = 1.17$
for the OIT). Effect sizes across all comparisons were also small, with Cohen's d effect
sizes ranging from 0.01 for accuracy in the scan task (difference between conditions), to
0.31 for the OIT.

2.4 Discussion

The aim of this study was to determine the utility and psychometric properties of
several measures of visual imagery when used with children, and to determine whether
each can be used to assess a distinct component of visual imagery ability. In relation to

the first aim, not all measures were found to be valid and reliable when used with this sample. Specifically, whereas the image maintenance, image scanning and image rotation tasks showed good reliability and validity, reliability of the object imagery task was less than adequate. It also appeared that the binocular rivalry measure could not be successfully administered to this sample.

When exploring these findings in more detail, the majority of the results pertaining to the image maintenance and scanning tasks were consistent with previous research. Firstly, in the maintenance task, participants responded faster and more accurately in the light load condition than the heavy load condition, reflecting the additional processing capacity needed to maintain mental images for longer time periods (Kosslyn et al., 1990). Secondly, in the scanning task, participants responded faster in the control than scanning trials. As with previous research (Dror & Kosslyn, 1994; Dror, Kosslyn, & Waag, 1993; Kosslyn et al., 1990), this is interpreted as an indication that scanning across the visual image occurred during the scan trials comparative to control.

Unlike previous research, however, no difference in accuracy between the two conditions of the scan task were evident, whereas Kosslyn et al. (1990) found that errors increased in trials where scanning was required. One reason for this discrepancy may be that Kosslyn et al's (1990) study included children as young as 5 years old, whereas the current study's sample consisted of 9- to 11-year-olds. It is thus possible that by this age, individuals have developed the skills necessary to complete this task accurately. Accordingly, it appears that this task may have been easily completed by children of this age group (as evident by the high mean accuracy in the scan condition of close to 90% accuracy rate). However, although Kosslyn et al (1990) note that *overall* error rates in imagery tasks decrease with age, the difference in errors between scan and no scan

trials generally remains significant across ages, and is evident in adults. Therefore, it could still be expected that there would be significantly more errors found in scan trials than no scan trials in the present sample.

Thus, high accuracy in the scan task may in fact be a reflection of the required scanning distance in the task used in the current study. Specifically, previous research suggests that accuracy rates decrease as a function of having to scan further distances (Dror et al., 1993; Dror & Kosslyn, 1994). The size of the grid stimuli used in this study was smaller than that used in previous research with children (i.e., Kosslyn et al., 1990), due to the use of a portable laptop computer that could be used for school administration. Efforts to overcome the effects of stimuli size were made by controlling for viewing distance, and consequently visual angle (i.e., the area of the visual field the stimuli occupied), yet it is still possible that more variability in accuracy for scanning trials would have been found if the required scanning distance was greater, as this would have increased task difficulty. Consequently, it appears that RT difference between conditions may be a better indication of individual differences in scanning ability, when shorter scanning distances are used.

The MRT also appeared to be particularly reliable when used with this sample, demonstrating high internal consistency, with item-total correlations indicating that all items discriminate between high and low scoring participants. The range of scores obtained on this measure was also wide (3.70% - 100%) and was normally distributed. Thus, reliability of this measure appears more than adequate, as does the ability to differentiate individuals by ability; however, issues of validity of this measure need to be addressed.

Firstly, due to using the spatial relations test as an indication of mental imagery, consideration needs to be given to whether objective measures of spatial ability indeed rely on visual imagery or form a separate factor from the imagery subcomponents described by Kosslyn. Although the earliest evidence of overlap between spatial ability and imagery, which was simply that individuals with high self-reported imagery performed better on spatial ability tasks than those with low self-reported imagery (Barratt, 1953), has largely been discredited (due to inabilities to replicate such findings, and the criticism that providing subjective ratings immediately following completion of spatial tasks may bias subsequent self-perceptions of imagery ability; McAvinue & Robertson, 2006) there does exist psychometric data that show tests of spatial ability correlate with the objective imagery measures devised by Kosslyn and colleagues (Dean & Morris, 2003; Poltrock & Brown, 1984). Further, both types of measures load on equivalent factors (Dean & Morris, 2003; Poltrock & Brown, 1984), thus supporting the notion that imagery processes such as high quality image maintenance, inspection and transformation underlie performance on tests of spatial ability (Dean & Morris, 2003; Poltrock & Brown, 1984).

However, results surrounding this have not always been consistent. For example, Burton and Fogarty (2003) administered a large battery of tests to 213 individuals, including 18 spatial ability tests, five self-report visual imagery measures, and seven objective measures of visual imagery based on Kosslyn's model, and found that spatial ability formed a separate factor from those produced by the imagery measures (which include quality, speed, and self-report). Reasons why spatial ability tasks do not always overlap with objective imagery tasks are, however, not always clear and may also lie in the conceptualisation of "spatial ability" itself. As highlighted in a review by McAvinue and Robertson (2006), the term "spatial ability" has been used as a catch-all phrase to

describe the variety of tasks that involve the manipulation or interpretation of visual and/or spatial representations. Additionally, it is now known that there is not a single spatial ability but several spatial abilities, yet it is unclear whether *all* of these involve visual imagery. Indeed, although extracted as separate factors, Burton and Fogarty (2003) did still find that, in comparison to other factors, there was substantial overlap between the VZ spatial factor (which includes the spatial relations test used in the current study) and imagery transformation.

Accordingly, when it comes to mental rotation specifically, there is substantial evidence that this ability requires the generation and transformation of imagery. The earliest acquired and the most often replicated evidence of this includes the linear increase in response times as a function of the angle of rotation, which is interpreted to represent the time required to visually rotate the mental representation on its axis (R. N. Shepard & Metzler, 1971), and has been found for the rotation of both 2-dimensional and 3-dimensional stimuli (L. A. Cooper, 1971; R. N. Shepard & Metzler, 1971; S. Shepard & Metzler, 1988). Importantly, this linear relationship between angle and speed of rotation, has been found to be related to performance on tasks of spatial ability (Poltrock & Brown, 1984), and a recent meta-analysis of neuroimaging data identified patterns of neural activation in the parietal cortex during mental rotation tasks that likely reflect simulation of analogue spatial representations (Zacks, 2008). In addition, in an extensive review of the literature, Harris, Hirsh-Pasek and Newcombe (2013) found neurological, cognitive and psychometric evidence that suggested although both 2D and 3D mental rotation tasks are susceptible to non-imagery based strategies, they are usually solved by transformation of visual imagery (J. Harris et al., 2013). Importantly, it appears that mental rotation tasks are less susceptible to non-imagery based strategies than other spatial ability tasks such as paper folding (J. Harris et al., 2013).

Consequently, it has been argued that spatial ability tests of rotation are appropriate as an indication of the transformation component as described by Kosslyn (D. G. Pearson, Deeprose, Wallace-Hadrill, Heyes, & Holmes, 2013).

However, the majority of research outlined has been conducted with adults, and the research regarding the strategies that children use to solve mental rotation tasks is less clear. Although a linear relationship between response time and angle of stimuli rotation has also been found in samples of children (Estes, 1998; Kail, Pellegrino, & Carter, 1980; Marmor, 1975; 1977) some researchers suggest that children may also rely on non-spatial strategies that do not require imagery when completing mental rotation. For example, Quaiser-Pohl, Rohe, and Amberger (2010) identified seven types of solution strategies used by young children when completing 3D mental rotation tasks, including three inappropriate strategies which would unlikely lead to successful performance (e.g., comparing object features rather than performing rotation), two semi-appropriate strategies (e.g., choosing stimuli that were facing the same direction as target stimuli), and two appropriate mental rotation strategies, including "the holistic approach" (mentally rotating the object as a whole) and "the analytic approach" (focusing and comparing parts of the target object to parts of the alternative stimuli; i.e., a "piecemeal" approach). In addition, a completely *verbal*-analytical approach has also been identified (i.e., describing the different direction of the stimulus parts in words, for example "the top part points to the left"; Geiser, Lehmann, & Eid, 2006; Geiser, Lehmann, Corth, & Eid, 2008).

However, although a less imagery-dependent strategy is available, it has been established that, by age 5, children are able to use appropriate spatial strategies such as the holistic approach, and the use of this strategy over inappropriate or non-spatial

strategies increases with age, and is developed by age 8 (the minimum age of children in the current study; Quaiser-Pohl et al., 2010). Based on this, it is plausible that children in the current study would have relied on an appropriate strategy if they had the capacity to do so. Therefore, any use of inappropriate and unsuccessful strategies is likely a reflection of poor ability to perform mental rotation. Indeed, it has been identified that use of the holistic-rotation approach is directly related to children's mental rotation performance of 2D stimuli (Estes, 1998; Tzuriel & Egozi, 2010), and children who use non-rotation strategies such as verbal-analytical tactics have been found to perform more poorly than those who utilise imagery-based rotation strategies (Geiser et al., 2006; 2008). Thus, if a great number of children were using this approach, it could be assumed that a floor effect would be evident for scores on this task, which was not found in the current study.

In addition, however, some may argue that the MRT captured VSWM processes, rather than visual imagery per se, as this task involved the manipulation of externally viewed stimuli, rather than described, or internally generated objects. Yet, the basis of this argument essentially lies within conceptualisations of how distinct visual imagery and VSWM actually are. This has remained a source of contradiction within the literature. One line of investigation suggests that VSWM and mental imagery are separate processing systems, with the evidence for this proposition most often provided through the use of the dynamic visual noise (DVN) technique. It has been demonstrated that while DVN (i.e., viewing a grid containing a display of dots that appear to flicker) interferes with the encoding of words learned via the imagery-based peg-word technique, it does not interfere with the short-term maintenance of these stimuli (Quinn & McConnell, 2006), or with other visual working memory tasks, including the recall of visually encoded stimuli such as Chinese characters and matrix patterns (Andrade,

Kemps, Werniers, May, & Szmalec, 2002). Consequently, it has been argued that DVN interferes selectively with visual imagery, but not VSWM, due to different mechanisms underlying each of these processes (Andrade et al., 2002; see also Borst, Ganis, Thompson, & Kosslyn, 2012, for a review).

On the other hand, contrary findings have also been established using the DVN technique (Borst et al., 2012) and it has been argued that any major distinction between VSWM and visual imagery is unfounded (Albers et al., 2013; Borst et al., 2012; Tong, 2013). For example, neuroimaging research indicates shared internal representations in the visual cortex for both tasks of VSWM (e.g., holding visual material in working memory) and visual mental imagery (e.g., internally generating a visual stimulus), and that these neural mechanisms are similar to those used in visual perception (Albers et al., 2013). Thus, it appears that both these processes rely on common representations that share the same depictive format. Subsequently, it has been argued that the early visual areas may act as a dynamic ''blackboard'' within the visual cortex, which supports information processing during both types of tasks; and thus, the primary visual cortex is not simply a gateway for subsequent processing in higher-order visual areas but rather itself a high-resolution buffer that is recruited for a variety of visual computations (Albers et al., 2013). Therefore, although the overlap between these two constructs may warrant further investigation, it currently does not make sense to completely separate these two processes in theory or practice.

In contrast to the MRT, findings regarding the object imagery task (OIT) indicated that the psychometric properties of this measure were questionable. Although the internal consistency of this measure could be considered adequate for a measure of ability, the low item-total correlations and poor split-half reliability are a cause for concern, and

may indicate poor internal validity. It is possible that the OIT is reliant on skills that are extraneous to imagery generation, for example executive functions such as inhibition. To elaborate, when making comparisons between the objects, inhibition may also have been required in order to control or ignore activation of irrelevant semantic information; for example, in the case of animals, additional semantic information may have included whether the animals were pets, farm animals, live indoors, have fur, and so on. It is possible that some children were unable to inhibit or ignore this information and based similarity decisions on these characteristics rather than visual appearance. This may have especially been the case for items in which more competing semantic information was available (i.e., in the case of animals, compared to tools) or the objects were highly familiar, thus increasing the likelihood that semantic information would be more automatically activated; or alternatively, where the objects were highly unfamiliar, thus increasing reliance on semantic information rather than visual appearance.

Indeed, although familiarity was controlled within lists using the Snodgrass and Vanderwart (1980) ratings, familiarity of an object also depends on an individual's background knowledge and life experience. Further, familiarity was not controlled for *across* items, as this would have been difficult to achieve and still retain a sufficient number of items for adequate reliability. It is therefore possible that some items on this measure may have been more reliant on these executive function capacities than others, which consequently lowered both the reliability and validity of this measure.

Lastly, the binocular rivalry tasks could not be successfully utilised with this sample. Reasons for this are explored here in detail. Firstly, the task instructions may have been too complex for this age group to understand or follow throughout the entire task (i.e., this task encompasses multiple task demands in the form of remembering prompts and

corresponding images, forming imagery, providing imagery ratings, and providing binocular rivalry responses). Similarly, it is possible that children of this age group do not have the attentional capacities required to maintain central fixation and visual attention for extended time periods (i.e., within each trial, and throughout the entire task).

Alternatively, variability across the testing environments may have reduced the ability to obtain perceptual balance and adequately measure visual imagery ability using the binocular rivalry task. Background light signals have been shown to interfere with imagery generation and maintenance (J. Pearson et al., 2008; Sherwood & Pearson, 2010), thus, testing was completed in a darkened room. However, due to testing taking place across different schools, it was difficult to adequately control for room luminance and ensure adequate reduction of light sources, including natural light from windows and artificial lighting from other areas. As such, luminance levels varied considerably across testing sessions and schools. Future research would benefit from using this measure with children in a controlled laboratory environment. This would be an important step in establishing whether this more objective measure of imagery vividness is suitable for use with children, although the practicality of this measure would unlikely extend to research conducted in school environments.

Correlations Between the Measures

The second aim of this study was to determine whether each imagery task captured a distinct component of imagery ability. Correlations between the measures revealed that the variables were not highly related, with only small to moderate effect sizes evident between the measures, and no significant correlations being found. Thus, it appears that performance on each of the measures used in this study reflects a distinct component of

the visual imagery system. This is consistent with previous research assessing these same subcomponents that found no significant correlations between measures designed to assess these imagery constructs (i.e., Kosslyn et al., 1990; 2004).

Limitations and Future Directions

It is acknowledged that this study was not without limitations. Although the sample size of the current study can be viewed as sufficient, it may be argued that only the minimum required number of participants was met for the analyses conducted (Nunnally, 1978; Tabachnick & Fidell, 2007). It is possible that a larger sample size may have altered the findings of the current study. Future research could increase the generalisability of the results of the current study by replicating these findings with a larger sample size. Furthermore, conducting research at educational institutions inevitably brings about changes in the testing environment between schools. As noted, this may have had particular consequences for the administration of the binocular rivalry measure. Future research may advance this area by administering the binocular rivalry measure to children in a more controlled environment. Additionally, although the results of the current study suggest that it is unlikely that participants were relying on a non-imagery strategy in order to complete the MRT, this cannot be completely ruled out without additional data; for example, the inclusion of chronometric data to determine whether there was a linear relationship between the degree of rotation and time taken to solve each item, or neurological data to determine whether there was corresponding activation of the visual cortex, or parietal areas implicated in the simulation of movement.

Further, relationships of these imagery measures with other measures of higher-level cognitive processes such as executive functioning are in need of further investigation.

This may help account for the potential variability these additional constructs provide to measures of visual imagery. Future research may also serve to further reconcile models of visual imagery with those of VSWM. Currently, there is a trend in measuring VSWM as a separate entity from visual imagery, with confusion surrounding the degree to which these constructs rely on common mechanisms. Although, as Tong (2013) highlights, there are obvious commonalities between the construct definitions of both visual working memory and visual mental imagery (i.e., in both cases, common definitions capture the ability of the individual to represent and manipulate visual information within in the mind) and, although research on both of these constructs emerged at similar time-points in the 1970s, the literature in these two areas has largely diverged into separate areas, with little cross-reference between the two. As such, it has been argued that there remains a lack of systematic investigation of the relationship between VWM and visual imagery (Borst et al., 2012), which is in need of further investigation.

In conclusion, this study demonstrated that visual imagery is not a unitary construct, but a set of skills that appear to be differentiated by 8 to 11 years of age. It also provides previously undetermined information about the utility and psychometric properties of several imagery measures, when used with younger populations. Specifically, the utility of three measures of imagery was established when used with this younger population; these included measures of image maintenance, image scanning, and image rotation. In contrast, the measures used to assess image generation and image vividness did not appear to capture these constructs well. This information may aid the choice of visual imagery measures in future studies with children, in order to make clear interpretations regarding the relationships between visual imagery and other developmental constructs.

In particular, it appears that clear classification of imagery components is crucial for valid and reliable research in this area.

Prelude to Study 2

In line with previous studies (i.e., Kosslyn et al., 1984; 1990; 2004), the results of Study 1 indicated that imagery is not a single construct, but can be defined and measured as several distinct processes, namely: image generation, maintenance, scanning and transformation.

As such, it is possible that each subtype of imagery is differentially related to reading comprehension. For example, if situation models represent meaning through visual simulation, simple imagery generation, while necessary for situation model construction, may not be sufficient for updating these representations as new information is encountered. Here, the ability to engage in more dynamic forms of imagery such as transformation or scanning may enhance representations of dynamic spatial relations and actions, and be beneficial for updating visual representations of story events.

Consequently, a second study was conducted, using multiple measures of visual imagery, to determine whether imagery is indeed related to reading comprehension, and whether subtypes of visual imagery predict reading comprehension to varying degrees. However, of the measures used in Study 1, not all appeared to be valid or reliable when used with children. Firstly, the object imagery task demonstrated poor reliability, possibly due to the influence of additional executive functions that are unrelated to the use of visual imagery. Secondly, the binocular rivalry measure could not be administered in a controlled manner, and children appeared to have difficulty maintaining the visual attention required to complete this task. Thus, only the mental rotation task, and scanning and maintenance tasks were included in Study 2.

Chapter 3. Study 2

3.1 Visual Imagery in Children's Reading Comprehension: A Multicomponent Approach

Improving the reading outcomes of children has been a major focus in developmental and educational research. However, an outcomes-based approach that largely focused on word-reading ability has led to a limited understanding of the *processes* that lead to successful reading comprehension. Theories from cognitive psychology have advanced research on reading comprehension by investigating a number of abilities that contribute to this overall construct. As such, it has been identified that comprehension is a multi-faceted and complex operation, requiring the execution and integration of several cognitive processes across word, sentence, and text levels (Perfetti et al., 2005). Yet, much remains in determining the extent to which each process contributes to comprehension and whether additional skills also play a role.

Broadly, reading comprehension has been defined as the process of extracting meaning from a text (Snow, 2002). As explored in Chapter 1, theoretical attempts to explain the cognitive processes that underlie successful reading comprehension have focused on how readers obtain meaning of a text via the construction of a coherent mental representation often referred to as a "situation model" (van Dijk & Kintsch, 1983; see Chapter 1, p. 18). These theories consequently explain reading comprehension as the result of differing levels of processing across word, sentence and text levels, with a focus on two predominant areas. The first of these is the "lower-level" processes, which include word-reading and basic processing abilities that allow a reader to translate written code into meaningful language and construct a representation of the text itself

(i.e., a "textbase representation"). For example, decoding, phonological processing, knowledge of grammar and syntax, and vocabulary knowledge are generally considered lower-level reading skills. In contrast, the "higher" or "message-level" processes include the cognitive processes that contribute to building a coherent representation of the meaning conveyed in a text (i.e., the situation model). These include the ability to: activate background knowledge and integrate it with information provided in the text to draw inferences; monitor and connect different ideas across the text; and identify and update relevant information in order to maintain global coherence (Cain et al., 2004a; Hannon & Daneman, 2001; Silva & Cain, 2015).

At the word-reading level, a vast amount of research suggests that comprehension depends on fluent reading skills. Fluency reflects the composition of both accurate and automatic word reading skills that result in fast, efficient and co-ordinated reading (Kuhn, Schwanenflugel, Meisinger, Levy, & Rasinski, 2010). According to the automaticity theory (LaBerge & Samuels, 1974) and the verbal efficiency theory (Perfetti, 1985), when decoding is effortful, the majority of a reader's resources remain dedicated to word-level processing. In contrast, when reading fluency is acquired and decoding becomes automatic, more cognitive resources are freed for comprehension (LaBerge & Samuels, 1974; Perfetti, 1985). Consistent with this, evidence shows that reading comprehension is compromised when decoding (and skills that support this process, including phonological awareness and accurate word identification) are poor (Gottardo et al., 1996; Nation & Snowling, 1998; Perfetti, 1985; Perfetti & Hart, 2001; Shankweiler et al., 1999; Storch & Whitehurst, 2002), especially in the earlier years of reading (Rupley et al., 1998; Willson & Rupley, 1997).

However, as Nation (2005) notes, adequate text reading accuracy alone does not necessarily indicate efficient word-level processing, as slow or inefficient reading may also inhibit comprehension even when decoding is accurate. Consequently, reading speed, or "rate" is often used as an indication of the automaticity of reading. It is assumed that faster and more automatic reading speed is an indication that children can devote fewer working memory resources to decoding and thus allocate their attention to the task of comprehension (Perfetti, 1985). Similarly, the use of reading speed as an index of processing speed may indicate the rate at which children are able to comprehend (Goff et al., 2005).

Accordingly, several studies have found that reading speed is a predictor of reading comprehension (Klauda & Guthrie, 2008; Riedel, 2007; Schwanenflugel et al., 2006; see also L.S. Fuchs, Fuchs, Hosp, & Jensen, 2001, for a review) and measures of fluency, measured as correct words per minute, correlate with measures of reading comprehension (Riedel, 2007; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008). In particular, text-reading fluency has been found to predict reading comprehension over and above list reading fluency (i.e., context-free word reading speed; Cutting, Materek, Cole, Levine, & Mahone, 2009; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Klauda & Guthrie, 2008), most likely because text reading fluency captures oral language processes in addition to word reading automaticity (Kim, Wagner, & Lopez, 2012).

Yet, the literature reviewed in Chapter 1 suggests that while the lower-level skills involved in reading fluency are a necessary component of text comprehension, it is now known that higher-level skills play an important role in comprehension beyond the contribution of lower-level skills. For example, regression analyses have revealed that

higher-level skills such as sensitivity to story structure, inference generation, information integration, comprehension monitoring, and working memory uniquely predict comprehension level when controlling for lower-level skills such as word reading accuracy, vocabulary knowledge, and verbal ability (Cain et al., 2004a). Also, correlations between lower-level skills and comprehension, although substantial, are not perfect (Nation, 2005).

Further evidence of this dissociation between word-level and comprehension skills comes from groups of children who have been identified as having reading comprehension levels well below what is predicted by their word reading ability and chronological age (Oakhill, 1994; Yuill & Oakhill, 1991; see also Perfetti, 2005, for a review). These children are often referred to as "poor comprehenders" or as having Specific Comprehension Deficit (SCD), or Specific Comprehension Impairment (SCI; Landi, 2010). Despite having age-appropriate word reading ability, compared to good comprehenders, poorer comprehenders demonstrate trouble with generating necessary inferences (Cain et al., 2001; 2004a; Oakhill et al., 2003; Oakhill & Cain, 2000), monitoring and maintaining coherence (Cain et al., 2004a; Oakhill et al., 2003; Oakhill, Hartt, & Samols, 2005b), and semantic processing (Nation et al., 1999; Nation & Snowling, 1998). Thus, it is now becoming apparent that the comprehension difficulties of these children are a result of impairments in higher-level cognitive systems (Cain, 2009; Cain et al., 2001; Kendeou et al., 2014; Nation, 2005). The current study focuses on the role of visual imagery in reading comprehension, while also acknowledging the role of verbal working memory.

To recapitulate on Chapter 1 (see Chapter 1.3.5.1), several studies have established a relationship between verbal working memory and reading comprehension in children

after controlling for word-reading and vocabulary skills (Cain et al., 2004a; Seigneuric et al., 2000; Sesma et al., 2009) and verbal working memory has been found to make an independent prediction to pre-schoolers' listening comprehension that goes over and above lower-level verbal skills (Florit et al., 2009). Further, in comparison to good comprehenders, poor comprehenders often demonstrate difficulties holding and using information in working memory (Cain et al., 2004a; Oakhill et al., 2003).

The role of working memory has been argued to be especially important in language comprehension as it supports situation model construction by enabling a reader to maintain relevant information so that it can be integrated with incoming information into the meaning-based model, and so more connections can be made between concepts in a text (Daneman & Carpenter, 1980; Just & Carpenter, 1992). Consistent with this, verbal working memory has been implicated in several of the higher-level processes involved in building a coherent representation of the meaning of a text. For example, verbal working memory has been implicated in both inference generation (Friedman & Miyake, 2000; Masson & Miller, 1983; Pérez et al., 2014; Singer et al., 1992; Singer & Ritchot, 1996) and maintaining global coherence of a narrative (Kim, 2014; Oakhill, Hartt, & Samols, 2005b; Orrantia et al., 2014; Whitney et al., 1991; Yuill et al., 1989). Thus, it appears that readers with limited working memory capacity are subject to constraints on how much information they can keep active as they read, which leads to poorer integration of information and less detailed situation model representations. Consistent with this, working memory tasks that require both storage and additional processing of information have more often been found to correlate with children's reading comprehension than tasks that assess passive storage capacity only (Daneman & Merikle, 1996).

In contrast to this, however, some studies suggest that the difficulties with verbal working memory tasks seen in poor comprehenders are a direct result of underlying reading or oral language deficits (Nation et al., 1999; Nation & Snowling, 1998). Specifically, it is possible that good readers perform better on measures of sentence span because decoding is not as effortful for them as for poor readers, therefore they can devote more resources to the memorisation component of the task (Goff et al., 2005). Similarly, it has been proposed that good comprehenders have advantages in semantic processing, which enhances encoding of words and sentences; thus, poor performance on these tasks may in fact be artefacts of difficulties with language and semantic skills, rather than poor working memory per se (Nation et al., 1999). Indeed, relationships between semantic memory and reading comprehension are often reported in the literature (Nation et al., 1999; Nation & Snowling, 1998; 1999; 2000). This is not surprising, as semantic memory is considered to be an individual's long-term representation of world knowledge (Tulving, 1972), thus comprises the main building blocks for situation model construction.

Deficits in semantic memory could, however, also be related to the use of visual imagery during the encoding of verbal stimuli. As outlined in Chapter 1 (see Chapter 1.2.1), dual coding theory proposes that both a verbal and a nonverbal (imagery) subsystem are involved in language processing and comprehension. Further, it is proposed that information that is coded in both these forms is strengthened in semantic memory, by increasing the associations between words and their attached meanings (Paivio, 1971; 1986). Consequently, due to additional encoding, this information becomes less prone to decay, thus enhancing its recall (Sadoski & Paivio, 2001).

Support for dual coding theory is provided by evidence of imageability effects on verbal recall. For example, abstract nouns (e.g., *truth*) are known to be less imageable than concrete nouns (e.g., *house*), as evidenced by: positive correlations between ratings of imageability and concreteness (Gullick, Mitra, & Coch, 2013); selective interference of concurrent visuospatial processing by writing definitions of concrete, but not abstract nouns (Kellogg, Olive, & Piolat, 2007); and neurophysiological evidence that shows hemodynamic and ERP differences in activation for concrete nouns compared to abstract nouns (see Gullick et al., 2013; Weiss, Mueller, Mertens, & Woermann, 2011, for further discussion). Consequently, abstract nouns are often recalled to a lesser degree than concrete nouns, presumably because abstract nouns are less amenable to dual coding (Gullick et al., 2013; Walker & Hulme, 1999).

As such, if poor visual imagery underlies deficits in memory encoding, it may be the case that poor comprehenders only activate a predominantly textbase, and inherently verbal representation when encoding words or sentences. This would explain why their recall is less than that of good comprehenders, who would be encoding words in both verbal and visual form (i.e., imagining pictures of the object represented by the word). In the same vein, these deficits in visual imagery could be a mediating factor in the relationship between semantic memory and reading comprehension.

Indeed, a main premise of the current thesis is that visual imagery may be an important contributor to higher-level comprehension processes and overall resulting level of comprehension. Specifically, not only is there is evidence that visual imagery is activated as part of the situation model representation (Bergen et al., 2007; Dijkstra et al., 2004; Engelen et al., 2011; Horton & Rapp, 2003; Stanfield & Zwaan, 2001; Zwaan et al., 2002; 2004; Zwaan & Pecher, 2012), but there is also evidence that imagery is

actually *required* for situation model construction (Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004; see Chapter 1, pp. 36-38 for further discussion of these studies). Furthermore, although it has been established that working memory aids comprehension via its effects on knowledge integration and coherence building (Kim, 2014), others have found that after controlling for word-level and verbal skills, the relationship between reading comprehension and both inference making and comprehension monitoring is only partially mediated by verbal working memory, thus additional processes must play a role in these higher-level skills (Cain et al., 2004a; Chrysochoou, Bablekou, & Tsigilis, 2011).

From an embodied cognition perspective (see Chapter 1.2.2), it is possible that visual and motor simulations may also aid a deeper experience and understanding of the situation described in a text, as it has been suggested that embodied simulation enables an individual to understand more deeply the events and behaviours portrayed by others (Fischer & Zwaan, 2008). In particular, perceptual symbols theory (PST; Barsalou, 1999) argues that cognition involves modal systems that utilise the same neural regions involved in actual perceptual experience to construct perceptual symbols that represent knowledge. Thus, it has been suggested that the situation models that readers construct in order to represent meaning are a perceptual and motor simulation of the situation described in a text that develops in a manner similar to a real-life physical scene, and which the reader experiences vicariously via the view of the protagonist (Barsalou, 1999; Glenberg, 1997; Johnson-Laird, 1983; Zwaan, 1999b). In addition, visual imagery during reading is proposed to lead to higher reading engagement (Green & Brock, 2002), and reading engagement has been found to be a significant predictor of reading comprehension (Guthrie & Wigfield, 2000; Wigfield et al., 2008). It is thus possible that both verbal skills and visual imagery skills are involved in reading

comprehension. Despite this, few studies have provided empirical evidence of a relationship between visual imagery processes and reading comprehension.

However, as explored in Chapter 1 (see Chapter 1.3.5.3), discrepancies in earlier research that explored the relationship between visual imagery and reading comprehension may have been due to not having a more detailed reading comprehension framework from which to build upon. Specifically, this may have limited the nature of imagery interventions as well as the way in which comprehension was measured in these studies. For example, many of these early studies relied on text recall as a measure of comprehension, rather than assessing the higher-level skills that contribute to building a coherent representation of the *meaning* of a text. As such, several of these studies also failed to account for how dynamic imagery may contribute to the construction and updating of this meaning-based representation, and thus simply measured imagery as an undifferentiated skill, often through the use of subjective measures of imagery. Consequently, early interventions too were designed with the aim of having participants simply *generate* visual imagery in response to textual input.

Yet, as outlined in the computational model presented by Kosslyn (Kosslyn, 1980; 1994; Kosslyn et al., 1984), and supported by the results of Study 1, visual imagery ability can be differentiated into at least four major components: image generation (formation of a visual image in the visual buffer), image maintenance (retaining the visual image), image inspection (interpreting object or spatial characteristics of the image in the visual buffer), and image transformation (manipulating or reorganising the image in some way; Kosslyn, 1980; 1983; Kosslyn et al., 1984). Despite this, models and measures that serve to delineate some of the key subskills of visual imagery are not often adopted in practice and researchers continue to attempt to measure visual imagery

as a unitary construct. It is proposed that moving forward, consideration of this entire visual imagery system is necessary in order to gain a greater understanding of the role of visual imagery in reading comprehension.

Along with the consideration of visual imagery measurement, definitions and measurement of comprehension are also important to consider when reviewing past literature. As mentioned, many earlier studies relied on text recall as an indication of comprehension, thus overlooking the importance of higher-level representations in comprehension, and studies investigating the relationship between reading comprehension and visuospatial working memory (VSWM; a similar construct to visual imagery) have mostly relied on standardised tests of reading comprehension. Indeed, several criticisms exist relating to the use of standardised comprehension tests in education and research, including that these measures are heavily dependent on word-level processing and do not assess many of the higher-level skills that are involved in situation model construction. Thus, the limitations of these measures may have hampered the findings of previous studies. These criticisms will be explored in more detail in the following section.

Criticisms of Current Measures of Reading Comprehension

Although the complex and multi-faceted nature of reading comprehension has now been captured by cognitive frameworks, an earlier focus on a single-component approach to reading comprehension and word-reading skills strongly influenced reading comprehension theory and measurement (see Hannon & Daneman, 2001, for a discussion). As such, many standardised measures of reading comprehension are often criticised for their heavy reliance on a reader's decoding skills (Francis et al., 2006; Keenan et al., 2008; Rowe et al., 2006; Spooner et al., 2004). This was highlighted by

Spooner et al. (2004) who found that combined measurement of accuracy and comprehension on the Neale Analysis of Reading Ability can result in comprehension scores that are heavily dependent on a reader's decoding ability. Specifically, in a sample of 7- and 8-year-old children, poorer decoders (defined by low accuracy scores on the Neale) attained Neale comprehension scores that were lower than what would be predicted by their age and listening comprehension level, whereas skilled decoders achieved comprehension scores higher than what would be predicted. In addition, the difference in Neale comprehension scores between these two groups was significant, with higher comprehension scores found in the group of skilled decoders, even though both groups had a similar level of listening comprehension ability (Spooner et al., 2004).

As the children in Spooner et al.'s (2004) study were matched for level of listening comprehension, rather than inferring that general comprehension ability is dependent on decoding ability, it was concluded that concurrent measurement of accuracy impairs comprehension during Neale administration (Spooner et al., 2004). Several reasons for this have been offered. Firstly, as reading errors are corrected during testing, frequent corrections may cause disruptions that do not allow children to sufficiently processes text information at a level that is required for comprehension, and amending narrative representations in response to corrections may be difficult or confusing for many children (Spooner et al., 2004). Similarly, weak decoding skills may compel a reader to focus on word reading and not actively engage in comprehension due to implicit testing pressures. Lastly, as the Neale includes a cut-off based on number of reading errors, the number of comprehension questions administered is dependent on level of reading accuracy. Therefore, due to early cut-off, children with low reading accuracy but unimpaired comprehension are not given the opportunity to answer all of the

comprehension questions which they could potentially have answered correctly (Spooner et al., 2004).

In contrast, however, there is also evidence that the Neale does assess some degree of higher-level comprehension. For example, Nation and Snowling (1997) found that the Neale was more dependent on listening comprehension than the Suffolk Reading Scale. In addition, Bowyer-Crane and Snowling (2005) concluded that the Neale was more heavily reliant on the generation of knowledge-based inferences than the Wechsler Objective Reading Dimensions Test of Reading Comprehension (WORD). However, it should be emphasised that both these studies only compared the Neale to one other measure, each which was arguably heavily dependent on reading accuracy itself. Indeed, in Nation and Snowling's (1997) study, listening comprehension only accounted for 16% of the variance in Neale comprehension scores after controlling for word-reading ability, and Neale comprehension scores loaded similarly onto a decoding factor (.62) and a comprehension factor (.67) that were extracted from various reading and comprehension tests. In Bowyer-Crane and Snowling's (2005) study, while approximately 29% of the questions in the Neale comprehension test were assessed as engaging knowledge-based inferences and another 5% tapped evaluative inference, the remainder of questions were found only to involve the generation of elaborative (4% of questions) or text connecting inferences (34% of questions). In addition, 14% of the questions on the Neale could be answered based on literal information found in the text passages, and another 14% were vocabulary dependent.

It is thus clear from these studies that the Neale is influenced by both word-level and higher-level reading skills. However, these studies also demonstrate how separate comprehension measures can differ quite substantially with regards to how much

variance each of these skill levels contribute. Other studies have come to similar conclusions. For example, when assessing the item difficulty of the seventh to ninth grade version of the Gates-MacGinitie Reading Test (GMRT), Rowe et al. (2006) found that item difficulty correlated with text passage features such as word frequency and sentence length, but not item characteristics such as whether an inference was required to answer the comprehension question. Rowe et al. (2006) thus argued that, although this version of the GMRT is utilised in schools and research as a measure of reading *comprehension*, it is rather a measure of reading *ability*, or at most a measure of basic comprehension. This finding with 7th- to 9th-grade students was replicated in a later study, which found that comprehension scores were primarily influenced by vocabulary difficulty, and other text-level features (Ozuru et al., 2008). In addition, Ozuru et al. (2008) found that there was less systematic influence of text-level variations on the 10th to 12th grade level version of the GMRT in comparison to the seventh to ninth grade level version, suggesting that these two versions of the measure may not be comparable in regard to the level of higher-level comprehension processes they tap into.

To further examine the validity of the assumption that comprehension measures can be used interchangeably, Cutting and Scarborough (2006) measured reading comprehension using three measures: the revised version of the GMRT (GMRT-R), the Gray Oral Reading Test (GORT), and the Wechsler Individual Achievement Test (WIAT), in a sample of 97 children aged 7.0 to 15.9 years. The comprehension measures were found to differ in their sensitivity to lower-level reading skills (measured as a composite of scores from two word-reading measures) and oral language skills (vocabulary skills and sentence processing). In addition, the two subcomponents of oral language skills provided unique contributions to each measure, suggesting that different measures of reading comprehension may make differential demands on vocabulary

knowledge and sentence-processing abilities, thus they should not be measured as a single component as done in prior research.

Similarly, Keenan et al. (2008) compared the scores of 510 8- to 18-year-olds on four measures of reading comprehension: the GORT, the Woodcock-Johnson Passage Comprehension (WJPC) subtest, the PIAT reading comprehension subtest, and the Qualitative Reading Inventory (QRI), along with two measures of listening comprehension and three measures of decoding. A factor analysis including all measures revealed two factors: a decoding factor and a comprehension factor. Although all the reading comprehension tests loaded onto the comprehension factor, they did so to varying degrees. In addition, two of these comprehension measures (the PIAT and the WJPC) also loaded highly onto the decoding factor, more so than they did on the comprehension factor. Regression analysis revealed a similar pattern of results: the measures were diverse in regards to how much variance decoding accounted for, and decoding accounted for more of the variance than listening comprehension, on both the PIAT and the WJPC scores with the reverse being true for the GORT and both QRI measures (Keenan et al., 2008). Intercorrelations between the comprehension measures were also variable, and mostly low (ranging from $r = .31$ to $r = .54$; with the exception of the correlation between the PIAT and WJPC, here $r = .70$), further suggesting these measures do not all tap the same component skills (Keenan et al., 2008).

The question of whether these tests measure skills that actually relate to comprehension has not only been examined by comparing tests, but also by looking at individual items within tests. Highlighting issues of test validity with an established reading measure, Keenan and Betjemann (2006) found that individuals can score above chance on the comprehension questions even when they do not read the passages. This suggests that

many of these questions may be answered using prior knowledge alone (i.e., are "passage independent") and that students are therefore likely to perform above their actual comprehension ability (Keenan & Betjemann, 2006). Further findings regarding the passage-independent items of the GORT also revealed important implications for educational applications. Firstly, there was no difference in performance on passage-independent items between children with a diagnosed reading disorder and a control group and performance on passage-independent items did not correlate with other reading or listening comprehension tests (Keenan & Betjemann, 2006). Thus, it appeared that these items cannot identify struggling readers and, rather than assessing comprehension, they gauge an additional variable, most likely, the level of general knowledge a reader brings to the task (Keenan & Betjemann, 2006). Secondly, as passage-*dependent* items *did* correlate with additional measures of reading and listening comprehension, it indicated that passage-dependent and passage-independent items differ in what they measure. For example, the cognitive processes tapped by passage-independent items are likely those involved in knowledge retrieval and do not overlap to any great extent with the higher-level comprehension processes required to answer passage-dependent items (e.g., integrating ideas; Keenan & Betjemann, 2006).

It is thus becoming apparent that commonly used tests of reading comprehension do not necessarily tap the same collection of cognitive processes. Therefore, different comprehension tests may identify different children as poor comprehenders depending on where skill deficits lie. According to theoretical models, reading comprehension goes beyond decoding and requires the integration and execution of several cognitive skills. However, as most standardised comprehension measures rely on offline questioning following reading of the text, it has been argued that they only provide an indication of the *product* of reading comprehension, rather than the *processes* that take place to

provide this outcome (S. E. Carlson, Seipel, & McMaster, 2014a; Rapp et al., 2007). It is therefore not surprising that the contribution of cognitive skills, such as IQ, verbal memory and attention, have been found to be subsumed by the contribution of more basic reading skills such as word-reading and vocabulary on several standardised comprehension measures (Cutting & Scarborough, 2006).

Yet, few measures have been designed to overcome the limitations of existing comprehension measures, and provide a standardised assessment of higher-level processes, which would be viable for use in educational settings. As outlined in Chapter 1, however, exceptions to this include a handful of measures built from cognitive theory, including The Diagnostic Assessment of Reading Comprehension (DARC; August et al., 2006), which aims to measure comprehension skills (i.e., text and knowledge integration) independently of decoding ability. The DARC was developed in recognition of the problem that the decoding and word recognition requirements of measurement tools can prevent accurate measurement of other cognitive processes necessary for comprehension (i.e., inferencing and accessing background knowledge). Thus, it was designed to measure comprehension while minimising the impact of lower-level reading abilities such as decoding, reading speed and vocabulary (August et al., 2006; Francis et al., 2006). Research with this measure shows that in comparison to a standardised measure of reading comprehension (the WJPC), the DARC is less influenced by word-reading skills and also more dependent on oral language and narrative skills, although both these measures are equally influenced by nonverbal reasoning (Francis et al., 2006). In fact, after accounting for the contributions of language skills and non-verbal reasoning, word-reading skills (decoding and fluency) were found to be significant predictors of scores on the WJPC, but not on the DARC (Francis et al., 2006).

The Current Study

The main aim of the current study was to examine the influence of visual imagery on reading comprehension. However, this study extended on previous research in two important ways. Firstly, in order to account for the multidimensional nature of visual imagery, several measures of this construct were included in the current study, each designed to tap a different subprocess. Secondly, in order to gain a more accurate picture of reading comprehension, two measures of reading comprehension were included in the current study: The Neale Analysis of Reading Ability, which is a traditional standardised measure of comprehension, and the Diagnostic Assessment of Reading Comprehension (DARC; August et al., 2006), a measure formulated from cognitive theory, which aims to measure the higher-level skills that contribute to reading comprehension (further details about these measures can be found in the procedure section; see pp. 119-121). Although this is not to say the DARC is completely unrelated to decoding (Francis et al., 2006), inclusion of this measure may provide a more accurate picture of reading comprehension while reducing the influence of word reading skills, allowing clearer interpretations of the relationship of imagery to higher-level comprehension. In addition, inclusion of this measure allowed an examination of how a more recently formulated measure of comprehension compared to one that is often used throughout research and practice.

It was hypothesised that, after controlling for word reading skills, fluid intelligence, and verbal working memory, imagery ability scores would be a predictor of comprehension scores on the DARC but not on the Neale; as the DARC provides a more valid measure of higher-level skills involved in reading comprehension, particularly inferencing, which previous research has shown may be reliant on visual imagery processes (i.e., Fincher-Kiefer & D'Agostino, 2004). In contrast, studies that have examined the role of

VSWM (a similar construct to visual imagery) in narrative comprehension using standardised measures of comprehension have not found a relationship. The investigation of which type of imagery is most predictive of reading comprehension is more exploratory in nature. However, it is likely that more complex forms of imagery (i.e., scanning and rotation) will be more predictive of reading comprehension, rather than simple image maintenance, as these imagery processes are closer to those that would take place during situation model construction and updating.

3.2 Method

Participants

One hundred and fifteen children in Grades 4 and 5 were screened for participation in the current study. All participants had normal or corrected-normal vision, were free from cognitive or diagnosed learning impairments and spoke English as their first language. To ensure no participants with undiagnosed reading disorders were included in the sample, children who had a Neale reading accuracy score within the range of "very low" for their age group (as per the norms provided in the Neale manual; Neale, 1999) were excluded from this sample. Seven participants did not meet the minimum criteria of reading accuracy and were therefore excluded from this study. Three additional participants were excluded due to failure to comply with task instructions. The final sample consisted of 105 children (56 female, and 49 male), with an age range of 8.24 to 10.91 years ($M = 9.58$, $SD = .57$), from seven primary schools in Perth, Western Australia. These schools represented a wide range of socioeconomic backgrounds (obtained using the Australian Index of Community Socio-Educational Advantage (ICSEA) ratings; the current schools ranged from of 885 – 1153 (nationwide, the median ICSEA score is 1000, and in Western Australia the total range

of ICSEA ratings across the metropolitan area is approximately 801-1211), and an effort was made to recruit comparable numbers of children from within each stratum of this range, in order to minimise sampling bias.

Measures

Reading Comprehension

The Neale Analysis of Reading Ability (Australian third edition; Neale, 1999)

The Neale is a standardised measure of reading rate, accuracy and comprehension, widely used in both education and research. This version of the Neale has Australian normative data, and demonstrates adequate internal consistency when used with this age group (Kuder-Richardson reliability coefficients [KR-21] of 0.94 for reading rate, 0.95 for reading accuracy, and 0.85 for reading comprehension [Year 4], and 0.95 for reading rate, 0.96 for reading accuracy, and 0.96 for reading comprehension [Year 5]; Neale, 1999). Inter-rater and test-retest reliability also appears to be high, with correlations of .95 (rate) .95 (accuracy) and .93 (comprehension), between teacher and assessor administration. Concurrent validity of this measure is adequate: raw scores of the Neale correlate with raw scores of the Dartmouth Advanced Reading Test ($r = .77$), and the Schonell Reading Test ($r = .76$ [rate], .95 [accuracy], .88 [comprehension]).

Materials

The Neale contains a written storybook with two practice stories and six test passages of increasing difficulty. A separate individual record form containing a copy of each story and the comprehension questions for each passage is used by the administrator for recording reading errors, reading rate and comprehension scores.

Procedure

Form 1 of the Neale was used in the current study, and administered and scored as per the standardised instructions in the manual (Neale, 1999). This involved children reading each passage out loud, and the administrator correcting any errors as they occurred. Following each story, comprehension questions were read aloud by the administrator to which the child provided a verbal response (open-ended). Accuracy scores were obtained by subtracting the number of reading errors from the highest possible score for each passage, fluency scores were calculated as words read per minute, and comprehension scores were scored as one point for each correct question.

The Diagnostic Assessment of Reading Comprehension (DARC; August et al., 2006)

Due to the criticisms that the Neale does not provide a valid assessment of all the skills involved in comprehension and is largely reliant on lower-level abilities such as decoding (Spooner et al., 2004), the DARC was included as an additional measure of reading comprehension in order to measure comprehension separate to the effects of word-reading ability.

Materials

The DARC consists of two versions: "Nan's Pets" and "Tom and Ren". Each version consists of a single story presented in text format, and is accompanied by 30 comprehension questions. The DARC controls for required level of decoding by using simple and highly decodable words in these texts, and requires inferencing and knowledge integration in order to answer the comprehension questions. This is achieved by presenting the reader with a story that describes the relations among a set of real entities (i.e., cats have fur) and artificial terms (i.e., culps are like cats) and statements that require a true or false response (i.e., culps have fur) in order to question the reader

on these relations (August et al., 2006). In addition, the DARC has four subscales of items that can be used to determine whether comprehension difficulties are due to poor memory for the text (text memory subscale; six items), not making inferences based on text information (text inferencing; five items), a lack of background knowledge (knowledge access; six items), or a failure to integrate this background knowledge with information presented in the text (knowledge integration; 13 items; August et al., 2006). Developers of this measure report adequate internal consistency for each version (α = .75 for Story 1 [Nan's Pets]; and α = .68 for Story 2 [Tom and Ren]), and reliability coefficients for the subtests within the range of .5 to .6.

Procedure

Due to slightly higher reliability, the passage "Nan's Pets" was chosen over "Tom and Ren" for use in the current study. Both the practice passage and story passage were administered to all participants, as per the standardised instructions. Participants read the test passage aloud in three separate parts and, after reading each part, answered a series of yes/no questions about the story (30 questions in total) for which they were scored one point for every correct answer, and zero points for incorrect answers.

Visual Imagery

Three measures of visual imagery were selected base on their apparent utility when used with children, as indicated by the results of Study 1 (see Chapter 2). These were the measures of image maintenance, image scanning, and image transformation.

Image Maintenance

The image maintenance task used in Study 1 was used in the current study to assess participants' ability to maintain an imaged pattern. This was a computer task adapted

from Kosslyn et al. (1984), which required participants to memorise a pattern contained within a grid, and then once the pattern was removed, decide whether two probes fell in grid cells that were previously filled by the pattern (see Chapter 2.2, pp. 77-81 for further details on this task).

The materials and procedures utilised for this task were identical to those of Study 1, with the exception that the total number of experimental trials was reduced from 40 to 28 trials (14 trials of each condition), in order to reduce the time taken to complete this task. This was done due to an increase in the overall testing time in the current study, which resulted from the inclusion of several additional measures. This task took approximately 10 minutes to complete.

Image Scanning

The image scanning task used in Study 1 was used in the current study to assess participants' ability to scan across a maintained visual image. This was a computer task adapted from Kosslyn et al. (1984), which required participants to memorise a pattern contained within a grid, and then once the pattern was removed, decide whether an X-shaped probe was placed in a cell that was previously filled by the pattern, or, if the probe was O shaped, decide whether the probe was placed in a cell *opposite* to a cell that was previously filled (see Chapter 2.2, pp. 81-83 for further details on this task).

The materials and procedures utilised for this task were identical to those of Study 1, again however, with the exception that the total number of experimental trials was reduced from 40 to 28 trials (14 trials of each condition), in order to reduce the time taken to complete this task. This task took approximately 10 minutes to complete.

Image Transformation: Mental Rotation Task (MRT)

The image transformation task that was used in Study 1 was used in the current study to assess participants' ability to transform a visual image. This task was The Primary Mental Abilities (PMA) Spatial Relations test (L. L. Thurstone & Thurstone, 1947). The materials and procedures utilised for this task were identical to that of Study 1 (see Chapter 2.2, pp. 83-84 for further details on this task). In the current study, no participant took longer than 12 minutes to complete the test items in this task.

Verbal Working Memory

Digit Span

Two verbal working memory measures were included in the current study: a simple span task (forward digit span) and a complex span task (backward digit span). Digit span tasks were chosen, as word or sentence span tasks may provide better readers with an additional advantage that is unrelated to working memory (see Nation et al., 1999). Further, digits are readily amenable to verbal coding, but would likely be harder than words to encode visually, as words are more susceptible to dual coding. Therefore, this increased the likelihood that the verbal working memory measures were distinct from the visual imagery measures.

Materials

Both the forward and backward span measures were administered using the digit span task from The Psychology Experiment Building Language (PEBL) test battery version 0.13 (S. T. Mueller, 2013), on the same laptop used for the imagery tasks.

Procedure

In the forward span task, a series of digits was displayed centrally on the computer screen, each for 1000ms with a 1500ms inter-stimulus interval. Following presentation

of the entire series of digits, participants recalled these in the same serial order they were presented, by typing their response on the laptop number pad. Participants received two trials of each span length, with an inter-trial interval of 5000ms. The task began with a length of three digits, and increased in length by one digit if a participant was correct on at least one of the two trials of the previous length. If a participant was incorrect on both trials the task was discontinued. The greatest number of digits recalled in the correct order was recorded as a participant's digit span. The procedure for the backward span task was identical to the forward span task, with the exception that participants were required to enter the digits in the reverse serial order to which they had been presented.

Fluid Intelligence (Non-Verbal Reasoning)

Raven's Standard Progressive Matrices (Raven, 1958).

To measure general intellectual ability, the 20-minute timed version of Raven's Standard Progressive Matrices (Raven, 1958) was used with the norms developed by the Australian Council for Educational Research (ACER) for use with Australian students (de Lemnos, 1989). Raven's is described as being a test of non-verbal reasoning ability that captures fluid intelligence, a component which underlies Spearman's *g* (general intelligence) factor (de Lemnos, 1989), and is considered to be one of the purest measures of *g* available (Carpenter, Just, & Shell, 1990).

Estimates of reliability of the timed version provided in the Australian manual also report adequate internal consistency for this age range: Kuder-Richardson 21 coefficients range from .80 (*SEM* = 3.60) for Year 5, to .85 (*SEM* = 3.65) for Year 4 (de Lemnos, 1989). Pearson's *r* correlations between the untimed and timed version of the SPM range from .76 (Year 5) to .85 (Year 4) for this age group, indicating good test-

retest reliability (de Lemnos, 1989). Validity of this test has also been demonstrated: moderate to strong correlations have been found between scores on the timed version of the SPM with performance on another test of non-verbal ability (Jenkins Non-Verbal Test, $r = .76$) and a test of general intellectual ability (ACER Test of Reasoning Ability, $r = .63$). Discriminant validity has also been demonstrated, with lower correlations found between the untimed version of the SPM and tests of verbal ability (ACER Word Knowledge Test Form E $r = .43$ and Form F $r = .49$) and with teacher ratings of performance (General Scholastic Ability $r = .43$).

Materials

This test consists of a test booklet containing 60 items separated into five sets (A-E) each containing 12 items, along with an answer sheet for recording responses.

Procedure

Raven's was administered to all participants according to the standardised instructions for timed group administration in the test manual (Raven, 1958). For each item, participants were required to select the correct missing piece of a large pattern from six or eight alternatives by shading in the number that corresponded to the selected alternative on the answer sheet. Participants completed as many items as possible within the 20-minute time limit.

General Procedure

Participation in this study took place at the child's school, during school hours, in a quiet area separate to the classroom. Children completed all tasks over three separate sessions, each lasting 30 minutes to 1 hour (including breaks between tasks), with no longer than three weeks between testing sessions. All participants completed these

sessions and tasks within each session in the same order. In the first session, which was completed individually, the order of tests was: the Neale, the DARC and finally the two digit span tasks. The second and third sessions were completed in groups of two or three participants in the following order, session two: the image maintenance task, the image scanning task and the mental rotation task; and in session three: Raven's Progressive Matrices.

3.3 Results

Data Screening and Reduction

For the image maintenance and scanning tasks, mean accuracy (percentage correct) and mean RT over correct trials (ms) was calculated for each participant for each condition. Trials in which a participant responded more than double their mean RT for that condition were considered to likely reflect a lapse in concentration, and were thus coded as errors along with incorrect responses. In order to increase the reliability of the data set, any participants who made more than 50% errors in any single condition, or over all conditions, were identified for exclusion from further analyses: no participants exceeded these criteria in the maintenance task, but data from five participants exceeded these criteria in the scan task.

Due to ceiling effects in the accuracy data for the scan and maintenance tasks, found here and in pilot work by the current author (see Chapter 2), RT was used as the primary indicator of scanning and maintenance ability. Paired-samples t-tests confirmed that in the maintenance task participants responded significantly faster in the light load condition ($M = 1371.99$ms, $SD = 349.81$) than the heavy load condition ($M = 1947.53$ms, $SD = 631.43$; $t(104) = 10.27$, $p < .001$, $d = 1.0$), and in the scan task

responded significantly faster in the control (M = 1632.13ms, SD = 355.16) than the scan (M = 2320.31ms, SD = 584.16) condition ($t(99)$ = 13.49, p < .001, d = 1.35) thus indicating scanning had occurred. Subsequently, each individual's mean difference in RT between conditions was calculated for both the maintenance and scanning task. For the scan task, this simply represented scanning time (i.e., a greater positive difference score indicated a longer time to scan), whereas in the maintenance task a longer RT in the heavy load condition as compared to the light load (i.e., a greater positive difference score) was used as an indication of greater difficulty in retrieving the maintained stimuli. Specifically, as the light load condition does not impose much memory load, being more likely accessible via a perceptual afterimage, a similarly fast RT in the heavy load condition would indicate maintenance of the visual mental image of the stimuli to a similar strength as still actually visually perceiving it. All analyses were conducted with the raw scores of the other measures.

Descriptive Statistics

Prior to analysis, all data were screened for multivariate outliers and to determine whether the assumptions of multivariate analysis were met. No significant violations of normality, linearity or homoscedasticity were detected using standard screening approaches (Tabachnick & Fidell, 2007). Using Mahalanobis distance of p < .001, seven multivariate outliers were detected, and one of these participants had also failed to complete the image scanning task with minimum accuracy. These seven cases were subsequently removed. A series of independent samples t-tests confirmed there were no significant gender differences on any of the independent or dependent variables. The means and standard deviations for the scores on each test for this sample are presented in Table 3.1.

Table 3.1

Mean, Standard Deviation, Range, and Possible Range, of all Variables (n = 94)

Measure	M	SD	Range	
			Actual	Possible
Age	9.63	.53	8.74–10.91	-
Raven's	39.97	5.87	19.00– 54.0	0–60
Neale Accuracy	71.56	17.47	35.0–98.0	0–100
Neale Rate	76.51	22.10	36.0–130.0	-
Digit Span Forwards	4.90	.86	3.0– 7.0	0–10
Digit Span Backwards	4.13	1.07	2.0– 7.0	0–10
Image Maintenance	650.43	454.84	-245.57–4581.02	-
Image Scanning	628.08	641.26	-290.56–2099.44	-
Mental Rotation	39.30	13.03	-5.0– 54.0	-66–54
Neale Comprehension	22.70	6.18	12.0–40.0	0–44
DARC	25.73	2.91	16.0–30.0	0–30

Selection of the Strongest Imagery Predictors

Two exploratory hierarchical regressions were conducted to determine the strongest predictors from the three visual imagery variables on each of the comprehension measures, using the 94 complete data sets (i.e., excluding the seven outliers, and the four additional participants who did not meet the minimum accuracy in the image scanning task). Due to multiple comparisons, a Holm-Bonferroni adjustment was applied to correlations from this analysis to control for Type I error. These correlations are shown in Table 3.2.

Table 3.2

Pearson's r Correlation Coefficients Between Reading Comprehension Measures and the Visual Imagery Measures (n = 94)

	Measure	1	2	3	4	5
1	Image Maintenance	-				
2	Image Scanning	.09	-			
3	Mental Rotation	.13	-.01	-		
4	Neale Comprehension	.16	-.03	.30*	-	
5	DARC Comprehension	.30*	.01	.25	.43*	-

* significant at adjusted alpha level following Holm-Bonferroni correction (original alpha = .05).

As can be seen in Table 3.2, while the MRT significantly correlated with comprehension on the Neale, RT in the image scanning and image maintenance tasks did not significantly correlate with either comprehension measure, with the exception of the correlation between the maintenance task and the DARC, which was significant, but in the opposite direction to what would be expected (i.e., a positive correlation, indicating a longer time to respond in the heavy load condition was related to higher comprehension performance).

As shown in Table 3.3, regression analysis revealed a similar finding: image scanning was not a significant predictor on the Neale or DARC after controlling for age and fluid intelligence. Image maintenance significantly contributed to scores on the DARC, however this was in the opposite direction to what would be expected. In contrast, mental rotation was a significant predictor on the Neale, and the variance accounted for by this measure was greater than that accounted for by the scan or maintenance task.

Table 3.3

Exploratory Hierarchical Regression Predicting Reading Comprehension Scores by Measures of Visual Imagery (n = 94)

Step	Predictors	Neale Analysis of Reading Ability (Comprehension)						The Diagnostic Assessment of Reading Comprehension					
		B	$SE\ B$	β	R^2	$R^2_{adjusted}$	ΔR^2	B	$SE\ B$	β	R^2	$R^2_{adjusted}$	ΔR^2
1	(Constant)	.61	11.80					16.32	5.48				
	Age	1.62	1.19	.14				.47	.56	.08			
	Raven's	.16	.11	.15	.05	.03	.05	.12	.05	.25*	.07	.05	.07*
2	(Constant)	-.17	11.58					15.13	5.32				
	Age	1.74	1.18	.15				.58	.54	.11			
	Raven's	.02	.12	.02				.08	.06	.15			
	Scan (RT)	.00	.00	-.02				.00	.00	.01			
	Maintenance (RT)	.00	.00	.14				.00	.00	.27**			
	MRT	.12	.05	.25*	.12	.07	.07	.03	.03	.14	.17	.12	.09*

* $p < .05$; ** $p < .01$; *** $p < .001$

Predictors of Reading Comprehension

Based on the results of the exploratory regression analysis, the image maintenance and scanning tasks were not included in the final regression analyses. Thus, to increase the power of the final analyses, the four remaining children who were excluded only on the basis of not having scan data were included. This resulted in a final sample of 98 children (52 female, 46 male), with an age range of 8.74 to 10.91 ($M = 9.61$, $SD = .54$). The flow of participants through each stage of the experiment is displayed in Figure 3.1 and the means and standard deviations for the scores on each measure for this final sample are presented in Table 3.4.
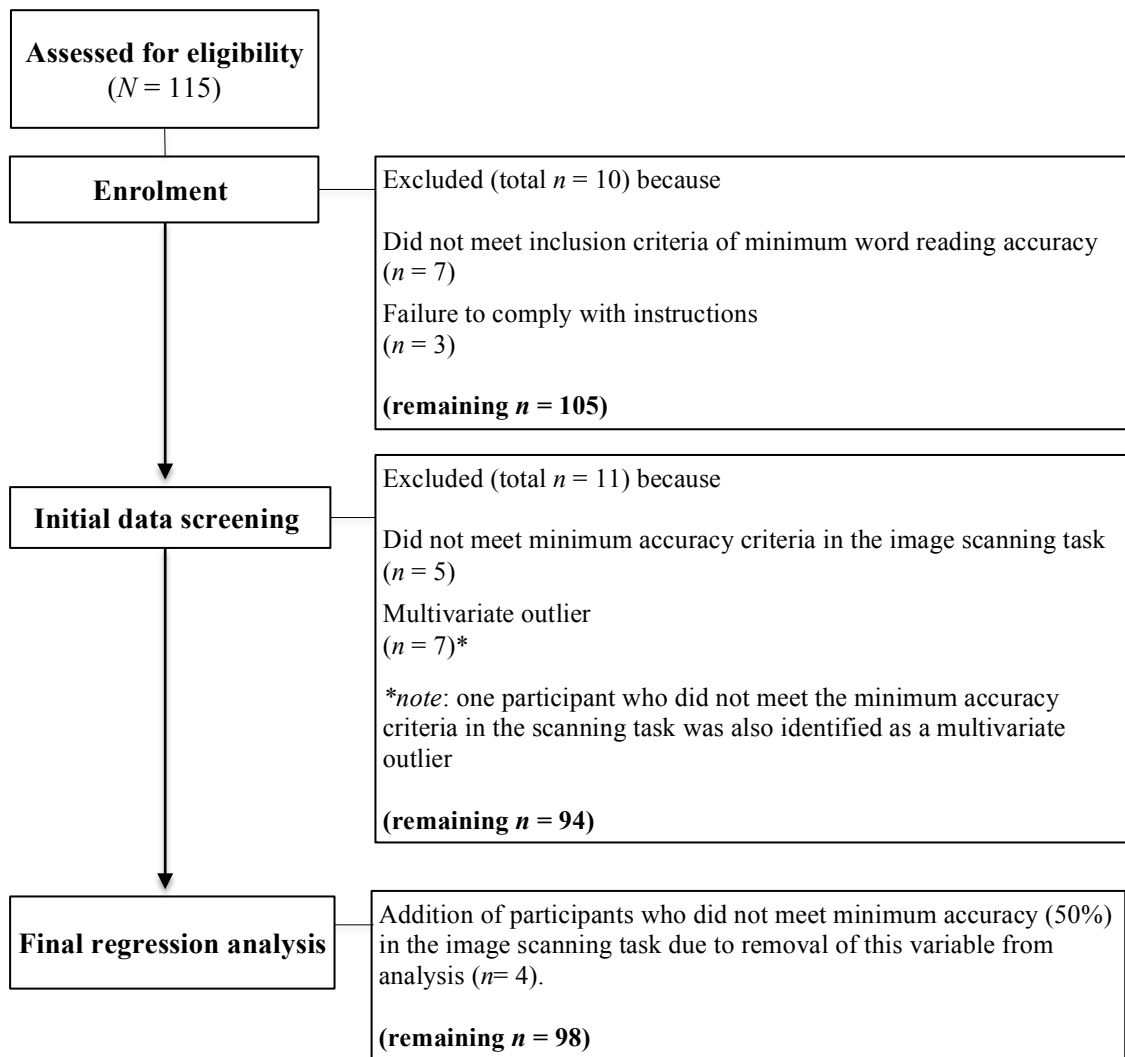
Figure 3.1. Flow of participants through each stage of the experiment.

Table 3.4

Mean, Standard Deviation, Range, and Possible Range, of all Variables (n = 98)

Measure	*M*	*SD*	Range Actual	Possible
Age	9.61	0.54	8.74 - 10.91	-
Raven's	39.65	6.16	19.0 - 54.0	0 – 60
Neale Accuracy	71.17	17.59	35.0 - 98.0	0 – 100
Neale Rate	76.34	21.98	36.0 - 130.0	-
Digit Span Forwards	4.89	0.86	3.0 - 7.0	0 – 10
Digit Span Backwards	4.10	1.06	2.0 - 7.0	0 – 10
Mental Rotation	38.51	13.63	-5.0 - 54.0	-66 – 54
Neale Comprehension	22.44	6.35	4.0 - 40.0	0 – 44
DARC Comprehension	25.62	2.98	16.0 - 30.0	0 – 30

Correlations

Correlations between each of the measures are shown in Table 3.5. Again, a Holm-Bonferroni adjustment was applied to correlations to control for Type I error due to multiple comparisons. The two comprehension measures showed a significant moderate correlation with each other. The Neale correlated more strongly with the word-reading measures (accuracy and rate) than with the other measures, and also correlated significantly with forward, but not backward digit span. In contrast to the Neale, the DARC showed much weaker correlations with the word-reading measures, and these correlations were not significant. Further, scores on the DARC did not correlate significantly with forward backward span, but did correlate significantly with backward span. The only other variable to correlate significantly with the DARC was scores on Raven's, which did not correlate significantly with the Neale. Although both comprehension measures correlated at a similar strength with mental rotation, this correlation was only significant for the Neale.

Table 3.5

Pearson's r Correlation Coefficients Between Reading Comprehension Measures and all Predictor Variables (n = 98)

	Measure	1	2	3	4	5	6	7	8	9
1	Age	-								
2	Raven's	.18	-							
3	Neale Accuracy	.29*	.23	-						
4	Neale Rate	.22	.20	.61*	-					
5	Forward Span	.19	.28	.31*	.38*	-				
6	Backward Span	.22	.26	.31*	.24	.44*	-			
7	Mental Rotation	.10	.52*	.34*	.10	.15	.30*	-		
8	Neale Comprehension	.19	.21	.45*	.47*	.31*	.24	.30*	-	
9	DARC Comprehension	.16	.31*	.26	.23	.23	.18*	.28	.47*	-

* significant at adjusted alpha level following Holm-Bonferroni correction (original alpha = .05).

Hierarchical Multiple Regressions

Two final hierarchical regressions were conducted in order to investigate the contribution of each of the predictors of reading comprehension using the two different comprehension measures as the outcome. Age and Raven's scores were entered in Step 1 as control variables. Step 2 contained the word-reading measures known to contribute to reading comprehension: accuracy and rate. The working memory measures (forward and backward digit span) were entered next in Step 3, as previous literature suggests a role of verbal memory in both propositional and higher-level comprehension. Finally, the visual imagery measure, mental rotation, was entered as Step 5. Thus, the order of the variables was consistent with the literature as to their theoretical contribution to reading comprehension, and enabled an assessment of whether imagery still contributed

to reading comprehension after the variance provided by previously established predictors of reading comprehension was accounted for.

As shown in Table 3.6, the results of the regression revealed that, after controlling for age and general ability, the highest predictors on the Neale were the lower-level reading skills; when combined, accuracy and rate accounted for 20% of the variance in Neale comprehension scores. Interestingly, reading rate ($sr = .24$) contributed to Neale comprehension scores more than reading accuracy ($sr = .18$). The verbal working memory measures and the visual imagery measure did not provide any additional significant variance to the Neale comprehension scores.

In contrast to the Neale, after controlling for age and general ability, word-reading skills did not provide any significant variance to DARC comprehension scores. However, similar to the Neale, verbal working memory and visual imagery provided no significant contribution to DARC comprehension scores. Thus, the only variable to predict any significant variance in the DARC scores was performance on Raven's; although, this variable did not remain a significant predictor in the final model.

Table 3.6.

Hierarchical Regression Analyses Predicting Reading Comprehension Scores by Age, General Intellectual Ability, Word Reading Ability, Verbal Working Memory, and Visual Imagery (n = 98)

Step	Predictors	Neale Analysis of Reading Ability (Comprehension)						The Diagnostic Assessment of Reading Comprehension					
		B	$SE\ B$	β	R^2	$R^2_{adjusted}$	ΔR^2	B	$SE\ B$	β	R^2	$R^2_{adjusted}$	ΔR^2
1	(Constant)	-3.22	11.41					14.50	5.25				
	Age	1.90	1.18	.16				.59	.55	.11			
	Raven's	.19	.104	.18	.07	.05	.07*	.14	.05	.28**	.10	.08	.10**
2	(Constant)	1.33	10.26					15.47	5.23				
	Age	.51	1.10	.04				.30	.56	.06			
	Raven's	.09	.10	.09				.12	.05	.24*			
	Neale Accuracy	.09	.04	.24*				.02	.02	.14			
	Neale Rate	.09	.03	.30**	.27	.24	.20***	.01	.17	.09	.14	.10	.04
3	(Constant)	.24	10.39					15.02	5.32				
	Age	.38	1.11	.03				.26	.57	.05			
	Raven's	.06	.10	.06				.11	.05	.23*			
	Neale Accuracy	.08	.04	.22				.02	.02	.13			
	Neale Rate	.08	.03	.27*				.01	.02	.07			
	Forward Span	.71	.78	.10				.28	.40	.08			
	Backward Span	.27	.61	.05	.28	.24	.01	.04	.31	.02	.15	.01	.01
4	(Constant)	.387	10.24					15.06	5.31				
	Age	.51	1.10	.04				.31	.57	.06			
	Raven's	-.04	.11	-.04				.08	.06	.16			
	Neale Accuracy	.05	.04	.15				-01	.02	.08			
	Neale Rate	.09	.03	.31*				.01	.02	.09			
	Forward Span	.83	.77	.11				.32	.40	.09			
	Backward Span	.07	.61	.01				-.02	.31	-.01			
	MRT	.10	.05	.21	.31	.26	.03	.03	.03	.14	.16	.09	.01

* $p < .05$; ** $p < .01$; *** $p < .001$

3.4 Discussion

The aim of this study was to determine whether visual imagery is a predictor of reading comprehension when measuring comprehension using a standardised test that may be more dependent on lower-level reading skills (the Neale), and a test that aims to measure higher-level comprehension while reducing the impact of word-reading ability (the DARC). In addition, this study included several measures of visual imagery, in order to determine whether some subtypes of this construct are more related to reading comprehension than others. It was found that two of the measures of visual imagery utilised in the current study (image maintenance and image scanning) were not related to comprehension scores on either measure. In contrast, the measure of image

transformation (mental rotation) significantly correlated with one of the measures of reading comprehension. Yet, contrary to predictions, this measure of comprehension was the Neale. When controlling for the variance accounted for by age, fluid intelligence, and verbal working memory, however, mental rotation no longer emerged as a predictor of scores on this comprehension measure. Thus, the hypothesis that visual imagery would be uniquely related to reading comprehension was not met.

The finding that mental rotation correlates with reading comprehension is consistent with earlier studies that found positive correlations between visual imagery and reading comprehension performance (e.g., Sadoski, 1983), and in line with intervention studies that found improvements in reading comprehension after visualisation training (i.e., F. L. Clark et al., 1984; Gambrell & Bales, 1986; Glenberg et al., 2004; Pressley, 1976). Interestingly, however, the finding of a significant positive correlation found between the Neale comprehension scores and the MRT is contradictory to previous research by Nyhout and O'Neill (2013) who did not find a correlation between mental rotation and Neale comprehension scores, and also mental rotation and performance on a spatial situation model task. This may, however, be due to differences in the mental rotation task used in Nyhout and O'Neill's (2013) study and the one used in the current study. In this previous study, the rotation task required children to examine two target shapes and decide what composite shape joining these objects could make, choosing from four alternatives. It is argued that this requires a greater degree of visual transference of stimuli onto one another than traditional mental rotation tasks like the one used in the current study.

 However, after controlling for additional variables known to influence reading comprehension, MRT no longer emerged as a predictor of Neale scores. This is in line

with several studies that have failed to find a relationship between visual imagery and reading comprehension. The current study hypothesised that this may be due to the use of existing standardised tests of reading comprehension in the vast majority of these studies. Such measures have been criticised on two major grounds. Firstly, these measures may be heavily influenced by lower-level word-reading skills and, thus, do not accurately measure many of the higher-level cognitive processes that lead to successful comprehension outcomes (Bowyer-Crane & Snowling, 2005; Francis et al., 2006; Keenan et al., 2008; Nation & Snowling, 1997; Rowe et al., 2006; Spooner et al., 2004; see also Hannon & Daneman, 2001); and secondly, these measures differ from one another with regards to the underlying skills they do actually measure (Cutting & Scarborough, 2006; Keenan et al., 2008; Keenan & Betjemann, 2006; Ozuru et al., 2008; Rowe et al., 2006). Indeed, the findings of the current study align with these criticisms, as most of the variance in Neale comprehension outcomes was accounted for by the contribution of verbal skills rather than integrative skills (i.e., those needed to complete the complex working memory task) and non-verbal skills such as visual imagery.

Specifically, the strongest predictor of Neale scores was performance in lower-level reading processes; combined, accuracy and reading rate accounted for 20% of the variance of Neale scores, and this contribution was significant. In comparison however, it was found that word-reading skills did not significantly contribute to comprehension when measured by the DARC: accuracy and reading rate only accounted for 4% of the variance of DARC scores after controlling for age and fluid intelligence. Although task-specific variance may have increased the variance that low-level skills contributed to the Neale comprehension scores in comparison to the DARC scores, these findings are still consistent with both the claims of the developers of the DARC (August et al., 2006;

Francis et al., 2006) and with the criticisms of the Neale (Spooner et al., 2004). In particular, this finding aligns with those of previous research, that the Neale is largely reliant on verbal skills such as decoding and that the accuracy and comprehension scales of the Neale cannot be appropriately separated (Spooner et al., 2004).

In addition, performance on Raven's was not a strong predictor of Neale scores in the regression analysis, whereas, in comparison, this was the only variable that significantly predicted scores on the DARC, prior to the inclusion of the visual imagery measures in the regression model. As Raven's places high demands on non-verbal and analogical reasoning (Carpenter et al., 1990), this potentially highlights the need for these skills in measures of comprehension that tap inferential over literal comprehension. Specifically, the DARC contains many questions aimed at engaging knowledge-based inference generation (i.e., 13 of the 30 DARC items tap knowledge integration). Further, although Raven's does not measure visual imagery per se, of relevance here is the substantial and significant correlation between Raven's and the mental rotation task that was found in the current study. Although some of this shared variance may be due to both these measures capturing the construct of g, this correlation is also a likely reflection of the visualisation skills and/or spatial ability skills that are necessary for successful performance on both these measures. Indeed, several researchers have identified a visuospatial factor that underlies performance on Raven's (Colom, Escorial, & Rebollo, 2004; DeShon, Chan, & Weissbein, 1995; Lynn, Allik, & Irwing, 2004). Consequently, this provides some indication that performance on the DARC may be more reliant on non-verbal, and possibly imagery-based, comprehension processes (i.e., those involved in situation modelling) than lower-level reading ability, whereas the opposite may be true for the Neale.

Yet, visual imagery did not predict comprehension scores on the DARC. This finding is difficult to make sense of in light of several theories of reading comprehension; for example, the event-indexing model of situation model theory (Zwaan, Langston, & Graesser, 1995a; Zwaan, Magliano, & Graesser, 1995b), perceptual symbols theory (Barsalou, 1999) and embodied (or, "grounded") cognition (Barsalou, 2008; Glenberg, 1997; Lakoff, 1987; Lakoff & Johnson, 1980), and also the findings of studies that have found visual imagery to be central to the higher-level comprehension skills that support situation model construction: for example, by demonstrating that reducing a reader's ability to engage in visuospatial imagery disrupts their ability to maintain global coherence (Fincher-Kiefer, 2001) and generate knowledge-based inferences (Fincher-Kiefer & D'Agostino, 2004).

This lack of relationship may therefore be due to how visual imagery was measured in the current study. In order to account for the multidimensionality of the visual imagery system, the current study provided a clearer differentiation of specific imagery processes. In contrast, most previous studies have used a single measure of visualisation, or focused on the concept of VSWM, and therefore used more general measures of spatial span, or span capacity with additional processing abilities (i.e., transformation or integration) within the visuospatial sketchpad. However, two of the imagery measures in the current study (image maintenance and image scanning) did not relate to reading comprehension significantly, or in the expected direction. The most likely interpretation here is that these tasks require a lesser degree of coordination, transformation, and inhibition compared to mental rotation, thus are less aligned with what takes place during comprehension. For example, during reading, a reader must both maintain attention and inhibit distractor information (Borella & de Ribaupierre, 2014; Cain, 2006; Pike et al., 2010) and transform and update representations based on

newly encountered information (Albrecht & O'Brien, 1993; Morrow et al., 1989; O'Brien et al., 1998). Thus, it may be that similar to reading comprehension, mental rotation also hinges on these executive processes, whereas simple image maintenance or scanning do not.

Although, while mental rotation appears to have some similarities with the metacognitive processes of reading comprehension, and it is plausible that in both these cases imagery is depictive and conscious, there are nevertheless arguable differences in the dynamic imagery of narrative scenes that takes place during comprehension and the imagery required for the purposeful manipulation of a single object. Specifically, transportation theory (Green & Brock, 2000; 2002) suggests that reading engagement is signified by a reader's experience of becoming immersed in a story and thus "transported" into a narrative world. Transportation is conceptualised as a distinct mental process that can be considered a guided form of mental simulation, vital to which is the ability to evoke visual imagery of the scene depicted (Green & Brock, 2002; Green & Donahue, 2008).

Further, in a model of narrative engagement and comprehension, Busselle and Bilandzic (2008) explain that transportation is realised by a reader via the construction of a situation model of the narrative world, and performing a "deictic shift" to centre their experience not in their current location but into this story world. This shift is motivated by deictic adverbs commonly found in narratives, such as *here*, *now*, and *today*, as these adverbs only make sense from the deictic centre of the story (Busselle & Bilandzic, 2008). Additionally, it has been argued that this deictic shift not only provides narrative engagement, but is also necessary for narrative *comprehension*, as it enables a reader to

view the story events and actions from the point of view of the protagonist and thus the centre of the story's meaning (Busselle & Bilandzic, 2008).

Indeed, imagining oneself within the story world is at the centre of the original theories that describe how meaning is acquired through the situation models that are constructed during reading comprehension (Zwaan, 1999a; 1999b), and is positioned with theories of embodied cognition which propose that, in order to understand another's behaviour, one must be able to simulate it (Fischer & Zwaan, 2008). Thus, this phenomenological experience that occurs during narrative comprehension is not entirely comparable to the visual imagery that is required for the mechanical manipulation of a single object, as in the mental rotation task. Future research is needed to compare comprehension level to visual imagery as it occurs during narrative comprehension; for example, by measuring dynamic visual imagery that is updated based on story information. This would therefore extend measurement of updating processes in reading comprehension from the verbal to the visual domain.

It is also surprising that verbal working memory was not predictive of reading comprehension scores, on either the Neale or the DARC. In fact, this is inconsistent with an extensive amount of previous literature (for a review, see Chapter 1.3.5.1). A possible explanation here is that, because fluid intelligence and working memory have been found to be separable, but highly related constructs (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Kane, & Engle, 2003; Engle & Kane, 2004; Engle, Tuholski, Laughlin, & Conway, 1999; Kane & Engle, 2002; Kane, Conway, Hambrick, & Engle, 2007; Kane, Hambrick, & Conway, 2005) Raven's accounted for all the reliable working memory variance in both regression models.

Thus, the correlations between comprehension and the working memory tasks will be interpreted in more detail. It was found that only forward, but not backward, digit span significantly correlated with the Neale, whereas the opposite pattern of correlations was found for the DARC. In comparison to forward span, the backward span task is considered to be a complex working memory task, as it requires transformation of information in conjunction with maintenance, in order to produce the output in a different format to which it was memorised (i.e., recalling digits in a reverse order; Conway, Kane, & Bunting, 2005). Thus, one possible interpretation of these findings is that performance on the Neale is more dependent on simple verbal working memory processes such as recalling maintained information, whereas, the DARC requires more complex working memory processes such as integration and updating of information.

To elaborate, despite some findings that simple storage tasks are not distinct from complex tasks in their prediction of comprehension (de Jonge & de Jong, 1996; Goff et al., 2005; Stothard & Hulme, 1992), it is often recognised that in relation to verbal processing, complex working memory tasks show more evidence of being related to comprehension than simple span tasks (i.e., see Daneman & Merikle, 1996, for a meta-analysis). It has consequently been suggested that central executive, rather than phonological loop, functions are more important for reading comprehension (Chrysochoou et al., 2011). This interpretation again makes sense when considering that successful comprehension depends not only the ability to maintain, but also the ability to update and integrate information in the mental representation as further information is encountered (Kintsch, 1988; Zwaan & Radvansky, 1998; Zwaan, Langston, & Graesser, 1995a).

Therefore, whether working memory tasks show a relationship with comprehension likely depends on how comprehension is defined and measured. For example, working memory measures that heavily involve executive processes such as manipulation and integration of verbal information may only appear as significant contributors when reading is assessed via higher-level comprehension processes such as inferencing. Comparatively, simple span or maintenance tasks would be stronger predictors when comprehension is measured offline via questions that simply require the retrieval of textbase information.

In support of this, working memory tasks which require updating processes have been shown to be particularly predictive of inferential, rather than literal, comprehension abilities (Potocki, Ecalle, & Magnan, 2013). Specifically, updating refers to the process of modifying the content of working memory to accommodate new input. As this requires the dynamic manipulation of working memory content, updating can broadly be considered an executive function (Carretti, Cornoldi, De Beni, & Romanò, 2005). In relation to comprehension, updating is relevant as it occurs when a reader compares and integrates incoming information with previous information and existing knowledge while the text is being processed (Carretti et al., 2005). Congruent with this, updating abilities have been found to be highly predictive of both reading comprehension (Carretti et al., 2005), and listening comprehension (Potocki et al., 2013).

Similarly, Chrysochoou et al. (2011) propose that the metacognitive processes that occur online during comprehension are similar to those that occur during the completion of complex working memory tasks that tap into central executive functions, such as updating; for example, co-ordination of storage and processing, strategy selection and operation, and the activation and manipulation of information in long-term memory.

Conversely, literal comprehension is considered to be more dependent on surface-level processing, such as the maintenance and integration of text-based information.

Hence, the finding that the complex span task correlated with DARC scores to a greater extent than with the Neale scores potentially highlights a disparity between these measures in the proportion of questions that simply assess one's ability to remember or combine literal information from the textbase, versus answer questions that require knowledge-based inferencing. In addition, these arguments may also help explain why Raven's predicted more variance on the DARC outcomes than the Neale. In conjunction with assessing reasoning and visuospatial skills, performance on Raven's is theorised to be more dependent on the central executive component of working memory, than simple storage facilities such as the phonological loop or visuospatial sketchpad (Carpenter et al., 1990).

The main overall implication of the findings of the current study, which supports those of previous research (i.e., Cutting & Scarborough, 2006; Keenan et al., 2008), is that different measures of reading comprehension tap different component skills to varying degrees. For example, there was a great divergence between the two measures with regards to the contribution of word-reading skills such as accuracy and reading rate. While recognising the influence of task-specific variance on the Neale, this highlights the impact that decoding skills can have on the measurement of comprehension when using a measure of comprehension that simultaneously measures lower-level reading skills.

Indeed, in addition to differences in the number of items that tap into higher-level cognitive processes such as inferencing, decoding requirements are often cited as a key cause of the disparities between comprehension measures. Furthermore, extraneous

variance that arises from different response formats of these tests may also play a role. For example, in comparison to the DARC, which is forced-choice, and therefore simply requires a statement of either true or false, the Neale requires open-ended responses, which some have argued are also dependent on skills external to reading comprehension such as expressive language (Spooner et al., 2004).

Consequently, it has been suggested that open-ended questions make heavy *output* demands, which is in conflict with the notion of comprehension being an *input* process that culminates in the construction of a situation model of the information presented (Spooner et al., 2004). Thus, output demands of open-ended questions confound measurement of this input process and should be minimised by using force-choice answers (Spooner et al., 2004). However, contrary to this argument is the point that multiple-choice questions can be problematic due to enabling performance at above chance levels based on guessing, recognition, or prior knowledge alone (S. Katz, Blackburn, & Lautenschlager, 1991; Keenan & Betjemann, 2006). Further, the picture becomes even more complex when considering additional response formats and other confounding variables related to test performance. For example, it has been suggested that 'cloze' tasks (i.e., sentence completion through filling in the missing word) may be more dependent on word-reading ability than open-ended questions (Nation & Snowling, 1997; Spear-Swerling, 2004). In addition, variations in passage-length, passage type (i.e., narrative versus expository) and definitions of comprehension, have also been identified as variables that may affect performance (Spear-Swerling, 2004). Furthermore, the differences in the skills that these tests measure may also be a function of developmental level (Keenan et al., 2008).

These issues further compound and exacerbate the fact that the questions used across different comprehension measures differ in their requirements for the reader to integrate and monitor information across sentences or with background knowledge, and highlight the importance of selecting and developing measures carefully and thoughtfully, depending on what underlying skills are to be assessed, or questions are being asked. Clearly, more work is also needed to develop measures that enable interpretations about comprehension performance to be made based on strengths and weaknesses of specific comprehension processes, rather than additional skills that are necessary for constructing responses.

Limitations and Future Directions

Due to practical considerations, the current study did not include measurement of other reading and cognitive skills previously found to be related to comprehension. Other foundational lower-level skills may have included vocabulary (word knowledge) and grammatical skills, including syntactic and morphological knowledge, and higher-level skills such as inference-generation. Indeed, overall, each model did not account for a large amount of variance in comprehension: the total amount of variance accounted for when predicting Neale scores was 31%, whereas when predicting DARC scores this was 16%. However, as the DARC is less reliable than the Neale, it is inherently less predictable (i.e., when taking into consideration that error variance cannot be adequately explained). Thus, consideration of prediction of only the reliable variance of these measures should be given.

Specifically, error variance of the Neale comprehension subscale is estimated to be approximately 9% (averaged across Year 4 and 5 students). Thus, the regression model in the current study accounted for 31% of the total variance of Neale scores, but 34% of

the reliable variance of this measure. Similarly, when taking into account the reliable variance of the DARC (which is estimated to be around 75%) the regression model in the current study accounted for 21% of this reliable variance. These percentages can be considered substantial. However, it is acknowledged that the inclusion of other higher-level skills such as inference generation could have accounted for a significant part of the variance in each model, particularly the DARC outcomes.

In addition, emerging research is establishing a role for executive processes such as attention and inhibition resources in reading comprehension and situation model construction (Borella & de Ribaupierre, 2014; Pike et al., 2010; see also Kendeou, 2014, for a review), which were also not examined in the current study. Although, it is possible that in the current study, fluid intelligence as measured by Raven's would have captured any relevant variance relating to these constructs. Future research could, however, extend these findings by including separate measures of lower and higher-level comprehension skills, to allow for a more in-depth examination of how these skills contribute to comprehension scores on the DARC, in comparison to other measures of reading comprehension, and other skills such as visual imagery.

Although few measures currently exist that can be administered easily and quickly to assess knowledge-based integration, this could be developed in the future to further test the model of the current study, by constructing a set of open-ended or forced-choice questions that require the reader to integrate their own knowledge with the information found in the text. However, care would need to be taken to ensure that incorrect answers were due to a failure to integrate this information, rather than the reader not having access to this knowledge in the first place. The DARC is able to somewhat control for this, as it minimises the need for extensive background knowledge, by limiting topics to

very familiar ones (e.g., pets) and introducing imaginary objects (e.g., "snerp") to represent completely novel relationships between concepts (e.g., a snerp is like a turtle, but slower than a turtle). True/false statements that require knowledge integration are thus based on these familiar objects and novel relationships (e.g., "a snerp has a shell"). In addition, the DARC includes a knowledge access subscale, which includes questions that ensure readers do in fact have knowledge about these familiar objects (e.g., "a turtle can live in water").

However, the DARC is also subject to its own limitations. Firstly, as this measure only uses a true/false response format, it may be more susceptible to a higher chance rate due to guessing than multiple-choice or open-ended answers (van Blerkom, 2009). Further, although the DARC assesses knowledge integration, Carlson et al. (2014a) point out it does not identify whether the reader builds a *coherent* representation of the text's content. Indeed, this measure focuses heavily on an individual's inferencing ability when it comes to assessing higher-level skills, and it is well established that the coherent representation that underlies text comprehension is not simply a collection of inferences. Future research could also assess visual imagery in relation to other situation modelling skills such as coherence monitoring. In addition, this measure has not been used extensively in research. As such, there is little information available regarding its utility or comparison with other measures. Thus, the current study provides an important contribution by increasing knowledge in this regard.

Additionally, it may be argued that the mental rotation task used in the current study assesses spatial *ability*, but not visual imagery per se (i.e., Burton & Fogarty, 2003). Similarly, some authors have argued that mental rotation hinges more on visuospatial working memory than visual imagery, as the stimuli for this task are not necessarily

encoded in depictive format (e.g., Quinn, 2008). However, as explored in the previous chapter (pp. 97-98), positive correlations have been found between performance on spatial ability tasks and self-reported imagery (Barratt, 1953), and spatial ability tasks and objective measures of the visual imagery components proposed by Kosslyn (Poltrock & Brown, 1984). Thus there is support for the hypothesis that spatial ability reflects the operation of imagery processes such as high quality image maintenance, inspection and transformation. Further, there is a considerable amount of both behavioural and neurological evidence that substantiates the notion that mental rotation requires at least some degree of depictive representation of the maintained visual information (see Chapter 2, pp. 97-101 for a review of these findings).

Despite this, it is acknowledged that some children have been found to rely on strategies that do not involve the use of visual imagery when completing tasks of mental rotation (Quaiser-Pohl et al., 2010), and hence do not show evidence of a linear relationship between the degree of rotation and time taken to solve each item (Waber, Carlson, & Mann, 1982). As the current study did not impose a time limit for completing the MRT, it is possible that children may have tried several different strategies before making their response, or even relied on slower but still successful strategies such as a piecemeal strategy, or even a verbal-analytical approach. Nevertheless, non-rotation strategies generally result in poorer performance than imagery-based rotation strategies (Quaiser-Pohl et al., 2010) and, as with Study 1, a floor effect was not evident for scores on the MRT in the current study. However, this longer time period may still have reduced a reliance on visual imagery for solving items, possibly explaining why no relationship was found between this measure and the DARC.

As noted in Chapter 2, due to the internal nature of visual imagery, it is a difficult construct to measure. Subjective measures of visual imagery have proven to be unreliable as individuals are often unaware of, or unable to adequately describe, these processes in order to provide these introspective reports (see McAvinue & Robertson, 2006, for a review). Yet, few objective measures of visual imagery exist. Future work is thus needed to develop more appropriate measures of imagery, especially for use with younger populations.

Collectively, the major theoretical and practical implications of this study pertain to the use of existing tests of reading comprehension, and how visual imagery is conceptualised and measured. Although many authors note that the psychometric properties of most standardised tests of reading comprehension are more than adequate, the validity of these measures is questionable when it comes to identifying which particular skills deficits are leading to poorer comprehension outcomes, and which groups of children are at risk. The current study adds to the literature that finds not all standardised tests are interchangeable in regards to the underlying skills that they measure, and thus careful consideration should be given when choosing measures of comprehension in research and practice. Additionally, the current study provides much needed information about how one newer measure of comprehension (the DARC) compares to a traditional standardised measure (the Neale). Although advances have recently been made in the development of multicomponent measures that aim not only to identify poor comprehenders, but also provide information about where specific skill deficits lie, these have yet to be used extensively in research and each is subject to its own limitations (see S. E. Carlson, Seipel, & McMaster, 2014a, for a review). Future research is needed to further investigate the utility of these tests and the variables to which they are sensitive.

With regards to the measurement of visual imagery, this study has provided further evidence that visual imagery is not a singular construct, and that different subcomponents of visual imagery may be differentially related to other skills, including reading comprehension. However, future research is needed to strengthen these findings by measuring the dynamic and pictorial imagery that occurs during narrative comprehension. This could be achieved in a number of ways. For example, examining the eye-movements of good and poor comprehenders during narrative listening comprehension, to determine whether both good and poor comprehenders eye-movements are consistent with the actions portrayed in the narrative; by inhibiting the use of narrative imagery during reading and examining the resulting effects on comprehension; or, by comparing good and poor comprehenders on a perceptual symbols task, for example, those developed by Zwaan and colleagues (Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012).

In conclusion, although visual imagery was not found to be a reliable predictor of reading comprehension, the findings of the current study contribute to previous research by demonstrating that not all measures of comprehension are interchangeable with regards to the underlying skills that they measure. In addition, the current study found that not all types of imagery are equally predictive of reading comprehension. Future research could extend on these findings by exploring additional measures of higher-level reading skills and visual imagery, including measures of imagery that assess this process as it occurs *during* reading. As Rowe et al. (2006) noted almost a decade ago "it is time to move away from an undifferentiated, a-theoretical approach of measuring reading ability" (p. 627); this is yet to be fully realised. The findings of the current study lend to the aim of improving reading comprehension measurement in research and practice through systematic investigations of the variations among reading

comprehension measures, both old and new, and are instrumental in the effort to uncover all of the skills that relate to reading comprehension and how they interact with one another.

Prelude to Study 3

The findings of Study 2 suggest that visual imagery is not a unique contributor to reading comprehension, even when comprehension is measured via higher-level processes such as inference generation. However, it was identified that a potential reason for this lack or relationship could be the use of a mental rotation task to measure visual imagery. Specifically, this task may also be reliant on spatial ability, working memory (i.e., central executive), or other non-imagery related processes. Furthermore, assuming imagery was taking place during this task, this type of imagery is likely quite different from that which is activated during narrative comprehension. Specifically, the imagery that takes place during narrative comprehension is hypothesised not only to be spatial and dynamic, as suggested by the event-indexing model (Zwaan, Langston, & Graesser, 1995a; Zwaan, Magliano, & Graesser, 1995b) but, from the view of embodied cognition and theories of transportation (see pp. 145-146), this imagery is also proposed to be extensive and pictorial, an experience similar to that of visually perceiving an actual scene.

Thus, potential ways to examine the dynamic and pictorial imagery that occurs during narrative comprehension were identified. These included: examining the eye-movements of good and poor comprehenders during narrative listening comprehension; comparing good and poor comprehenders on a perceptual symbols task (i.e., Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012); or inhibiting the use of narrative imagery during reading and examining the resulting effects on comprehension. The third study in this thesis focuses on the latter methodology.

Specifically, if depictive imagery is central to situation model construction, preventing its use during reading should impair situation model related processes such as

knowledge-integration. Indeed, this is what Fincher-Kiefer and D'Agostino (2004) appear to have found, when participants who were required to maintain an unrelated visual image no longer showed evidence of drawing predictive inferences from text. Thus, based on dual-logic theory (Baddeley, 1992), which proposes that a disruption of a cognitive process will occur if the resources it requires are simultaneously employed by a separate task, it was concluded that the failure to draw predictive inferences was due to the perceptual resources required for situation model construction being utilised in the visual memory task.

Indeed, theories of visual imagery, such as that presented by Kosslyn and colleagues (Kosslyn, 1980; 1983; 1994; Kosslyn et al., 1984) argue that all conscious imagery takes place within an equivalent medium (i.e., a visual "buffer"). Thus, if this medium is already being utilised for one visual task (i.e., maintenance of an image), then it will be significantly less efficient at simultaneously generating additional imagery, including the imagery that is hypothesised to occur during reading (note however, processing imagery from two conflicting sources is not to be confused with the process that occurs when one adds to, or manipulates, an already generated and thus previously existing image).

However, although there is evidence that inhibiting the use of visual imagery during reading can disrupt situation model processes such as knowledge-based inference generation, this research has not been extended to include explicit measurement of comprehension level. Specifically, as it has been proposed that knowledge-based inference generation is required for complete comprehension of a text, good comprehenders should perform better than poor comprehenders on a task that assesses the activation of these types of inferences. However, the introduction of a visual load

should impair good comprehenders ability to draw these inferences, thus reducing their performance to a level similar to that of the poor comprehenders. This would demonstrate that visuospatial information is not only required to draw inferences, but it is likely a contributing factor to differences in comprehension ability. Conversely, this disruption should not occur when good comprehenders are required to hold a verbal load, as knowledge-based inferencing takes place at the level of the situation model, which is hypothesised to be less reliant on verbal processing (i.e., in comparison to the construction of a textbase representation).

Thus, an additional study was conducted, using an experimental design based on the above premises, in order to determine whether a causal relationship between visual imagery and higher-level comprehension exists. In particular, the aim of Study 3 was to investigate whether good and poor comprehenders are differentiated specifically by their use of visual imagery when making predictive inferences, thus providing further evidence that comprehension is related to the construction of an imagery-rich situation model.

Chapter 4. Study 3

4.1 The Involvement of Visuospatial Imagery in Children's Predictive Inference Generation and Reading Comprehension

As highlighted in Chapter 3, reading intervention and assessment research has historically focused on lower-level reading skills such as decoding, fluency and vocabulary, while overlooking the higher-level cognitive abilities that lead to successful reading comprehension. However, a renewed interest in a multi-component approach to reading research has identified the importance of several higher-level skills such as inference generation and coherence monitoring, in overall reading comprehension. Indeed, the current thesis has built on the premise that reading comprehension is no longer seen as a single construct but rather the consequence of several interacting processes that result in a coherent mental representation of the situation described by a text, often referred to as a "mental model" (Johnson-Laird, 1983), or "situation model" (van Dijk & Kintsch, 1983) throughout the cognitive psychology literature.

As noted, situation models were first proposed to explain how comprehension goes beyond word-level processing (i.e., the construction of a textbase representation), and involves the processing of higher-level semantic and pragmatic information that contributes to a coherent representation of the *meaning* conveyed in a text (van Dijk & Kintsch, 1983). Specifically, it has been proposed that situation models are formed and continually updated via integrative processes that allow a reader to combine information found in the textbase with implicit information stored in long-term memory, such as a reader's background knowledge (i.e., knowledge gained from previous experiences and previous textbases; Kintsch, 1988; van Dijk & Kintsch, 1983). Thus, situation models

contain extensive information that goes beyond that described in a text, and is connected along several dimensions, including: space, time, protagonist, causation and intentionality (Zwaan, Langston, & Graesser, 1995a; Zwaan, Magliano, & Graesser, 1995b).

It is not surprising then, that variability in the quality of children's situation model constructions appears to be a predictor of reading comprehension (Barnes et al., 2014; Nyhout & O'Neill, 2013; van der Schoot et al., 2009; 2010; 2011 see Chapter 1.3.4). Indeed, as outlined earlier (see Chapter 1.3.2.2) the process of constructing a complete and rich situation model of a text's meaning depends on a reader's ability to go beyond explicitly stated information, and draw knowledge-based inferences by combining information found in the textbase with background knowledge stored in long-term memory. Consequently, this component of situation modelling has been largely investigated in relation to children's narrative comprehension and several studies have supported this relationship (Cain et al., 2001; Cain & Oakhill, 1999; Elbro & Buch-Iversen, 2013; Oakhill, 1984; Tompkins et al., 2013), whether a story is to be comprehended via oral, picture, or text presentation (Kendeou et al., 2008). Such research has thus led to suggestions that, although poor comprehenders build adequate textbase representations to maintain local coherence, they do not build as rich and elaborate situation models as good comprehenders and, therefore, tend to rely on textbase representations to obtain meaning.

The importance of knowledge-based inference generation in reading comprehension also appears to endure across ages, with evidence of this relationship in children as young as 4 years old (Kendeou et al., 2008; Tompkins et al., 2013) through to older children (Cain et al., 2004a) and adults (Mellard, Fall, & Woods, 2010). Further, this

relationship is evidently causal; longitudinally, inference making skills in earlier years have been found to uniquely contribute to narrative comprehension at a later age (Kendeou et al., 2008; Lepola et al., 2012; Oakhill & Cain, 2012; Silva & Cain, 2015), and intervention studies show that instruction aimed at improving the spontaneous generation of children's inferences through engagement with background knowledge can improve reading comprehension (Bos, De Koning, Wassenburg, & van der Schoot, 2016; Elbro & Buch-Iversen, 2013; A. H. Paris & Paris, 2007; Yuill & Oakhill, 1988).

It has thus been suggested that knowledge-based inferences enhance comprehension by adding extended information about several narrative features to a representation of a text's meaning (Graesser et al., 1994). Indeed, several studies that show readers make online inferences in regards to a number of narrative features, including: the causes and consequences of events (Kuperberg et al., 2011); expectations about future events (Fincher-Kiefer, 1993); properties of objects (Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan & Pecher, 2012); spatial relationships among entities (Rinck et al., 1996; Tversky, 1993); and the characteristics of protagonists, such as their knowledge and beliefs, traits and emotions (Gernsbacher et al., 1992; Oakhill, Garnham, & Reynolds, 2005a), as well as the goals and plans that motivate their actions (Graesser et al., 1994; D. L. Long & Golding, 1993).

However, a notion that has not yet been explored in the current thesis is that certain types of knowledge-based inferences may be more relevant to narrative comprehension than others. For example, Tompkins et al. (2013) found that, although story comprehension was related to the total number of inferences preschool children made online (e.g., during self-narration of a picture book), when analysing each inference type separately, only three types of inferences were found to be predictors of

comprehension: inferences about characters goals, inferences about actions that achieved character goals, and inferences about character states. These three inference types remained predictors of story comprehension even after controlling for age and expressive vocabulary, although the authors note that children made these types of inferences relatively infrequently.

Similarly, Kendeou et al. (2008) found that, overall, children who made more inferences had a significantly greater level of story comprehension compared to those who made fewer inferences, yet inferences about goals, actions and causal antecedents were the greatest contributors to reading comprehension. It is thought that these types of inferences are central to comprehension because they explain characters' actions throughout the story and advance the story's causal sequence (Kendeou et al., 2008; Lynch & van den Broek, 2007; Tompkins et al., 2013). Indeed, based on the causal-network model of comprehension (Trabasso et al., 1989; Trabasso & Sperry, 1985), several authors have argued that understanding the causal structure of a narrative is central to reading comprehension, and therefore causal inferences are one of the most important types of inference for aiding situation model construction, if not *the* most critical (S. E. Carlson, Seipel, & McMaster, 2014a; Trabasso & Suh, 1993; van Kleeck, 2008).

Causal inferences are defined as those that connect the causes and consequences of an event depicted in a narrative and can relate to several story dimensions, including initiating an event, action or problem; potential solutions to problems; consequences of events and actions; resulting internal/emotional responses of a character; and desires and goals of characters (Graesser, Bertus, & Magliano, 1995). It is hypothesised that understanding the causal structure of a narrative allows children to obtain a greater

meaning of the "how and why" of the events described (Kendeou et al., 2005).

Inevitability, this understanding will be deeper if children can go beyond what is

mentioned in a text, and infer these causal connections, as often they are not explicitly

stated (Kuperberg et al., 2011).

Several studies have supported the proposition that understanding the causal structure of

a narrative is important for both comprehension and memory of a text (Trabasso et al.,

1989; Trabasso & Sperry, 1985). For example, van den Broek, Lorch, and Thurlow

(1996) found that recall of narrative events by young children is much greater when a

narrative has more causal connections. Similarly, Lynch et al. (2008) found 4- and 6-

year-olds' correct recall of aural or televised narratives, as well as answers to

comprehension questions, was related to their sensitivity to the causal structure of the

narratives. As such, several researchers have found that children with adequate low-

level reading skills, but poor comprehension skills (i.e., poor comprehenders) often fail

to make causal inferences while reading (Cain & Oakhill, 1999; 2006; McMaster et al.,

2012). In line with this, training aimed at improving causal inferencing skills has been

found to result in an improvement of children's general reading comprehension ability

(Bos et al., 2016).

Comprehension failure may, therefore, result from insufficient activation of appropriate

background knowledge to form necessary inferences (Cain et al., 2001; Elbro & Buch-

Iversen, 2013; Recht & Leslie, 1988). Specifically, this may be more likely when a

reader is unable to draw inferences about the causal sequence of events in a narrative in

order to update their situation model. This may occur when the reader does not possess

the required knowledge to form the inferences necessary to comprehend the text (Recht

& Leslie, 1988), has incorrect background knowledge which leads to erroneous

inferences (Kendeou & van den Broek, 2007), or possesses the necessary background

information but does not use it to form inferences (Cain et al., 2001; Oakhill & Cain,

2007).

However, as outlined in Chapter 1, another possible source of disruption to this class of

inferences has been identified by Fincher-Kiefer and D'Agostino (2004), who presented

a group of adult participants with short texts designed to elicit either a predictive

inference (experimental condition) or no inference (control condition), under one of two

between-group conditions: (i) while holding either a visuospatial memory load (an array

of five dots within a 4 x 4 grid), or (ii) a verbal memory load (a string of six letters). It

was found that participants given a verbal memory load showed the typical facilitation

in reaction time to predicted inference targets; however, participants given a

visuospatial memory load displayed a reduced facilitation effect. Based on dual-logic

theory (Baddeley, 1992), which proposes that a disruption of a cognitive process will

occur if the resources it requires are simultaneously employed by a separate task, it was

concluded that the failure to draw predictive inferences was due to the perceptual

resources required for this being utilised in the visual memory task. In addition,

Fincher-Kiefer and D'Agostino (2004) found that a concurrent visuospatial load did not

disrupt inferencing when the experimental passages were designed to elicit *bridging*

inferences, rather than predictive inferences. Bridging, or textbase inferences, differ

from knowledge-based inferences as they are used to maintain local coherence of a

narrative at the textbase level (i.e., to make links between premises in a text), rather than

requiring the integration of background knowledge (Fincher-Kiefer & D'Agostino,

2004). For example, when reading the passage "The man was eating his soup when the

train screeched to a halt. He jumped up and wiped of his pants" a reader may make a

bridging inference that "the soup spilled" in order to maintain coherence between the

two sentences. In contrast, in the absence of the second sentence, a reader may still infer that the soup had spilled by drawing on background knowledge of similarly encountered scenarios.

It thus appears that predictive inferences were disrupted, because they occur at the level of the situation model, which requires visuospatial resources for complete construction (Fincher-Kiefer & D'Agostino, 2004). These findings align with several theories that propose a role for visual imagery in reading comprehension (i.e., dual coding theory; Pavio, 1971, and embodied or "grounded" cognition; Barsalou, 2008; Glenberg, 1997; Johnson, 1987; Lakoff, 1987; Lakoff & Johnson, 1980), and the conceptualisation of situation models as a perceptual simulation of the events described in a text (Zwaan, 1999a; 1999b). Also in accordance with this, various studies have supported the notion that visual imagery is activated at the level of the situation model to represent, not only explicitly stated, but *inferred* information. For example, evidence from perceptual-mismatch studies shows that both adult and child readers simulate several implied visual features of the objects described in texts, even when these features are not mentioned explicitly in the textbase (Connell, 2007; Dijkstra et al., 2004; Engelen et al., 2011; Stanfield & Zwaan, 2001; Zwaan et al., 2002; 2004; Zwaan & Pecher, 2012; for further description of these studies, refer to Chapter 1.3.3, pp. 30-32). Similarly, the availability of perceptual information has been found to change as a function of the narrative, even when perceptual availability is implied, rather than explicitly stated (Horton & Rapp, 2003; see Chapter 1.3.3, pp. 32-33).

Thus, as visual imagery appears to be an important component of situation model construction, there is a rationale for expecting this imagery to also be related to an individual's ability to comprehend what they are reading. Yet, although there is mounting evidence that children construct situation models to represent text meaning

(Barnes et al., 2014; O'Neill & Shultis, 2007; Pyykkönen & Järvikivi, 2012; Rall & Harris, 2000; Uttal et al., 2006; van der Schoot et al., 2011; Ziegler et al., 2005), and that these representations are dynamic (Fecica & O'Neill, 2010) and contain spatial (Barnes et al., 2014; Ziegler & Acquah, 2013) and perceptual (Engelen et al., 2011; Nyhout & O'Neill, 2013; Rall & Harris, 2000; Ziegler et al., 2005) information, studies that have directly assessed whether variability in children's overall comprehension level is related to the perceptual quality of their situation model constructions have only recently begun to emerge (see Chapter 1.3.4).

In addition, criticisms of these studies exist. For example, many of these studies use texts or additional methods that put an emphasis on spatial information and spatial relationships and therefore may encourage participants to activate a greater amount of visuospatial information during situation model construction than what might occur in other reading situations. Furthermore, although situation model instruction centred on imagining story content has been shown to result in an increase of correct answers to comprehension questions and a redistribution of resources from textbase processing to situation modelling (as indicated by slower reading times and eye-fixation on situation model versus textbase variables; van der Schoot et al., 2010), it is still unclear from these studies to which aspect of situation model construction it is that imagery-based instruction specifically contributes (i.e., inference making, updating or integration; van der Schoot et al., 2010).

Thus, more information is clearly needed to determine whether an individual's reading comprehension level is related to the utilisation of perceptual information during specific situation model processes, and under reading conditions that do not explicitly emphasise spatial information. The aim of the current study was, therefore, to determine

whether children who are good comprehenders rely more on visual imagery when constructing knowledge-based predictive inferences than children who are poor comprehenders while comprehending language that does not necessarily emphasise spatial or perceptual information.

Using an experimental design adapted from Fincher-Kiefer and D'Agostino (2004), the current study utilised a computer-based lexical decision task (word/non-word) to present children with short text passages designed to elicit either a predictive inference, or no inference. Reaction times to following target words were measured as an indication of facilitation in the inference condition compared to the no inference (control) condition. This measure of inferencing was used in order to capture the inference making process as it occurred online, as offline measures may be more dependent on additional processes such as retrieval of information from long-term memory.

A concurrent visuospatial load task was also included in the lexical decision task, in order to disrupt imagery during reading of the text passages. The robustness of the dual-task paradigm has been established in previous comprehension research (Bergen et al., 2007; Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004). Specifically, visual imagery is proposed to take place within a "visual buffer", a cognitive medium for holding short-term visual information, which makes use of the same neural resources required for actual visual perception (Kosslyn, 2005; Kosslyn, Ganis, & Thompson, 2001). Further, despite the argument that imagery processes can be considered somewhat distinct (e.g., generation, maintenance, scanning, and transformation; Kosslyn et al., 1990; 2004): the generation of any type of conscious imagery is still proposed to require the visual buffer. Thus, if this medium is already being utilised for

one visual task (i.e., maintenance of an image), then it will be significantly less efficient at simultaneously generating other types of imagery, including imagery that may be activated from textual descriptions. A verbal load task was also included to ensure any interference effects were specifically due to an overload of visual resources and not just additional task demands in general. Thus, in the current study, texts were read under three conditions: without any additional cognitive load, with an additional visual load, and with an additional verbal load.

Performance on the lexical decision task was compared between groups of good and poor comprehenders, created based on comprehension scores on the Neale (Neale, 1999). However, due to existing criticisms of the Neale (along with other standardised tests of comprehension), that comprehension scores obtained do not assess skills beyond lower-level text recall and integration processes (Francis et al., 2006; Keenan et al., 2008; Rowe et al., 2006; Spooner et al., 2004; see also Chapter 3), the DARC was again included as an additional measure of reading comprehension (which in addition to text inferencing skills, aims to also capture higher-level knowledge-based inferencing without being dependent on word-reading ability). Thus, inclusion of the Neale allowed groups to be defined based on existing age-related norms, whereas the inclusion of the DARC allowed for additional interpretations of findings regarding inferencing ability in the lexical decision task.

Consequently, it was hypothesised that because reading comprehension goes beyond the construction of a textbase representation of a narrative and requires the use of background information to draw inferences, good comprehenders would show greater facilitation to predictive inference targets than poor comprehenders when no additional task load was present for either group. Secondly, because predictive inferences take

place at the level of the situation model, it was hypothesised that when given an additional visuospatial load, good comprehenders would no longer show greater facilitation to predictive inference targets than poor comprehenders, as the visuospatial resources they require for situation model construction would no longer be available due to utilisation in the load task; however, when given an additional verbal load, good comprehenders would still show greater facilitation to predictive inference targets than poor comprehenders, thus providing an indication that it was not simply task complexity that reduced the facilitation effect in the visuospatial condition.

4.2 Method

Design

This study utilised a 2x3x2 mixed variable design to manipulate the within groups independent variables of (i) inference type (predictive or control), (ii) load type (no load, visuospatial load, or verbal load), and the between groups independent variable of (iii) comprehension group (poor or good). The dependent variable, reaction time in a lexical decision task, was used as an indication of inference generation. More specifically, reaction time was used as evidence of facilitation to target words that represent inference concepts. For example, if participants activate inference concepts during reading, correct responses to target words representing the inference should be facilitated because lexical access will be primed by inference activation. Therefore, correct response times should be shorter for target words following contexts that induce predictive inferences than for target words that follow contexts that are unlikely to elicit inferences (i.e., the control condition). Hence, the difference in RT between the control and inference condition can be used as an indication of facilitation, and thus the greater

the difference between these conditions (i.e., the facilitation effect) the stronger the inference.

Participants

Seventy-seven participants (39 female) in Grades 4 and 5 (aged 8.35 - 10.91 years) were screened for participation in this study. This age group was chosen rather than a younger sample, as the discrepancy between lower-level skills and comprehension becomes more apparent as children grow older and develop fluent decoding skills (Catts & Weismer, 2006). Consistent with previous research (Cain & Oakhill, 2006), children with word reading accuracy scores on the Neale that were six months or more below their chronological age were excluded from the sample; this resulted in 23 participants being excluded from further testing. Five participants were also excluded due to failure to comply with task instructions. Similar to previous research (i.e., Cain et al., 2001; Cain & Oakhill, 2006; Nation & Snowling, 1999), good comprehenders were then selected on the basis that they had Neale reading comprehension scores either at or above those predicted by their word reading accuracy scores. The poor comprehender group was made up of those who had normal for age accuracy, but a discrepancy of at least six months between their chronological age and their reading comprehension age, and also between their word reading accuracy and comprehension age (i.e., their comprehension was lower than predicted by their age and accuracy scores). This ensured participants all had age-appropriate word reading skills, but either good or poor comprehension.

The resulting sample consisted of 16 poor (12 female) and 16 good (7 female) comprehenders, from six primary schools of varying socio-educational advantage in Perth, Western Australia. This sample size is consistent with other studies that have

specifically recruited poor comprehenders (Cain et al., 2001; Nation & Snowling, 1997; 1998; Weekes, Hamilton, Oakhill, & Holliday, 2008), including those that have compared groups on a repeated measures lexical decision task (Nation & Snowling, 1999), and is a reflection of the estimated 10-15% of poor comprehenders that exist within the general population (Stothard & Hulme, 1995). The age range of the final sample ($N = 32$, 19 female) was 8.77 to 10.91 years ($M = 9.69$, $SD = .59$). All participants had normal or corrected-normal vision, were free from cognitive impairment or diagnosed learning impairments and spoke English as their first language.

General Procedure

Children who participated in this study did so as part of a larger research project that assessed the relationship between visual imagery and reading comprehension, and which also encompassed Study 2 of the current thesis. All sessions took place at the child's school, during school hours, in a quiet area separate from the classroom. Participants completed all tasks over five separate sessions, each lasting 30 minutes to one hour, with no longer than three weeks between testing sessions. The same order of task administration was followed for all participants. In the first session, which was completed individually, the order of tests was: the Neale, the DARC and finally two working memory tasks. The second and third sessions were completed in groups of two or three participants, but measures completed in these sessions were not included in the current study. Session two consisted of completing three imagery tasks and in session three participants completed Raven's Standard Progressive Matrices (Raven, 1958). In the fourth session participants individually completed the no load version of the lexical decision task, and in the fifth session they completed both the visuospatial and verbal load tasks.

Measures

Reading Comprehension

The measures of reading comprehension that were used in the current study were identical to those used in Study 2. Firstly, The Neale Analysis of Reading Ability (Australian third edition; Neale, 1999) was included as a standardised measure of reading rate, accuracy and comprehension, which has been widely used in both education and research. Secondly, due to the criticisms that the Neale does not provide a valid assessment of all the skills involved in comprehension and is largely reliant on lower-level abilities such as decoding (Spooner et al., 2004), the DARC (August et al., 2006) was included as an additional measure of reading comprehension that measures comprehension separately from the effects of word-reading ability.

As with Study 2, Form 1 of the Neale was used in the current study, and administered and scored as per the standardised instructions in the manual (Neale, 1999; see Chapter 3.2, pp. 125-126 for more details on this task), and the practice passage and the story "Nan's Pets" from the DARC was administered to all participants, as per the standardised instructions (see Chapter 3.2, pp. 126-127 for more details on this task).

Verbal Working Memory

The measures of verbal working memory measures that were used in Study 2 were also included in the current study. These included both the simple span task (forward digit span) and the complex span task (backward digit span; see Chapter 3.2, pp. 129-130). Again, digit span tasks were chosen rather than word or sentence span, as the latter may provide better readers with an additional advantage that is unrelated to working memory ability, but a result of additional linguistic skills (i.e., see Nation et al., 1999). Both these measures were administered using the digit span task from The Psychology

Experiment Building Language (PEBL) test battery version 0.13 (S. T. Mueller, 2013)

following the same procedure as in Study 2.

Inference Generation Task

To measure the generation of predictive inferences, a computer-based lexical decision

task was adapted from the one used by Fincher-Kiefer and D'Agostino (2004). Three

conditions of this task were used in this study: a no-load task, a visuospatial load task,

and a verbal load task.

Materials

Stimuli for the lexical decision task were presented on a Toshiba Satellite C660

notebook with the monitor set at 1280 x 720 screen resolution, 32-bit colour and 85

Hertz refresher rate using DirectRT version 2010 software (Jarvis, 2006) run on an Intel

Core i3 processor with a Windows XP operating system and 2 GB Ram. A DirectIN

(Empirisoft Corporation) 305mm x 75mm response box was connected to the laptop via

USB cable. The response box had nine buttons on it (corresponding to numbers 1-9 on

the computer keypad), however, only two were labelled and could be used to provide

responses in this study: the far left (1) button was labelled "yes" and the far right button

(9) was labelled "no".

Text stimuli consisted of three types of narrative passages: passages designed to elicit

either a predictive or a bridging inference and control passages designed not to elicit an

inference (i.e., the experimental text conditions) and filler passages. However, bridging

inferences were not investigated in the current study, therefore will not be discussed in

detail. The majority of these experimental texts have been used in prior research (e.g.,

Fincher-Kiefer, 1993; 1995; 1996; Fincher-Kiefer & D'Agostino, 2004),[2] with the

exception of three of the 30 predictive texts, 22 of the 30 control texts, and 37 of the 70

filler texts, which were designed for use in the current study in order to have enough

stimuli in each condition so that no version of a story was read more than once in a

within-groups design. Additionally, all stimuli that were obtained from Fincher-Kiefer

and colleagues were adapted to make them an appropriate reading level for the age of

the current sample, and American spelling, vocabulary, and place names were replaced

with those that would be more familiar to an Australian sample. The resulting average

Flesch-Kincaid grade level of readability was 3.8 (U.S. grade level) across all passages

(calculated based on both word length and sentence length of the passages; Kincaid,

Fishburne, Rogers, & Chissom, 1975). This grade level is equivalent to Grade 4.8 in

Western Australia; therefore passages were of an appropriate reading level for the

current sample.

All experimental text passages were three lines long, and written so that the final

sentence either elicited a predictive inference (predictive inference condition) or did not

(control condition), and were followed by a target word. Predictive inference texts were

"causal consequence inferences" (Graesser et al., 1994), because they involved events

that were immediate reactions to or consequences of an action or event described in the

final sentences of the text (see Table 4.1). These inferences require readers to go beyond

information presented in the textbase but are readily available from general knowledge;

are relevant to comprehension as they can facilitate subsequent text processing and

allow a reader to construct an understanding of a narrative that goes beyond textbase

information (Allbritton, 2004; Estevez & Calvo, 2000); and pilot work demonstrated

that readers predicted only one consequence when these texts were read (see also

---

[2] The author thanks Professor Rebecca Fincher-Kiefer for graciously providing these texts.

Fincher-Kiefer, 1993; Fincher-Kiefer & D'Agostino, 2004). Control passages were designed not to elicit an inference about the consequence of the action or event described in the last line of the text.

The target word presented after each predictive inference text represented the inference elicited by the experimental sentence. In contrast, control target words were not related to any inference that could have been made earlier. Care was also taken to ensure that both narrative and control texts had a similar number of words that were semantically related to the target word (see Keenan, Golding, Potts, Jennings, & Aman, 1990). Thus, the only difference between target words presented in the inference compared to control condition was that they were related to the most likely consequent of the event described in the text. No target word was used more than once. In addition to the experimental texts, filler texts were constructed so that their sentences did not elicit any inferences. The target presented after the filler text was always a non-word. These non-words were orthographically and phonemically legal and equated on word length (number of letters) with the word targets. Example passages and target words for each condition can be found in Table 4.1.

Table 4.1

Example of the Narrative Texts and Target Words Used in the Lexical-Decision Task

Condition	Text	Target Word
Predictive	The salesman was sitting in the dining car of a train. The waitress brought him a bowl of soup. Suddenly, the train screeched to a stop.	SPILL
Control	James loved playing baseball. When it was his turn, he swung the bat hard. The ball flew over the fence and way out of the park.	CATCH
Filler	The lifeguard spotted the dangerous shark fin. Before he could act, the swimmers started yelling. Soon there was panic on the beach.	BOLZE

Texts were randomly distributed so that the no load task consisted of 10 predictive inference texts, 10 bridging inference texts, 10 control texts, and 30 filler texts; the visuospatial and verbal load tasks each consisted of 10 predictive inference texts, 10 control texts, and 20 filler texts.

Visuospatial Load Task

The visuospatial memory load manipulation was constructed based on those used in previous dual-task experiments (Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004; Kruley et al., 1994; Sims & Hegarty, 1997). In this task, each text was preceded by a 4 x 4 grid (220 x 220mm/600 x 600 pixels), with an array of five solid black dots placed within separate squares to form 40 unique patterns, with the constraint that the dots could not fall in a recognisable shape (such as a letter; see Figure 4.1). Following presentation of the target word, another array of dots was displayed. Half of these were the *same* as the pattern that preceded the narrative, and the other half were *different* from the one which preceded the narrative (the pattern was changed so that one dot was

relocated to an adjacent grid cell). All participants received identical pattern and story pairings.

Verbal load task

The verbal memory load manipulation is also similar to that used in previous research (Fincher-Kiefer, 2001; Fincher-Kiefer & D'Agostino, 2004; Sims & Hegarty, 1997), and has been shown to be equal in difficulty to the visuospatial memory load task (Sims & Hegarty, 1997). In this task, each text was preceded by a string of five letters, all consonants, presented in a horizontal line in the centre of the computer screen. A string of five letters was presented rather than the string of six used in previous research with adults (i.e., Fincher-Kiefer & D'Agostino, 2004) to ensure the task was not too difficult for the age group of the current study, as research indicates that the average memory span of children is closer to five than the adult average of seven (Chi, 1976; Dempster, 1978). Following each target word a single letter was presented in the centre of the screen: half of these letters had appeared in the initial letter string and half had not. All participants received the same letter string and story pairings.
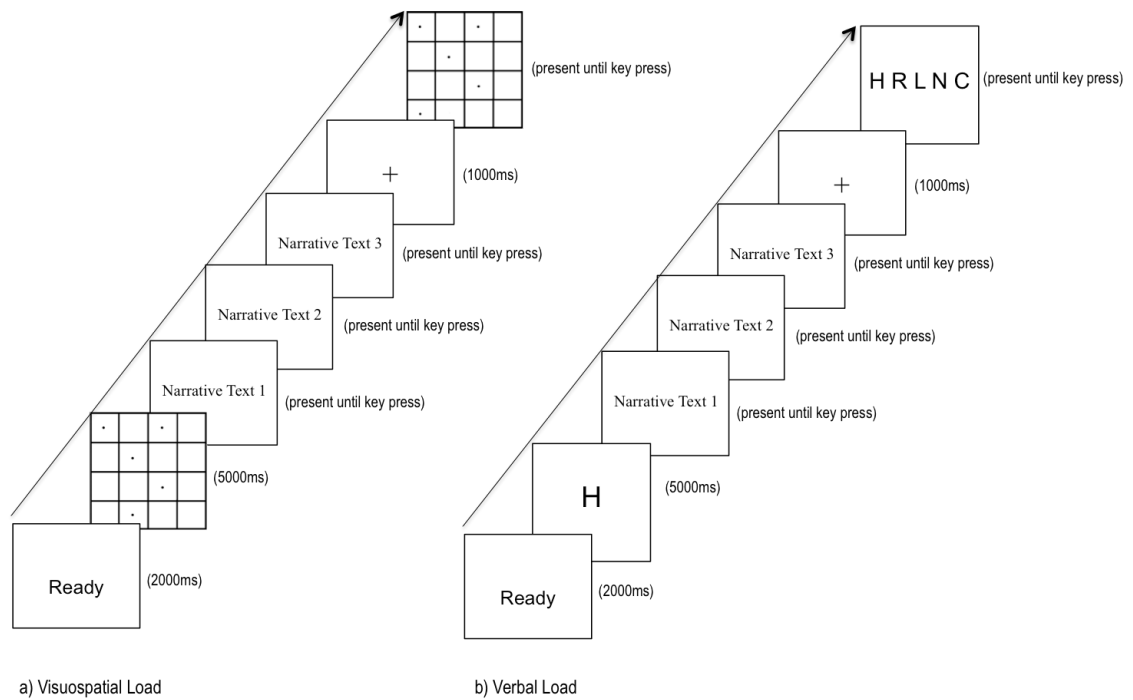
(present until key press)

(1000ms)

Narrative Text 3 (present until key press)

Narrative Text 2 (present until key press)

Narrative Text 1 (present until key press)

(5000ms)

Ready (2000ms)

a) Visuospatial Load

HRLNC (present until key press)

(1000ms)

Narrative Text 3 (present until key press)

Narrative Text 2 (present until key press)

Narrative Text 1 (present until key press)

H (5000ms)

Ready (2000ms)

b) Verbal Load

Figure 4.1. Example of the stimuli presentation and timing in the visuospatial load and verbal load tasks.

Procedure

Instructions for the lexical decision task were given to participants verbally, and examples of yes and no trials for the lexical decisions were shown to participants using a printed version of stimuli not utilised in the actual task. Prior to the load tasks, participants were also shown example trials of yes and no responses for the dot array/letter string decisions. Participants were encouraged to respond to target words as quickly as possible while still being accurate. During the task, each trial began with the word "Ready" presented in the centre of the laptop screen in black 20-point Arial Black font for 2000ms. In the no load task, this signal immediately preceded the first sentence of the narrative text, whereas in the load tasks, the "Ready" signal was immediately followed by either the dot array, presented for 5000ms, or the letter string, presented for 5000ms, before the first sentence of the narrative appeared. Each narrative sentence appeared centrally in black 24-point Times New Roman font. A self-timed reading

procedure was used to account for the variability in reading rates found in children:

therefore, once the participant had read each sentence, they pressed any of the centre

buttons (2 through to 8) on the response box to receive the next sentence, until they had

read all three sentences of the narrative.

Upon reading the third sentence of the text, participants again pressed any of the

designated response buttons. The screen then cleared and a central fixation cross was

presented on the screen for 1000ms, in order to prepare the participant for the target

word or non-word. Once the target word appeared, participants responded by pressing

the "yes" button if the target was a word, or the "no" button if the target was a non-

word. The target remained on the screen until the participant made a response,

following which the screen went blank for 2000ms; in the no-load task this was

followed by the "Get Ready" signal to indicate the next trial was about to start; in the

visuospatial load task this was followed by the second array of dots and the participant

indicated whether the dots in the second array were in the same place as before the story

by pressing either "yes" or "no" on the response box; in the verbal load task, the single

letter followed the response to the target word, and the participant indicated whether

this letter had been included in the string presented prior to the story by pressing "yes"

or "no" on the response box. No feedback regarding decisions to the word or the load

task was given (i.e., correct/incorrect). See Figure 4.1 for a visual depiction of the steps

taken by participants during completion of a trial in this task.

In all three of the task versions, participants completed four practice trials prior to

beginning the experimental trials. All experimental trials were presented in random

order, and rest breaks were offered to participants via a message on the computer screen

after completion of every 10 trials. To ensure participants were attending to the story

passages, 30% of the text passages were followed by a simple comprehension question about the story they had just read (based on text recall only; as an example, the question that followed the predictive inference passage displayed in Table 4.1 was "what did the waitress bring the man?"). Thus, within every block of 10 trials, three stories had accompanying comprehension questions. All participants received the same story/question combinations, however, as stories within each block appeared in random order, the timing of stories that were accompanied by a question varied across participants. Answers to questions were given verbally to the researcher, who made a note of correct responses, and then pressed any key on the computer keyboard once the participant was ready to continue. On average, the no load task took each participant 30-40 minutes to complete, including rest breaks, and as outlined was completed in a single session. Whereas the visuospatial and verbal load tasks each took approximately 20-25 minutes to complete, including rest breaks, and were both completed in the same session; thus, this session took approximately 50 minutes to one hour to complete, including a rest break in between load tasks.

4.3 Results

Data Screening and Reduction

Participants' mean reaction times (RTs) for correct trials were calculated for each condition, separately for each load task. Trials in which a participant responded +/-2.5 standard deviations from the mean for that condition were considered to likely reflect a lapse in concentration, thus were coded as errors along with incorrect responses. In order to increase the reliability of the data set, any participants who obtained less than 50% correct on the comprehension questions, or obtained more than 50% incorrect trials in the lexical decision task (overall or within a single condition) were excluded from the

analysis for that load task. No participant's data were removed from the no load condition based on this criteria, however, three (two poor, one good comprehension) participants' data were removed from the visuospatial load data, and eight participants' (six poor, two good comprehension) data were removed from the verbal load data. In addition, one participant's data from the low comprehender group were not included in the load tasks due to failure to comply with task instructions. Three participants' data (two poor, one good comprehension) were also removed from the verbal load task due to a computer error that resulted in missing data. Due to apparent difficulty experienced by participants in completing the verbal load tasks, data collection of this task ceased prior to all participants being tested for ethical reasons in terms of minimising participant stress, given their young age. The flow of participants through each stage of the experiment and resulting sample size for each load condition is displayed in Figure 4.2.
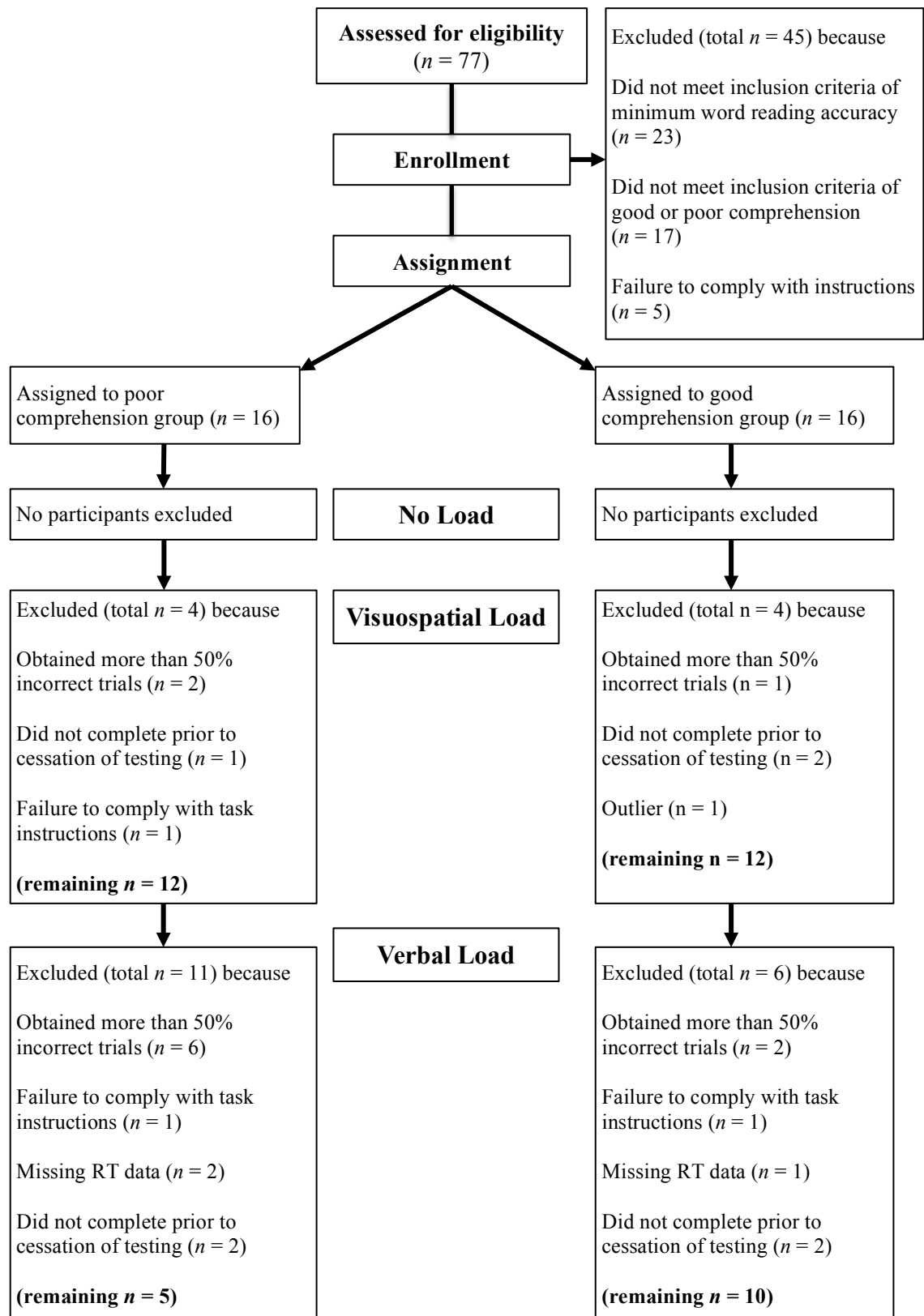
| Assessed for eligibility ($n = 77$) | | Excluded (total $n = 45$) because

Did not meet inclusion criteria of minimum word reading accuracy ($n = 23$)

Did not meet inclusion criteria of good or poor comprehension ($n = 17$)

Failure to comply with instructions ($n = 5$) |

Enrollment

Assignment

| Assigned to poor comprehension group ($n = 16$) | | | Assigned to good comprehension group ($n = 16$) |

| No participants excluded | **No Load** | No participants excluded |

| Excluded (total $n = 4$) because

Obtained more than 50% incorrect trials ($n = 2$)

Did not complete prior to cessation of testing ($n = 1$)

Failure to comply with task instructions ($n = 1$)

(remaining $n = 12$) | **Visuospatial Load** | Excluded (total n = 4) because

Obtained more than 50% incorrect trials (n = 1)

Did not complete prior to cessation of testing (n = 2)

Outlier (n = 1)

(remaining n = 12) |

| Excluded (total $n = 11$) because

Obtained more than 50% incorrect trials ($n = 6$)

Failure to comply with task instructions ($n = 1$)

Missing RT data ($n = 2$)

Did not complete prior to cessation of testing ($n = 2$)

(remaining $n = 5$) | **Verbal Load** | Excluded (total $n = 6$) because

Obtained more than 50% incorrect trials ($n = 2$)

Failure to comply with task instructions ($n = 1$)

Missing RT data ($n = 1$)

Did not complete prior to cessation of testing ($n = 2$)

(remaining $n = 10$) |

Figure 4.2. Flow of participants through each stage of the experiment.

To calculate facilitation to predictive inference targets, the mean difference in RT to target words between the control and predictive inference conditions was calculated for each participant, for all three of the load variations of the lexical decision task. Data from all variables were then screened for outliers using the deletion criteria of +/-3SDs from the mean. This resulted in RT data from one participant being removed from the visuospatial load facilitation and the visuospatial load predictive inference condition. No outliers were detected on any other variables. Following the removal of outliers, no significant deviations from normality were detected.

Group Descriptives

Independent samples *t*-tests confirmed that there was no significant difference between the groups on reading accuracy, rate or verbal working memory, but that the groups differed significantly on comprehension when measured by either the Neale or the DARC (see Table 4.2).

Table 4.2

Mean Values (Standard Deviations Shown in Parentheses) for Age, Verbal Working Memory, and the Reading Measures for Each Comprehension Group, and Group Comparisons

Measure	Group Mean (*SD*)		Group Comparisons		
	Poor (*n* = 16)	Good (*n* = 16)	*t*(df)	*p*	Cohen's *d*
Age	9.75 (.57)	9.63 (.63)	.56(30)	.58	0.20
Reading accuracy	77.63 (11.58)	72.75 (15.74)	1.00(30)	.33	0.35
Reading rate	75.94 (18.28)	84.63 (26.73)	1.07(30)	.29	0.38
Comprehension (Neale)	17.00 (3.31)	30.56 (4.37)	9.91(30)	<.001	3.50
Comprehension (DARC)[a]	25.19 (3.19)	27.56 (1.71)	2.63(22.98)	.015	0.93
Digit span forward	5.06 (1.20)	5.19 (.97)	.36(30)	.72	0.12
Digit span backward	3.81 (1.17)	4.50 (.97)	1.82(30)	.08	0.64

[a]*equal variances not assumed.*

Inference Generation Task

Reaction time

Initial paired-samples *t*-test revealed that across both groups, RT (ms) was significantly faster in the predictive inference condition (*M* = 1394.83, *SD* = 334.71) than the control condition (*M* = 1466.35, *SD* = 360.99; *t*(31) = 2.30, *p* =. 028, *d* = 0.41) in the no load condition, indicating that overall facilitation to predictive inference targets had occurred within the entire sample. In contrast, in the visuospatial load condition (*n* = 24), paired-samples *t*-tests revealed that overall, RT (ms) was not significantly faster in the predictive inference condition (*M* = 1460.86, *SD* = 381.38) than the control condition (*M* = 1542.16, *SD* = 527.00; *t(*23) = 1.28, *p* = 214, *d* = 0.26), thus facilitation to

predictive inference targets had not occurred within the entire sample. Similarly, in the verbal load condition ($n = 15$), there was no difference in RT (ms) between the predictive inference condition ($M = 1228.56$, $SD = 282.38$) and the control condition ($M = 1378.80$, $SD = 388.23$; $t(14) = 1.81$, $p = .092$, $d = 0.47$) across the entire sample.

As there were missing data in each of the load tasks, in order to maintain a sufficient level of power to determine significant effects, data from each load type were analysed separately. Firstly, to examine group differences in facilitation in the no load task, a 2 x 2 mixed model ANOVA with the within-groups factor of stimulus type (control or predictive inference) and between-groups factor of group (good or poor comprehenders) was conducted. This revealed a significant main effect for stimulus type ($F(1, 30) = 5.16$, $p = .03$, $\eta_p^2 = .15$) and group type ($F(1, 30) = 5.29$, $p = .03$, $\eta_p^2 = .15$) but no interaction between group and stimulus type ($F(1, 30) = .220$, $p = .64$, $\eta_p^2 = .01$; see Figure 4.3).

Follow-up paired samples t-tests revealed that RT (ms) was not significantly faster in the predictive condition than the control condition within either the group of poor comprehenders ($t(15) = 1.67$, $p = .12$, $d = 0.42$), or the good comprehenders ($t(15) = 1.58$, $p = .14$, $d = 0.40$; see Figure 4.3). Independent samples t-tests revealed that, although the good comprehenders responded significantly faster than the poor comprehenders to both control ($t(30) = 2.27$, $p = .031$, $d = 0.80$) and predictive ($t(30) = 2.16$, $p = .039$, $d = 0.76$) target words than the poor comprehenders (see Figure 4.3), there was no difference in the amount of facilitation to predictive targets between the good comprehenders ($M = 56.74$ms, $SD = 143.79$) and poor comprehenders ($M = 86.29$ms, $SD = 206.71$; $t(30) = .47$, $p = .64$, $d = 0.17$).
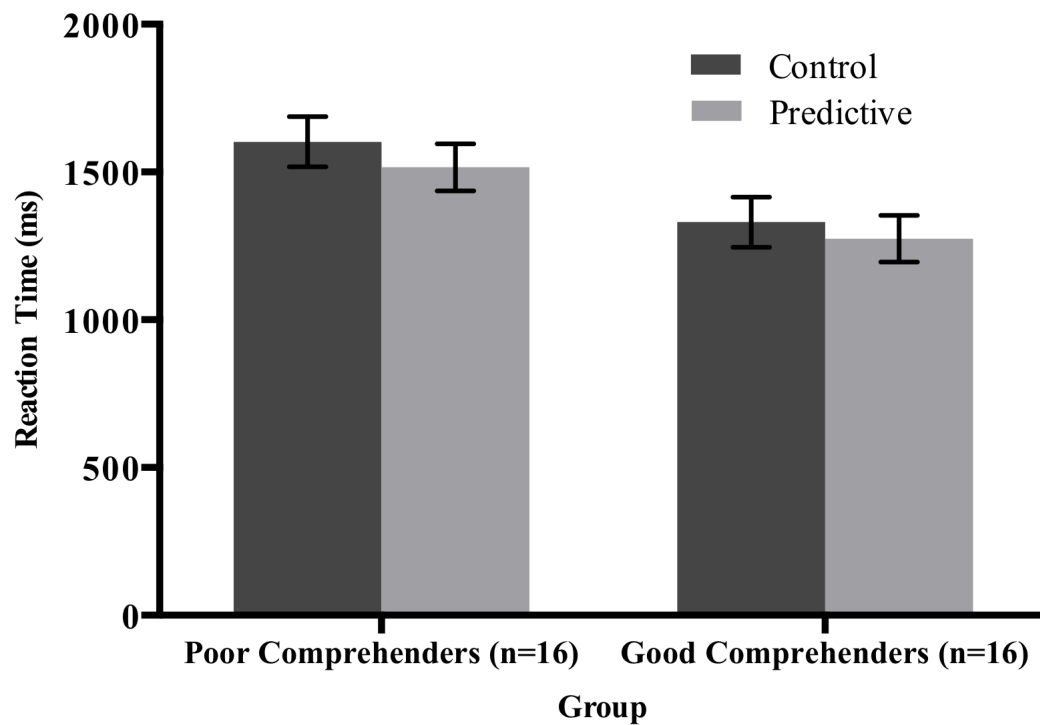
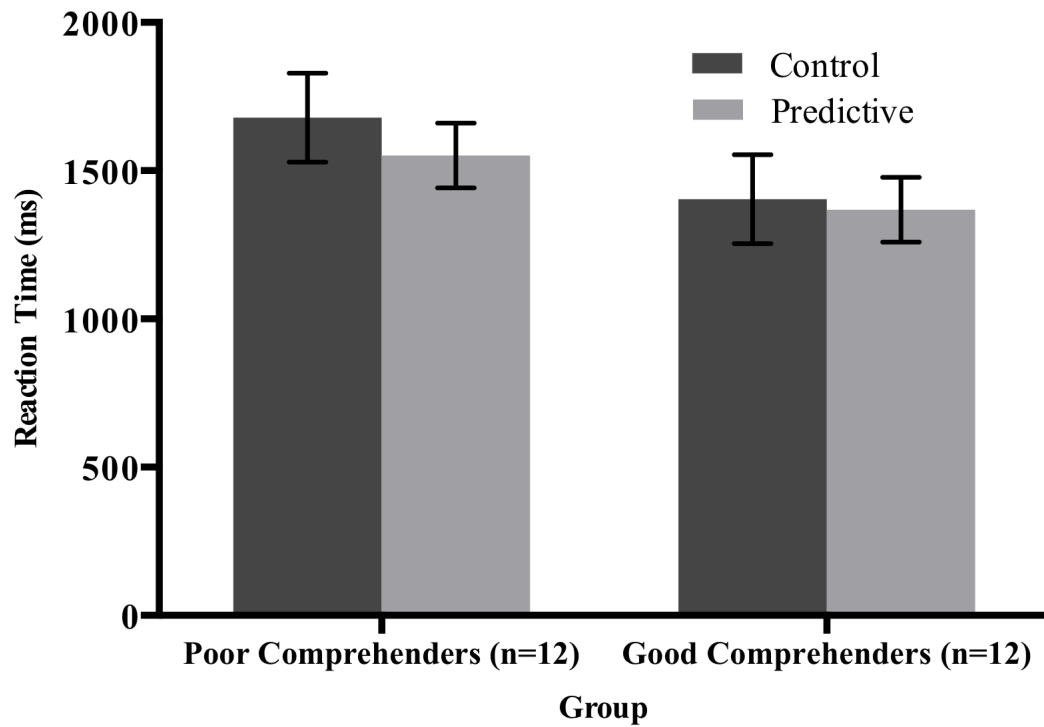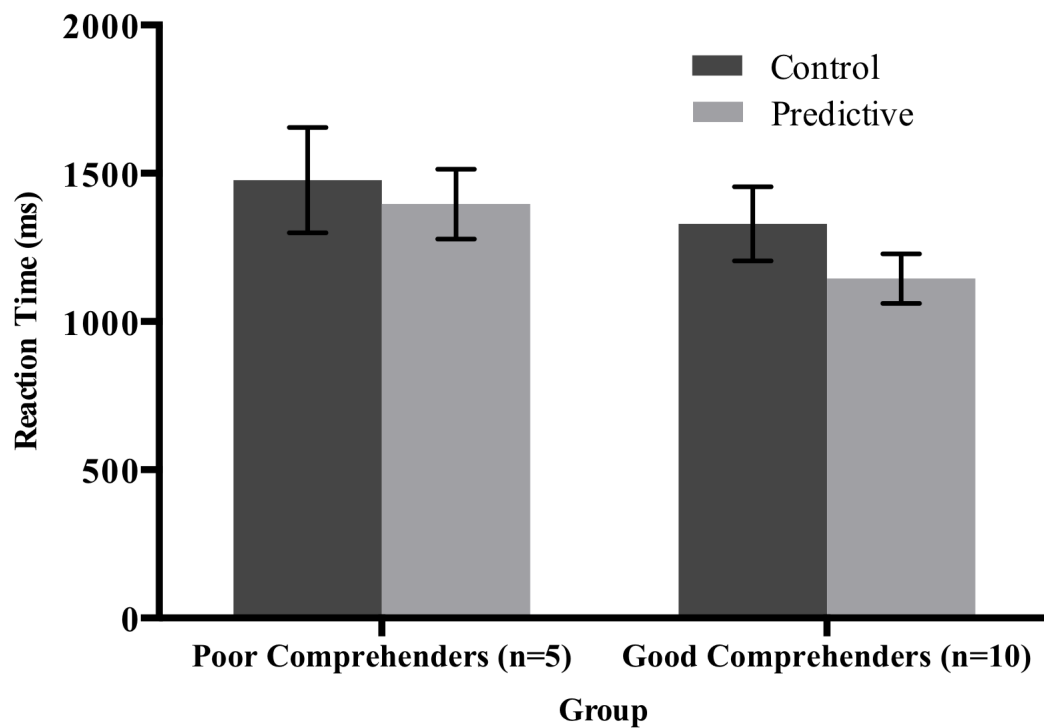*Figure 4.3.* Mean reaction time by condition and group in the no load task. Error bars represent +/-1 SE of the mean.

To examine group differences in facilitation under each load condition, separate 2 x 2 mixed-model ANOVAs were conducted for the data from each of the load types. For each load type, there were no significant main effects for stimulus type or group type, or any interactions between group and stimulus type (highest $F = 2.16$, for the effect of stimulus type in the verbal load condition). See Figure 4.4 for mean reaction time values by group and stimulus type in the visuospatial load task, and Figure 4.5 for mean reaction time values by group and stimulus type in the verbal load task.

Figure 4.4. Mean reaction time by condition and group in the visuospatial load task. Error bars represent +/-1 SE of the mean.

Figure 4.5. Mean reaction time by condition and group in the verbal load task. Error bars represent +/-1 SE of the mean.

Lastly, a series of paired samples *t*-tests was conducted separately for each group, to determine whether there were any differences in facilitation between the load types, within either group. No differences in facilitation to predictive inferences targets was found across any of the three load conditions within either group of comprehenders (highest $t(9) = 1.39$, $p = .19$) for the difference between the visuospatial load and verbal load conditions in the good comprehension group), see Table 4.3.

Table 4.3

Comparisons of RT Facilitation (ms) Across Conditions for Each Group

Group	Comparison	
Poor	No Load (*M* = 84.77, *SD* = 119.45)	VS Load (*M* = 126.84, *SD* = 285.03)
	No Load (*M* = 61.11, *SD* = 132.82)	Verbal Load (*M* = 80.92, *SD* = 376.96)
	VS Load (*M* = 142.55, *SD* = 389.64)	Verbal Load (*M* = 80.92, *SD* = 376.96)
Good	No Load (*M* = 48.93, *SD* = 149.93)	VS Load (*M* = 35.76, *SD* = 342.75)
	No Load (*M* = 44.21, *SD* = 149.75)	Verbal Load (*M* = 184.89, *SD* = 306.94)
	VS Load (*M* = -12.99, *SD* = 280.08)	Verbal Load (*M* = 184.89, *SD* = 306.94)

Note. VS = Visuospatial. A larger RT reflects greater facilitation to predictive inference targets.

Accuracy Rates

Due to unexpected findings in the reaction time data and the difficulties children displayed with completing the load tasks, post hoc analyses were conducted on the accuracy data. Data of all children who were included in the reaction time analyses (see Figure 4.2) were included in the accuracy analyses. In addition, the children who obtained less than 50% accuracy (overall or within a single condition) of the load tasks were also included in these analyses to obtain a complete picture of task difficulty.

However, six children's data (three poor, two good comprehenders) were removed from the verbal load task due to a computer error that resulted in missing accuracy data. The resulting sample size for each load condition is displayed in Table 4.4.

Table 4.4

Sample Sizes of Accuracy Data by Group and Load Task Condition

Load Task	*n* (Poor Comprehenders)	(Good Comprehenders)	Total *n*
No Load	16	16	32
VS Load	14	16	30
Verbal Load	8	10	18

Note. VS = Visuospatial.

Mean accuracy across all trials (% correct) was then calculated for each of the load tasks (no load, visuospatial load and verbal load). Data from these variables were screened for outliers using the deletion criteria of +/-3SDs from the mean. No outliers were detected. However, a series of Shapiro-Wilk tests revealed the assumption of normality was not met on several variables (see Table 4.5). Thus, non-parametric tests were used to analyse the accuracy data from the lexical decision task where appropriate.

Table 4.5

Descriptive Statistics and Shapiro-Wilks Tests for the Accuracy Data in each Load Task Condition

Variable	Mean (SD) (% correct)	Shapiro-Wilks Test
No load: Total trials		
Poor comprehenders	92.81 (5.40)	$W(16) = .95, p = .46, ns$
Good comprehenders	93.65 (6.03)	$W(16) = .88, p = .04$
Overall	93.23 (5.65)	$W(32) = .93, p = .04$
VS load: Total trials		
Poor comprehenders	74.11 (10.22)	$W(14) = .90, p = .10, ns$
Good comprehenders	75.71 (10.63)	$W(14) = .89, p = .07, ns$
Overall	74.91 (10.26)	$W(28) = .91, p = .02$
Verbal load: Total trials		
Poor comprehenders	53.21 (10.87)	$W(7) = .94, p = .64, ns$
Good comprehenders	60.00 (8.02)	$W(8) = .70, p = .002$
Overall	56.83 (9.75)	$W(15) = .90, p = .08, ns$

Note. VS = Visuospatial.

A series of Wilcoxon signed-rank tests were conducted to analyse the difference in accuracy across the load conditions. Median scores for these analyses can be found in Figure 4.6. It was revealed that across all participants accuracy was significantly higher in the no load task than the visuospatial task, $T = 0$, p $< .001$, $r = -.86$, and also significantly higher in the visuospatial task than the verbal load task, $T = 0, p =. 001, r = -88$.
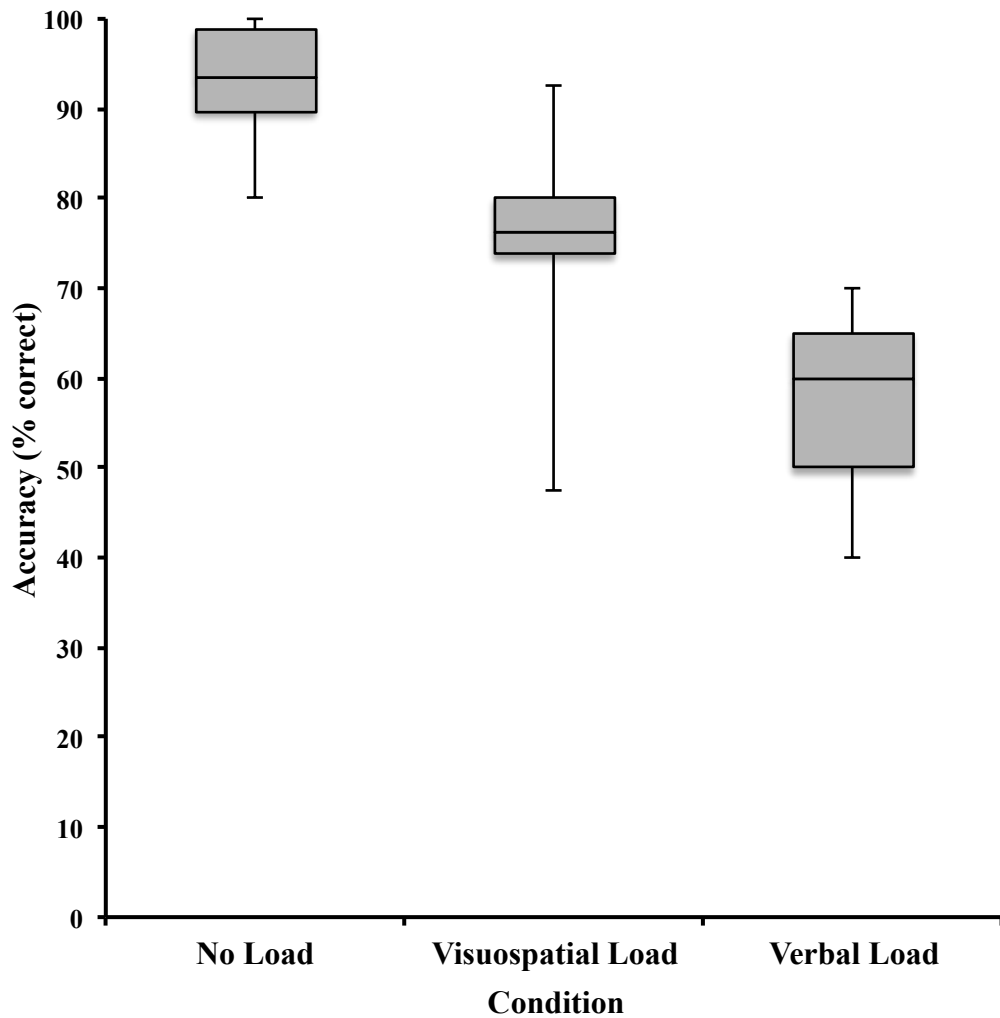
Figure 4.6. Boxplot of overall accuracy rates across the three load task conditions.

Differences between the groups' performance on each of the three load tasks were also examined. Firstly, a Mann-Whitney U test revealed no significant difference between poor (M*dn* = 93.33%) and good (M*dn* = 94.17%) comprehenders overall performance in the no load task, $U = 141.50, p = .62, r = .09$. An independent samples *t*-test revealed no significant difference between poor and good comprehenders overall performance in the visuospatial load task, $t(26) = .41, p = .69, d = 0.15$; see Table 4.5. Finally, a Mann-Whitney U test revealed no significant difference between poor (M*dn* = 55.0%) and good (M*dn* = 65.0%) comprehenders overall performance in the verbal load task, $U = 40.0, p = .19, r = .37$.

4.4 Discussion

The aim of the current study was to determine whether the utilisation of perceptual information is a factor that enables good comprehenders to construct more knowledge-based predictive inferences than poor comprehenders. However, the first hypothesis, that good comprehenders would show more evidence of facilitation to predictive inference targets than poor comprehenders, was not met; although when no additional task load was given, facilitation to predictive targets was evident across the entire sample, there was no difference in the magnitude of this facilitation between the good and poor comprehenders. Further, the facilitation effect within each group was not strong; when analysing each group's results separately, there was no longer a significant difference in reaction time between the control and predictive inference conditions. The hypotheses regarding the load conditions were, therefore, also not met; as although good comprehenders appeared to show some reduction in facilitation in the visuospatial load condition compared to the no load condition and the verbal load condition, these differences between conditions were not significant.

The finding that good comprehenders do not make predictive inferences during reading goes against the vast amount of literature that suggests good comprehenders are more aware of the causal structure of narratives (Lynch & van den Broek, 2007), and make more knowledge-based causal inferences (Cain & Oakhill, 1999; 2006; Kendeou et al., 2008; Tompkins et al., 2013) than poor comprehenders. Therefore, to ensure firstly that these results were not simply due to poor validity of the Neale comprehension scores (i.e., it has been suggested that these scores largely reflect lower-level reading and integration abilities than higher level comprehension (Bowyer-Crane & Snowling, 2005; Nation & Snowling, 1997; Spooner et al., 2004), groups' DARC scores were also compared. Here it was found that the poor comprehenders obtained significantly lower

DARC scores than the good comprehenders, giving further indication that these groups did in fact differ on higher-level comprehension skills and not simply lower-level reading ability. Thus, it can be assumed that any discrepancies between the results of the current study and previous research are more likely a reflection of the online measure of inferencing that was utilised in the current study.

At first glance, these results therefore seem to suggest that the generation of predictive inferences is not relevant to narrative comprehension. Indeed, in many cases it has been argued that predictive inferences are not necessary for comprehension, but are simply elaborative, serving to embellish a text rather than explain it (Magliano, Baggett, Johnson, & Graesser, 1993; McKoon & Ratcliff, 1992; Potts, Keenan, & Golding, 1988; Singer & Ferreira, 1983). However, subsequent studies investigating the specific conditions that lead to the activation of predictive inferences has resulted in a revision of this view. It is now recognised that although predictive inferences are not *routinely* made online during comprehension, they do occur under certain circumstances: for example, when easily accessible (i.e., if the predicted outcome is highly constrained by the preceding context; Casteel, 2007; Cook, Limber, & O'Brien, 2001; Lassonde & O'Brien, 2009; Murray, Klin, & Myers, 1993), when induced by the reading purpose or strategy (i.e., reading for study versus reading for entertainment; Allbritton, 2004; Calvo, Castillo, & Schmalhofer, 2006; van den Broek, Lorch, Linderholm, & Gustafson, 2001), or when explicitly required for comprehension (i.e., to maintain local or global coherence, particularly referential or causal coherence; Klin, Guzmán, & Levine, 1999a; Klin, Murray, Levine, & Guzmán, 1999b); in which case it can be argued that these predictive inferences may become *necessary causal inferences* rather than *elaborative predictive inferences*.

In the current study, each of the texts used was highly constrained to a single and specific likely consequence, in order to increase the likelihood that the inference would be made, assuming the reader was tracking the causal sequence of the narrative, and was able to access and integrate the relevant background knowledge accordingly (i.e., they were constructing a situation model). Admittedly, however, these inferences were not required to maintain on-going global coherence, or for later comprehension. Indeed, the only requirement for comprehension was the answering of intermittent questions used to ensure all participants were paying attention to the narrative texts. As these questions required only the recollection of textbase information (i.e., in order not to disadvantage poor comprehenders who may be reading the text, but unable to engage in deeper processing) they may have inadvertently influenced the demands of the reading task in a way that reduced the participants' motivation to engage in deeper constructionist comprehension processes such as monitoring relevant causal information and drawing inferences.

Specifically, previous research has shown that explicit instructions to recall rather than comprehend a text can prompt readers to allocate more resources towards processing textbase variables and hence only construct a surface form or textbase representation rather than a meaning-based situation model (Aaronson & Ferres, 1983; 1986; Stine-Morrow, Milinder, Pullara, & Herman, 2001; Zwaan, Magliano, & Graesser, 1995b). Further, instructions to memorise a text can result in less sensitivity to causal discontinuities in a narrative (Zwaan, Magliano, & Graesser, 1995b). Accordingly, Van den Broek and colleagues (van den Broek et al., 2001; 2005) propose that readers may adopt differing standards of coherence depending on the demands of the reading task and their motivation, which consequently dictates inferential activities. Correspondingly, in comparison to when participants are required to read as normal, or

for recall, instructions to read for study, or to evaluate, have been found to increase the coherence-building activities readers engage in (Magliano & Trabasso, 1999; Narvaez, van den Broek, & Ruiz, 1999; Rapp & Kendeou, 2007; van den Broek et al., 2001), including an increase in the number of predictive inferences that are made (Allbritton, 2004; Calvo et al., 2006; Magliano & Trabasso, 1999; van den Broek et al., 2001), even when reader strategies are induced by the requirements of the experimental task itself, rather than by explicit task instructions (Allbritton, 2004).

As such, it is possible that the questioning that occurred in the task induced participants to allocate more processing towards simply remembering textbase information, rather than engage in deeper causal monitoring and inferential processes, which could explain why facilitation was limited. Additionally, this may also explain why participants had trouble with the verbal load task, as verbal working memory resources were likely being depleted by the maintenance of a mostly textbase rather than situation model representation, thus the additional verbal information became too difficult to hold. Indeed, it was found that overall, participants made significantly more errors in the verbal load condition than the visuospatial load condition, suggesting that the verbal load task was in fact the most difficult. Nevertheless, it cannot be entirely ascertained from these results whether this difficulty was indeed due to children in the current study allocating more resources to processing textbase information or, simply, that the verbal load task was more difficult than the visuospatial load task. Indeed, it is noted that the verbal load of five items utilised in the current study is the average maximum capacity for children of this age group, whereas the load of six items used in previous research of this nature with adults (i.e., Sims & Hegarty, 1997) is sub-capacity to the average maximum of seven items.

Further interpretations of the load tasks are difficult, however, due to the fact that facilitation to predictive inferences was not significant within the group of good comprehenders alone, and task difficulty resulted in small sample sizes in the load task manipulations. Thus, the results regarding visual imagery are to be interpreted with caution. Over the entire sample, it was found that both load tasks disrupted predictive inferencing. This may be an indication that both verbal and visuospatial resources are needed for comprehension (i.e., the premise put forth in dual coding theory; Paivio, 1986), or simply a reflection of the difficulty children of this age group have with undertaking dual-load tasks. The latter may be a more likely interpretation due to the findings of previous research that a verbal load does not disrupt predictive inferencing in adults (Fincher-Kiefer & D'Agostino, 2004), in addition to the increased variance in reaction time displayed by both groups in the load conditions, the increase in number of errors in the load conditions compared to the no load condition, and the number of children who were unable to complete the load tasks successfully.

Yet, although predictions regarding the good comprehenders were not substantiated in this study, an interesting pattern of results emerged in the results from the load task manipulations, especially in the poor comprehension group. Firstly, participants in this group had more difficulty with the verbal load task in comparison to the visuospatial load task, and in comparison to the good comprehenders. Specifically, 42.85% of the poor comprehenders who completed the verbal load task did not meet the minimum accuracy criteria of 50% correct trials either overall, or within one of the inference conditions, compared to only 14.29% of the good comprehenders who did not meet this criteria. In contrast, only 13.3% of the poor comprehenders and 7.14% of good comprehenders who completed the visuospatial did not meet the minimum accuracy criteria. Although this could simply be a reflection of the poorer performance on verbal

working memory tasks that is often found within groups of poor comprehenders (Cain et al., 2004a; Cain & Oakhill, 1999; Oakhill et al., 2003; Oakhill, Hartt, & Samols, 2005b), the poor comprehenders in the current study were not found to differ on either a simple or complex verbal working memory task. This may therefore be an indication that poor comprehenders' have a greater reliance on text-level processing rather than situation modelling during reading, hence their inability to process any additional verbal information.

However, it should also be noted that, when the difference in accuracy between good and poor comprehenders was analysed in more detail, it was found that although good comprehenders made fewer errors in the verbal load task than the poor comprehenders, this difference between the groups was not significant. Despite this, a moderate effect size was detected; therefore, this non-significant finding could be due to the small sample size in the verbal load task. Future research is thus needed with an adequate sample size, to clarify whether poor comprehenders perform worse on this type of dual-load task than good comprehenders, and whether this poorer performance is due to a reliance on textbase processing, rather than poor verbal working memory per se.

For the most part, however, the current study highlights the importance of ensuring task demands encourage participants to go beyond textbase processing, and also engage in predictive inferences that are *necessary* for comprehension when exploring how these integrative processes vary between different groups. Yet, it is also undeniable that under normal reading conditions, comprehension involves the processing and coordination of both necessary *and elaborative* information. Specifically, because elaborations, including forward predictions, are constructed without the guidance of complete context (i.e., there is only the anticipation they may be useful), many of these become redundant

once further context has been clarified, and therefore need to be left out of the situation model (Keefe & McDaniel, 1993; Kintsch, 1988). In contrast, fully encoding predictive inferences that then turn out to be incorrect or unnecessary for comprehension would be a highly inefficient process that would heavily tax working memory resources, and could even result in the need for backward corrections and revisions to the situation model.

Accordingly, studies using lexical decision and naming times have revealed that the activation of predictive inferences decays unless supported by following text content (Calvo & Castillo, 1996; Casteel, 2007; Fincher-Kiefer, 1996; Keefe & McDaniel, 1993; Whitney, Ritchie, & Crane, 1992). Further, what is initially activated or encoded may not be a complete or specific inference (Casteel, 2007; Cook et al., 2001). Thus, it appears that predictive inferences are only minimally activated, and these inferences are only maintained until subsequent context supports them, at which point they are fully encoded or, otherwise, deactivated from working memory.

It has, therefore, been suggested that once word reading is acquired, in order to develop comprehension, children must become proficient in adjusting their reading standards efficiently to accommodate inferences that are necessary, while decreasing resources spent on those that are merely elaborative (Cain et al., 2001). Supporting this proposition, a line of recent studies has identified two subgroups of poor comprehenders: "elaborators", who are identified as children who generate knowledge-based inferences, but also make connections to background knowledge that are not appropriately related to the context of the text; and "paraphrasers", children who mostly repeat the ideas presented explicitly in the text, but show minimal evidence of inferencing (S. E. Carlson, Seipel, & McMaster, 2014a; McMaster et al., 2012;

McMaster, Espin, & van den Broek, 2014; Rapp et al., 2007). Importantly, it has been found that elaborators do not differ from good or average comprehenders in the *overall* number of knowledge-based inferences they make during reading, however, they make more *unnecessary* elaborative inferences and *incorrect* predictive inferences than readers with good comprehension, whereas good comprehenders only make inferences that are *required* for comprehension (S. E. Carlson, van den Broek, McMaster, Rapp, et al., 2014b; McMaster et al., 2012).

It thus seems that consideration of both elaborative and necessary predictive inferences would be of use in future research. Furthermore, as readers defined as elaborators demonstrate the ability to draw knowledge-based inferences, this highlights the possibility that inclusion of these individuals within a group of poor comprehenders can lead to an increase in the overall evidence of inferencing that is found within this group, especially if these inferences are highly contextually constrained, reducing the risk of an incorrect inference being drawn. This is of relevance to the current study, as it may explain why significant activation to predictive inference targets was found within the entire sample in the current study, despite the fact that these inferences were not necessary for the comprehension task. Yet, this also implies that some of the good comprehenders too were allocating resources to making these unnecessary inferences. However, in this case, rather than monitoring the causal structure of the narratives regardless of comprehension requirements and task demands, these individuals may have automatised certain comprehension processes, which allows them to attend to multiple task demands at once (i.e., memorisation of textbase information, along with the construction of a situation model).

It is currently difficult to make interpretations about the conditions under which good and poor comprehenders activate predictive inferences, as this has remained largely unexplored. Beyond the current study, it appears few studies have examined how groups of comprehenders differ in their generation of predictive or causal inferences, using online measures of this process (for two exceptions with adults see Binder, Chace, & Manning, 2007, and Murray & Burke, 2003, but note that groups in these studies were defined based on *reading ability* rather than comprehension per se). Indeed, most previous studies have relied on offline questioning or recall measures of inference generation when assessing how knowledge-based inferencing contributes to comprehension. Yet, while these studies demonstrate the importance of inference generation to comprehension outcomes, they do little to provide information about the process that occurs during reading; for example, at what point inference concepts are activated and subsequently encoded into the long-term representation in order to contribute to offline comprehension.

Similarly, think-aloud procedures, which have largely been used to identify subgroups of comprehenders (e.g., S. E. Carlson, Seipel, & McMaster, 2014a; McMaster et al., 2012; Rapp et al., 2007), although often reported as an "online" measure of comprehension, are not sensitive to the time-course of activation of inferences, and thus are subject to many of the same limitations as offline measures. Further, these procedures may alter comprehension processing in several ways, particularly they may promote inferencing and attention to the causal structure of narratives (S. E. Carlson, van den Broek, McMaster, Rapp, et al., 2014b; Rapp et al., 2007), which could induce poor comprehenders to provide unnecessary or invalid inferences due to the implicit task demands of having to provide a response. In addition, they are subject to the influence of extraneous skills that do not occur during normal reading, such as

expressive language, and readers may not always be explicitly aware of or adequately able to describe all of the processes that occurred during reading (Rapp et al., 2007). Therefore, the role of activation and inhibition of predictive inferences in children's comprehension is in need of further investigation with the use of more sensitive and ecologically valid measures of these processes.

Limitations and Future Research

The main limitation of the current study pertains to the fact that the short narrative texts used, combined with the nature of comprehension questions, led to implicit task demands in which the construction of predictive inferences were merely optional, or elaborative, rather than necessary for comprehension. Thus, it cannot be concluded from the results of the current study whether good and poor comprehenders differ in their immediate activation of predictive inferences in other situations, including normal reading environments, or other experimental conditions.

In addition, other limitations may have affected the results of this study. Firstly, the use of a lexical decision task means that retrospective context checking can occur when participants are responding to the target word. Context checking refers to the situation where a reader attempts to compare the target with the context of the preceding text (Potts et al., 1988). With respect to the current study, rather than generating predictive inferences online *during* comprehension of the passages, the presentation of a topic word could have led to a backward association between the target word and the preceding narrative context. Although, it cannot be ascertained if this occurred in the current study: it is both possible, due to the unlimited length of time participants were given to respond, and the large amount of variance evident in the reaction time measure (thus indicating participants were possibly relying on varying strategies), yet also

unlikely, as inferences were not required for overall comprehension and, consequently, participants need not have engaged in any strategy for making inferences, including context checking.

A more pertinent limitation is that the dual-task paradigm may have been too difficult for children of this age group, particularly with the added demands of having to remember textbase information in order to answer the comprehension questions. Further investigation of the role of visuospatial imagery in comprehension may however be warranted, due to the differential difficulties that were evident between the two groups. Specifically, it appeared that during reading, good comprehenders were able to better manage additional information in either verbal or visuospatial format, whereas poor comprehenders had more difficulty handling additional verbal information. Thus, inefficient reading strategies may lead poor comprehenders to struggle with a highly loaded verbal working memory. Yet, whether this is specifically due to an overall reliance on remembering textbase information, rather than engaging in integrative situation modelling processes, is in need of further clarification. Here, it may also be important to consider the strategies used by different subgroups of comprehenders; for example, while some readers may be prone to rely simply on textbase information (i.e., paraphrasers), others, who are prone to elaboration, may experience an overload in working memory that is due to attempts to maintain information that is incorrect, or not necessary for comprehension (i.e., that serves to simply embellish a text).

Accordingly, as it is possible that some poor comprehenders fail to suppress merely elaborative or incorrect inferences in working memory in order to free resources for on-going text content as it is encountered, further consideration of this inhibition process in situation model construction may be warranted. Indeed, a role for inhibition in reading

comprehension has largely been outlined by Gernsbacher and colleagues, who suggest that skilled comprehenders are better able to suppress contextually irrelevant information, and thus more effectively process relevant information (Gernsbacher, 1997; Gernsbacher & Faust, 1991; Gernsbacher, Varner, & Faust, 1990). Correspondingly, several studies have found that tasks assessing inhibitory functions predict children's reading comprehension level (Borella & de Ribaupierre, 2014; Borella, Carretti, & Pelegrina, 2010; Carretti et al., 2009), and poor comprehenders have been found to perform worse than good comprehenders on tasks that assess their ability to inhibit distractor information (Borella et al., 2010; Cain, 2006; De Beni & Palladino, 2000; Palladino, Cornoldi, De Beni, & Pazzaglia, 2001). However, while suppression of irrelevant information *during* reading processes has been investigated largely in relation to good and poor comprehenders' suppression of ambiguous word meanings (Gernsbacher et al., 1990; Gernsbacher & Faust, 1991), or younger and older readers' incorrect inferences about objects (Lorsbach & Reimer, 1997; Lorsbach, Katz, & Cupak, 1998), there is less information regarding the suppression of unnecessary or incorrect *predictive* inferences, specifically in relation to comprehension level.

Thus, it appears that examination of individual differences in the extended time-course of activation and subsequent decay (or encoding) of both necessary and elaborative predictive inferences between good and poor comprehenders would be useful. Similar to previous studies (i.e., Calvo & Castillo, 1996; D. L. Long, Oppy, & Seely, 1994), this could be achieved by varying the stimulus onset asynchrony (SOA) of target words following presentation of the inference eliciting context, to determine at which time points activation of the inference concept is evident, but with the extension that this research then compares these patterns between different groups of developing comprehenders. Further, in order to improve interpretations, more ecologically valid

measures of online inference generation should be used to overcome the limitations highlighted by the current study, and those of previously used think-aloud and questioning procedures.

Indeed, as highlighted by the current study, the use of lexical decision and naming tasks can lead to an inherently non-natural reading environment, thus limiting the types of generalisations that can be made. In contrast, methodologies that incorporate eye-tracking and electroencephalography (EEG) may prove more useful in future research, as they have the potential to provide measurement of comprehension processes as they occur and with minimal task demands or interruptions. For example, readers' eye movements can be measured online, under the basic assumption that changes in processing time or fixation patterns indicate increased processing demands (Raney, Campbell, & Bovee, 2014; Rayner, Chace, Slattery, & Ashby, 2006). Specifically, longer reading times and more frequent regressions often indicate a reader's difficulty with integrating information into the preceding passage, or existing situation model representation. Accordingly, when target information is predictable from the preceding context, reading times are shorter, indicating ease of processing (Hand, Miellet, O'Donnell, & Sereno, 2010), including when this information is made predictable through inference (Calvo, Meseguer, & Carreiras, 2001; Ingram, Hand, & Moxey, 2014; O'Brien, Shank, Myers, & Rayner, 1988).

Advantages of recording eye movements over other online comprehension measures are clearly evident. Firstly, as measurement occurs while the participant reads naturally there are minimal disruptions to comprehension such as button presses or providing verbal responses. This also provides valuable temporal information about the precise moment a manipulated variable has an effect, which can be divided into more fine-

grained components, each reflecting a different stage in the processing time course (Rayner et al., 2006). As monitoring eye movements also provides an indication of whether participants are focusing on the text, this could displace the need for comprehension questions that may inadvertently alter task demands and thus level of text processing.

Similarly, measurement of event-related potentials (ERPs) via EEG can provide another online measure of semantic activation that does not require an overt response from the participant, and provides superior temporal resolution. This is achieved by recording neural activity through electrodes placed on the scalp, and time-locking multiple trials of EEG traces, which are averaged to remove spontaneous EEG activity and maintain specific event-related neural activity (Baretta, Tomitch, Lim, & Waldie, 2012). Thus, ERPs provide a measure of neural activity that occurs in response to a specific external event or stimuli (Baretta et al., 2012). In particular, the N400 component, has been found to represent the ease with which a stimulus is integrated into a given context (Kutas & Hillyard, 1984; see also Baretta et al., 2012, for a review), and appears to be particularly bound to language comprehension, as violations in other domains (i.e., musical melodies) do not elicit an N400 component (Besson & Macar, 1987). Thus, in relation to inference generation, if a reader encounters a word or sentence that is consistent with the inference they had made, this explicit statement would be redundant; resulting in a small N400 component in comparison to a condition in which the prior inference had not previously been made. As such, several studies have investigated the N400 as an online measure of the different types of text-based and knowledge-based inferences that are made during reading (i.e., Baetens, Van der Cruyssen, Achtziger, Vandekerckhove, & Van Overwalle, 2011; Baretta, Tomitch, MacNair, Lim, & Waldie, 2009; St George, Mannes, & Hoffman, 1997), including predictions (i.e., van Berkum,

Brown, & Zwitserlood, 2005). This could also be extended to provide a more detailed picture of the types of inferences that are made online by good and poor comprehenders.

In summary, it is recognised that reading comprehension is a complex operation that requires dynamic and possibly strategic processing. It thus appears that the importance of inference generation may not be fully realised by examining only activation and integration processes but by also taking into account other executive functions such as inhibition during reading. It is apparent that the task used in this study did not capture the complexity of the numerous causal relations that readers must infer during reading, which often extend over long distances in the text, require the coordination of multiple pieces of information, and are not necessarily obvious. Further, although often overlooked in previous research and theory, it is possible that children with poor comprehension difficulties do not form a homogeneous group. This should also be taken into consideration in future research.

To conclude, the results of the current study do not provide evidence that good and poor comprehenders differ in their online activation of predictive inference constructs. However, this could be due to limitations of the lexical decision task used to measure inference generation. The results of the current study were therefore also inconclusive regarding the role of visual imagery in predictive inference generation and comprehension. Several proposals for how future research can address these limitations have been explored, with the intention that these could provide clearer interpretations regarding the role of inferencing in reading comprehension and the contribution of both visual and verbal resources to this process.

Chapter 5. General Discussion

5.1 Summary of Findings

Several important findings emerged from the results from the current thesis. Firstly, Study 1 supported the theory that visual imagery ability is not a single undifferentiated construct, but rather can be conceptualised as several distinct subprocesses, namely: image generation, maintenance, scanning and transformation. Furthermore, extending on previous research, the current thesis provided evidence that this distinction exists within child populations; thus, imagery skills may become differentiated at a fairly young age.

In regards to the main aim of the current thesis, which was to determine whether a relationship exists between visual imagery and reading comprehension, the results were varied and somewhat ambiguous. Firstly, although a correlation between the mental rotation task and reading comprehension was found in Study 2, mental rotation ability did not predict reading comprehension after controlling for the variance provided by age, fluid intelligence, lower-level reading skills and verbal working memory. Thus, this suggests that there is no unique relationship between visual imagery and reading comprehension. This finding was particularly unexpected when measuring comprehension using the DARC, which aims to measure higher-level comprehension independent of lower-level reading ability as, based on theories of situation modelling, it could be expected that visual imagery would emerge as a stronger predictor when measurement is focused on these higher-level skills.

Regardless, performance on Raven's was a stronger predictor of DARC scores than Neale scores, which may highlight a potential role for non-verbal skills in higher-level

reading comprehension. In particular, performance on Raven's is recognised to be somewhat dependent on visuospatial skills (Lynn et al., 2004), and in line with this, Raven's was found to significantly correlate with the MRT task. It is thus possible that the MRT was not a significant predictor of comprehension in the current study, as all of the reliable variance in visualisation skill that was captured by this measure was accounted for by performance on Raven's.

In addition, it is possible that the type of imagery assessed by the MRT may not be completely analogous to the imagery that takes place during narrative comprehension. Specifically, while both comprehension and completion of the MRT likely require imagery that is dynamic and depictive, in contrast to the MRT, the imagery that occurs during reading is likely to be less purposeful and more resemblant of the visual simulation of an entire narrative scene. Indeed, the findings of several studies suggest that the imagery activated from verbal descriptions is largely unconscious and automatic (Bergen et al., 2007; Just et al., 2004; Speer et al., 2009; see also Kosslyn & Moulton, 2009). Further, recent neuroimaging research indicates that the neural regions that are activated during simulation of language only partially overlap with those that are activated during tasks of visual imagery (Hartung, Hagoort, & Willems, 2015). Thus, these two constructs of simulation and visual imagery may not be entirely similar, hence the weak contribution of this task to comprehension scores.

While Study 3 aimed to overcome this limitation, by preventing the use of any type of imagery during reading, and examining the resulting effects on comprehension, the results were not conclusive. Specifically, the aim of Study 3 was to determine whether group differences in comprehension were a result of the use of visual imagery in a key process required for situation model construction (the generation of knowledge-based

inferences), by preventing readers from generating visual imagery during reading through the use of a dual-load task. However, although inferences that connect the causal sequence of a narrative were measured in Study 3 (due to theoretical and empirical evidence of their importance in reading comprehension; S. E. Carlson, Seipel, & McMaster, 2014a; Kendeou et al., 2008; Lynch & van den Broek, 2007; Tompkins et al., 2013; Trabasso & Suh, 1993; van Kleeck, 2008), it appears that methodological limitations inherent in Study 3 precluded these inferences from being necessary for comprehension and, thus, provided the reader with merely elaborative information. As such, there was limited evidence that these knowledge-based inferences were being drawn within either the group of good or poor comprehenders.

Despite this, the load task manipulations in Study 3 did reveal an interesting pattern of findings, which suggest that the role of visual imagery in reading comprehension may be worthy of further investigation. Specifically, participants in the poor comprehension group had more difficulty than good comprehenders with the verbal load task, and also more difficulty with the verbal load task than the visuospatial load task. Further, this difficulty was evident in spite of the fact that poor comprehenders did not perform more poorly than the good comprehenders on an extraneous complex verbal working memory task that also tapped storage and additional processing capacity. Thus, rather than being a reflection of the poorer performance on verbal working memory tasks that is often found within groups of poor comprehenders (Cain et al., 2004a; Cain & Oakhill, 1999; Oakhill et al., 2003; Oakhill, Hartt, & Samols, 2005b), this finding may be an indication that poor comprehenders rely more on text-level information during reading, rather than constructing imagery-rich situation model representations; hence their inability to process any additional verbal information. However, further research is needed with a

larger sample size, in order draw clearer conclusions as to whether these differences in accuracy between the two groups are indeed significant.

5.2 Theoretical Interpretations and Implications

5.2.1 Implications for Measurement and Theory

5.2.1.1 Visual Imagery

The current thesis lends support to theories of visual imagery that suggest that it is not a singular construct but, rather, can be differentiated into distinct subprocesses, including image generation, image maintenance, image scanning and image transformation. Specifically, Study 1 found low correlations between separate measures of visual imagery, each which was designed to tap a different subprocess, suggesting that, while there may be some overlap in these processes, each can generally be considered distinct. These findings align with the computational model of imagery proposed by Kosslyn (Kosslyn, 1980; 1983; 1994; Kosslyn et al., 1984) and subsequent research that suggest visual imagery is supported by several differing subprocesses (Kosslyn et al., 1984; 1990; 2004; Poltrock & Brown, 1984). The finding that imagery is best represented as separate subcomponents was also supported by the results of Study 2, which found that separate measures of imagery, each designed to tap a separate subprocess, differentially predicted variance in reading comprehension. Extending on the findings of the majority of previous research in this area, the current thesis also found that this separation of imagery processes is evident in young children; thus, this differentiation of subprocesses may appear at an early age. As such, the current thesis suggests that imagery should not be measured as a single construct, even in child populations.

It is also possible that imagery processes beyond those measured in the current thesis exist. For example, the image transformation component described by Kosslyn (Kosslyn, Holtzman, Farah, & Gazzaniga, 1985) may not only encompass processes such as rotation, but also the ability to add and/or subtract details from a visual image, engage in tasks such as mental paper folding, and zoom in and out on a mental image. Here, further research is needed in order to determine how these processes overlap with those described in the current thesis.

5.2.1.2 Reading Comprehension

In regards to the measurement of reading comprehension, the findings of the current study are consistent with previous research that suggests different measures of reading comprehension do not assess the same underlying constructs (i.e., Cutting & Scarborough, 2006; Keenan et al., 2008; Keenan & Betjemann, 2006; Ozuru et al., 2008; Rowe et al., 2006). Specifically, Study 2 revealed that lower-level skills such as reading rate and accuracy were stronger predictors of Neale comprehension scores than constructs relevant to higher-level comprehension such as working memory. In contrast, accuracy and reading rate were strong predictors of scores on the DARC, suggesting that skills beyond lower-level text processes account for the variance on this measure. While it is acknowledged that the relationship between these lower level skills and comprehension would be greater for the Neale than the DARC for task specific reasons (i.e., as accuracy and rate were measured in combination with comprehension on the Neale) this still demonstrates the importance of controlling for lower level skills in order to get an accurate assessment of comprehension, as poor decoding skills may influence comprehension scores on the Neale, regardless of an individual's actual comprehension level.

Indeed, Study 3 showed that, when controlling for lower-level reading skills (i.e., by ensuring all participants had age-appropriate word reading ability) groups of comprehenders as differentiated by the Neale also differed significantly on the DARC. Thus, it appears that when lower-level reading skills are sufficient, the Neale may provide a more accurate picture of reading comprehension. This makes sense in light of previously discussed reasons for how combined measurement of accuracy affects Neale comprehension scores; for example, frequent corrections of reading errors during testing may disrupt text processing at a level that impairs comprehension; poor decoding skills may compel a reader to focus on word reading, rather than engage in comprehension; and, as the number of comprehension questions administered is dependent on level of reading accuracy, children with low reading accuracy but unimpaired comprehension are not given the opportunity to answer all of the comprehension questions they could potentially have answered correctly due to early cut-off (Spooner et al., 2004). Ideally, however, beyond controlling for lower-level reading ability, care should be taken to measure comprehension separately from lower-level reading skills, as combined measurement may still induce some readers to allocate more resources to the task of reading accurately, rather than focusing on comprehension. This is especially so, as the conditions that are necessary for this combined measurement require text to be read aloud. In addition, readers are often aware of the fact that their reading is being assessed. Thus, future research could extend on the findings of the current study by using assessments of decoding and fluency that are measured completely independently of the measurement of comprehension.

Regardless, the current thesis has demonstrated that a measure of comprehension that is based on theories of reading comprehension that emphasise the importance of higher-level cognition (i.e., those that propose comprehension is a consequence of the

construction of a coherent situation model) is less influenced by lower-level skills than a traditional standardised measure. This is a noteworthy finding, as few studies have explicitly compared scores across these two types of measures and, as such, there is limited empirical evidence for this proposition. One exception to this, however, is work by the authors of the DARC, which showed that after accounting for the contributions of language skills and non-verbal reasoning, word-reading skills (decoding and fluency) were significant predictors of scores on another standardised test of reading comprehension (the Woodcock-Johnson Passage Comprehension subtest; WJPC), but not on the DARC (Francis et al., 2006). This finding thus demonstrated that the DARC is less influenced by word-reading skills in comparison to the WPJC (Francis et al., 2006). The findings of the current thesis corroborate this previous research and add to these authors' claims that the DARC measures comprehension while minimising the impact of lower-level reading abilities such as decoding and reading speed (August et al., 2006; Francis et al., 2006).

In addition, the current thesis has also highlighted several implications for measuring comprehension via the use of experimental tasks. It appears that while measures such as the lexical decision task may be useful for measuring specific comprehension processes, it is important to also consider how task demands and text-reader interactions may differ under these highly controlled conditions in comparison to reading in more natural environments. Furthermore, consideration should be given to subgroups of comprehenders beyond "good" and "poor" (a distinction that is most often made in the literature). For example, previous research has made a distinction between groups of poor comprehenders who are prone to elaboration (i.e., engage in inappropriate or unnecessary higher-level comprehension processes) versus those who could be defined as paraphrasers (i.e., those who simply rephrase the information found explicitly in a

text; S. E. Carlson, Seipel, & McMaster, 2014a; McMaster et al., 2012; 2014; Rapp et al., 2007). The current thesis further calls attention to how these potential differences may impact on the assessment of situation model construction in these groups (and the quality of these representations), especially as deficits in key skill areas such as inferencing may not be evident (see Chapter 4.4, pp. 199-200).

Overall, the findings of the current thesis support the notion of investigating multiple skills at varying levels when determining what differentiates good from poor comprehension. Particularly, identification and intervention are unlikely to be successful if the focus is on a sole predictor of reading comprehension.

5.2.2 Understanding of the Relationship Between Visual Imagery and Reading Comprehension

Despite not finding a clear relationship between visual imagery and reading comprehension, the current thesis offers several important theoretical implications in relation to this aim. Firstly, it is apparent that a clear conceptualisation of the concept of visual imagery and the underlying system that supports this construct should be adopted when researching the relationship between comprehension and visual imagery ability. The current thesis indicates that imagery is not a unitary construct, but rather can be differentiated into various subprocesses. Thus, it is important to consider exactly what types of visual imagery are relevant to reading comprehension.

In particular, the results of Study 2 indicated that transformation of visual imagery, a specific imagery process that takes place within the visual buffer, may be of more relevance to reading comprehension than other types of imagery processes (i.e., simple image maintenance and scanning across a maintained image), which did not show a

positive correlation with reading comprehension. It is theorised that this may be due to more complex types of imagery being required in order to transform and update dynamic situation model representations based on newly encountered information. However, the contributions of these imagery variables were not significant; thus, going beyond specific subprocesses of visual imagery, measurement of the resulting phenomenological experience that is activated via the input of language would likely reveal additional findings regarding the contribution of visual imagery to reading comprehension. Indeed, this would extend on research findings that language that denoting perceptual input or movement activates neural areas similar to those involved in both actual perception and bodily action (Aziz-Zadeh et al., 2006; R. F. Goldberg, Perfetti, & Schneider, 2006a; 2006b; Hauk et al., 2004; Hauk & Pulvermuller, 2004; Just et al., 2004; Pulvermuller, 2005; Speer et al., 2009). Hence, these studies suggest that embodiment is an important part of the process of language comprehension.

Thus, to further investigate this proposition, the aim of Study 3 was to disrupt imagery during reading in order to determine how this affects comprehension, and indeed, the results of this study did implicate both verbal and visuospatial resources as necessary for one higher-level comprehension process. Specifically, it was found that both a verbal and visuospatial load disrupted predictive inferencing during reading, thus supporting the underlying premises of dual coding theory (Paivio, 1986). In addition, the results of this study provided some evidence that poor comprehenders may rely more on verbal information during text comprehension than good comprehenders, as poor comprehenders appeared to have additional difficulties maintaining concurrent verbal information during reading.

However, in stating this, clear interpretations regarding the findings of Study 3 are not entirely possible. Firstly, it could not be ascertained from Study 3 whether the disruption of inferencing that occurred within the entire sample was indeed due to a depletion of resources required for situation model construction or, rather, whether children of this age group have difficulty with dual-tasks in general. Further, the difficulty in completing the verbal load task that was evident within the group of poor comprehenders may also be due to the general difficulties in complex working memory tasks that are often reported in this group (Cain et al., 2004a; Cain & Oakhill, 1999; Oakhill et al., 2003; Oakhill, Hartt, & Samols, 2005b). In addition, as recent research has identified a role for central executive functions such as attention allocation in higher-level comprehension (see Kendeou et al., 2014, for a review of this research), it may be that poor comprehenders lack the attentional capacities to complete this cognitively demanding task accurately. Further research is thus needed to clarify which interpretation here is the most plausible.

In addition, due to limitations of the lexical decision task used in Study 3, evidence of inference generation was minimal. Thus, additional research is needed to clarify the role of visual imagery in higher-level comprehension processes. Indeed, emerging research findings suggests that exploration is warranted. For example, Francey and Cain (2015) recently found that visual imagery training can aid poor comprehenders' resolution of which character a pronoun (e.g., "he" or "she") referred to (i.e., when gender cues were absent from these pronouns). Thus, these results suggest that imagery may indeed aid integrative processes such as the generation of inferences (in this case, inferences about the referent of stated pronouns). Nevertheless, in order to further investigate similar propositions, as highlighted by the findings of Study 3, it is important not only to consider specific types of higher-level processes when measuring comprehension, but

also the conditions under which these processes are executed. Specifically, although knowledge-based inferences may be of vital importance to comprehension, under some circumstances these inferences may be unnecessary and thus even hinder comprehension. These points should be taken into account in future research that aims to investigate the factors that lead to successful reading comprehension.

5.2.3 The Contribution of Additional Skills to Reading Comprehension (Verbal and Non-Verbal).

5.2.3.1 Working Memory

In contrast to a large amount of previous research (see Chapter 1.3.5.1), the current thesis did not find evidence of a strong relationship between verbal working memory and reading comprehension. Specifically, neither simple nor complex verbal working memory emerged as significant independent predictor of reading comprehension on either the Neale or DARC in Study 2. However, as discussed (see Chapter 3.5, p. 147), this may have been due to Raven's accounting for all the reliable working memory variance, due to an overlap between the constructs of working memory and fluid intelligence.

Despite this, the correlations between comprehension measures and the working memory tasks did reveal some potentially important findings. In particular, it was found that forward digit span significantly correlated with the Neale but not the DARC whereas backward digit span correlated with the DARC but not the Neale. As backwards span is considered the more complex of these two working memory tasks (due to the requirement of both maintenance and transformation of information) this may indicate that many questions on the Neale may simply require maintenance and recall of information, whereas the DARC is more dependent on additional working

memory processes such as integration. Indeed, this interpretation makes sense when considering that the DARC was designed specifically to tap into higher-level comprehension processes such as knowledge integration. Thus, whether working memory tasks show a relationship with comprehension likely depends on how comprehension is defined and measured.

In contrast, however, Study 3 did not reveal any differences between good and poor comprehenders' verbal working memory, when measured using either a simple or complex verbal working memory task. It is possible that the lack of a difference in simple working memory span is a result of the fact that all the children in Study 3 had proficient lower-level reading ability. To elaborate, it has been suggested that poor decoding skills can deplete the verbal and cognitive resources required for the maintenance component of verbal working memory tasks (Goff et al., 2005). Similarly, it is possible that the use of digits can reduce any positive influence that semantic memory has on memory performance (i.e., as digits are less amenable to dual-coding; see Chapter 3; see also Nation et al., 1999). Thus, as all children in Study 3 had a similar level of reading ability, it may have negated any impact that these skills normally have on the encoding and recall of information in working memory recall.

However, the lack of a difference between these groups on the complex working memory task is more difficult to explain. Here, it is possible that the backward span task utilised was not complex enough to capture many of the integrative processes required for comprehension (e.g., updating the contents of working memory via the activation and integration of semantic information in long-term memory). Indeed, only a moderate effect size was found for the correlation between this measure and the DARC in Study 2 ($r = .31$). It is possible that utilisation of a measure of verbal working memory that

assessed an even greater level of executive processing would have yielded different results. This would align with previous research centred on the measurement of the central executive component of verbal working memory, which has revealed that updating abilities are predictive of higher-level comprehension processes such as inferencing (Potocki et al., 2013).

Alternatively, it may be that verbal working memory is less uniquely important for reading comprehension than previously thought, as information can be maintained and integrated into the situation model by encoding information in visual format. In accordance with the proposal of others (i.e., Chrysochoou et al., 2011), it is suggested that future research may benefit by shifting the focus to the central executive rather than the phonological loop component of working memory, when examining the relationship between this construct and reading comprehension.

Indeed, as outlined, a possible reason why mental rotation did not predict comprehension in the current study could be due to this measure assessing a type of imagery that is likely more effortful than the imagery that is automatically activated from verbal descriptions. Thus, the MRT is likely not only dependent on visualisation ability, but also central executive functions. Specifically, during imagery manipulations (such as rotation) high demands may be placed on executive functions due to the strong interference between the external stimuli and the internal representations of those items: for example, participants need to retain an active internal image of a figure while resisting interference from the external visual stimuli during performance of the mental manipulation (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). This may provide another explanation as to why Raven's, which is also argued to assess executive functioning capacity (Carpenter et al., 1990), may have largely accounted for the

variance provided by performance on the MRT. Further investigation is needed, however, to clarify whether information in both formats (i.e., visual and verbal) is likely to be governed by the central executive in order for situation model construction, and whether this differs as a function of the reading condition.

To guide this research, clearer conceptualisation and understanding of how visual imagery is activated from verbal descriptions may be required. Specifically, although dual coding theory puts forth the notion of information being coded in both verbal and visual format, several questions still exist regarding how information that goes through a verbal mechanism such as verbal working memory is then translated into a visual (or visuospatial) representation. Similarly, little is known regarding how these representations are stored in long-term memory for later use.

Certainly, earlier attempts have been made to resolve this question via computer simulation models that propose spatial information is initially encoded as propositions that are then integrated into a spatial representation that is perhaps held in visuospatial working memory (Glenberg, Kruley, & Langston, 1994; Haenggi et al., 1995). However, although these models initiate some conception of how this operation occurs, they do not provide a detailed specification of this translation process. Further, these models largely focus on how spatial language is represented in working memory, but do not explain how other types of textual information may be represented in a spatial or visual format. In particular, if situation models are used to provide a perceptual representation of narrative texts, then it would be useful to determine whether these representations are indeed constructed within visuospatial working memory, the visual buffer of the visual imagery system, or rather, some alternative mechanism unrelated to these structures. More recently, it appears that few attempts have been made to explain

how this translation process works; thus, several questions remain regarding exactly how verbal and visual components of working memory interact during in reading comprehension.

5.2.3.2 Inhibition and Attention Allocation

In addition to integration and manipulation of information in working memory, emerging research is establishing a role for other executive processes such as attentional resources and inhibition in reading comprehension and situation model construction (Borella & de Ribaupierre, 2014; Pike et al., 2010; see also Kendeou et al., 2014). Although the current thesis has somewhat supported this notion by demonstrating that a relationship between Raven's and the DARC exists, the inclusion of separate measures of these constructs would likely reveal more specific findings regarding the degree at which separate executive functions are involved in reading comprehension.

Indeed, as discussed in Study 3, inhibition may play a particularly important role in reading comprehension. Specifically, learning to use working memory resources in an efficient manner in order to accommodate and integrate information that is necessary for comprehension, while inhibiting that which is merely elaborative, may be a key component in the development of successful comprehension. As such, it is possible that children with poor inhibition skills may have been included in the group of poor comprehenders in Study 3. These children may be particularly prone to elaboration that is detrimental to comprehension (i.e., as although they are able to engage in knowledge-based inferences, they also make connections to background knowledge that is not related to the context of the text, or unnecessary for the purpose of coherence). Consequently, this may have increased the amount of facilitation in the lexical decision task that was found in the group of poor comprehenders, as inferencing was not

necessarily required for comprehension in this task. Thus, the role of both inhibition and elaboration should be taken into consideration in future research, particularly when drawing conclusions regarding the how inference generation contributes to reading comprehension.

5.3 Limitations and Future Research Directions

As discussed, the results of the current thesis highlight the complexity of reading comprehension, particularly how this process may vary depending not only on higher-level cognitive ability but also task demands and other reader characteristics. Thus, a multitude of skills and extraneous variables may need to be considered when seeking to understand what constitutes successful reading comprehension. Future research should explore this question through careful consideration of the measures used to assess these processes. Previously, a diverse range of measures has been used to identify the higher-level processes that occur during children's comprehension. These include both offline measures (i.e., those that measure inference generation and understanding of a story via retrieval of this information after comprehension has occurred), and online measures (i.e., those that measure inference generation at the time of encoding). Thus, offline measures generally involve cued recall, usually by asking participants a set of forced-choice, or open-ended comprehension questions, to determine whether the correct inference has been drawn and level of meaning obtained. However, this technique has been criticised on the basis that cued recall cannot distinguish between whether the inferences occurred due to encoding and integration of information during reading, or at retrieval (Keenan et al., 1990) and therefore provide little information about the *process* that occurs during reading.

It is important to distinguish between the product and processes of reading comprehension, because it is through the combination of these processes that the quality of the product is determined (Kendeou et al., 2014). Therefore, in order to provide adequate interventions, it is important to understand where specific cognitive processes may fail and how they can be positively influenced. Thus, a variety of online measures have also been used in order to try to establish an assessment of inference and coherence processes as they occur during reading. These include interspersing questions throughout a text to determine if the reader's developing situation model contains implicit information; for example, predictions about what is going to happen next in a narrative, or inferences about why story events occurred. However, this method may still induce some level of prompting; for example, the questioning method itself may lead participants to engage in elaboration and inference generation that would not have occurred otherwise, or only occurred at the point of questioning but not during reading. Thus, think-aloud protocols, during which the participant articulates their thoughts throughout or at selected points during reading (and which are then coded for a variety of comprehension-related processes; e.g., number of knowledge-based inferences drawn) are assumed to provide a more authentic reflection of the reader's developing mental representation and understanding of the story, as this method involves providing free-recall without responding to specific questions or prompts (Lynch & van den Broek, 2007).

Yet, although often reported as an *online* measure of comprehension, think-aloud procedures are not sensitive to the immediate activation of particular comprehension processes, so they too are subject to many of the same limitations as offline measures. Further, these procedures may alter comprehension processing in several ways. For example, they may promote inference generation and attention to the causal structure of

narratives (S. E. Carlson, van den Broek, McMaster, Rapp, et al., 2014b; Rapp et al., 2007) and, thus, provide implicit cues to elaborate on information and, hence, lead to the generation of unnecessary or incorrect inferences. In addition, responses may be dependent on extraneous skills that are not utilised during normal reading, such as expressive language, and readers may not always be explicitly aware of or adequately able to describe all of the processes that occurred during reading (Rapp et al., 2007).

Therefore, alternative online measures, including that used in Study 3 of the current thesis, which involve assessing the level of activation of the inference concept while the inference is being drawn (or immediately following) have been employed. These generally involve using methods of recognition, lexical decision, Stroop-interference, or concept naming in response to highly controlled conditions. For example, in Study 3, the texts were designed to provide conditions that either induced an inference or did not. As another example, text conditions can be manipulated to include a break in global coherence, compared to those in which coherence is maintained. Thus, on-line measures enable a more detailed analysis of the characteristics of specific comprehension processes, such as their time-course, specificity (i.e., whether readers infer a specific outcome, or rather, that the context could indicate a range of consequences), and the conditions under which they occur or are inhibited.

However, these methods are also subject to criticisms. Particularly, they require a participant to provide a response (i.e., in order to measure naming, reading or reaction times); thus, they may disrupt or alter comprehension process to varying degrees. Therefore, developments in cognitive measurement have led to the use of eye-tracking and ERP analysis, as these methods may be able to provide online measurement of comprehension processes, while overcoming several limitations of other methods, as

they are non-invasive and require minimal task demands. Specifically, because eye movements and changes in EEG fluctuations occur naturally in response to the reading process, concurrent tasks are not needed to draw conclusions about reading comprehension and task engagement. As an example, longer eye fixations and ERPS (particularly, the N400 component) can be used as an indication that a reader is having trouble integrating encountered information as they read, and can suggest whether an inference has, or has not, been drawn (see Chapter 4.4, pp 205-206).

Thus, measures of eye movements and ERPs can provide superior temporal resolution in comparison to reaction-time and self-paced reading paradigms, while also minimising task demands that create an inherently unnatural reading environment. This is not to say however, that these measures are indeed faultless. For example, these methods still often require the use of short and highly controlled text passages that may be devoid of complete context; thus, they may not capture all of the processes involved in constructing a more complex and dynamic situation model. Further research is also needed to establish the utility of these tools for the measurement of inferences that require the activation and integration of background knowledge (i.e., in comparison to text connecting inferences), as these measures have been used less extensively for this purpose, especially in comparison to lexical decision tasks employing reaction time or naming time as the dependent variable. Furthermore, although some studies have used eye-tracking to measure higher-level comprehension processes such as inference generation and comprehension monitoring, there are few studies that have done so in the context of how these processes relate to children's overall level of reading comprehension.

With regards to the measurement of reading comprehension using standardised tests, in line with previous research, the current study found that different measures of comprehension do not tap the same component skills. Moving forward, further validation of more recently developed measures of comprehension that are based on cognitive theory should be undertaken in order to establish their utility. This is particularly important as these measures have the advantage over traditional school-based reading assessments, due to their diagnostic qualities that allow specific skill deficits to be identified in order to provide targeted interventions. However, these recently constructed measures are not without their limitations. For example, although the DARC is useful for identifying strengths and weaknesses in specific types of comprehension processes, it also does not assess important skills, such as whether a reader maintains causal or referential coherence of a narrative. Indeed, it may be argued that this measure focuses heavily on inferencing, without taking into account other cognitive skills that may influence comprehension. Other newly formed measures of comprehension have also faced similar criticisms (for a discussion of the criticisms, see S. E. Carlson, Seipel, & McMaster, 2014a). Thus, further development or extensions of these measures may also be warranted.

Indeed, although the current study sought to establish the multiple skills involved in reading comprehension, measurement of higher-level comprehension was largely confined to integration processes such as the generation of knowledge-based inferences. It is acknowledged that the higher-level skills involved in comprehension go beyond inference generation. Indeed, as Zwaan and Radvansky note (1998), although inferences are made in the process of situation model construction, situation models themselves are not simply a collection of inferences. Thus, future research should extend the findings

of the current study by exploring imagery in relation to other comprehension outcomes, such as global coherence, for example.

Accordingly, in order to clearly establish visual imagery's role in comprehension and consequently refine imagery-based interventions, more information is needed to disentangle for exactly which comprehension skills visual imagery is important. For example, research is needed to determine whether visual imagery is directly related to specific skills such as inference generation and/or comprehension monitoring or, alternatively, has a greater influence on comprehension via reading engagement and embodiment in meaning. It is plausible that a combination of both explanations are correct; for example, imagery may play a role in both the maintenance and integration of story information, while also allowing a reader to become more immersed in the story experience, increasing engagement. Yet, this may vary depending on text type and coherence demands also.

While the current study focused on narrative comprehension, exploration of the role of imagery in comprehension could also be extended to examine other types of texts; for example, scientific or expository texts. Although, here it is possible that different types of imagery may differentially predict comprehension for varying types of texts. Specifically, as noted in Chapter 1, VSWM possibly aligns more closely with the processing of spatial texts such as route descriptions (see Chapter 1.3.5.2). This is supported by previous research that often finds a relationship between these types of texts and measures of VSWM, whereas few studies have found this relationship with narrative texts. Thus imagery processes that overlap with the construct of VSWM (i.e., maintenance of visual information; particularly, spatial layouts or sequences) or even spatial ability (i.e., the ability to accurately conduct spatial manipulations using visual

imagery and to perceive spatial relationships) could be more relevant to the processing of expository or explicitly spatial texts. However, route descriptions are a very specific type of text, which focus on procedural information regarding navigation through a spatial environment, and it appears that fewer studies have investigated the role of visual imagery or VSWM in other types of expository or scientific texts. Although, in relation to this, a role for both dynamic visual imagery (Sanchez & Wiley, 2014) and maintenance of visuospatial information (Kruley et al., 1994) has been indicated. Thus, as with narrative comprehension, it is possible that a situation model that is rich with depictive imagery is also important for understanding scientific or expository texts.

Specifically, similar to narrative comprehension, the phenomenological experience of "seeing" what is being described may also guide the comprehension of expository texts, as it provides another mechanism for a reader to keep track of events and efficiently update representations of meaning based on dynamic information, and even possibly increases engagement with text content. Unlike narrative comprehension, however, imagery may be less important for extracting meaning via embodiment and transportation into the narrative situation; here, underlying imagery processes such as manipulation may be more, or equally, as important as this phenomenological experience. To elaborate, many scientific texts describe phenomena with elements that move, interact and change across time and space; thus, understanding of these topics may require manipulations of spatially-based mental representations, as such concepts are unlikely to be easily represented as verbal propositions (Sanchez & Wiley, 2014).

Indeed, a potential limitation of the current study is that in trying to measure different aspects of imagery in order to provide a complete assessment of this construct, it was unsuccessful in examining the phenomenological experience of imagery that results as

an amalgamation of these underlying processes. Thus, the full embodied experience, which may be key to children's narrative comprehension, was not assessed. Particularly, when conceptualising the imagery that takes place in response to narrative input, it is unlikely that purposeful manipulation of several separate images occurs within a situation model, as such a process would clearly be too time-consuming and inefficient to represent the automaticity of the reading comprehension process. Furthermore, it could be argued that this pictorial experience is not, in fact, a result of the efficient interaction of the separate processes measured in the current study, but rather a separate type of imagery that occurs more automatically in response to linguistic input. This may also explain why in previous research no relationship has been found between measures that assess the vividness of one's experience of imagery (i.e., which require a reader to imagine a scene from a verbal description and then rate the strength of this imagery), and more objective measures of spatial ability and imagery processes (i.e., in which the object to be imaged, maintained and manipulated is often provided; Dean & Morris, 2003; Durndell & Wetherick, 1976; Ernest, 1977; Lequerica et al., 2002; Poltrock & Brown, 1984; Richardson, 1977). This may especially be the case in light of findings that this lack of correlation is not simply due to an inability to accurately introspect on imagery (Dean & Morris, 2003).

Unfortunately, measurement of the concept of vividness was not successful in the current research, as the measure used to tap this construct (the binocular rivalry task) did not prove reliable when used with children in a school setting. This measure is also subject to other criticisms. In particular, despite inclusion of catch trials, response bias may be present in which a participant chooses the imagined stimuli for reasons other than dominance (i.e., implicit bias due to task demands). Subsequently, in Study 3, attempts were made to block the generation of any type of imagery, including the

phenomenological experience that may occur as a result of situation model construction. To achieve this, an additional visual load was introduced during a lexical decision task that aimed to capture knowledge-based inference generation (a key component of situation model construction). However, due to limitations of the lexical decision task used, this measure appears to have been limited in its ability to capture this construct. Accordingly, few conclusions could be drawn regarding the role of imagery in situation modelling from this study.

Thus, when examining how imagery is relevant to comprehension, future research could benefit from also aiming to capture the dynamic visual imagery of scenes as they develop *during* reading or language comprehension, in order to gauge the resulting phenomenon of imagery, rather than each individual process. One possible way to achieve this may be by extending on the visual world method (R. M. Cooper, 1974; Tanenhaus, Spivey-Knowlton, & Eberhard, 1995), which has been used to investigate the role of visual information in comprehension processes, by recording participants' eye movements to visual scenes while they listen to orally presented narratives.

Using the visual world method it has been determined that individuals make saccadic eye movements towards explicitly mentioned, or anaphoric pronouns of mentioned entities, during comprehension, and has proven useful for investigating a variety of comprehension related processes; for example, good and poor comprehenders' resolution of anaphoric inferences (Engelen, Bouwmeester, de Bruin, & Zwaan, 2014). Further, similar patterns of eye movements have been found using a blank screen paradigm, in which the visual scene was presented prior to, but then absent during, spoken language comprehension (Altmann, 2004). Extending on this, using the blank screen paradigm, children's eye movements could be tracked during listening

comprehension, to determine whether their eye movements are consistent with the described movement and location of the characters and events both explicitly and implicitly presented in the narrative. This would help determine whether comprehenders keep track of narrative events through *visual imagery* of the scene portrayed. Comparisons could then be made to determine whether good comprehenders' eye movements are more consistent with the events portrayed in the narrative than poor comprehenders'.

In addition, neurological data may also be useful when seeking to determine whether good comprehenders activate more visual imagery during reading than poor comprehenders. For example, neuroimaging studies have revealed neural activity that is consistent with the activation of visual imagery during language comprehension, including an overlap of the neural substrates involved in actual bodily movement with those that are activated while reading words, or extended passages, that denote the perceptual input or movement (Aziz-Zadeh et al., 2006; R. F. Goldberg, Perfetti, & Schneider, 2006a; 2006b; Hauk et al., 2004; Hauk & Pulvermuller, 2004; Just et al., 2004; Pulvermuller, 2005; Speer et al., 2009), such as the manual manipulation of objects, and navigation of spatial environments (Speer et al., 2009). Thus, future research could extend this research to determine whether good and poor comprehenders differ in the level of activation found in these regions during comprehension tasks.

Furthermore, other types of imagery may also be important to investigate in order to determine whether comprehension is related to the embodied experience that occurs during reading. For example, in order to fully construct a perceptual "scene" of what is described in a story, simulations of sound and olfaction may be necessary to complement visual simulations, and in line with this, activation of brain regions

involved in sensory perception has been found in response to phrases describing experiencing these senses (Olivetti Belardinelli et al., 2009; Palmiero et al., 2009). However, these domains have been of less focus in research on situation models, thus further investigation as to how this entire perceptual experience is relevant to comprehension could also be beneficial.

5.4 Conclusion

To summarise, the current thesis adds to the literature that suggests not all reading comprehension measures are interchangeable in regards to the underlying skills they measure. Furthermore, the current thesis provides some evidence that good and poor comprehenders may differ in their use of textbase versus visual representations during reading comprehension; specifically, it appears that poor comprehenders may rely more on textual information, which is a possible reflection of their difficulty in constructing an imagery-rich situation model of the events described in a text in order to aid meaning generation.

However, although visual imagery may be relevant to reading comprehension, it is likely that this relationship will be further established through careful conceptualisation and measurement of visual imagery versus visual simulation. In particular, it appears that visual imagery goes beyond being a singular construct, even when measured in younger populations. Yet, how these processes overlap with those involved in visual simulation generated from textual descriptions is still to be determined.

In conclusion, the findings of the current study have clear implications regarding the use of existing comprehension measures in research and practice. As measures of comprehension may not be interchangeable, or accurately measure all of the skills

involved in situation model construction, further research is needed to develop and establish the validity of assessments of comprehension by utilising cognitive models that explain how an individual obtains meaning from written language. The current thesis may aid future research with this purpose, particularly that which seeks to further investigate the role of visual imagery in higher-level comprehension processes.

References

Aaronson, D., & Ferres, S. (1983). Lexical categories and reading tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(5), 675–699. http://doi.org/10.1037/0096-1523.9.5.675

Aaronson, D., & Ferres, S. (1986). Reading strategies for children and adults: A quantitative model. *Psychological Review*, *93*, 89–112. http://doi.org/10.1037/0033-295x.93.1.89

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431. http://doi.org/10.1016/j.cub.2013.05.065

Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology*, *21*(6), 1459–1468. http://doi.org/10.1037/0278-7393.21.6.1459

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061–1070. http://doi.org/10.1037/0278-7393.19.5.1061

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes*, *38*(3), 309–322. http://doi.org/10.1207/s15326950dp3803_2

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm." *Cognition*, *93*(2), B79–B87. http://doi.org/10.1016/j.cognition.2004.02.005

Andrade, J., Kemps, E., Werniers, Y., May, J., & Szmalec, A. (2002). Insensitivity of visual short-term memory to irrelevant visual information. *The Quarterly Journal of Experimental Psychology*, *55*(3), 753–774. http://doi.org/10.1080/02724980143000541

Aristotle (1961) *De anima, Books II and III,* (D. W. Hamlyn, Trans). Oxford, UK: Oxford University Press. (Original work published in 4th century B.C.).

August, D., Francis, D. J., Hsu, H. Y. A., & Snow, C. E. (2006). Assessing reading comprehension in bilinguals. *The Elementary School Journal*, *107*(2), 221–238. http://doi.org/10.1086/510656

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*(18), 1818–1823. http://doi.org/10.1016/j.cub.2006.07.060

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Scientific Studies of Reading*, *255*(5044), 556–559. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1736359

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*(1), 126–145. http://doi.org/10.1037//0096-3445.129.1.126

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–90). New York, NY: Academic Press.

Baetens, K., Van der Cruyssen, L., Achtziger, A., Vandekerckhove, M., & Van Overwalle, F. (2011). N400 and LPP in spontaneous trait inferences. *Brain Research*, *1418*(C), 83–92. http://doi.org/10.1016/j.brainres.2011.08.067

Baretta, L., Tomitch, L. M. B., Lim, V. K., & Waldie, K. E. (2012). Investigating reading comprehension through EEG. *Ilha Do Desterro a Journal of English Language, Literatures in English and Cultural Studies*, *0*(63), 69–100. http://doi.org/10.5007/2175-8026.2012n63p69

Baretta, L., Tomitch, L. M. B., MacNair, N., Lim, V. K., & Waldie, K. E. (2009). Inference making while reading narrative and expository texts: An ERP study. *Psychology & Neuroscience*, *2*(2), 137–145. http://doi.org/10.3922/j.psns.2009.2.005

Barnes, M. A., Raghubar, K. P., Faulkner, H., & Denton, C. A. (2014). The construction of visual-spatial situation models in children's reading and their relation to reading comprehension. *Journal of Experimental Child Psychology*, *119*(C), 101–111. http://doi.org/10.1016/j.jecp.2013.10.011

Barratt, P. E. (1953). Imagery and thinking. *Australian Journal of Psychology*, *5*(2), 154–164. http://doi.org/10.1080/00049535308256068

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. http://doi.org/10.1017/s0140525x99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. http://doi.org/10.1146/annurev.psych.59.103006.093639

Bayliss, D. M., Jarrold, C., Baddeley, A. D., & Gunn, D. M. (2005). The relationship between short-term memory and working memory: Complex span made simple? *Memory*, *13*(3-4), 414–421. http://doi.org/10.1080/09658210344000332

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92. http://doi.org/ 10.1037/0096-3445.132.1.71

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*(5), 733–764. http://doi.org/10.1080/03640210701530748

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*(1), 14–25. http://doi.org/10.1111/j.1469-8986.1987.tb01853.x

Betts, G. H. (1909). *The distribution and functions of mental imagery*. New York, NY: Teachers College, Columbia University.

Binder, K. S., Chace, K. H., & Manning, M. C. (2007). Sentential and discourse context effects: adults who are learning to read compared with skilled readers. *Journal of Research in Reading*, *30*(4), 360–378. http://doi.org/10.1111/j.1467-9817.2007.00349.x

Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension, memory and production. *Journal of Verbal Learning and Verbal Behavior*, *18*(2), 187–198. http://doi.org/10.1016/s0022-5371(79)90118-x

Bloom, C. P., Fletcher, C. R., van den Broek, P., Reitz, L., & Shapiro, B. P. (1990). An on-line assessment of causal reasoning during comprehension. *Memory & Cognition*, *18*(1), 65–71. http://doi.org/10.3758/BF03202647

Borella, E., & de Ribaupierre, A. (2014). The role of working memory, inhibition, and processing speed in text comprehension in children. *Learning and Individual Differences*, *34*(C), 86–92. http://doi.org/10.1016/j.lindif.2014.05.001

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, *43*(6), 541–552. http://doi.org/10.1177/0022219410371676

Borst, G., Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2012). Representations in mental imagery and working memory: Evidence from different types of visual masks. *Memory & Cognition*, *40*(2), 204–217. http://doi.org/10.3758/s13421-011-0143-7

Bos, L. T., De Koning, B. B., Wassenburg, S. I., & van der Schoot, M. (2016). Training inference making skills using a situation model approach improves reading comprehension. *Frontiers in Psychology*, *7*(116), 1–13. http://doi.org/10.1037/0033-2909.123.2.162

Boulenger, V., Hauk, O., & Pulvermuller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*(8), 1905–1914. http://doi.org/10.1093/cercor/bhn217

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Scientific Studies of Reading*, *247*(4938), 44–48. http://doi.org/10.1126/science.2403694

Bowyer-Crane, C. A., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology*, *75*(2), 189–201. http://doi.org/10.1348/000709904x22674

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897x00357

Brunyé, T. T., Ditman, T., Mahoney, C. R., Walters, E. K., & Taylor, H. A. (2010). You heard it here first: Readers mentally simulate described sounds. *Acta Psychologica*, *135*(2), 209–215. http://doi.org/10.1016/j.actpsy.2010.06.008

Burton, L. J., & Fogarty, G. J. (2003). The factor structure of visual imagery and spatial abilities. *Intelligence*, *31*(3), 289–318. http://doi.org/10.1016/s0160-2896(02)00139-3

Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory*, *18*(2), 255–280. http://doi.org/10.1111/j.1468-2885.2008.00322.x

Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory*, *14*(5), 553–569. http://doi.org/10.1080/09658210600624481

Cain, K. (2009). Making sense of text: Skills that support text comprehension and its development. *Perspectives on Language and Literacy*, *35*(2), 11–14. Retrieved from http://search.proquest.com/openview/d7fbc53dabdeb8fd7877cb0d7606e0d5/1

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*(5-6), 489–503. Retrieved from http://link.springer.com/article/10.1023/A:1008084120205#page-1

Cain, K., & Oakhill, J. V. (2006). Profiles of children with specific reading comprehension difficulties. *Journal of Educational Psychology*, *76*(4), 683–696. http://doi.org/10.1348/000709905x67610

Cain, K., & Oakhill, J. V. (2007). Reading comprehension difficulties: Correlates, causes, and consequences. In K. Cain & J. V. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41–75). New York, NY: Guilford Press.

Cain, K., Oakhill, J. V., & Bryant, P. E. (2004a). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31–42. http://doi.org/10.1037/0022-0663.96.1.31

Cain, K., Oakhill, J. V., & Lemmon, K. (2004b). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*(4), 671–681. http://doi.org/10.1037/0022-0663.96.4.671

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, *29*(6), 850–859. http://doi.org/10.3758/bf03196414

Calvo, M. G., & Castillo, M. D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes*, *22*(1), 57–78. http://doi.org/10.1080/01638539609544966

Calvo, M. G., Castillo, M. D., & Schmalhofer, F. (2006). Strategic influence on the time course of predictive inferences in reading. *Memory & Cognition*, *34*(1), 68–77. http://doi.org/10.3758/bf03193387

Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research*, *65*(3), 158–169. http://doi.org/10.1007/s004260000050

Carlson, S. E., Seipel, B., & McMaster, K. L. (2014a). Development of a new reading comprehension assessment: Identifying comprehension differences among readers. *Learning and Individual Differences*, *32*, 40–53. http://doi.org/10.1016/j.lindif.2014.03.003

Carlson, S. E., van den Broek, P., McMaster, K. L., Rapp, D. N., Bohn-Gettler, C. M., Kendeou, P., & White, M. J. (2014b). Effects of comprehension skill on inference generation during reading. *International Journal of Disability, Development and Education*, *61*(3), 258–274. http://doi.org/10.1080/1034912X.2014.934004

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. http://doi.org/10.1037/0033-295x.97.3.404

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246–251. http://doi.org/10.1016/j.lindif.2008.10.002

Carretti, B., Cornoldi, C., De Beni, R., & Romanò, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology*, *91*(1), 45–66. http://doi.org/10.1016/j.jecp.2005.01.005

Casteel, M. A. (2007). Contextual support and predictive inferences: What do readers generate and keep available for use? *Discourse Processes*, *44*(1), 51–72. http://doi.org/10.1080/01638530701285572

Cataldo, M. G., & Oakhill, J. V. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology*, *92*(4), 791–799. http://doi.org/10.1037/A»22-0663.92.4.791

Catts, H. W., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech*, *49*(2), 278–293. http://doi.org/10.1044/1092-4388(2006/023)

Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The effect of visual imagery training on the reading and listening comprehension of low listening comprehenders in Year 2. *Journal of Research in Reading*, *22*(3), 241–256. http://doi.org/10.1111/1467-9817.00088

Chi, M. T. (1976). Short-term memory limitations in children: Capacity or processing deficits? *Memory & Cognition*, *4*(5), 559–572. http://doi.org/10.3758/bf03213219

Chrysochoou, E., Bablekou, Z., & Tsigilis, N. (2011). Working memory contributions to reading comprehension components in middle childhood children. *The American Journal of Psychology*, *124*(3), 275–289. http://doi.org/10.5406/amerjpsyc.124.3.0275

Clark, F. L., Deshler, D. D., Schumaker, J. B., Alley, G. R., & Warner, M. M. (1984). Visual imagery and self-questioning strategies to improve comprehension of written material. *Journal of Learning Disabilities*, *17*(3), 145–149. Retrieved from http://10.1177/002221948401700304

Colom, R., Escorial, S., & Rebollo, I. (2004). Sex differences on the Progressive Matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences*, *37*(6), 1289–1293. http://doi.org/10.1016/j.paid.2003.12.014

Connell, L. (2005). Colour and stability in embodied representations (pp. 1–6). Presented at the Twenty-seventh Annual Conference of the Cognitive Science Society, Stresa, Italy. Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2005/docs/p482.pdf

Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*(3), 476–485. http://doi.org/10.1016/j.cognition.2006.02.009

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 168–183. http://doi.org/10.1016/s0160-2896(01)00096-4

Conway, A. R. A., Kane, M. J., & Bunting, M. F. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. http://doi.org/10.3758/bf03196772

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. http://doi.org/10.1016/j.tics.2003.10.005

Cook, A., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language*, *44*(2), 220–234. http://doi.org/10.1006/jmla.2000.2744

Cooper, L. A. (1971). Mental rotation of random two-dimensional shapes. *Cognitive Processing*, *7*(1), 20–43. http://doi.org/10.1016/0010-0285(75)90003-1

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Processing*, *6*(1), 84–107. http://doi.org/10.1016/0010-0285(74)90005-x

Coulbeau, L., Royer, P., Brouziyne, M., Dosseville, F., & Molinaro, C. (2008). Development of children's mental representations: Effects of age, sex, and school experience. *Perceptual and Motor Skills*, *106*(1), 241–250. http://doi.org/10.2466/pms.106.1.241-250

Cramer, E. H. (1980). Mental imagery, reading attitude, and comprehension. *Reading Improvement*, *17*(2), 135–139. Retrieved from http://search.proquest.com/openview/a7bf007b3a0e1e577d18ac0bcbf45415/1?pq-origsite=gscholar&cbl=2030479

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, *10*(3), 277–299. http://doi.org/10.1207/s1532799xssr1003_5

Cutting, L. E., Materek, A., Cole, C. A. S., Levine, T. M., & Mahone, E. M. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia*, *59*(1), 34–54. http://doi.org/10.1007/s11881-009-0022-0

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. http://doi.org/10.1016/s0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. http://doi.org/10.3758/bf03214546

De Beni, R., & Palladino, P. (2000). Intrusion errors in working memory tasks: Are they related to reading comprehension ability? *Learning and Individual Differences*, *12*(2), 131–143. http://doi.org/10.1016/s1041-6080(01)00033-4

De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology*, *17*(1), 77–95. http://doi.org/10.1080/09541440340000529

de Jonge, P., & de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. *Personality and Individual Differences*, *21*(6), 1007–1020. http://doi.org/10.1016/s0191-8869(96)00161-4

De Koning, B. B., & van der Schoot, M. (2013). Becoming part of the story! Refueling the interest in visualization strategies for reading comprehension. *Educational Psychology Review*, *25*(2), 261–287. http://doi.org/10.1007/s10648-013-9222-6

de Lemnos, M. M. (1989). *Standard Progressive Matrices: Australian manual.* Hawthorn, Victoria: Australian Council for Academic Research (ACER).

Dean, G. M., & Morris, P. E. (1991). Imagery and spatial ability: When introspective reports predict performance. In R. H. Logie & M. Denis (Eds.), *Mental images in human cognition* (pp. 331–347). New York, NY: Elsevier.

Dean, G. M., & Morris, P. E. (2003). The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology*, *94*(2), 245–271. http://doi.org/10.1348/000712603321661912

Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*(10), 1189–1199. http://doi.org/10.1016/S0028-3932(98)00159-6

Dempster, F. N. (1978). Memory span and short-term memory capacity: A developmental study. *Journal of Experimental Child Psychology*, *26*(3), 419–431. http://doi.org/10.1016/0022-0965(78)90122-4

Denis, M., & Cocude, M. (1997). On the metric properties of visual images generated from verbal descriptions: Evidence for the robustness of the mental scanning effect. *European Journal of Cognitive Psychology*, *9*(4), 353–379. http://doi.org/10.1080/713752568

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence*, *21*(2), 135–155. http://doi.org/10.1016/0160-2896(95)90023-3

Dijkstra, K., Yaxley, R. H., Madden, C. J., & Zwaan, R. A. (2004). The role of age and perceptual symbols in language comprehension. *Psychology and Aging*, *19*(2), 352–356. http://doi.org/10.1037/0882-7974.19.2.352

Dror, I. E., & Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and Aging*, *9*(1), 90–102. http://doi.org/10.1037/0882-7974.9.1.90

Dror, I. E., Kosslyn, S. M., & Waag, W. L. (1993). Visual-spatial abilities of pilots. *Journal of Applied Psychology*, *78*(5), 763–773. http://doi.org/10.1037//0021-9010.78.5.763

Durndell, A. J., & Wetherick, N. E. (1976). The relation of reported imagery to cognitive performance. *British Journal of Psychology*, *67*(4), 501–506. http://doi.org/10.1111/j.2044-8295.1976.tb01538.x

Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading*, *17*(6), 435–452. http://doi.org/10.1080/10888438.2013.774005

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*(4), 659–675. http://doi.org/10.1016/j.jecp.2011.06.009

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2014). Eye movements reveal differences in children's referential processing during narrative comprehension. *Journal of Experimental Child Psychology*, *118*(C), 57–77. http://doi.org/10.1016/j.jecp.2013.09.005

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). London, UK: Elsevier. http://doi.org/10.1016/s0079-7421(03)44005-x

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. http://doi.org/10.1037/0096-3445.128.3.309

Ernest, C. H. (1977). Imagery ability and cognition: A critical review. *Journal of Mental Imagery*, *2*(1), 181–216. Retrieved from http://psycnet.apa.org/psycinfo/1978-30907-001

Estes, D. (1998). Young children's awareness of their mental activity: The case of mental rotation. *Child Development, 69*(5), 1345–1360. http://doi.org/10.2307/1132270

Estevez, A., & Calvo, M. G. (2000). Working memory capacity and time course of predictive inferences. *Memory, 8*(1), 51–61. http://doi.org/10.1080/096582100387704

Fecica, A. M., & O'Neill, D. K. (2010). A step at a time: Preliterate children's simulation of narrative movement during story comprehension. *Cognition, 116*(3), 368–381. http://doi.org/10.1016/j.cognition.2010.05.014

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16*(1-2), 99–124. http://doi.org/10.1080/01638539309544831

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 981–995. http://doi.org/10.1037/0278-7393.21.4.981

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes, 22*(3), 225–246. http://doi.org/10.1080/01638539609544974

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition, 29*(2), 336–343. http://doi.org/10.3758/bf03194928

Fincher-Kiefer, R., & D'Agostino, P. R. (2004). The role of visuospatial resources in generating predictive and bridging inferences. *Discourse Processes, 37*(3), 205–224. http://doi.org/10.1207/s15326950dp3703_2

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology, 61*(6), 825–850. http://doi.org/10.1080/17470210701623605

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language, 27*(3), 235–244. http://doi.org/10.1016/0749-596x(88)90052-6

Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes, 48*(2), 119–138. http://doi.org/10.1080/0163853X.2010.494244

Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology, 27*(4), 935–951. http://doi.org/10.1348/026151008x397189

Francey, G., & Cain, K. (2015). Effect of imagery training on children's comprehension of pronouns. *The Journal of Educational Research, 108*(1), 1–9. http://doi.org/10.1080/00220671.2013.824869

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, *10*(3), 301–322. http://doi.org/10.1207/s1532799xssr1003_6

Frick, A., Daum, M. M., Walser, S., & Mast, F. W. (2009). Motor processes in children's mental rotation. *Journal of Cognition and Development*, *10*(1-2), 18–40. http://doi.org/10.1080/15248370902966719

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, *129*(1), 61–83. http://doi.org/10.1037/0096-3445.129.1.61

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256. http://doi.org/10.1207/ S1532799XSSR0503_3

Funk, M., Brugger, P., & Wilkening, F. (2005). Motor processes in children's imagery: The case of mental rotation of hands. *Developmental Science*, *8*(5), 402–408. http://doi.org/10.1111/j.1467-7687.2005.00428.x

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3-4), 455–479. http://doi.org/10.1080/02643290442000310

Galton, F. (1883). *Inquiries into human faculty and its development*. London, UK: MacMillon.

Gambrell, L. B., & Bales, R. J. (1986). Mental imagery and the comprehension-monitoring performance of fourth-and fifth-grade poor readers. *Reading Research Quarterly*, *21*(4), 454–464. http://doi.org/10.2307/747616

Gambrell, L. B., & Jawitz, P. B. (1993). Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly*, *28*(3), 265–276. http://doi.org/10.2307/747998

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating "rotators" from "nonrotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research*, *41*(3), 261–293. http://doi.org/10.1207/s15327906mbr4103_2

Geiser, C., Lehmann, W., Corth, M., & Eid, M. (2008). Quantitative and qualitative change in children's mental rotation performance. *Learning and Individual Differences*, *18*(4), 419–429. http://doi.org/10.1016/j.lindif.2007.09.001

Gernsbacher, M. A. (1990). *Language as structure building*. Hillsdale, New Jersey: Lawrence Erlbaum & Associates.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, *23*(3), 265–304. http://doi.org/10.1080/01638539709544994

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 245–262. http://doi.org/10.1037//0278-7393.17.2.245

Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. W. (1992). Do readers mentally represent characters emotional states? *Cognition & Emotion*, *6*(2), 89–111. http://doi.org/10.1080/02699939208411061

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 430–445. http://doi.org/10.1037/0278-7393.16.3.430

Giesen, C., & Peeck, J. (1984). Effects of imagery instruction on reading and retaining a literary text. *Journal of Mental Imagery*, *8*(2), 79–90. Retrieved from http://psycnet.apa.org/psycinfo/1985-29914-001

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1–55. http://doi.org/10.1017/s0140525x97000010

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. http://doi.org/10.3758/bf03196313

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*(1), 1–26. http://doi.org/10.1080/01638539909545067

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*(3), 379–401. http://doi.org/10.1006/jmla.2000.2714

Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2011). Improving early reading comprehension using embodied CAI. *Instructional Science*, *39*(1), 27–39. http://doi.org/10.1007/s11251-009-9096-7

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*(3), 424–436. http://doi.org/10.1037/0022-0663.96.3.424

Glenberg, A. M., Kruley, P., & Langston, W. (1994). Analogical processes in comprehension. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 609–640). San Diego, CA: Academic Press.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*(1), 69–83. http://doi.org/10.1016/0749-596X(87)90063-5

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing*, *18*(7-9), 583–616. http://doi.org/10.1007/s11145-004-7109-0

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006a). Distinct and common cortical activations for multimodal semantic categories. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(3), 214–222. http://doi.org/10.3758/cabn.6.3.214

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006b). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*, *26*(18), 4917–4921. http://doi.org/10.1523/jneurosci.5389-05.2006

Gordon, R. (1949). An investigation into some of the factors that favour the formation of stereotyped images. *British Journal of Psychology*, *39*(3), 156–167. http://doi.org/10.1111/j.2044-8295.1949.tb00215.x

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology*, *63*(3), 563–582. http://doi.org/10.1006/jecp.1996.0062

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

Graesser, A. C., Bertus, E. L., & Magliano, J. P. (1995). Inference generation during the comprehension of narrative text. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 295–320). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. Mahwah, NJ.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. http://doi.org/10.1037/0033-295x.101.3.371

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. http://doi.org/10.1037/0022-3514.79.5.701

Green, M. C., & Brock, T. C. (2002). In the mind's eye: Transportation-imagery model of narrative persuasion. In M. C. Green & T. C. Strange (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 315–341). Mahwah, NJ: Lawrence Erlbaum & Associates.

Green, M. C., & Donahue, J. K. (2008). Simulated worlds: Transportation into narratives. In K. D. Markman, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 241–254). New York, NY: Psychology Press.

Gullick, M. M., Mitra, P., & Coch, D. (2013). Imagining the truth and the moon: An electrophysiological study of abstract and concrete word processing. *Psychophysiology*, *50*(5), 431–440. http://doi.org/10.1111/psyp.12033

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (3rd ed., pp. 403–422). New York: Longman.

Gyselinck, V., De Beni, R., Pazzaglia, F., Meneghetti, C., & Mondoloni, A. (2007). Working memory components and imagery instructions in the elaboration of a spatial mental model. *Psychological Research*, *71*(3), 373–382. http://doi.org/10.1007/s00426-006-0091-1

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology*, *22*(3), 353–374. http://doi.org/10.1002/acp.1411

Gyselinck, V., Meneghetti, C., De Beni, R., & Pazzaglia, F. (2009). The role of working memory in spatial text processing: What benefit of imagery strategy and visuospatial abilities? *Learning and Individual Differences*, *19*(1), 12–20. http://doi.org/10.1016/j.lindif.2008.08.002

Haenggi, D., Kintsch, W., & Gernsbacher, M. A. (1995). Spatial situation models and text comprehension. *Discourse Processes*, *19*(2), 173–199. http://doi.org/10.1080/01638539509544913

Hakala, C. M. (1999). Accessibility of spatial information in a situation model. *Discourse Processes*, *27*(3), 261–279. http://doi.org/10.1080/01638539909545063

Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1294–1313. http://doi.org/10.1037/a0020363

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*(1), 103–128. http://doi.org/10.1037//0022-0663.93.1.103

Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: Similarities and differences between mental rotation and mental folding. *Cognitive Processing*, *14*(2), 105–115. http://doi.org/10.1007/s10339-013-0544-6

Hartung, F., Hagoort, P., & Willems, R. (2015). Simulation versus mental imagery: commonalities and differences (pp. xxiii–xxiv). Presented at the 8th annual Conference on Embodied and Situated Language Processing (ESLP), Lyon, France. Retrieved from http://pubman.mpdl.mpg.de/pubman/item/escidoc:2227163:2/component/escidoc:2227169/Hartung_Hagoort_Willems_2015_Lyon.pdf

Hauk, O., & Pulvermuller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191–201. http://doi.org/10.1002/hbm.10157

Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307. http://doi.org/10.1016/s0896-6273(03)00838-9

Hecker, R., & Mapperson, B. (1997). Dissociation of visual and spatial processing in working memory. *Neuropsychologia*, *35*(5), 599–603. http://doi.org/10.1016/s0028-3932(96)00106-6

Hiscock, M. (1978). Imagery assessment through self-report: What do imagery questionnaires measure? *Journal of Consulting and Clinical Psychology*, *46*(2), 223–230. http://doi.org/10.1037/0022-006x.46.2.223

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*(1), 104–110. http://doi.org/10.3758/bf03196473

Huitema, J. S., Dopkins, S., Klin, C. M., & Myers, J. L. (1993). Connecting goals and actions during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1053–1060. http://doi.org/10.1007/s11145-013-9437-4

Ingram, J., Hand, C. J., & Moxey, L. M. (2014). Processing inferences drawn from the logically equivalent frames half full and half empty. *Journal of Cognitive Psychology*, *26*(7), 799–817. http://doi.org/10.1080/20445911.2014.956747

Isaac, A. R., & Marks, D. F. (1994). Individual differences in mental imagery experience: Developmental changes and specialization. *British Journal of Psychology*, *85*(4), 479–500. http://doi.org/10.1111/j.2044-8295.1994.tb02536.x

Isaac, A. R., Marks, D. F., & Russell, D. G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire (VMIQ). *Journal of Mental Imagery*, *10*(4), 23–30. http://doi.org/10.1037/t07980-000

Jansen, P., Schmelter, A., Quaiser-Pohl, C., Neuburger, S., & Heil, M. (2013). Mental rotation performance in primary school age children: Are there gender differences in chronometric tests? *Cognitive Development*, *28*(1), 51–62. http://doi.org/10.1016/j.cogdev.2012.08.005

Jarvis, B. (2006). DirectRT (Version 2002) [Computer software]. New York, NY: Empirisoft.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C. A., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, *95*(4), 719–729. http://doi.org/10.1037/0022-0663.95.4.719

Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., & Binkofski, F. (2010). Grasping language – A short story on embodiment. *Consciousness and Cognition*, *19*(3), 711–720. http://doi.org/10.1016/j.concog.2010.06.020

Johnson, M. (1987). *The body and the mind: The bodily basis of meaning, imagination, and reason*. Chicago, IL: University of Chicago Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. http://doi.org/10.1037//0033-295x.99.1.122

Just, M. A., Newman, S. D., Keller, T. A., McEleney, A., & Carpenter, P. A. (2004). Imagery in sentence comprehension: an fMRI study. *NeuroImage*, *21*(1), 112–124. http://doi.org/10.1016/j.neuroimage.2003.08.042

Kail, R., Pellegrino, J., & Carter, P. (1980). Developmental changes in mental rotation. *Journal of Experimental Child Psychology*, *29*(1), 102–116. http://doi.org/10.1016/0022-0965(80)90094-6

Kamphaus, R. W., & Pleiss, K. L. (1992). Draw-a-Person techniques: Tests in search of a construct. *Journal of School Psychology*, *29*(4), 395–401. http://doi.org/10.1016/0022-4405(91)90026-n

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. http://doi.org/10.3758/BF03196323

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York, NY: Oxford University Press.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. http://doi.org/10.1037/0033-2909.131.1.66

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M. E., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*(3), B79–B89. http://doi.org/10.1016/j.cognition.2004.06.005

Katz, S., Blackburn, A. B., & Lautenschlager, G. J. (1991). Answering reading comprehension items without passages on the SAT when items are quasi-randomized. *Psychological Science*, *51*(3). http://doi.org/10.1177/0013164491513026

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language*, *32*(4), 446–463. http://doi.org/10.1006/jmla.1993.1024

Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading*, *10*(4), 363–380. http://doi.org/10.1207/s1532799xssr1004_2

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*(3), 281–300. http://doi.org/10.1080/10888430802132279

Keenan, J. M., Golding, J. M., Potts, G. R., Jennings, T. M., & Aman, C. J. (1990). Methodological issues in evaluating the occurrence of inferences. *Psychology of Learning and Motivation*, *Volume 25*, 295–312. http://doi.org/10.1016/s0079-7421(08)60261-3

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica*, *124*(3), 382–397. http://doi.org/10.1016/j.actpsy.2006.02.005

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1–11. http://doi.org/10.3758/bf03193491

Kendeou, P., Bohn-Gettler, C. M., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*(3), 259–272. http://doi.org/10.1111/j.1467-9817.2008.00370.x

Kendeou, P., Lynch, J. S., van den Broek, P., Espin, C. A., White, M. J., & Kremer, K. E. (2005). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal*, *33*(2), 91–98. http://doi.org/10.1007/s10643-005-0030-6

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction*, *22*(5), 354–367. http://doi.org/10.1016/j.learninstruc.2012.02.001

Kendeou, P., Savage, R., & van den Broek, P. (2009a). Revisiting the simple view of reading. *British Journal of Educational Psychology*, *79*(2), 353–370. http://doi.org/10.1348/978185408X369020

Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, *29*(1), 10–16. http://doi.org/10.1111/ldrp.12025

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009b). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*(4), 765–778. http://doi.org/10.1037/a0015956

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PLoS ONE*, *6*(12), e29221. http://doi.org/10.1371/journal.pone.0029221.g003

Kim, Y. S. (2014). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, *86*(1), 128–144. http://doi.org/10.1111/cdev.12293

Kim, Y. S., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of Experimental Child Psychology*, *113*(1), 93–111. http://doi.org/10.1016/j.jecp.2012.03.002

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy enlisted personnel. Research Branch report 8-75. Chief of Naval technical training: Naval air station Memphis. *Scientific Studies of Reading*, *12*, 281–300. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a006655.pdf

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. http://doi.org/10.1037/0033-295x.95.2.163

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. http://doi.org/10.1037/0033-295x.85.5.363

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, *100*(2), 310–321. http://doi.org/10.1037/0022-0663.100.2.310

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General*, *133*(3), 355–381. http://doi.org/10.1037/0096-3445.133.3.355

Klin, C. M., & Drumm, A. M. (2010). Seeing what they read and hearing what they say: Readers' representation of the story characters' world. *Psychonomic Bulletin & Review*, *17*(2), 231–236. http://doi.org/10.3758/PBR.17.2.231

Klin, C. M., Guzmán, A. E., & Levine, W. H. (1999a). Prevalence and persistence of predictive inferences. *Journal of Memory and Language*, *40*(4), 593–604. http://doi.org/10.1006/jmla.1998.2628

Klin, C. M., Murray, J. D., Levine, W. H., & Guzmán, A. E. (1999b). Forward inferences: From activation to long-term memory. *Discourse Processes*, *27*(3), 241–260. http://doi.org/10.1080/01638539909545062

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Processing*, *7*(3), 341–370. http://doi.org/10.1016/0010-0285(75)90015-8

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1983). *Ghosts in the mind's machine: Creating and using images in the brain*. London, England: Norton.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Boston, MA: MIT Press.

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, *22*(3-4), 333–347. http://doi.org/10.1080/02643290442000130

Kosslyn, S. M., Brunn, J., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition*, *18*(1), 195–243. http://doi.org/10.1016/0010-0277(84)90025-8

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642. http://doi.org/10.1038/35090055

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2003). Mental imagery: Against the nihilistic hypothesis. *Trends in Cognitive Sciences*, *7*(3), 109–111. http://doi.org/10.1016/S1364-6613(03)00025-1

Kosslyn, S. M., Holtzman, J. D., Farah, M. J., & Gazzaniga, M. S. (1985). A computational analysis of mental image generation: Evidence from functional dissociations in split-brain patients. *Journal of Experimental Psychology: General*, *114*(3), 311–341. http://doi.org/10.1037/0096-3445.114.3.311

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age differences in imagery abilities. *Child Development*, *61*(4), 995–1010. http://doi.org/10.2307/1130871

Kosslyn, S. M., & Moulton, S. T. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B, 364,* 1273–1280. http://doi.org10.1098/rstb.2008.0314

Kosslyn, S. M., Thompson, W. L., Shephard, J., Ganis, G., Bell, D., Danovitch, J., et al. (2004). Brain rCBF and performance in visual imagery tasks: Common and distinct processes. *European Journal of Cognitive Psychology*, *16*(5), 696–716. http://doi.org/10.1080/09541440340000475

Kosslyn, S. M., Thompson, W. L., Sukel, K. E., & Alpert, N. M. (2005). Two types of image generation: Evidence from PET. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 41–53. http://doi.org/10.3758/cabn.5.1.41

Kruley, P., Sciama, S. C., & Glenberg, A. M. (1994). On-line processing of textual illustrations in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition*, *22*(3), 261–272. http://doi.org/10.3758/bf03200853

Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, *45*(2), 230–251. http://doi.org/10.1598/RRQ.45.2.4

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–1246. http://doi.org/10.1162/jocn.2010.21452

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. http://doi.org/10.1038/307161a0

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Processing*, *6*(2), 293–323. http://doi.org/10.1016/0010-0285(74)90015-2

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing*, *23*(6), 707–717. http://doi.org/10.1007/s11145-009-9180-z

Langston, W., Kramer, D., & Glenberg, A. M. (1998). The representation of space in mental models derived from text. *Memory & Cognition*, *26*(2), 247–262. http://doi.org/10.3758/bf03201137

Lassonde, K. A., & O'Brien, E. J. (2009). Contextual specificity in the activation of predictive inferences. *Discourse Processes*, *46*(5), 426–438. http://doi.org/10.1080/01638530902959620

Lepola, J., Lynch, J. S., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Reading Research Quarterly*, *47*(3), 258–282. http://doi.org/10.1002/RRQ.020

Lequerica, A., Rapport, L., Axelrod, B. N., Telmet, K., & Whitman, R. D. (2002). Subjective and objective assessment methods of mental imagery control: Construct validations of self-report measures. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1103–1116. http://doi.org/10.1076/jcen.24.8.1103.8370

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Lawrence Erlbaum & Associates.

Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1424–1429. http://doi.org/10.1037/0278-7393.27.6.1424

Long, D. L., & Golding, J. M. (1993). Subordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes*, *16*(1-2), 55–73. http://doi.org/10.1080/01638539309544829

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology*, *20*(6), 1456–1470. http://doi.org/10.1037/0278-7393.20.6.1456

Long, S. A., Winograd, P. N., & Bridge, C. A. (1989). The effects of reader and text characteristics on imagery reported during and after reading. *Reading Research Quarterly*, *24*(3), 353–372. http://doi.org/10.2307/747774

Lorsbach, T. C., & Reimer, J. F. (1997). Developmental changes in the inhibition of previously relevant information. *Journal of Experimental Child Psychology*, *64*(3), 317–342. http://doi.org/10.1006/jecp.1996.2350

Lorsbach, T. C., Katz, G. A., & Cupak, A. J. (1998). Developmental differences in the ability to inhibit the initial misinterpretation of garden path passages. *Journal of Experimental Child Psychology*, *71*(3), 275–296. http://doi.org/10.1006/jecp.1998.2462

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372. http://doi.org/10.1111/j.1467-8624.2004.00745.x

Lynch, J. S., & van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on- and off-line inferences about characters' goals. *Cognitive Development*, *22*(3), 323–340. http://doi.org/10.1016/j.cogdev.2007.02.002

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, *29*(4), 327–365. http://doi.org/10.1080/02702710802165416

Lynn, R., Allik, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's Standard Progressive Matrices. *Intelligence*, *32*(4), 411–424. http://doi.org/10.1016/j.intell.2004.06.007

Magliano, J. P., & Trabasso, T. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*(4), 615–629. http://doi.org/10.1037//0022-0663.91.4.615

Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes*, *16*(1-2), 35–53. http://doi.org/10.1080/01638539309544828

Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning*, *6*(2), 131–154. http://doi.org/10.1007/s11409-010-9064-2

Maher, J. H., & Sullivan, H. (1982). Effects of mental imagery and oral and print stimuli on prose learning of intermediate grade children. *Educational Technology Research and Development, 30*(3), 175–183. Retrieved from http://link.springer.com/article/10.1007%2FBF02766598#page-1

Mammarella, I. C., Pazzaglia, F., & Cornoldi, C. (2006). The assessment of imagery and visuo-spatial working memory functions in children and adults. In T. Vecchi & G. Bottini (Eds.), *Imagery and spatial cognition: Methods, models and cognitive assessment* (pp. 28–51). Amsterdam, Netherlands: John Benjamins Publishing.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24. http://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery*, *19*(1), 153–167. Retrieved from http://psycnet.apa.org/psycinfo/1996-29150-001

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. *Contemporary Educational Psychology*, *32*(3), 537–550. http://doi.org/10.1016/j.cedpsych.2007.03.003

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2010). What cognitive benefits does an activity-based reading strategy afford young Native American readers? *The Journal of Experimental Education*, *78*(3), 395–417. http://doi.org/10.1080/00220970903548061

Marmolejo-Ramos, F., de Juan, M. R. E., Gyax, P., Madden, C. J., & Roa, S. M. (2009). Reading between the lines: The activation of background knowledge during text comprehension. *Pragmatics & Cognition*, *17*(1), 77–107. http://doi.org/10.1075/pc.17.1.03mar

Marmor, G. S. (1975). Development of kinetic images: When does the child first represent movement in mental images? *Cognitive Processing*, *7*(4), 548–559. http://doi.org/10.1016/0010-0285(75)90022-5

Marmor, G. S. (1977). Mental rotation and number conservation: Are they related? *Developmental Psychology*, *13*(4), 320. http://doi.org/10.1037//0012-1649.13.4.320

Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, *75*(2), 314–318. http://doi.org/10.1037/0022-0663.75.2.314

McAvinue, L. P., & Robertson, I. H. (2006). Measuring visual imagery ability: A review. *Imagination, Cognition and Personality*, *26*(3), 191–211. http://doi.org/10.2190/3515-8169-24j8-7157

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–446. http://doi.org/10.1037/0033-295x.99.3.440

McMaster, K. L., Espin, C. A., & van den Broek, P. (2014). Making connections: Linking cognitive psychology and intervention research to improve comprehension of struggling readers. *Learning Disabilities Research & Practice*, *29*(1), 17–24. http://doi.org/10.1111/ldrp.12026

McMaster, K. L., van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., et al. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, *22*(1), 100–111. http://doi.org/10.1016/j.lindif.2011.11.017

McNamara, D. S., & Magliano, J. P. (2009). Toward a comprehensive model of comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51). Burlington, MA: Academic Press.

Mehta, Z., Newcombe, F., & De Haan, E. (1992). Selective loss of imagery in a case of visual agnosia. *Neuropsychologia*, *30*(7), 645–655. http://doi.org/10.1016/0028-3932(92)90069-x

Mellard, D. F., Fall, E., & Woods, K. L. (2010). A path analysis of reading comprehension for adults with low literacy. *Journal of Learning Disabilities*, *43*(2), 154–165. http://doi.org/10.1177/0022219409359345

Meneghetti, C., De Beni, R., Gyselinck, V., & Pazzaglia, F. (2013). The joint role of spatial ability and imagery strategy in sustaining the learning of spatial descriptions under spatial interference. *Learning and Individual Differences*, *24*(C), 32–41. http://doi.org/10.1016/j.lindif.2012.12.021

Meneghetti, C., Gyselinck, V., Pazzaglia, F., & De Beni, R. (2009). Individual differences in spatial text processing: High spatial ability can compensate for spatial working memory interference. *Learning and Individual Differences*, *19*(4), 577–589. http://doi.org/10.1016/j.lindif.2009.07.007

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621–640. http://doi.org/10.1037//0096-3445.130.4.621

Moore, P. J., & Kirby, J. R. (1988). Comprehension training and reading performance. *Reading*, *22*(2), 126–136. http://doi.org/10.1111/j.1467-9345.1988.tb00665.x

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, *28*(3), 292–312. http://doi.org/10.1016/0749-596x(89)90035-1

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, *26*(2), 165–187. http://doi.org/10.1016/0749-596x(87)90122-7

Mueller, S. T. (2013). The Psychology Experiment Building Language (Version 0.13) [Computer software]. Retrieved from http://pebl. sourceforge.net

Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes*, *35*(2), 81–102. http://doi.org/10.1207/S15326950DP3502_1

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language*, *32*(4), 464–473. http://doi.org/10.1006/jmla.1993.1025

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*(2-3), 131–157. http://doi.org/10.1080/01638539809545042

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *4*(20), 876–886. http://doi.org/10.1037/0278-7393.20.4.876

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, *91*(3), 488–496. http://doi.org/10.1037//0022-0663.91.3.488

Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Malden, MA: Blackwell Publishing.

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, *67*(3), 359–370. http://doi.org/10.1111/j.2044-8279.1997.tb01250.x

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*(1), 85–101. http://doi.org/10.1006/jmla.1998.2564

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition*, *70*(1), B1–B13. http://doi.org/10.1016/s0010-0277(99)00004-9

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics*, *21*(1), 229–241. http://doi.org/10.1017/s0142716400002046

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, *73*(2), 139–158. http://doi.org/10.1006/jecp.1999.2498

Neale, M. D. (1999). *Neale analysis of reading ability: Manual*. Melbourne, Australia: Australian Council for Academic Research (ACER).

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Nyhout, A., & O'Neill, D. K. (2013). Constructing spatial representations from narratives and non-narrative descriptions: Evidence from 7-year-olds. *Workshop on Computational Models of Narrative*, 158–165. http://doi.org/10.4230/OASIcs.CMN.2013.158

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 777–784. http://doi.org/10.1037/0278-7393.18.4.777

O'Brien, E. J., & Myers, J. L. (1999). Text comprehension: A view from the bottom up. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 35–53). Mahwah, New Jersey: Lawrence Erlbaum & Associates.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1200–1210. http://doi.org/10.1037/0278-7393.24.5.1200

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 410–420. http://doi.org/10.1037//0278-7393.14.3.410

O'Neill, D. K., & Shultis, R. M. (2007). The emergence of the ability to track a character's mental perspective in narrative. *Developmental Psychology*, *43*(4), 1032–1037. http://doi.org/10.1037/0012-1649.43.4.1032

Oakhill, J. V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, *54*(1), 31–39. http://doi.org/10.1111/j.2044-8279.1984.tb00842.x

Oakhill, J. V. (1994). Individual differences in children's text comprehension. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 821–848). San Diego, California: Academic Press.

Oakhill, J. V., & Cain, K. (2000). Children's difficulties in text comprehension: Assessing causal issues. *Journal of Deaf Studies and Deaf Education*, *5*(1), 51–59. http://doi.org/10.1093/deafed/5.1.51

Oakhill, J. V., & Cain, K. (2007). Issues of causality in children's reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: theories, interventions, and technologies* (pp. 47–72). Mahwah, NJ: Lawrence Erlbaum & Associates.

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91–121. http://doi.org/10.1080/10888438.2010.529219

Oakhill, J. V., & Patel, S. (1991). Can imagery training help children who have comprehension problems? *Journal of Research in Reading*, *14*(2), 106–115. http://doi.org/10.1111/j.1467-9817.1991.tb00012.x

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, *18*(4), 443–468. http://doi.org/10.1080/01690960344000008

Oakhill, J. V., Garnham, A., & Reynolds, D. (2005a). Immediate activation of stereotypical gender activation. *Memory & Cognition*, *33*(6), 972–983. http://doi.org/10.3758/bf03193206

Oakhill, J. V., Hartt, J., & Samols, D. (2005b). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*(7-9), 657–686. http://doi.org/10.1007/s11145-005-3355-z

OECD. (2001). Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy. OECD Publishing. Retrieved from http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33692793.pdf

OECD. (2014). PISA 2012 Results: What students know and can do: Student performance in mathematics, reading and science (Vol. 1). OECD Publishing. http://doi.org/10.1787/9789264201118-en

Olivetti Belardinelli, M., Palmiero, M., Sestieri, C., Nardo, D., Di Matteo, R., Londei, A., et al. (2009). An fMRI investigation on image generation in different sensory modalities: The influence of vividness. *Acta Psychologica*, *132*(2), 190–200. http://doi.org/10.1016/j.actpsy.2009.06.009

Orrantia, J., Múñez, D., & Tarín, J. (2014). Connecting goals and actions during reading: the role of illustrations. *Reading and Writing*, *27*(1), 153–170. http://doi.org/10.1007/s11145-013-9437-4

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods*, *40*(4), 1001–1015. http://doi.org/10.3758/BRM.40.4.1001

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart, & Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Paivio, A. (2007). *Mind and its evolution: A dual coding approach*. New Jersey, NJ: Lawrence Erlbaum & Associates.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition*, *29*(2), 344–354. http://doi.org/10.3758/bf03194929

Palmiero, M., Olivetti Belardinelli, M., Nardo, D., Sestieri, C., Di Matteo, R., D'Ausilio, A., & Romani, G. L. (2009). Mental imagery generation in different modalities activates sensory-motor areas. *Cognitive Processing*, *10*(S2), 268–271. http://doi.org/10.1007/s10339-009-0324-5

Paris, A. H., & Paris, S. G. (2007). Teaching narrative comprehension strategies to first graders. *Cognition and Instruction*, *25*(1), 1–2. http://doi.org/10.1080/07370000709336701

Pearson, D. G., Deeprose, C., Wallace-Hadrill, S. M. A., Heyes, S. B., & Holmes, E. A. (2013). Assessing mental imagery in clinical psychology: A review of imagery measures and a guiding framework. *Clinical Psychology Review*, *33*(1), 1–23. http://doi.org/10.1016/j.cpr.2012.09.001

Pearson, D. G., Logie, R. H., & Gilhooly, K. J. (1999). Verbal representations and spatial manipulation during mental synthesis. *European Journal of Cognitive Psychology*, *11*(3), 295–314. http://doi.org/10.1080/713752317

Pearson, J., & Clifford, C. W. (2005). Mechanisms selectively engaged in rivalry: Normal vision habituates, rivalrous vision primes. *Vision Research*, *45*(6), 707–714. http://doi.org/10.1016/j.visres.2004.09.040

Pearson, J., Clifford, C. W., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology*, *18*(13), 982–986. http://doi.org/10.1016/j.cub.2008.05.048

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the mind's eye: The metacognition of visual imagery. *Psychological Science*, *22*(12), 1535–1542. http://doi.org/10.1177/0956797611417134

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. American Psychological Association.

Perfetti, C. A., Landi, N., & Oakhill, J. V. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Malden, MA: Blackwell Publishing.

Pérez, A. I., Paolieri, D., Macizo, P., & Bajo, T. (2014). The role of working memory in inferential sentence comprehension. *Cognitive Processing*, *15*(3), 405–413. http://doi.org/10.1007/s10339-014-0611-7

Pham, A. V., & Hasson, R. M. (2014). Verbal and visuospatial working memory as predictors of children's reading ability. *Archives of Clinical Neuropsychology*, *29*(5), 467–477. http://doi.org/10.1093/arclin/acu024

Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). Journal of Experimental Child Psychology. *Journal of Experimental Child Psychology*, *105*(3), 243–255. http://doi.org/10.1016/j.jecp.2009.10.006

Poltrock, S. E., & Brown, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence*, *8*(2), 93–138. http://doi.org/10.1016/0160-2896(84)90019-9

Potocki, A., Ecalle, J., & Magnan, A. (2013). Narrative comprehension skills in 5-year-old children: Correlational analysis and comprehender profiles. *The Journal of Educational Research*, *106*(1), 14–26. http://doi.org/10.1080/00220671.2012.667013

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, *27*(4), 399–415. http://doi.org/10.1016/0749-596x(88)90064-2

Pressley, G. M. (1976). Mental imagery helps eight-year-olds remember what they read. *Journal of Educational Psychology*, *68*(3), 355–359. http://doi.org/10.1037//0022-0663.68.3.355

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nature*, *6*(7), 576–582. http://doi.org/10.1038/nrn1706

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*(1), 1–24. http://doi.org/10.1037/h0034650

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*(2), 157–182. http://doi.org/10.1017/s0140525x02000043

Pylyshyn, Z. W. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences*, *7*(3), 113–118. http://doi.org/10.1016/S1364-6613(03)00003-2

Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology*, *48*(2), 521–529. http://doi.org/10.1037/a0025526

Quaiser-Pohl, C., Rohe, A. M., & Amberger, T. (2010). The solution strategy as an indicator of the developmental stage of preschool children's mental-rotation ability. *Journal of Individual Differences*, *31*(2), 95–100. http://doi.org/10.1027/1614-0001/a000017

Quinn, J. G. (2008). Movement and visual coding: The structure of visuo-spatial working memory. *Cognitive Processing*, *9*(1), 35–43. http://doi.org/10.1007/s10339-007-0184-9

Quinn, J. G., & McConnell, J. (2006). The interval for interference in conscious visual imagery. *Memory*, *14*(2), 241–252. http://doi.org/10.1080/09658210500210019

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition*, *34*(5), 1150–1156. http://doi.org/10.3758/bf03193261

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review*, *14*(6), 1027–1042. http://doi.org/10.3758/bf03193088

Rall, J., & Harris, P. L. (2000). In Cinderella's slippers? Story comprehension from the protagonist's point of view. *Developmental Psychology*, *36*(2), 202–208. http://doi.org/10.1037//0012-1649.36.2.202

Raney, G. E., Campbell, S. J., & Bovee, J. C. (2014). Using eye movements to evaluate the cognitive processes involved in text comprehension. *Journal of Visualized Experiments*, *83*, 1–7. http://doi.org/10.3791/50780

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396. http://doi.org/10.1016/j.neuropsychologia.2008.09.017

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, *35*(8), 2019–2032. http://doi.org/10.3758/bf03192934

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, *11*(4), 289–312. http://doi.org/10.1080/10888430701530417

Raven, J. C. (1958). *Standard Progressive Matrices*. Hawthorne, Victoria: Australia Council for Academic Research (ACER).

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255. http://doi.org/10.1207/s1532799xssr1003_3

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, *80*(1), 16–20. http://doi.org/10.1037/0022-0663.80.1.16

Richardson, A. (1977). The meaning and measurement of memory imagery. *British Journal of Psychology*, *68*(1), 29–43. http://doi.org/10.1111/j.2044-8295.1977.tb01556.x

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, *42*(4), 546–567. http://doi.org/10.1598/RRQ.42.4.5

Rinck, M., Bower, G. H., & Wolf, K. (1998). Distance effects in surface structures and situation models. *Scientific Studies of Reading*, *2*(3), 221–246. http://doi.org/10.1207/s1532799xssr0203_3

Rinck, M., Williams, P., Bower, G. H., & Becker, E. S. (1996). Spatial situation models and narrative understanding: Some generalizations and extensions. *Discourse Processes*, *21*(1), 23–55. http://doi.org/10.1080/01638539609544948

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology*, *46*(3), 343–366. http://doi.org/10.1016/j.jsp.2007.06.006

Rose, M. C., Cundick, B. P., & Higbee, K. L. (1983). Verbal rehearsal and visual imagery mnemonic aids for learning-disabled children. *Journal of Learning Disabilities*, *16*(6), 352–354. http://doi.org/10.1177/002221948301600610

Rowe, M., Ozuru, Y., & McNamara, D. S. (2006). An analysis of standardized reading ability tests: what do questions actually measure? In S. A. Barab, K. E. Kay, & D. T. Hickey (Eds.), *Proceedings of the 7th International Conference on Learning Sciences* (pp. 627–633). Mahway, NJ: Lawrence Erlbaum & Associates. Retrieved from http://129.219.222.66/Publish/pdf/ROM_an%20analysis%20of%20standardized_2006.pdf

Rupley, W. H., Willson, V. L., & Nichols, W. D. (1998). Exploration of the developmental components contributing to elementary school children's reading comprehension. *Scientific Studies of Reading*, *2*(2), 143–158. http://doi.org/10.1207/s1532799xssr0202_3

Sadoski, M. (1983). An exploratory study of the relationships between reported imagery and the comprehension and recall of a story. *Reading Research Quarterly*, *19*(1), 110–123. http://doi.org/10.2307/747341

Sadoski, M. (1985). The natural use of imagery in story comprehension and recall: Replication and extension. *Reading Research Quarterly*, *20*(5), 658–667. http://doi.org/10.2307/747949

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwa, NJ: Lawrence Erlbaum & Associates.

Sadoski, M., & Pavio, A. (2004). A dual coding theoretical model of reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5 ed., pp. 1329–1362). Newark, DE: International Reading Association.

Sadoski, M., Goetz, E. T., & Fritz, J. B. (1993). Impact of concreteness on comprehensibility, interest, and memory for text: implications for dual coding theory and text design. *Journal of Educational Psychology*, *85*(2), 291–304. http://doi.org/10.1037/0022-0663.85.2.291

Sanchez, C. A., & Wiley, J. (2014). The role of dynamic spatial ability in geoscience text comprehension. *Learning and Instruction*, *31*(C), 33–45. http://doi.org/10.1016/j.learninstruc.2013.12.007

Schuil, K. D. I., Smits, M., & Zwaan, R. A. (2013). Sentential context modulates the involvement of the motor cortex in action language processing: An FMRI study. *Frontiers in Human Neuroscience*, *7*(100), 1–13. http://doi.org/10.3389/fnhum.2013.00100

Schwanenflugel, P. J., Meisinger, E. B., Wisenbaker, J. M., Kuhn, M. R., Strauss, G. P., & Morris, R. D. (2006). Becoming a fluent and automatic reader in the early elementary school years. *Reading Research Quarterly*, *41*(4), 496–522. http://doi.org/10.1598/RRQ.41.4.4

Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing*, *18*(7-9), 617–656. http://doi.org/10.1007/s11145-005-2038-0

Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing*, *13*(7-9), 81–103. http://doi.org/10.1007/s11145-005-2038-0

Sesma, H. W., Mahone, E. M., Levine, T. M., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology*, *15*(3), 232–246. http://doi.org/10.1080/09297040802220029

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., et al. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading*, *3*(1), 69–94. http://doi.org/10.1207/s1532799xssr0301_4

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Scientific Studies of Reading*, *171*(3972), 701–703. http://doi.org/10.1016/b978-1-4832-1446-7.50050-9

Shepard, S., & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 3–11. http://doi.org/10.1037/0096-1523.14.1.3

Sherwood, R., & Pearson, J. (2010). Closing the mind's eye: Incoming luminance signals disrupt visual imagery. *PLoS ONE*, *5*(12), e15217. http://doi.org/10.1371/journal.pone.0015217

Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*, *107*(2), 321–331. http://doi.org/10.1037/a0037769

Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition*, *25*(3), 321–332. http://doi.org/10.3758/bf03211288

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, *22*(4), 437–448. http://doi.org/10.1016/s0022-5371(83)90282-7

Singer, M., & Ritchot, K. F. M. (1996). The role of working memory capacity and knowledge access in text inference processing. *Memory & Cognition*, *24*(6), 733–743. http://doi.org/10.3758/BF03201098

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. *Memory & Cognition*, *20*(5), 539–548. http://doi.org/10.3758/bf03199586

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. http://doi.org/10.1037/0278-7393.6.2.174

Snow, C. E. (2002). Reading for understanding: Toward a research and development program in reading comprehension. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1465.pdf

Spear-Swerling, L. (2004). Fourth graders' performance on a state-mandated assessment involving two different measures of reading comprehension. *Reading Psychology*, *25*(2), 121–148. http://doi.org/10.1080/02702710490435727

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, *20*(8), 989–999. http://doi.org/10.1111/j.1467-9280.2009.02397.x

Spooner, A. L. R., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology*, *74*(2), 187–204. http://doi.org/10.1348/000709904773839833

St George, M. S., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, *9*(6), 776–787. http://doi.org/10.1162/jocn.1997.9.6.776

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. http://doi.org/10.1111/1467-9280.00326

Stine-Morrow, E. A. L., Milinder, L. A., Pullara, O., & Herman, B. (2001). Patterns of resource allocation are reliable among younger and older readers. *Psychology and Aging*, *1*(16), 69–84. http://doi.org/10.1037//0882-7974.16-1.69

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*(6), 934–947. http://doi.org/10.1037//0012-1649.38.6.934

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language and working memory skills. *Reading and Writing*, *4*(3), 245–256. http://doi.org/10.1007/bf01027150

Stothard, S. E., & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology and Psychiatry*, *36*(3), 399–408. http://doi.org/10.1111/j.1469-7610.1995.tb01298.x

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence*, *21*(1), 83–108. http://doi.org/10.1016/0160-2896(95)90040-3

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5 ed.). Boston, MA: Pearson Education.

Tanenhaus, M. K., Spivey-Knowlton, M. J., & Eberhard, K. M. (1995). Integration of visual and linguistic information in spoken language comprehension. *Scientific Studies of Reading*, *268*(5217), 1632–1634. http://doi.org/10.1126/science.7777863

Thurstone, L. L., & Thurstone, T. G. (1947). *SRA primary mental abilities*. Chicago, IL: Science Research Associates.

Tompkins, V., Guo, Y., & Justice, L. M. (2013). Inference generation, story comprehension, and language skills in the preschool years. *Reading and Writing*, *26*(3), 403–429. http://doi.org/10.1007/s11145-012-9374-7

Tong, F. (2013). Imagery and visual working memory: One and the same? *Trends in Cognitive Sciences*, *17*(10), 489–490. http://doi.org/10.1016/j.tics.2013.08.005

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, *24*(1), 595–611. http://doi.org/10.1016/0749-596x(85)90048-8

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes*, *16*(1-2), 3–34. http://doi.org/10.1080/01638539309544827

Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, *12*(1), 1–25. http://doi.org/10.1080/01638538909544717

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). New York: Academic Press.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank & I. Campari (Eds.), *Spatial Information Theory: A theoretical basis for GIS, proceedings COSIT '93.* (pp. 14–24). Berlin, Germany: Springer.

Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the landscape model: Modeling inferential processes and memory representations of text comprehension. *Behavioral Research Methods, Instruments, and Computers*, *37*(2), 277–286. http://doi.org/10.3758/bf03192695

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development*, *81*(5), 1417–1430. http://doi.org/10.1111/j.1467-8624.2010.01482.x

Uttal, D. H., Fisher, J. A., & Taylor, H. A. (2006). Words and maps: Developmental changes in mental models of spatial information acquired from descriptions and depictions. *Developmental Science*, *9*(2), 221–235. http://doi.org/10.1111/j.1467-7687.2006.00481.x

van Berkum, J. J. A., Brown, C. M., & Zwitserlood, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology*, *31*(3), 443–467. http://doi.org/10.1037/0278-7393.31.3.443

van Blerkom, M. L. (2009). *Measurement and statistics for teachers*. New York, NY: Taylor & Francis.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081–1087. http://doi.org/10.3758/bf03206376

van den Broek, P., Pugzles-Lorch, E., & Thurlow, R. (1996). Childrens' and adults' memory for television stories: The roles of causal factors, story-grammar categories, and hierarchical level. *Child Development*, *67*(6), 3010–3028. http://doi.org/10.2307/1131764

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, *39*(2-3), 299–316. http://doi.org/10.1080/0163853x.2005.9651685

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, New Jersey: Lawrence Erlbaum & Associates.

van der Schoot, M., Horsley, T. M., & van Lieshout, E. C. D. M. (2010). The effects of instruction on situation model construction: An eye fixation study on text comprehension in primary school children. *Educational Psychology*, *30*(7), 817–835. http://doi.org/10.1080/01443410.2010.510600

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2011). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*(7), 1665–1690. Advance online publication. http://doi.org/10.1007/s11145-011-9337-4

van der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A., & van Lieshout, E. C. D. M. (2009). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*, *101*(1), 21–36. http://doi.org/10.1037/a0013382

van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

van Kleeck, A. (2008). Providing preschool foundations for later reading comprehension: The importance of and ideas for targeting inferencing in storybook-sharing interventions, *Psychology in the Schools 45*(7), 627–643. http://doi.org/10.1002/pits.20314

Waber, D. P., Carlson, D., & Mann, M. (1982). Developmental and differential aspects of mental rotation in early adolescence. *Child Development*, *53*(6), 1614–1621. http://doi.org/10.1016/s0079-7421(03)44005-x

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256–1271. http://doi.org/10.1037/0278-7393.25.5.1256

Weekes, B. S., Hamilton, S., Oakhill, J. V., & Holliday, R. E. (2008). False recollection in children with reading comprehension difficulties. *Cognition*, *106*(1), 222–233. http://doi.org/10.1016/j.cognition.2007.01.005

Weiss, S., Mueller, H. M., Mertens, M., & Woermann, F. G. (2011). "Tooth and truth": Brain activation during passive listening to concrete and abstract nouns. *The Open Behavioral Science Journal*, *5*, 37–47. http://doi.org/10.2174/1874230001105010037

White, K. D., & Ashton, R. (1977). Visual imagery control: One dimension or four? *Journal of Mental Imagery*, *1*(2), 245–251. Retrieved from http://psycnet.apa.org/psycinfo/1978-30801-001

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, *14*(2), 133–145. http://doi.org/10.1080/01638539109544779

Whitney, P., Ritchie, B. G., & Crane, R. S. (1992). The effect of foregrounding on readers' use of predictive inferences. *Memory & Cognition*, *20*(4), 424–432. http://doi.org/10.3758/BF03210926

Wigfield, A., Perencevich, K. C., Taboda, A., Lutz Klauda, S., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes, *45*(5), 432–445. http://doi.org/10.1002/pits.20307

Willson, V. L., & Rupley, W. H. (1997). A structural equation model for reading comprehension based on background, phonemic, and strategy knowledge. *Scientific Studies of Reading*, *1*(1), 45–63. http://doi.org/10.1207/s1532799xssr0101_3

Yuill, N. M., & Oakhill, J. V. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, *2*(1), 33–45. http://doi.org/10.1002/acp.2350020105

Yuill, N. M., & Oakhill, J. V. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, UK: Cambridge University Press.

Yuill, N. M., Oakhill, J. V., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, *80*(3), 351–361. http://doi.org/10.1111/j.2044-8295.1989.tb02325.x

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Neuroscience*, *20*(1), 1–19. http://doi.org/10.1162/jocn.2008.20013

Ziegler, F., & Acquah, D. K. (2013). Stepping into someone else"s shoes: Children create spatial mental models from the protagonist"s point of view. *European Journal of Developmental Psychology*, *10*(5), 546–562. http://doi.org/10.1080/17405629.2012.744689

Ziegler, F., Mitchell, P., & Currie, G. (2005). How does narrative cue children's perspective taking? *Developmental Psychology*, *41*(1), 115–123. http://doi.org/10.1037/0012-1649.41.1.115

Zwaan, R. A. (1999a). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*(1), 81–88. http://doi.org/10.1080/01638539909545070

Zwaan, R. A. (1999b). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, *8*(1), 15–18. http://doi.org/10.1111/1467-8721.00004

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE*, *7*(12), 1–10. http://doi.org/10.1371/journal.pone.0051382.t001

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Science*, *123*(2), 162–185. http://doi.org/10.1037//0033-2909.123.2.162

Zwaan, R. A., & van Oostendorp, H. (1993). Do readers construct spatial representations in naturalistic story comprehension? *Discourse Processes*, *16*(1-2), 125–143. http://doi.org/10.1080/01638539309544832

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995a). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297. http://doi.org/10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, *28*(4), 611–619. http://doi.org/10.1016/j.cogsci.2004.03.004

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995b). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386–397. http://doi.org/10.1037/0278-7393.21.2.386

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, *2*(3), 199–220. http://doi.org/10.1207/s1532799xssr0203_2

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171. http://doi.org/10.1111/1467-9280.00430

Appendix A

Examples of the Lists Used in the Object Imagery Task (Study 1)

List 1
Bus
Bicycle
Train

List 2
Banana
Apple
Orange

List 3
Nail
Screwdriver
Scissors

List 4
Goat
Pig
Donkey

Appendix B

Stimuli Used in the Binocular Rivalry Task (Study 1)