DOCTORAL THESIS

A longitudinal investigation into the relationship between early theory of mind and later listening and reading comprehension
exploring broad metacognition and maternal mental state talk

Jackson, Sophie

Award date:
2020

Awarding institution:
University of Roehampton

A longitudinal investigation into the relationship between early theory of mind and later listening and reading comprehension: exploring broad metacognition and maternal mental state talk

by

Sophie Jackson BSc, MSc

A thesis submitted in partial fulfilment of the requirements for the degree of PhD

Department of Psychology

University of Roehampton

2019

Abstract

Recent research suggests that ToM facilitates both LC (Kim, 2017) and RC (Atkinson et al., 2017; Kim, 2017) but, until now, no study has examined this longitudinally using a large sample to test direct and indirect models, or asked what it is about the nature of ToM which is important for LC and RC development. This thesis used the DIET and DIER models of LC and RC (Kim, 2015; Kim, 2016; Kim, 2017) as theoretical framework and employed a longitudinal design whereby a sample of 204 children's development in language, social and cognitive skills was tracked from age three to six years. Longitudinally DIET and DIER models were tested, and other non-social types of metacognition were included within the models to assess if the social nature of ToM is vital for LC and RC, or if the broad metacognitive nature is important. Lastly, the role of mental state talk as a facilitator of both ToM and LC was investigated to further address the question over why and how ToM helps LC.

Findings showed that a concurrent DIET model of LC fitted at six year olds, with ToM making a direct contribution to LC. Findings also showed that this model fitted well longitudinally for skills at the age of four predicting LC at the age of five with ToM again making a direct contribution. However, there was no evidence that ToM contributes to LC further across time. When comparing ToM to other forms of metacognition, findings suggested overall ToM was a slightly better predictor of LC than other forms of metacognition, as concurrently at the age of six and longitudinally across 12 months (from four years until five years) the fit of a DIET model including ToM rather than a broad metacognition latent variable, was better.

Regarding RC, both concurrently (aged six) and longitudinally (aged four to aged six) findings did not support past work that ToM directly predicts RC (Atkinson et al., 2017) as a DIET model of RC did not show ToM to make a direct contribution to RC. Instead concurrent findings aged six back-up a DIER model of RC in which ToM make an indirect contribution to RC. This however, was not supported longitudinally as ToM did not make an in-direct contribution to later RC (via LC). However, the model including ToM was a better fit when compared to one including broad metacognition which, as with LC, suggests that the social specificity of ToM is important for comprehension. When considering the home environment, findings showed that maternal mental state talk did not predict LC directly or indirectly (via ToM) either at four year old, five years old or longitudinally. These results were mirrored for children's own mental state talk when measured through live conversations, however, when measured through mothers' self-report of their child's mental state talk, longitudinally only, an indirect effect of child mental state talk on LC via ToM was found. Overall, although findings were not consistent across time points; they add to the growing body of research that demonstrates that ToM is important for LC and RC in the early years and provide some partial evidence that this is because of the social specificity of ToM in that it can help with understanding social information within a story.

Table of Contents

List of figures

List of tables

Acknowledgements

Foremost, I would like to thank my supervisors Lance Slade and Sam McCormick for your knowledge, guidance and encouragement throughout my doctoral studies. I would also like to thank my director of studies Joe Levy for your insightful comments especially in the final stages of this thesis. Also, with thanks to Rebecca Lucas for your support with data collection in the very early days when I was not quite sure what I had let myself in for yet! My study would also not have been possible without the support of the Vice Chancellor's Scholarship.

I would like to thank all the children, parents and staff involved in the research. This includes everyone at Dorchester Primary, Robin Hood Infants and Winnington Park Primary, and all the families in Cohort 2 who let me visit them in their homes. Thank you all for taking such an interest in my research and being so helpful and welcoming. I have thoroughly enjoyed watching the children grow and develop over the last four years. Also, with thanks to the charity Learn 2 Love 2 Read and Ronald Ross Primary School for your invaluable help with recruitment.

On a personal note, thank you to my mum who has always encouraged me to "*go the extra mile*" with everything in life, I doubt you ever imagined that mile would be a four year PhD? Thank you to my dad for, quite possibly, reading every word I have ever written?!? For that I can only apologise. To Lucy…. for just for being you, and for your vital artistic contribution to this research (see page 135). Lastly, but very importantly, thank you to Jonathan for being my biggest supporter in this journey but also my harshest critic (by far).

"*I have this belief that children become readers before they can read. They become hooked on books because they were read aloud to as a child*"

Dame Jacqueline Wilson, Chancellor of University of Roehampton and my childhood hero.

Ethical approval

The research for this project was submitted for ethics consideration under the reference PSYC 16/ 210 in the Department of Psychology and was approved under the procedures of the University of Roehampton's Ethics Committee on 07.04.16 and 06.10.16.

1 Literature Review

1.1 Reading and listening comprehension

Reading comprehension (RC) is the ability to read text, process it and take meaning from it (Snowling & Hulme, 2008). Comprehension of written text is an essential skill and one of the fundamental aims of primary school education (Florit & Cain, 2011), but it is a very complex task (Tobia & Bonifacci, 2015). It has importance not only for educational and employment purposes, but also for participation in many cultural and social activities throughout life (Florit & Cain, 2011). Given the far-reaching effects of RC it is important that the factors central to its development are understood.

Reading is a relatively new human phenomenon as for most of human existence experiences, information, and knowledge which facilitate participation in culture were passed on via word-of-mouth (Johnson, 2015). This is listening comprehension (LC), using lexical information to achieve sentence and discourse interpretations (Hoover & Gough, 1990), or more simply, the ability to understand what is heard (Hogan, Adlof, & Alonzo, 2014). LC is important for RC (and is often cited as component skill of RC, see Section 1.2.3), but LC is an important skill in its own right, and it is important that that the factors central to the development of LC are also understood. Therefore, this thesis is concerned with the development of both LC and RC. This chapter will describe and critique the literature concerned with early development of RC and LC. The chapter will discuss language and cognitive factors which underpin both LC and RC in relation to the Simple View of Reading (Gough

& Tunmer, 1986), the situation model (Kintsch, 1988; Van Dijk, Kintsch, & Van Dijk, 1983), and the recent cognitive models developed by Kim; the Direct and Indirect Effect Model of Text Comprehension (DIET; Kim, 2016; Kim, 2017), and the Direct and Indirect Effect Model of Reading (DIER; Kim, 2015; Kim, 2017).

1.1.1 Defining listening comprehension, oral language and linguistic comprehension

Listening comprehension is an oral language skill concerned with the processing and understanding of prose received orally (Gough & Tunmer, 1986; Hogan et al., 2014). In the literature, it is also referred to as "story comprehension", "verbal comprehension", "comprehension of spoken text" or "narrative comprehension", with these phrases often used interchangeably (Cain, 2017). Here, as it appears to be the most common phrase used (Cain, 2017) listening comprehension (LC) will be used to describe this competency.

Measures of LC vary in their content and these different ways of assessing LC have led to confusion as to what it entails. For example, in the "understanding spoken paragraphs" subset of the Clinical Evaluation of Language Fundamentals – 5th edition (CELF-5;Wiig, Semel & Secord, 2015) and when the Neale Analysis of Reading Ability – Second Revised British Edition (NARA; Neale, 1999) has been administered orally (e.g. Cain & Bignell, 2014; Nation, Cocksey, Taylor, & Bishop, 2010), children are read aloud paragraphs and answer questions on the main ideas and details from the passage. By contrast, in the listening comprehension subset of the Oral and Written Language Scales – second edition (OWLS-II; Carrow-

Woolfolk, 2011) children's comprehension of single words, phrases and sentences is assessed using a picture pointing task. The variation between these measures with their opposing task demands demonstrates how LC can be defined and measured in diverse ways by different researchers (Nation & Snowling, 1997).

The relationship between LC, other oral language skills and linguistic comprehension is also viewed inconsistently by the literature (Cain, 2017). Some propose that LC is separate to all other oral language skills. For example, in their literacy model, Juel, Griffith, and Gough (1986) make a distinction between children's performance on component oral language skill measures, such as vocabulary, and LC. On the other hand, others state that oral language skills, such as vocabulary and knowledge of syntax and grammar, are separate to LC but make contributions towards it (e.g. Hogan et al., 2014; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012). Lastly, others treat LC as part of a larger oral language construct. For example, Kendeou, Van den Broek, White, and Lynch (2009), and Catts, Herrera, Nielsen, and Bridges (2015) include LC under the same classification as vocabulary and call this construct oral language. Likewise, Foorman, Herrera, Petscher, Mitchell, and Truckenmiller (2015) include LC, syntactic knowledge, and vocabulary in their oral language construct. In support of this Cain (2017) concluded LC, vocabulary and grammatical knowledge were best categorised under the same construct when using confirmatory factor analysis. This is also a similar view taken by the work of Kim (2015; 2016; 2017) and will be explored further by this thesis (See Section 1.5). Kim suggests a hierarchical structure in which other oral language skills of syntax and vocabulary contribute to LC to create the construct of LC. It should be noted that Kim also posits that cognitive skills are part of this construct too; this will be explored later in this chapter.

Regarding the distinction between the phrases “linguistic comprehension” and “listening comprehension”, even studies which try to differentiate these are still not consistent. For example, in a study by Cain (2017), with the aim of determining whether listening comprehension and oral language are the same construct, the terms listening comprehension and linguistic comprehension are used synonymously. This is also the case too in the seminal work of Gough and Tunmer (1986) and their Simple View of Reading model (See Section 1.3), and in subsequent work discussing this influential model, as here the phrases are used interchangeably without clarification (e.g. Cain, 2017). For example, in their follow up work on the simple view, although Hoover and Gough (1990) refer to the competency as linguistic comprehension in their discussion of the model, they only use a listening comprehension measure to assess it. Given this, Hogan et al., (2014) propose that over time linguistic comprehension has been referred to as listening comprehension, suggesting they are the same concept.

More recent work has taken the view that the linguistic comprehension construct comprises many subskills, one of which is listening comprehension (e.g. Foorman et al., 2015). Under this suggestion linguistic comprehension and listening comprehension are not synonymous, as listening comprehension is just one component of linguistic comprehension. This seems to be confirmed in a recent review of the Simple View of Reading model by Hoover and Tunmer (2018), who state that under the model this component is often referred to as linguistic comprehension or listening comprehension interchangeably and although the terms are often used equivalent in meaning, LC is used to denote a particular way of assessing the more general construct of linguistic comprehension. Hoover and Tunmer (2018) also describe recent studies (e.g. Lonigan, Burgess, & Schatschneider, 2018) which have measured linguistic comprehension using a latent variable consisting of measures of; listening comprehension, receptive vocabulary, expressive vocabulary, receptive syntax, and expressive syntax. This is similar, and linked to, the argument discussed earlier that LC is part of a large oral language construct but here the argument is taken a step further to suggest that this construct can be referred to as linguistic comprehension. This perspective has also been taken by other recent work (e.g. Atkinson, Slade, Powell, & Levy, 2017; Kim, 2015; Kim, 2017), where the construct of linguistic comprehension consists of a composite of a vocabulary measure, language skills measures (linguistic concepts and recalling sentences in context), and a narrative comprehension measure (Atkinson et al., 2017). Other recent work goes further, proposing that the components are linked in particular (hierarchical) ways (Kim, 2015; Kim, 2017). As the most recent work views linguistic comprehension as consisting of component skills including LC, this will also be the approach taken by this thesis.

1.2 The acquisition of reading and listening comprehension

The following section will discuss the literature relating to early development of both LC and RC. It will consider the skills underpinning both types of comprehension and describe the relationship the two have with one another.

1.2.1 Early predictors of listening and reading comprehension

Research suggests the same common language and cognitive skills underlie and precede both RC and LC (Florit, Roch, & Levorato, 2011; Oakhill & Cain, 2007). Given this, both types of comprehension will be discussed together. Emergent literacy is a term used to describe knowledge and skills a child possesses before they learn to read which will aid them in reading (Whitehurst & Lonigan, 1998). These skills include those relevant to both LC and RC (e.g. oral language skills and cognitive skills such as working memory). They also include code related skills such as letter sound knowledge which are needed only for RC. The early contribution of these skills to both types of comprehension is described below.

1.2.2 Oral language skills for listening and reading comprehension

Oral language skills underlie and precede both LC and RC. These include vocabulary and syntactic knowledge. The contribution of these oral language skills to LC and RC is discussed below.

1.2.2.1 Vocabulary

Vocabulary is an oral language skill defined as the body of familiar words in a person’s language (Schmitt, 2000). In children vocabulary can be measured using tasks such as the British Picture Vocabulary Scale (BPVS; Dunn, Dunn & National Foundation for educational Research, 2009) or the Peabody Picture Vocabulary test (PPVT; Dunn & Dunn, 2007), which require the child to point to pictures which relate to words spoken. The relationship between vocabulary and LC in the preschool years is strong as shown by cross-sectional studies in Finnish and Italian speaking children aged four to six years even when controlling for working memory (Florit et al., 2011; Florit & Levorato, 2012; Lepola et al., 2012). Further, longitudinal studies show that vocabulary aged four predicts LC aged five (Florit, Roch, Altoè, & Levorato, 2009; Kendeou, Bohn-Gettler, White, & Van Den Broek, 2008) for both Italian and English speaking children when using the PPVT as a measure of vocabulary, even when controlling for non-verbal ability (Sénéchal, Pagan, Lever, & Ouellette, 2008).

There is also evidence that vocabulary relates strongly to RC. Tannenbaum, Torgesen and Wagner (2006) reported that the correlation between RC and vocabulary varied between $r = .3$ and $r = .8$ in children aged from eight to 11 years. Moreover, early vocabulary is shown to predict later RC. Findings from longitudinal studies show that vocabulary at the age of four predicts RC at six years old (Silva & Cain, 2015) in an English speaking UK sample when measuring vocabulary using the BPVS and controlling for non-verbal ability. Research even shows that vocabulary at the age of two years predicts RC up to five years later, accounting for 18% of variance (Duff, Reen, Plunkett, & Nation, 2015).

Further evidence for the importance of vocabulary in RC comes from children with English as an Additional Language (EAL). Vocabulary is a weakness of these children (Mahon & Crutchley, 2006) and this is suggested to be a main cause of inferior RC compared to monolingual peers (Bowyer-Crane, Fricke, Schaefer, Lervåg, & Hulme, 2017). When these EAL children begin to read, they may be encountering both the written and spoken forms of a new word simultaneously (Bowyer-Crane et al., 2017) with effects on their RC.

1.2.2.2 Syntactic knowledge

Syntactic knowledge, also called grammatical knowledge, is concerned with word order and grammatical rules (Tomasello, 2000). In young children syntactic knowledge can be measured using picture pointing tasks in which children are read aloud sentences of increasing syntactic complexity (e.g. embedded relative clauses) and are required to point to the corresponding picture. For example, they are asked to point to "*mum showed the dog the cat*" when shown several pictures of a woman, dog and a cat in various positions. Such measures include the sentence structure subset of the Clinical Evaluation of Language Fundamentals (CELF: Semel, Wiig, & Secord, 2004; Wiig, Secord, & Semel, 2006) and the Test for Reception of Grammar (Bishop, 1989).

Evidence suggests that for children aged five and six, LC is directly predicted by concurrent syntactic knowledge (Kim, 2015; Potocki, Ecalle, & Magnan, 2013) in Korean and French speaking children. Longitudinal work has found that syntax at the age of five predicts LC aged eight (Alonzo, Yeomans-Maldonado, Murphy, & Bevens, 2016) in an English speaking US sample when using the CELF to measure syntactic knowledge. Regarding RC, knowledge of syntax has been shown to predict RC across two years from four years to six years in both English and French speaking children (M. Bianco et al., 2012; Muter, Hulme, Snowling, & Stevenson, 2004). These relationships are independent of vocabulary, as research has found that syntax predicts LC (Potocki et al., 2013) over and above the effect of vocabulary concurrently at the age of five years. This is also seen in RC (Gottardo, Mirza, Koh, Ferreira, & Javier, 2018; Mokhtari & Niederhauser, 2012) for children aged nine to 13 years, again over and above the effect of vocabulary. The same has been demonstrated longitudinally for RC from syntax aged four years to RC at the age of six years (Silva & Cain, 2015).

1.2.3 The relationship between reading and listening comprehension

LC is an oral language skill. Although LC is an important competency in its own right (Hogan et al., 2014), it has been robustly shown that RC is dependent on LC and in particular that LC predicts both concurrent and later RC. For example, de Jong and van der Leij (2002) found that LC aged six explained unique variance (after controlling for decoding) in both concurrent RC, and RC two years later. Additionally, Roth, Speece, and Cooper (2002) found that LC at the age of five predicted RC aged seven in an English speaking US sample, and Kendeou et al., (2009) demonstrated similar findings for LC measured at the ages of four to six, predicting RC aged six to eight. Parallel findings can be seen in younger children as Bianco et al. (2012) found that LC at the age of four predicted RC two years later and that it explained twice as much variance than phonological awareness did in a sample of French speaking children. Moreover, the NICHD study which tracked 1,137 typically developing children in the US from three until seven years old, found that LC at four years significantly related to RC aged six (NICHD Early Child Care Research Network, 2005). This predictive relationship is also seen further across time, with LC aged six predicting RC aged 11 (Ouellette & Beers, 2010). These findings are replicated by more recent studies such as that by Lervåg, Hulme, and Melby-Lervåg (2018) who found that LC, as measured by the Neale Analysis of Reading Ability (NARA; Neale, 1999), at the age of seven years predicted RC aged 12 years.

1.2.4 Code related skills and reading comprehension

In addition to oral language skills, code related skills which help children to 'break the code' and to acquire early alphabetical principles (Cabell, Justice, Konold, & McGinty, 2011), are important for RC. Code-related skills facilitate children's ability to become accurate and fluent decoders of text (Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2013). Phonological decoding is the process of sounding out and blending together printed letters to form spoken words (Gough & Tunmer, 1986; Metsala & Ehri, 2013). The purest measure of decoding is single word reading of non-words where participants are required to read aloud single non-words, because this requires children to pronounce letter-strings devoid of lexical content (Nation & Snowling, 1997). Decoding skills are measured in numerous standardised reading assessments and those standardised for a UK population include The Diagnostic Test of Word Reading Processes (Forum for Research into Language and Literacy, 2012). These standardised measures often include a combination of words – those which can be decoded (regular words which can be sounded out and blended such as '*it*') and those that cannot be decoded and instead require visual word recognition of learned whole word forms (high frequency irregular words such as '*of*') as well as non-words (such as '*Wup*').

Decoding plays an important role in emergent RC because a child first needs to be able to read individual words before they can comprehend the words, sentences and paragraphs. Indeed, a meta-analysis of 110 studies found the average concurrent correlation between RC and decoding to be $r = .74$ across ages and different assessments (García & Cain, 2014). However, the meta-analysis found that the strength of this relationship decreases with age, with $r = .86$ across studies with participants younger than seven years, and r = .41 for studies with participants older than 16.

Longitudinal studies show that earlier decoding predicts later RC. For example, Kendeou et al., (2009) found that decoding in children aged 4-6 years directly predicted RC aged 6-8 years, over and above the oral language skills of vocabulary and LC. This predictive relationship can be seen further across time with Verhoeven and van Leeuwe (2008) showing that in a sample of 2,143 Dutch children decoding aged six predicted RC up to five years later when children were 11 years old (after controlling for oral language skills). These findings are replicated by more recent studies such as that by Lervåg et al., (2018) who found that decoding at the age of seven years (as measured using a Norwegian translation of the Test of Word Reading Efficiency; TOWRE) predicted RC aged 12 years in a group of Norwegian speaking children, again over and above oral language skills.

Whereas oral language skills such as vocabulary are typically weak in EAL children, code-related skills are a strength for these bilingual readers with findings showing that they outperform their monolingual peers in such tasks (Campbell & Sais, 1995; Kang, 2012; Marinova-Todd, Zhao, & Bernhardt, 2010; McBride–Chang & Kail, 2002). Given that these children still have at least adequate RC skills (Bowyer-Crane et al., 2017), this acts as evidence that both code-related skills and oral language skills are required for proficient RC. The developing role of decoding for RC in relation to the Simple View of Reading model is discussed in the following section (Section 1.3).

1.2.4.1 Early predictors of decoding

Two key predictors of decoding are the sound-related skills of letter sound knowledge (knowledge of the letters or groups of letters which represent the individual speech sounds in language; Hulme et al, 2009) and phonological awareness (a broad skill that includes identifying and manipulating units of oral language such as words and syllables; Stahl & Murray, 1994). For example, at the age of six phonological awareness is shown to correlate strongly with decoding as measured by single word reading (Swank & Catts, 1994). Moreover, both letter sound knowledge and phonological awareness at the age of three years were found to be unique predictors of decoding aged six (Lonigan, Burgess, & Anthony, 2000). This predictive relationship is also seen in older children with letter sound knowledge and phonological awareness aged five shown to predict decoding both at seven years old and nine years old (Hogan, Catts, & Little, 2005). These findings are consistent with more recent research which has found that both letter sound knowledge and phonological awareness at the age of five predicts decoding aged seven years (Catts et al., 2015). Given this, the skills of letter sound knowledge and phonological awareness are often used in longitudinal reading studies with very young children (non-readers) as a measure of precursory decoding (e.g. Atkinson et al., 2017; Cain & Chiu, 2018; Kendeou et al., 2009).

Overall, there is clear evidence that early decoding is important for later RC, and that both letter sound knowledge and phonological awareness are precursors of decoding. Research also suggests that oral language skills are of key importance for the development of both RC and LC. The notion that oral language skills and decoding are the most important component skills of RC is held by the Simple View of

Reading model (Gough & Tunmer, 1986). This model is described and evaluated in the subsequent section of this chapter.

1.3 The Simple View of Reading

The Simple View of Reading (SVR) proposed by Gough and Tunmer (1986) is, to date, the most influential model of learning to read and can be used at all levels of RC from non-readers to highly proficient adults (Hoover & Tunmer, 2018). The model views reading (RC) as a product of the two separate unrelated components, linguistic comprehension (C) and decoding (D).

The model can be expressed as:

RC = C x D

The model defines decoding as the ability to read isolated words quickly, accurately and silently (Gough & Tunmer, 1986), but it is nearly always measured using tests of accuracy (Florit & Cain, 2011). The linguistic comprehension component is defined as the ability to interpret sentences and discourse presented orally (Gough & Tunmer, 1986). As discussed in Section 1.1.1, in the past the linguistic comprehension component was often measured only using a listening comprehension measure (e.g. Johnston & Kirby, 2006; Joshi & Aaron, 2000; Ouellette & Beers, 2010), but more recent research has measured linguistic comprehension using a construct consisting of several oral language skills (e.g. Atkinson et al., 2017; Catts et al., 2015; Foorman et al., 2015; Lonigan et al., 2018) such as listening comprehension, vocabulary and syntactic knowledge. This is argued by Hoover and Tunmer (2018) as the most accurate way to measure linguistic comprehension within the SVR.

Under the model, Gough and Tunmer (1986) are clear that for proficient RC one component is not sufficient by itself. Put simply by Catts and Weismer (2006), decoding allows for the translation of print into a spoken form (the reading element), and the linguistic comprehension component makes sense of it (the comprehension element). Gough and Tunmer see the SVR as a product model rather than an additive model (i.e. it is not RC = C + D). The authors state that it is the interaction between the two components which is important, and that the effect of an increase in either depends upon the level of the other. Figure 1.1 shows the theoretical relationship between the variables in the SVR.

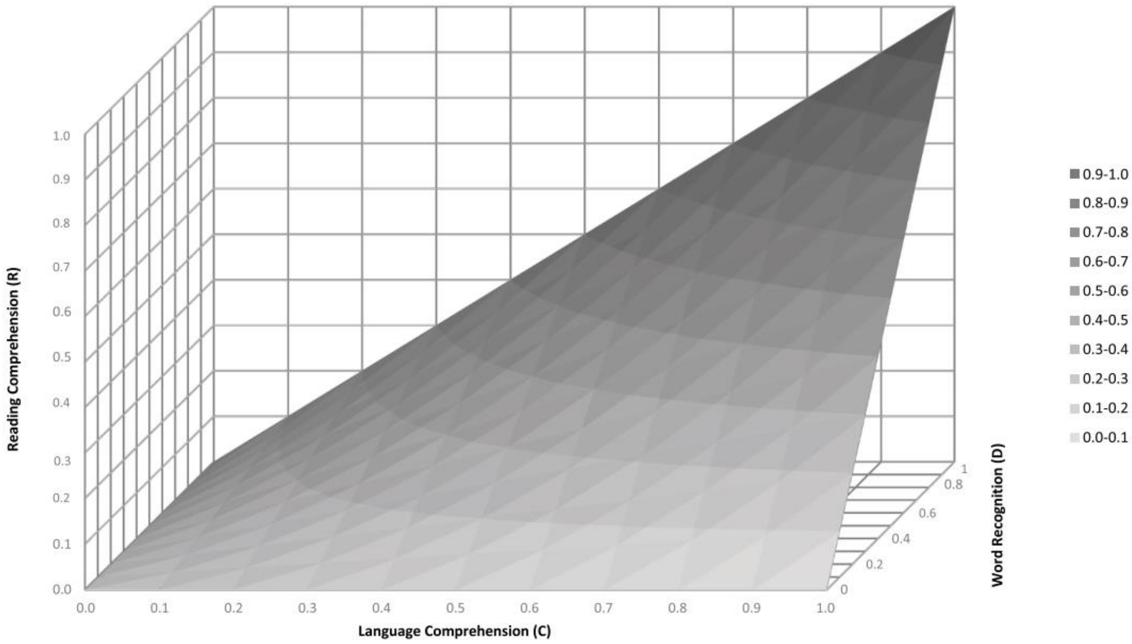


Figure 1.1: The theoretical relationships between the three variables of the SVR as suggested by Hoover and Tunmer (2018), where reading comprehension (R) is the product of word recognition (D for decoding) and language (or linguistic) comprehension (C), with each variable ranging in value from 0 (no skill) to 1 (perfect skill). Source: Graphic from Hoover and Tunmer (2018).

There is much empirical evidence to support the SVR model across all age ranges. For example the ages of: five years (Catts et al., 2015), six years (Tiu, Thompson, & Lewis, 2003), seven years (Cutting, & Scarborough, 2006), eight years (Vellutino, Tunmer, Jaccard, & Chen, 2007), nine and 10 years (Storch & Whitehurst, 2002), 11 and 12 years (Chen & Vellutino, 1997), and 13-16 years (Savage, 2006). These findings are supported by very recent studies (Cain & Chiu, 2018; Hoover & Tunmer, 2018; Massonnié, Bianco, Lima, & Bressoux, 2018). Estimates of variance explained by the model range from 45 to 85%, depending on the population assessed and the measures used (Conners, 2009; Hoover & Gough, 1990).

More recent research employing a latent variable approach confirms that decoding and linguistic comprehension explains a large proportion of the variance in RC for children aged 5-14 years (Foorman et al., 2015; Kim, 2015; Kim & Wagner, 2015; Language and Reading Research Consortium, 2015). The model is also shown to be a good predictor of future RC as findings show that earlier decoding and linguistic comprehension robustly predict later RC. This is seen for decoding and linguistic comprehension at the of age four to six years old predicting RC aged six to eight years (Kendeou et al., 2009), for these skills at age six predicting RC both at seven (Massonnié et al., 2018) and 10 years old (Juel, 1988), these skills at age eight years predicting RC aged 10 years (Dreyer & Katz, 1992), and skills aged 10 years predicting RC aged 13 years (van Wingerden, Segers, van Balkom, & Verhoeven, 2018).

The SVR model has also been validated cross-culturally by research seeking to corroborate the SVR in other languages beyond English. Evidence suggests that the model holds for French (Massonnié et al., 2018), Finnish (Müller & Brady, 2001), Dutch (de Jong & van der Leij, 2002), Norwegian (Hagtvet, 2003), Italian (Florit, Levorato, & Roch, 2008; Tobia & Bonifacci, 2015), Greek (Kendeou, Papadopoulos, & Kotzapoulou, 2013), Mandarin Chinese (Ho, Chow, Wong, Waye, & Bishop, 2012), Arabic (Asadi, 2018) and Korean (Kim, 2015) speaking children.

As a result of the wealth of evidence supporting the fit of the SVR for predicting reading from six years onwards across languages, the model has become extremely influential in the fields of both psychology and education. The SVR has been applied to reading policy (Hoover & Tunmer, 2018) particularly in the UK. Following an independent review by Rose (2005) the SVR became the theoretical basis for the UK literacy framework, the agenda which informs teachers on how best to teach literacy. Additionally, most researchers in the field of reading use the SVR as a theoretical basis for their research. However, an often-cited criticism of the SVR is that it is too simple (e.g. Gottardo et al., 2018; Kim, 2017; Kirby & Savage, 2008; Lervåg et al., 2018; Ouellette & Beers, 2010).

1.3.1 Is the Simple View of Reading too simple?

Some argue that the Simple View of Reading is just that, too simple, and suggest that it is time for a more extensive model (e.g. Adlof, Catts, & Little, 2006; Cain, 2015; Conners, 2009; Gottardo et al., 2018; Hoien-tengesdal, 2010; Joshi & Aaron, 2000; Kirby & Savage, 2008; Lervåg et al., 2018; Ouellette & Beers, 2010). Those of this opinion argue that reading is an incredibly complex skill that such a parsimonious model cannot explain the full story (Kirby & Savage, 2008). A scientific model, of any kind, attempts to simplify the process to reduce it to its core (Kirby & Savage, 2008). The issue here is, whether this reduction is useful or whether it misses out key aspects of the phenomenon of reading making the model inadequate (Kirby & Savage, 2008).

Yet, it must be noted that the SVR was never proposed as a complete theory of the cognitive processes involved in reading, but as a simple view with the aim of providing a framework (Kirby & Savage, 2008). The authors do not deny that other skills are important for reading (Gough, Hoover, Peterson, Cornoldi, & Oakhill, 1996a; Hoover & Gough, 1990). Indeed, Hoover and Gough (1990) write that “The simple view does not deny the complexity of reading, but asserts that such complexities are restricted to either of the two components” (p.150). More recently Hoover and Tunmer (2018) state that “the SVR does not claim that reading is simple. Both word recognition [or decoding] and language [or linguistic] comprehension are highly complex, and because of that, reading is complex. The SVR simply separates the complexity of reading into two component parts” (p. 306).

There are two alternative approaches to research to assess whether the SVR is too simple. The first has sought to find an additional component that could be added to the SVR to explain variance over and above the existing model (e.g. Conners, 2009; Kirby & Savage, 2008; Ouellette & Beers, 2010). This research is concerned with skills which make a direct contribution to RC after controlling for decoding and linguistic comprehension. The second avenue of research is interested in unpacking the two skills of the SVR (decoding and linguistic comprehension) to find their component skills (e.g. Kim, 2017; Massonnié et al., 2018). This research is concerned with skills which make an indirect contribution to RC via decoding or linguistic comprehension.

1.3.1.1 Contenders for an additional component of the simple view

Research concerned with finding an additional component of the SVR has suggested several skills which could explain variance over and above the existing model. To be a sufficient contender for an additional component these factors must not be sub-skills of the two main dimensions of the model (i.e. they must make a direct contribution to RC after controlling for decoding and linguistic comprehension) and must account for variance unexplained by the current model. Suggested contenders fit into three sub-types; speeded processes, non-verbal abilities and executive function skills.

1.3.1.1.1 Speeded processes

One suggested additional factor to be added to the SVR is reading fluency or reading with speed (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Fluency is measured by asking children to read as many words as possible within a time limit (Adlof et al., 2006). Although the SVR defines decoding as the accurate, quick and silent reading of isolated words (Gough & Tunmer, 1986), studies that support the SVR have often only used accuracy as a measure of decoding (Florit & Cain, 2011). Kirby and Savage (2008) suggest that for successful reading, decoding must be both accurate and fast, in other words fluid. They therefore argue decoding fluency is also vital. Several studies have found a strong correlation between reading fluency and RC (e.g. Braze, Tabor, Shankweiler, & Mencl, 2007; M. Wolf & Katzir-Cohen, 2001). Yet, research which accounted for linguistic comprehension and decoding accuracy, found that fluency did not account for any unique variance in RC (Adlof et al., 2006). In fact, results showed that few participants had problems in fluency separate from problems in linguistic comprehension and accurate decoding, and thus there is no clear case for arguing that fluency should be added to the SVR as an additional factor. These earlier conclusions are supported by more recent findings which have shown that in children aged seven to nine, reading fluency did not add unique variance to the SVR model (Cadime et al., 2017).

Naming speed, measured by the speed at which an individual can name sets of visual stimuli (Joshi & Aaron, 2000), has also been suggested. Some have found that naming speed still accounts for variance in RC after decoding and linguistic comprehension are controlled for, with unique variance suggested to be between 2 - 10% (Catts et al., 2015; Johnston & Kirby, 2006; Joshi & Aaron, 2000). However, Georgiou, Das, and Hayward (2009) found no unique contribution of naming speed to RC. In addition, Wolff (2014) found that naming speed only predicted reading speed and not RC for children aged nine. Given that naming speed has not be reliably replicated as a direct predictor of RC there is no clear case for adding naming speed to the SVR.

Processing speed more generally has also been suggested as an additional component to the SVR. This is the ability to automatically and fluently perform cognitive tasks (Hale, 1990). Joshi and Aaron (2000) claim that processing speed explains 48% of the total variance for RC, and 10% once decoding and linguistic comprehension are considered. This prompted Aaron and colleagues to suggest a revised SVR model which they call the component model (Aaron, Joshi, Gooden, & Bentum, 2008; Joshi & Aaron, 2000) with the addition of processing speed. Support for this component model comes from Tiu et al. (2003) who also indicate that processing speed explains a significant amount of variance in RC over that accounted for by the SVR. Yet, evidence for the component model has not been fully replicated, with others finding that processing speed gives no unique contributions to the SVR model (Borella & De Ribaupierre, 2014; Oh, 2016; Van Gelderen et al., 2004). Therefore, processing speed is another factor which has not been reliably replicated to add unique variance to the SVR.

1.3.1.1.2 Non-verbal abilities

Non-verbal IQ which assesses an individual's visuospatial intellectual ability by using batteries such as The Wechsler Intelligence Scales for Children (Wechsler, 2014), has also been suggested as an additional component. This has been found to add significant variance to the SVR by Tiu et al. (2003), Cutting and Scarborough (2006) and Oakhill and Cain (2012), when using the *Wechsler Intelligence Scale for Children* as a measure. Catts, Gillispie, Leonard, Kail and Miller (2002) suggest that the link between RC and non-verbal IQ might be underpinned by visual-spatial and analytic skills or that it might be mediated by higher-order verbal skills (e.g. verbal reasoning).

However, as with many of these suggested additional influences, IQ is not reliably supported as a direct predictor with Conners (2009), and Gustafson, Samuelsson, Johansson, and Wallmann (2013) finding that it did not contribute significant variance to the prediction of RC over and above the components of the SVR, despite using the same measure of IQ. Therefore, due to conflicting findings IQ should not be added to the SVR.

Despite this, many reading studies have controlled for non-verbal intelligence when assessing RC (e.g. Cain, Oakhill, & Bryant, 2004; Cain & Oakhill, 2011; Cartwright, Marshall, Dandy, & Isaac, 2010; Luoni et al., 2015; Sénéchal et al., 2008; Silva & Cain, 2015) with the argument that reading requires at least some general intellectual ability. The work of Ricketts and colleagues into RC and those with developmental disabilities, rationalises this by stating that it is important to consider non-verbal IQ when assessing RC, but it should not necessarily be included into models of RC (Ricketts, 2011; Ricketts, Jones, Happé, & Charman, 2013). Ricketts et al., (2013) ran their analysis with and without controlling for non-verbal ability, finding that models did not change, and that IQ did not predict significant unique variance in RC in any model. Therefore, although due to conflicting findings IQ should not be added to the SVR model, it will be controlled for in this thesis to be consistent with other research such as that by Ricketts and colleagues.

1.3.1.1.3 Executive functioning skills

Conners (2009) suggests attentional control, the ability to supress irrelevant information and concentrate only on the important material (MacLeod, 2007), as an additional factor. In both adults and children this is assessed in a variety of ways, but common measures include: set-shifting card sorting tasks, flanker tests, stop signal tasks, star counting tests, Stroop tasks, and go/no-go tasks (Diamond, 2013). The rationale for including this in relation to reading is that this ability may help in coordinating decoding and linguistic comprehension during reading. Others have discussed the role of attentional control in reading (e.g. Guajardo & Cartwright, 2016; Reynolds, 2000; Walczyk, 2000), hypothesizing that during reading when a word is decoded its meaning may be automatically triggered, but if this meaning does not fit with the context, it should be supressed and replaced with a meaning which does. Conners (2009) found that in eight year olds attentional control (as measured using a star counting test) did contribute significant variance in RC after controlling for the components of the SVR, with the amount of unique variance explained ranging from 5 to 10%. In support of this a more recent longitudinal study found that both attentional control (as measured using a card sorting task) at three to five years old (Time 1) and three years later (Time 2) accounted for unique variance in RC at the latter time point after accounting for oral language and decoding (Guajardo & Cartwright, 2016).

However, there are contradictory findings as Atkinson, Slade, Powell, and Levy (2017) found that an executive functioning composite consisting of attentional control (as measured using a card sorting task) and a measure of working memory measured when children were three years old, did not make a significant unique contribution to RC aged six. Similarly, for children aged 8-16 years attentional control (measured using a stop signal task) added no unique variance to RC (Borella & De Ribaupierre, 2014; Christopher et al., 2012). Therefore, as attention is not reliably replicated as adding unique variance to the SVR it should not be added to the model separate from linguistic comprehension and decoding.

It has been well established that working memory and RC are related (Cain, Oakhill, & Lemmon, 2004; Cain et al., 2004; Cain, 2006; Nouwens, Groen, & Verhoeven, 2017). Working memory is the cognitive system responsible for temporarily processing and holding information (Miyake & Shah, 1999). Memory may be important for RC because it allows for information to be retrieved, such as that needed to make inferences about the text, while other information from the text is held (Cook, Halleran, & O'Brien, 1998). Some suggest that working memory contributes direct unique significance to the prediction of RC (Cain et al., 2004; Conners, 2009; Goff, Pratt, & Ong, 2005), with variance ranging from 1 to 6.9% across these studies. However, contrary studies do exist, with findings that working memory does not add any unique variance (Logan, 2017; Parrila, Kirby, & McQuarrie, 2004). Therefore, working memory has not been added to the SVR as an additional factor.

Overall, the research concerned with finding an additional component for the SVR has inconsistent findings and there is, as of yet, no reliably replicated evidence for one single factor. As previously mentioned, other research evaluating the SVR has instead concentrated on expanding the SVR by unpacking its component skills (e.g. Kim, 2017; Massonnié et al., 2018). This type of research is concerned with factors which make an indirect contribution to RC via the two components.

1.3.1.2 Unpacking the Simple View of Reading

Research concerned with unpacking the SVR and looking at skills which make an indirect contribution to RC via linguistic comprehension has predominantly come from one research group (Kim, 2015; Kim, 2017). This research suggests that the SVR is not simplistic because the two components are complex and include many sub-skills (Kim, 2017). This is consistent with the views of Hoover and Gough (1990) when they express that "The simple view does not deny the complexity of reading, but asserts that such complexities are restricted to either of the two components" (p.150). Kim states that it is important to understand the sub-components of linguistic comprehension as under the SVR linguistic comprehension is "underspecified in terms of processes and structural relations" (Kim, 2017, p. 15).

Work by Kim aimed at unpacking the SVR (e.g. Kim, 2015; Kim, 2016; Kim & Petscher, 2016; Kim, 2017) hypothesizes that the linguistic comprehension component of the SVR is hierarchical and consists of language and cognitive skills where lower level processes feed into higher level processes. According to Kim these skills include both low level and high-order skills and may include vocabulary, grammatical knowledge, working memory, attentional control, inference making, comprehension monitoring and theory of mind (Kim, 2017). This research has also concentrated on unpacking LC as well as RC (e.g. Kim, 2015; Kim, 2016; Kim, 2017). The following section of this chapter will review the influences which these high-order skills have on both LC and RC.

1.4 Influence of high-order skills on listening and reading comprehension

High-order skills are those which involve complex cognition such as critical thinking and problem solving (Anderson et al., 2001). Such skills which have been linked to LC and RC are: inference making, comprehension monitoring, working memory and attentional control.

1.4.1 Inference making

Language is not completely explicit, instead inferences are needed to bridge elements and make it coherent (Perfetti, Landi, & Oakhill, 2005). There are two main types of inference; text connecting inferences (or local inferences) which integrate information from different parts of the text, and gap filling inferences (or global inferences) which are needed to fill in missing details not stated within the text, and come from the outside knowledge of the reader (Cain & Oakhill, 1999). Measures of inference are not commonly administered before the age of six and often require children to make both types of inference in questions asked of them about a passage they have read or have had read to them (e.g. Cain & Oakhill, 1999; Cain, Oakhill, Barnes, & Bryant, 2001; Freed & Cain, 2017). For example, within a passage used by Freed and Cain (2017) children are told that "*Mum told them to put on some suncream, so that they didn't burn*" and are later asked "*What was the weather like*?" in order to answer this question correctly children need to make the global inference that the weather was hot/sunny.

Evidence suggests that inference making supports successful RC (Elbro & Buch-Iversen, 2013; Oakhill & Cain, 2012). Oakhill and Cain (2012) found that both concurrent and earlier inference making skills (at the age of seven to eight years) predicted RC aged 10-11 years. Further evidence of the importance of inference making comes from training studies which show that participation in inference interventions improves RC for typical readers (Bos, De Koning, Wassenburg, & van der Schoot, 2016; E. M. Carr, Dewitz, & Patberg, 1983). Early readers use inferences to aid comprehension, with research showing that children as young as six are capable of drawing inferences from text they have read (Casteel & Simpson, 1991).

Inference making has also been found to be important for LC. For preschool children aged four to five years old the total number of inferences, as measured by inferences made (as defined as conclusions drawn from the story that were not explicitly stated) by a child whilst narrating a wordless book , was related to their LC of the story (Tompkins, Guo, & Justice, 2013). Similar results are found by others across children aged four to eight year olds (Cain et al., 2001; Kendeou et al., 2008; Kim, 2016; Lepola et al., 2012) in UK, US and Finnish samples. Longitudinal work also shows that inference making aged five predicts LC aged eight (Alonzo et al., 2016) for a US sample.

1.4.2 Comprehension monitoring

Comprehension monitoring is the process of reflecting and evaluating one's comprehension to judge the quality of understanding, and to determine that the correct perspective of the text has been gained (L. Baker, 1984a; Pitts, 1983). Comprehension monitoring techniques include re-reading and repair strategies when understanding is unclear (Pitts, 1983). As with inference making, comprehension monitoring is not usually assessed before the age of six because the task demands are too complex[1]. Measures of comprehension monitoring assess a reader's ability to recognise inconsistencies in the text, such as contradictory sentences, as this requires evaluation of the understanding of the text (Cain et al., 2004).

Evidence suggests that comprehension monitoring is important for successful RC, as findings show that comprehension monitoring in eight to 11 year olds explains unique variance in RC (Cain et al., 2004). Additionally, longitudinally

---

[1] Although very recently some LC studies have begun to administer comprehension monitoring measures to children as young as four years old e.g. Strasser and Río (2014). See later paragraphs for more information concerning this.

comprehension monitoring at the age of seven has been shown to predict RC at 11 years old (Oakhill, Cain, & Bryant, 2003; Oakhill & Cain, 2012). This is supported by more recent longitudinal studies which find that comprehension monitoring aged seven makes a significant contribution to RC aged nine, even after controlling for decoding, vocabulary and working memory (Cain & Yeomans-Maldonado, 2017), and that comprehension monitoring at the age of 10 predicts RC a year later (Muijselaar et al., 2017).

Comprehension monitoring during RC can be seen even in beginner readers as young as six. Findings show that six year olds are capable of comprehension monitoring, and that their level of decoding and LC is related to their comprehension monitoring ability (Kinnunen, Vauras, & Niemi, 1998). Comprehension monitoring ability increases with age (L. Baker, 1984b; Cain & Yeomans-Maldonado, 2017), with nine year olds shown to perform significantly better than seven year olds (Cain & Yeomans-Maldonado, 2017).

Comprehension monitoring is also important for LC. Comprehension monitoring (as measured by a task where children were required to identify parts of an oral story which "*did not make sense*") has been found to make a significant contribution to LC in typically developing children aged seven and eight years (Kim & Phillips, 2014; Kim, 2015). The relationship has even been shown in children as young as four when comprehension monitoring was measured with simplified stories and examiners were trained to recognise verbal and nonverbal signals that demonstrated that a child believed the story did not make sense (Strasser & Río, 2014).

Longitudinal work also shows that comprehension monitoring aged five predicts LC aged eight, again using a similar measure of comprehension monitoring where stories that did and did not "make sense" were read to the children (Alonzo et al., 2016). These recent studies also suggest that comprehension monitoring can be measured as early as four years, something previously not done before the age of six due to task demands. This said, reliability issues are raised by Strasser and Rio's interpretation of children's nonverbal signals for identifying their monitoring of a story. Overall, there is clear evidence from cross-sectional, longitudinal and training studies in both typical and atypical populations, that both comprehension monitoring and inference making are important for comprehension, and that without these processes comprehension is poor.

1.4.3 Working memory

As discussed earlier in relation to the search for additional factors of SVR (Section 1.3.1.1), studies have shown there is a relationship between RC and working memory. Although research has not reliably replicated working memory as an additional component to the SVR it is plausible that working memory could indirectly predict RC as there is evidence for a relationship between LC and working memory. Florit et al., (2009) found that working memory aged four and five was a significant predictor of current LC over and above vocabulary and verbal IQ. Longitudinal work has found that working memory at the age of five predicts LC aged eight (Alonzo et al., 2016). In a similar way to RC, they suggest this is because it allows children to hold information they have heard while making inferences and links with existing knowledge and integrating meaning. In support of this, a more recent study found that working memory directly predicted concurrent LC in six and seven year olds (Kim, 2016).

1.4.4 Inhibition and attentional control

Again, as discussed earlier in relation to the SVR and additional components (Section 1.3.1.1) a relationship can be seen between RC and inhibition/attentional control. Although research has not reliably replicated inhibition/attentional control as an additional component to the SVR, evidence still suggests the two are linked (e.g. Conners, 2009; Savage, Cornish, Manly, & Hollis, 2006). The relationship may instead be indirect via oral language as demonstrated by Kieffer, Vukovic, and Berry (2013) who showed a significant indirect association for both inhibition and attention shifting to RC via language comprehension with nine year olds when using path analysis.

More recently, LC has also been linked to both inhibition and attention for six and seven year olds (Kim & Phillips, 2014; Kim, 2016). Kim and Phillips (2014) found that in six year olds inhibitory control (as measured using a modified version of the day/night stroop task) uniquely contributed to LC, after accounting for age and vocabulary. Longitudinally, inhibition (again as measured by the day/night Stroop task) at the age of four years predicted LC seven months later (Scrimin, Patron, Florit, Palomba, & Mason, 2017).

In summary, the research described in the previous sections establishes that high-order skills are of importance for both LC and RC from preschool onwards. These skills may make an indirect contribution to both LC and RC rather than a direct one. This is developed more precisely by Kim (2015; 2016; 2017) in models of LC and RC.

1.5 The Direct and Indirect Effect Models of Reading and Text Comprehension

Following work with Korean and American six to eight year olds (Kim, 2015; Kim, 2016; Kim, 2017), Kim proposes two models: a model of LC (Figure 1.2); the Direct and Indirect Effect Model of Text Comprehension (DIET), and a model of RC (Figure 1.3); the Direct and Indirect Effect Model of Reading (DIER). The DIET model refers to LC and RC together as text comprehension as Kim (2016) states that LC is the comprehension of oral and written text whereas RC is the comprehension of written text and that both include the same processes so together should be referred to as text comprehension (Kim, 2016, p. 102).

These models are hierarchical such that low level skills of vocabulary, syntactic knowledge and working memory, predict high-order skills of comprehension monitoring and theory of mind (ToM), which in turn predict LC. These lower order skills also make a direct contribution to LC as well as an indirect one. The model of RC includes the addition of decoding to the model of LC to predict RC, but here in line with the SVR only linguistic comprehension and decoding make a direct contribution to RC. Across studies (Kim, 2015; Kim, 2016; Kim, 2017) these models have changed slightly, the figures below are the best fitting models from the most recent paper (Kim, 2017) and so reflect most recent thinking.

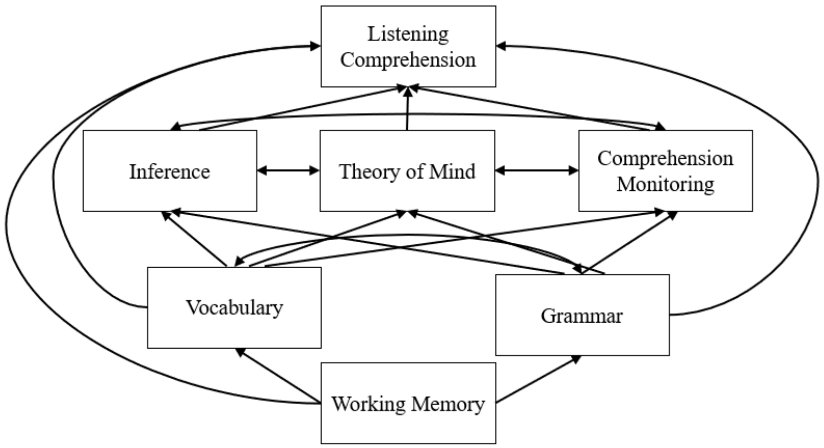


Figure 1.2: The DIET model of listening comprehension according to Kim (2017). *Note.* Two-sided arrows represent covariances.

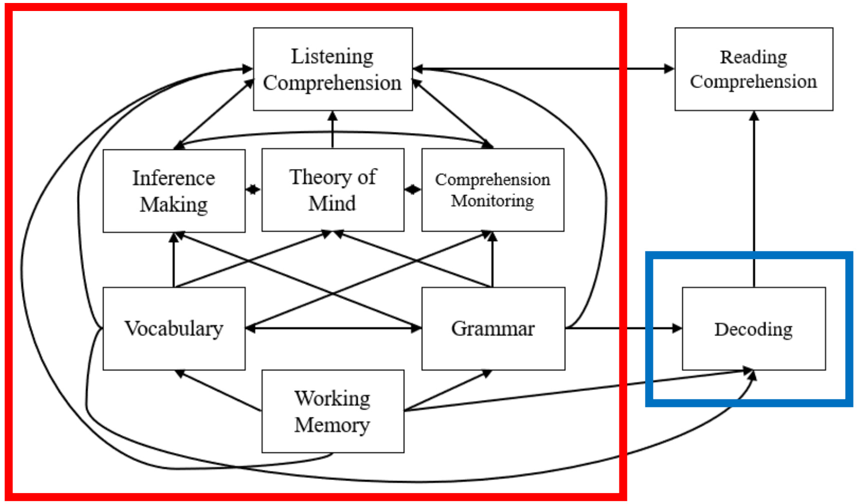


Figure 1.3: The DIER model of reading comprehension according to Kim (2017). *Note.* Two-sided arrows represent covariances. The coloured boxes represent the two components of reading comprehension as suggested by the Simple View of Reading. Within the blue box is the decoding component. Within the red box is the linguistic comprehension component containing all the sub-components of linguistic comprehension including listening comprehension at the top of the hierarchy

1.5.1 The situation model and the DIET and DIER models

In addition to the SVR, the DIET and DIER models are theoretically underpinned by the situation model (Kintsch, 1988; Van Dijk et al., 1983). The situation model differs from the SVR in that the SVR is a model of reading development and how new readers learn to read, whereas the situation model is a framework of text comprehension and describes how the process occurs (Kim, 2017). For this reason, the models are not alternatives to one another and do not contradict each other, rather during RC the two models occur in parallel in that the SVR gives the skills which are important for the process, whereas the situation model describes in detail how this process occurs. In this way Ricketts et al., (2013) describe their findings in relation to both the SVR and the situation model, and Kim (2017) states that her models support both the SVR and the situation model.

The situation model states that successful comprehension is ultimately achieved when an accurate, rich and elaborate mental picture of the situation portrayed in text or oral language is obtained (Kim, 2016; Kim, 2017). During comprehension one must build a mental representation of the message in the text (oral or written) in order to take meaning from it (Kim, 2016; Kim, 2017; Paris & Stahl, 2005; Van Dijk et al., 1983; Zwaan, 2016). When successful, comprehenders construct a micro-world in which they create representations of characters, events, goals and actions described in a story (Zwaan, Langston, & Graesser, 1995). This goes beyond the text (oral or written) and uses prior knowledge to form a mental picture (Paris & Stahl, 2005).

There are a number of variations of such a model, the most influential of which is the construction–integration model by Van Dijk and colleagues (Kintsch, 1988; Van Dijk et al., 1983), with the origins of this model evolving from an earlier model by the same authors (e.g. Kintsch & Van Dijk, 1978). The model has three levels: (a) the surface level (b) the textbase level (c) the situation model. As depicted in Figure 1.4, these three levels are said to be hierarchical, whereby the situation level is built upon the textbase representation, which is built upon the surface level. In the surface level the reader extracts key words and phrases from the text. In the textbase level the reader takes literal meaning from the text. In the situation level the reader integrates this literal meaning with their prior knowledge to create a mental image of the text (Kintsch, 1994). Whilst here this model has been described in the context of comprehending text during RC, this model can also be used to explain LC in the DIET model.

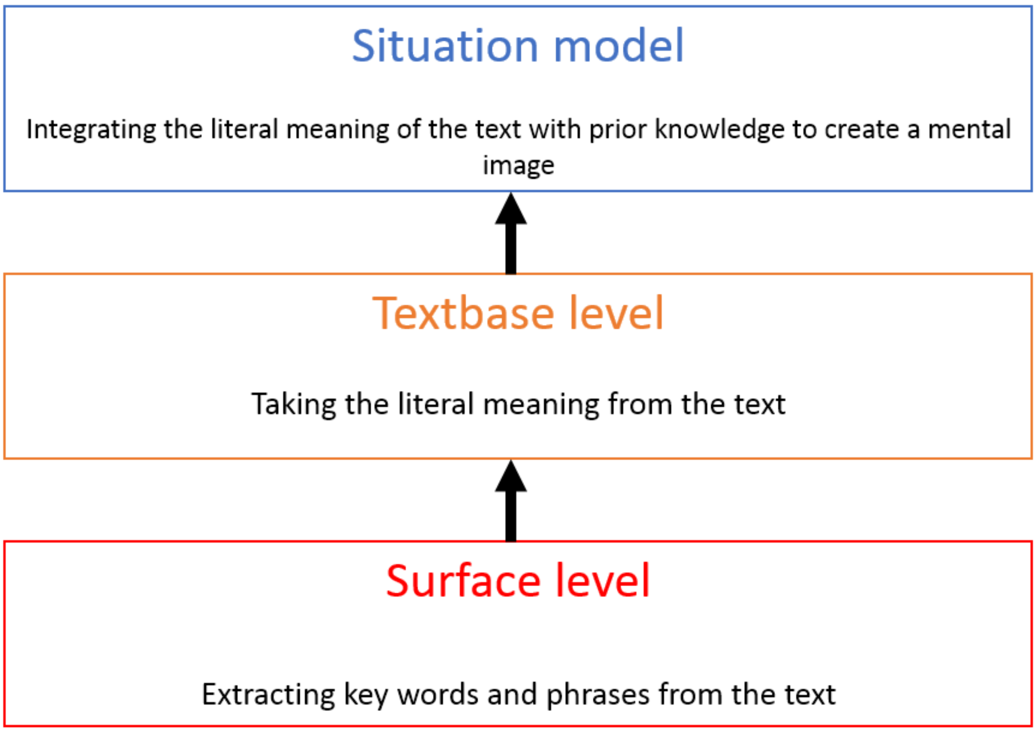


Figure 1.4: The three levels of text representation proposed by Van Dijk et al., (1983) and Kintsch (1988).

The DIET and DIER models (Kim, 2017) hypothesize that each level of the situation model requires a different set of language and cognitive skills and that these language and cognitive skills have hierarchical structural relations in that lower level processes feed into higher level processes.

1.5.2 A direct only model of reading comprehension

As already stated, the DIER model of RC is an alternative perspective to the work concerned with searching for an additional component of the SVR (e.g. Conners, 2009; Kirby & Savage, 2008; Ouellette & Beers, 2010), as it instead seeks to explain the complexity of reading by unpacking the linguistic comprehension element of the SVR. However, Kim (2017) also describes a direct model of RC in which working memory, vocabulary, syntactic knowledge, inference making, theory of mind and comprehension monitoring make a direct contribution to RC (see Figure 1.5). Like the DIER model (see Figure 1.3), this direct only model was shown by Kim (2017) to have a good fit for 350 children aged seven years. This leaves uncertainty over whether there are factors which could be added to the SVR to add extra variance. Despite not including the decoding component of the SVR this direct model explained 66% of total variance in RC (Kim, 2017).

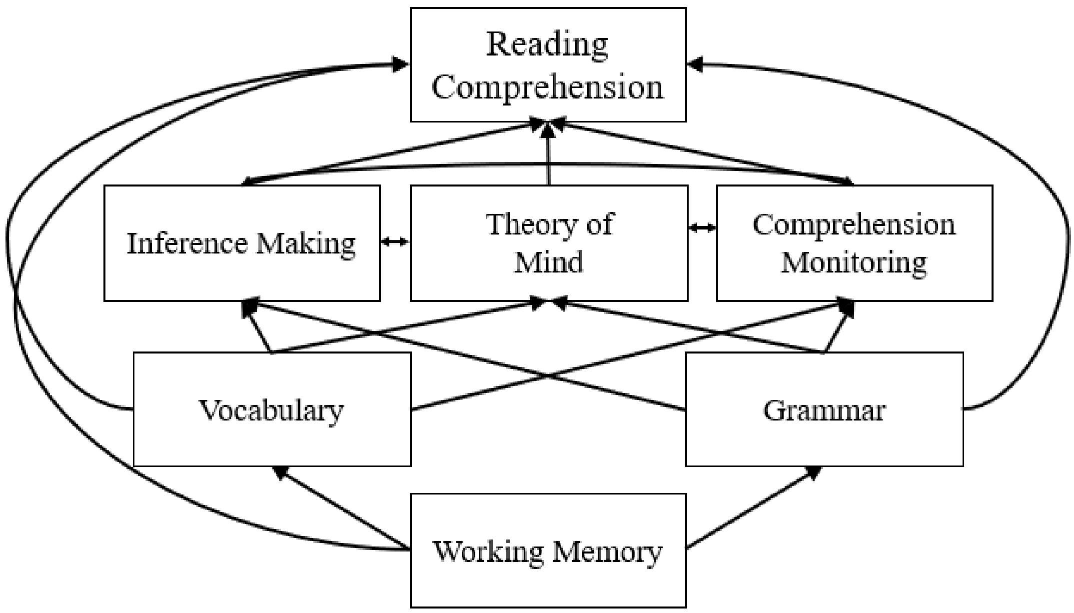


Figure 1.5: A direct only model of reading comprehension as described by Kim (2017). *Note.* Two-sided arrows represent covariances.

As well as proposing that the well-established skills (vocabulary, working memory, syntactic knowledge and comprehension monitoring) contribute to RC and LC, the models of Kim (2015;2016;2017) also suggest that ToM is needed for proficient comprehension. ToM is a cognitive skill which has not until recently been considered in relation to RC and LC. The subsequent section of this chapter explains why ToM can be seen as important for early LC and RC development.

1.6 Theory of mind

Theory of mind (ToM), a phrase coined by primatologists Premack and Woodruff (1978), is defined as the ability to impute mental states to oneself and to others. Mental states include intentions, desires, beliefs and perspectives, and those with a proficient ToM use these mental states to predict and explain behaviour (Astington, 2001). ToM is widely assessed using false belief tasks in which the individual must attribute a false belief to a character or a past false belief to themselves (Wellman, Cross, & Watson, 2001). For example, a classic unexpected transfer task (seen in Figure 1.6) concerning the location of a moved object. In this task character 1 places an object in location A before leaving. In character 1's absence, character 2 moves the object from location A to location B. An understanding of ToM can be attributed to the child if they expect character 1 to look in location A on their return, where they originally placed the object, rather than location B, where the child knows object is (Baron-Cohen, 1989; Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983).

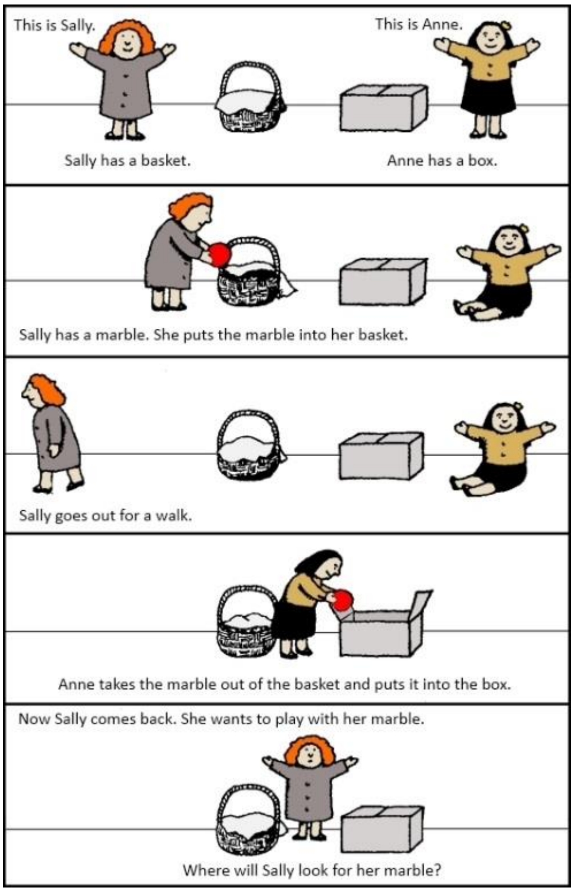


Figure 1.6: The unexpected transfer task by Baron-Cohen et al. (1985).

There is a crucial milestone around the age of four when children acquire this false belief understanding. This is indicated in performance on the above false belief task as demonstrated in the original study by Wimmer and Perner (1983) who showed that no 3-4 year olds, 57% of 4-6-year olds, and 86% of 6-9 year olds correctly stated that the protagonist would look for the object in the original location. This was later confirmed by a meta-analysis of 178 separate studies (Wellman et al., 2001). This age milestone is seen cross-culturally, as demonstrated by Callaghan et al. (2005) who showed that children from Canada, India, Peru, Samoa, and Thailand all passed the same false belief test at a similar age. Perner (1991) argues that passing this task reflects a developmental shift in representational understanding of the mind in that children are now able to represent (think) how someone represents (thinks about) something. Perner and colleagues describe the possession of a ToM as mental model building (Johnson-Laird, 1983; Perner, 1991) as a belief is a mental model of the world. They suggest that children fail a false belief test because they do not understand that mental models or representation can differ from true reality (Lillard & Flavell, 1992; Perner, 1991).

1.6.1 Theory of mind and language

Numerous studies, including correlational and longitudinal studies show that language ability is related to ToM understanding. For example, Jenkins and Astington (1996) found a correlation between false belief and general language ability (using broad measures of language e.g. vocabulary and grammar) in typically developing three to five year olds, a finding supported by others (e.g. Cutting, & Dunn, 1999; Hughes & Dunn, 1997). Indeed, a meta-analysis of 104 studies indicated a moderate to large correlation between language and false belief understanding independent of age (Milligan, Astington, & Dack, 2007). Moreover, this correlation can be seen across different language skills including lexical knowledge (or vocabulary), semantics, syntactic knowledge, and pragmatics (Milligan et al., 2007).

Longitudinal studies also support a relationship between ToM and language, showing that language ability aged two predicts false belief understanding at the age of four (Farrar & Maag, 2002; Watson, Painter, & Bornstein, 2001), and that vocabulary at the age of four correlates with later ToM understanding (Hughes, 1998b). Longitudinal studies also show that general language (including semantics and syntax) at the age of three predicts later false belief understanding measured at several subsequent time points (Astington & Jenkins, 1999; Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003).

There is a strong and a weak version of the explanatory role that language plays in ToM (Astington & Jenkins, 1999). The weak version states that the relationship between language and ToM reflects the task demands of a false belief task in that a certain level of language ability is needed in order to pass the tasks. This version of the explanation does not state that language is necessarily unimportant for ToM, just that methods for testing it limit the argument (Astington & Jenkins, 1999). Indeed when language was simplified in a classic false belief test by including temporal markers (e.g. "Where will Sally look *first*"), both Lewis and Osborne (1990) and Siegal and Beattie (1991) showed that even three year old children could pass the test. This highlights that the language ability needed to pass a false belief could confound findings which show that language is associated to ToM. On the other hand, the strong version of the explanation maintains that language is fundamental to the conceptual development of ToM. This strong version is argued by Ruffman et al. (2003) and supported by auto-regressive longitudinal studies which show that language predicts later false belief understanding even when controlling for earlier false belief understanding (Astington & Jenkins, 1999).

In order to address the language demands of a false belief test, some have used implicit (non-verbal) false belief tests. For example, Onishi and Baillargeon (2005) used a violation of expectation paradigm with 15 month olds. After infants were familiarised with a researcher hiding a toy in one of two locations, then returning later to retrieve the object from that location, they were shown scenes where the toy was moved without the researcher's knowledge. Infants were then presented with the researcher searching for the hidden toy either (a) where the researcher falsely believed it to be, or (b) where it was actually located. Infants looked longer at the 'unexpected' event, which Onishi and Baillargeon (2005) suggested to be evidence that infants expected the researcher to search for the toy where they believed it to be located, thus the infants had attributed a false belief to the researcher. Similar findings have been found by more recent studies such as Yott and Poulin-Dubois (2012) who replicated the findings with 18 months olds, and Kovacs, Teglas, and Endress (2010) who showed that even seven month olds looked longer at unexpected events when viewing a scene involving a Smurf and a false belief about a ball behind a barrier, a finding also paralleled with adult participants. In addition to providing evidence for the strong version of the explanatory role that language plays in ToM, these studies also make the suggestion that a ToM understanding develops earlier than first thought. Yet research into the relationship between children's language and their ability to pass an implicit false belief test is still to be conducted (Milligan et al., 2007).

There is also an important debate over whether the relationship between language and ToM is unidirectional or bidirectional. Astington and Jenkins (1999) originally argued that it is a unidirectional relationship in which earlier language abilities predict later ToM performance, but earlier ToM does not predict later language skills (after controlling for earlier language ability). However, Slade and Ruffman (2005) challenged this as their findings from their cross-lagged longitudinal study suggest both that earlier performance on measures of language ability predicts later false belief task performance, as well as the reverse, that false belief performance has an effect on later language. More recent meta-analytic evaluation of this by Milligan et al., (2007) showed that though stronger for earlier language predicting later ToM, there was clear evidence for a bidirectional effect. This is important for indirect models of reading discussed previously which include both language and ToM.

1.6.1.1 Theory of mind and emergent literacy

Evidence suggests that ToM understanding remains important for language development as children progress to more demanding language challenges, with a relationship found between false belief understanding and emergent literacy. Storytelling has been linked to ToM understanding as it is suggested that for successful comprehension of a story children must be able to consider the mental states of characters, such as their emotions, intentions and beliefs, to make inferences about the reasons for their actions (Emery, 1996; Shanahan & Shanahan, 1997). Indeed, it has been found that young children's ability to make inferences about characters' goals, actions that achieved those goals, and characters' mental states, significantly predicts their story comprehension while narrating a wordless book (Tompkins et al., 2013). Parallel findings come from Pelletier and Astington (2004), who showed that young children with a less developed ToM (as indexed by failing false belief tasks) fared worse at retelling a story because they were unable to describe the thoughts of characters and their reasons for holding such thoughts. Similarly, when told stories with plots revolving around false belief occurrences, children who passed the false belief questions in these stories were also able to retell the story better because they could explicitly articulate the false belief (Riggio & Cassidy, 2009).

1.6.1.2 Theory of mind and reading comprehension

As research shows that ToM and emergent literacy involving storytelling are closely linked, it is worth considering the role it may play in RC. Indeed, a recent review has suggested that ToM may be the "hidden factor" in RC (Dore, Amendum, Golinkoff, & Hirsh-Pasek, 2018) as it is relatively unexplored. Indeed, to date only a limited number of researchers have focused on the relation between ToM and RC.

The first to investigate the role of ToM in RC were Ricketts, Jones, Happé, and Charman (2013). The authors were motivated by the difficulties with reading seen in those with Autism Spectrum Disorder (ASD) and explored the relationship in a group of adolescents. Consistent with the SVR they found that both decoding and linguistic comprehension were unique predictors of RC. However, they also found, importantly, that mental state understanding (as indexed by their performance on advanced theory of mind tasks) was a significant predictor of RC after accounting for linguistic comprehension and decoding. The authors conclude that there are factors which contribute to RC for those with ASD which are not captured by the SVR. Ricketts et al., (2013) discuss these findings in the context of the situation model. They hypothesise that readers with ASD have difficulties constructing a reliable situation model. They suggest that a meaning based representation of text that is integrated with prior knowledge and experience is not fully formed by autistic individuals. This is because these readers fail to understand social norms, and in particular difficulties in mentalising may thwart their ability to form an extensive situation model of the text. However, this research needs to be extended to a typically developing population of readers, if ToM is to be added to a model of reading.

The link between RC and ToM in typically developing children was first explored by Kim (2015) whose later model (Kim, 2015; Kim, 2017) is examined in Section 1.5. Despite being the first to propose a model of RC (see Figure 1.3) which included a role for ToM, the author acknowledged the need for longitudinal studies to determine directional and causal nature of the relations between ToM and RC proposed in the model which was developed with cross-sectional data. However, to date, very few longitudinal studies have been published. Guajardo and Cartwright (2016) conducted a small-sample longitudinal study with 31 three to five year olds (Time 1), following them up when they were six to nine years (Time 2). Their findings showed that false belief understanding at Time 1 contributed to children's phrase and sentence comprehension also at Time 1, and to later (Time 2) reading awareness, assessed by questions regarding thinking about reading habits and RC e.g. "*How can you tell which sentences are the most important ones in a story*?" and "*If the teacher told you to read a story to remember the general meaning, what would you do*?". However, in contrast to their predictions false belief understanding at Time 1 did not contribute uniquely to RC at Time 2. Although correlation between false belief understanding and RC approached significance, false belief understanding did account for unique variance in RC. On the other hand, cognitive flexibility and counterfactual reasoning did account for unique variance in RC. Despite these findings Guajardo and Cartwright (2016) suggest that because ToM was shown to contribute to later reading awareness this highlights the importance of ToM for the development of RC. Additionally, the authors note that a primary limitation of their work was the small sample size (n = 31), suggesting that results, which were approaching significance, may have been significant if a larger sample had been used.

The most recent study to assess the role of ToM in RC is that of Atkinson et al. (2017). This study was longitudinal and tracked the development of skills over two and a half years. Additionally, it had a larger sample than that of Guajardo and Cartwright (2016), with 80 children followed from preschool to Year 1. Their mediation analysis showed two things, firstly that ToM when children were around four years old indirectly predicted RC via language ability at the later time point when they were six years old. Secondly, that ToM at the first time point also directly predicted RC over and above the skills in SVR at the later time point. This is the first study to show a direct longitudinal relation from preschool ToM to emerging RC when children are aged six.

1.6.1.3 Theory of mind and listening comprehension

As with RC it is possible that ToM plays a role in LC. The rationale for this is the same as that for RC i.e. that a child who is more aware of the intentions of a character in a story or the intention of the speaker will be able to form a more comprehensive situation model and thus comprehend the story better (Dore et al., 2018). Indeed, correlational studies show that LC is correlated with ToM for 5-8 year olds (Kim & Phillips, 2014), and structural equation modelling shows that LC is directly predicted by concurrent ToM for 6-7 year olds (Kim, 2015; Kim, 2016). This relationship is seen across ages as Pelletier and Beatty (2015) show that for both four year olds and twelve year olds ToM understanding predicted children's understanding of fable stories read to them.

However, other research has found that although the two correlated, ToM does not contribute to unique variance in LC for four to six year olds (Strasser & Río, 2014). Due to this limited and mixed evidence, and because there is no longitudinal research, this thesis also explored the role played by ToM in LC. Additionally, Kim (2015) and Kim (2017) make the case that research should explore the contribution of language and cognitive skills to RC and LC in parallel not separately. Kim states the theoretical framework (the situation model and the SVR) does not differentiate between oral versus written text comprehension and so it is important to examine the contribution of language and cognitive skills to LC and RC together for generalisability of theoretical models. This perspective is reinforced by Hoover and Tunmer (2018) who state that both reading and listening comprehension require the same cognitive capacities just the point of accesses differ, one is through print and the other through speech.

1.6.1.4 Summary of the role of theory of mind in listening and reading comprehension

To summarise, only a limited amount of research has explored the role of ToM in RC and even less its role in LC. As well as reaching somewhat contradictory findings, these studies have suffered from small sample sizes and/or were not longitudinal in nature. To claim that ToM should be added to a model of RC or LC, future research needs to be longitudinal, measuring all related variables before children experience formal reading education and track their development over several years as they progress into Year 1, when they will begin to become proficient comprehenders of text. Additionally, in terms of a model of RC it is still unclear if additional factors which add extra variance to the SVR model should be searched for, or if future work should concentrate on unpacking the component skills of linguistic comprehension and look at indirect predictors of RC. Therefore, in addition to addressing the above limitations of the past work, the intention of this thesis was to compare both the direct and indirect contribution of ToM to RC in models.

Moreover, to fully assess the involvement of ToM in LC and RC, research must explore the specific element of ToM which contributes towards comprehension.

Specifically, since ToM is a metacognitive skill (Flavell, 2000), this thesis aimed to assess whether it is the general metacogntive nature of ToM that is important or the more socially relevant aspect more specifically. This is novel and will allow a deeper understanding of the influence ToM has to RC and LC. The following section of this chapter will discuss this.

1.7 Metacognition

Metacognition, a notion coined by Flavell (1976; 1979) is defined as any knowledge about cognitive phenomena. Essentially, it is the skill of knowing about knowledge, or thinking about thinking. Under this definition, metacognition can be seen as a broad umbrella term encompassing a variety of skills which all require the ability to understand cognition. For example, Flavell describes a task in which children are asked to study a set of items until they can perfectly recall them (Flavell, Friedrichs, & Hoyt, 1970; Flavell, 1979). Findings showed that older children usually correctly perceived when they had studied the items long enough to recall them perfectly, whereas younger children did not (Flavell et al., 1970). This task is metacognitive as the children are required to think about their own cognition (and memory capabilities) to recognise when they had learnt the items proficiently. This type of metacognition is called metamemory, knowledge of one's own memory capabilities and strategies that can aid memory (Lockl & Schneider, 2007).

Other metacognitive skills seen in young children include metalinguistic awareness and source monitoring. Metalinguistic awareness is the ability to reflect on the use of language (Doherty & Perner, 1998; Doherty, 2000), and in young children is often measured using homonym judgement tasks in which children must recognise that some words have two or more distinct, unrelated, meanings (e.g. bat as a flying mammal and bat as sports equipment), and synonym judgement tasks in which children must recognise that some things can have more than one name or label (e.g. a lady can also be called a woman). These tasks are metacognitive as they require the child to think and reason about the use of language. Source monitoring, on the other hand, is the understanding of the source of one's own knowledge (Bright-Paul, Jarrold, & Wright, 2008; O'Neill & Gopnik, 1991). In young children it can be assessed by asking the child to recognise the source of their knowledge of the identification of objects. For example, they know that the object is a ball because they felt it, or they know that it is a toy car because somebody told them so, or they know that it is a box of crayons because they saw them (O'Neill & Gopnik, 1991). This source monitoring task is metacognitive because children must think and reason about their own knowledge. These metacogntive abilities, including ToM, all show significant development over the preschool and primary school ages (Doherty, 2000; Lockl & Schneider, 2007; O'Neill & Gopnik, 1991).

1.7.1 Metacognition and theory of mind

ToM can be viewed as a metacognitive skill because it involves thinking about the mental states of oneself and others (Premack & Woodruff, 1978). The ability to pass a false belief test is undoubtedly metacognitive because it involves thinking about how the protagonist will think (Flavell, 2000). Despite being metacognitive, ToM research has been historically relatively distinct and unconnected to research into other types of metacognition e.g. metamemory (Lockl & Schneider, 2007). A reason for these separate research paths could be due to the uniqueness of ToM as a socially-specific skill as it is concerned with others and social information, whereas other types of metacognition are concerned with cognition more generally. In this way ToM can be said to be socially-specific metacognition, whereas other types of metacognition (e.g. metamemory) are more broad in nature.

More recently ToM has been linked to these other types of metacognition such as metamemory (Lecce, Bianco, Demicheli, & Cavallini, 2014; Lockl & Schneider, 2007), metalinguistic awareness (Doherty & Perner, 1998; Doherty, 2000) and source monitoring (Bright-Paul et al., 2008). This relationship between ToM and other non-social types of metacognition can be viewed in two ways. Either the two are related because they tap the same underlying concept (Iao, Leekam, Perner, & McConachie, 2011; Perner, 1991) i.e. they draw on common ability to understand how something can be thought about (represented) in different ways. Or, alternatively the socially specialised ability of ToM facilitates these broader metacognitive skills (Lockl & Schneider, 2007; Ricketts et al., 2013). In support of this some studies have shown that ToM facilitates some aspect of metacognition including metamemory, but not the reverse (Ebert, 2015; Lockl & Schneider, 2007) suggesting that ToM is a specific ability somewhat separate to these other aspects of metacognition.

1.7.2 Metacognition and language

Given the conceptual similarities between metacognition and ToM, a link between metacognition and language seems likely. Indeed, longitudinal studies show a concurrent and longitudinal correlation between vocabulary and metaknowledge about reading in children aged 6-9 years (Lecce, Zocchi, Pagnin, Palladino, & Taumoepeau, 2010) and a correlation between metamemory at five years old and earlier language competencies (Lockl & Schneider, 2007).

Regarding LC, Annevirta, Laakkonen, Kinnunen, and Vauras (2007) describe a high correlation between metacognition (assessed using children's understanding of the cognitive processes of remembering, understanding, and learning) and LC in five year old children. These findings are supported by work with older children showing that the use of metacognitive strategies correlates with LC ability in nine to 12 year olds (Vandergrift, 2002), and training studies which show that teaching of metacognitive strategies successfully improve LC in nine to 12 year olds (Brand-Gruwel, Aarnoutse, & Van Den Bos, 1998; Goh & Taib, 2006).

There is also research concerned with metacognition and RC but this has primarily concentrated on metacognitive strategies to help older readers (aged eight onwards). Two such metacognitive skills are comprehension monitoring and inference making and have been previously discussed in Sections 1.5.1 and 1.5.2. Comprehension monitoring and inference making are metacognitive in nature because they both require awareness and understanding of one's own thought processes (L. Baker & Brown, 1984; Flavell, 1979; Halpern, 1998; Kinnunen et al., 1998; Pitts, 1983).

Comprehension monitoring calls upon the reader to think about their own understanding of text (Dabarera, Renandya, & Zhang, 2014; Kinnunen et al., 1998; Paris & Myers, 1981; Pitts, 1983). Likewise, during inference making the reader must think about their own knowledge to gain a deeper understanding (Cain & Oakhill, 1999; Cain et al., 2001; Cain et al., 2004).

Yet, metacognitive reading strategies are not observed in preschoolers or early readers, when reading is largely confined to very simple texts. To date, research has not addressed whether very early metacognitive skills are associated with the development of RC. Given both that metacognition is important for RC in older readers, and that ToM has shown to be related to early RC, a link with broad metacognition and RC is plausible. Moreover, Atkinson et al., (2017) make the argument that research such as theirs, which has found a direct contribution of ToM to RC, does not indicate whether it is the social element of ToM which promotes RC, or instead whether it is the broad metacognition nature of ToM assisting RC. This links back to the debate in Section 1.7.1 regarding the relationship between ToM and other forms of metacognition. If ToM and other forms of metacognition tap the same general meta-representational ability, we would expect measures of metacognition (such as metamemory, metalinguistic awareness and source monitoring) to have similar effects on LC and RC as that shown by ToM. Alternatively, if it is that ToM facilitates other forms of metacognition because of its social nature, we would expect that measures of metacognition will not have comparable effects on LC and RC in comparison to ToM. Examining these skills at this earlier age, in relation to later RC, is therefore novel and has the potential to contribute to understanding both of the role of metacognition in reading, and the relationship between ToM and other forms of metacognition.

There is a logical argument that it is the social component of ToM which assists RC and LC, because if a child has a better awareness of the perspectives of a character in a story, or indeed the perspective of the author or speaker of the story, then they are also more likely to have a better understanding of the story itself (Atkinson et al., 2017; Dore et al., 2018). However, there is an alternative explanation, instead it could be that it is the broad metacognitive nature that ToM taps which assists RC because it informs knowledge, actions, and understandings, not necessarily related with story characters or oneself as a reader (Atkinson et al., 2017). Exploring the role of social-specific versus broad metacognition was the key purpose of this thesis.

In summary, recent research has suggested a link from ToM to both LC and RC (Atkinson et al., 2017; Guajardo & Cartwright, 2016; Kim & Phillips, 2014; Kim, 2015; Kim, 2016; Ricketts et al., 2013). However, no research to date has explored the role of broader metacognition in early RC and LC development. Following discussion by Atkinson et al., (2017) this thesis aimed to determine if it is specifically the social nature of ToM which supports RC and LC, or whether it is the general metacognitive nature that is important.

1.8 Mental state talk and theory of mind

A significant correlate of children's developing ToM is maternal mental state talk (Ruffman, Slade, & Crowe, 2002). This is where a mother has conversations involving mental states, such as cognition (e.g. "think" "know"), emotions (e.g. "sad" "happy") and desires (e.g. "want" "dislike") with her child. Over twenty five years ago Dunn, Brown, Slomkowski, Tesla, and Youngblade (1991) conducted their seminal study showing that the frequency of maternal mental state talk predicted individual difference in a child's ToM understanding and their ability to pass a false belief test. This created a growth in work into family influence of ToM (See Hughes & Devine, 2015 for a review), with research looking at the effects of socio-economic status (e.g. Hughes et al., 2005; Pears & Moses, 2003; Shahaeian, 2015), family size (e.g. Cutting, & Dunn, 1999; Jenkins & Astington, 1996; Perner, Ruffman, & Leekam, 1994) and parental mind-mindedness, parents' ability to treat their child as an individual with a mind capable of thoughts and feelings, (e.g. Ereky-Stevens, 2008; Lundy, 2013; Meins et al., 2002; Meins, Fernyhough, Arnott, Leekam, & Rosnay, 2013) as well as maternal talk. Together these findings underpin a social account of ToM (e.g. Carpendale & Lewis, 2006; Heyes & Frith, 2014) which suggests that as well as children's own cognitive development (such as their language ability), social understanding is driven by socially mediated processes (Nelson, 2004). This was also the interpretation of Dunn et al. (1991) who proposed that through talk about mental states, thoughts and memories, children's attention is focused on mental states which facilitates their developing ToM, as such this can be viewed within a Vygotskian framework (Taumoepeau & Ruffman, 2006; Taumoepeau & Ruffman, 2008).

A recent meta-analysis (Devine & Hughes, 2016a) identified 28 studies looking at the relationship between maternal mental state talk and ToM. Some have examined the relation using self-report methods where mothers gave details on their mental state talk language usage (Farrant, Maybery, & Fletcher, 2012; C. Peterson & Slaughter, 2003) but most studies have focused on observing mothers and their children in different settings. These studies have examined the link across children aged 3-5 years under various different circumstances, such as the language used during interaction with a picture book or pictures (Adrian, Clemente, Villanueva, & Rieffe, 2005; Adrián, Clemente, & Villanueva, 2007; LaBounty, Wellman, Olson, Lagattuta, & Liu, 2008; Ruffman et al., 2002; Slaughter, Peterson, & Mackintosh, 2007), conversations during joint play (Howard, Mayeux, & Naigles, 2008; Laranjo, Bernier, Meins, & Carlson, 2014; Moeller & Schick, 2006), talk during storytelling (Symons, Peterson, Slaughter, Roche, & Doyle, 2005) and dialogue during other family activities such as meal times (Ensor, Devine, Marks, & Hughes, 2014; Howard et al., 2008). With the exception of one study, that by Ontai and Thompson (2008), all these studies found evidence for a link between maternal mental state talk and child ToM understanding. In fact, the meta-analysis of Devine and Hughes (2016) shows that across these studies and 1,914 two to five year olds there was a modest but statically significant relationship between mental state talk and the ability to pass a false belief test which held even when verbal ability was accounted for.

Studies have also attempted to explore different elements of mental state talk. These studies have found that the content of mothers' mental state talk changes over time. Mothers use a higher frequency of desire language than cognitive language and talk about emotion when children are two years old (Taumoepeau & Ruffman, 2008), but a shift occurs so that by the age of six mothers are making twice as many cognitive references than emotion and desire references (Ensor et al., 2014; Jenkins, Turrell, Kogushi, Lollis, & Ross, 2003). This change over time also affects the relationship between mental state talk and ToM. At 15 months talk about the child's own desires is the best predictor of ToM at 24 months (Taumoepeau & Ruffman, 2006), whereas talk about other people's desires and cognitive talk is a weaker correlate. Later on though (at 3 years old), talk about cognition is the best predictor of ToM at 5 years old (Adrián et al., 2007). This shift is thought to help scaffold the development of ToM (Ruffman, 2014).

The relationship between mental state talk and ToM also persists across time with longitudinal studies finding that earlier maternal mental state talk is associated with later ToM understanding. For example, Adrián et al. (2007) showed that mothers' earlier use of mental state talk in picture book reading (children had a mean age of four years seven months) correlated with children's later ToM understanding (mean age five years nine months). Likewise, Ensor et al., (2014) found that mothers' talk when their children were two years old predicted ToM understanding at the ages of both six years and ten years old. This relationship still holds under highly controlled conditions as demonstrated by Ruffman et al. (2002), who collected data on mental state talk and ToM across three time points, finding that mothers' mental state talk at Time 1 (mean age 3.01 years) and Time 2 (mean age 3.41 years) correlated with the children's ToM during Times 2 and 3 (mean age 4.04 years). This association held even when potential mediators were accounted for, including children's Time 1 and 2 language ability and ToM. Yet a reciprocal relationship was not found, that is children's earlier ToM did not predict mothers' later mental state talk. Ruffman et al.

(2002) argue that this unidirectional relationship is consistent with the idea that mental state talk causally facilitates children's developing ToM.

1.8.1 Mental state talk, theory of mind and listening comprehension

There is strong evidence for a link between maternal mental state talk and ToM, whereby mothers' mental state talk predicts child ToM understanding (Devine & Hughes, 2016a). Additionally, as demonstrated in Section 1.6.1 there is also strong evidence for a link between ToM and language ability including LC. However, no study to date has explored these three factors together. Given that shared book reading is a great opportunity for talk about mental states (Symons et al., 2005) it is possible that maternal mental state talk during this type of activity facilitates ToM understanding which in turn improves children's comprehension of stories read to them. Investigation into ToM as a mediating factor in the relationship between mental state talk and LC will give a deeper understanding into relationship between ToM and LC[2].

1.9 Chapter summary

RC and LC are complex processes underpinned by a multitude of factors. The SVR, the most influential model of RC, sees RC as a product of decoding and linguistic comprehension (Gough & Tunmer, 1986; Hoover & Tunmer, 2018). There is much empirical support for this SVR model across all age ranges and languages. However, a number of critics suggest that the SVR may be too simple and have therefore

[2] Note that within this thesis only the relation between these skills and LC, not RC, was explored due to time constraints.

attempted to expand the model. Researchers concerned with expanding the SVR take two alternative perspectives. One perspective is to look for an additional component of the SVR to explain extra variance in the prediction of RC, although as of yet no additional component has been reliably replicated. The other position is concerned with unpacking the existing components of the SVR to find their sub-skills. Both perspectives are taken by the work of Kim (2017). Kim (2017) proposes models of LC (the DIET model) and RC (the DIER model) which attempt to unpack LC and RC and the SVR. The models suggest that linguistic comprehension comprises sub-skills which include both language and cognitive skills. Kim (2017) also tests a fully direct model of RC which sees skills beyond the components of the SVR making a direct contribution to RC. All these models include a place for ToM, a higher-order cognitive skill relatively unexplored in relation to LC and RC. However, crucially the DIET and DIER models are not longitudinal and therefore cannot determine directional and causal relations fully. Therefore, this thesis aimed to extend Kim's models longitudinally.

Moreover, past research has not examined the element of ToM which is important for LC and RC. This thesis aimed to assess whether it is the social relevant aspect of ToM which assists the development of RC and LC, or if instead it is the broad metacognitive nature of ToM. This has not been explored before and will ensure a deeper understanding of the influence ToM has on LC and RC.

Given the importance of ToM for facilitating LC (and potentially later RC), it is crucial to examine the factors that, in turn, may in turn facilitate ToM. There is a well-established link between maternal mental state talk and ToM (Ruffman et al., 2002; Taumoepeau & Ruffman, 2006), and recent evidence suggests a relationship between ToM and LC (Kim, 2015; Kim, 2016; Kim, 2017). However, no study has looked at these three factors together to examine whether maternal mental state talk is linked to LC, and if so if this is mediated by ToM. The applications of this work are important for future approaches to promoting (and interventions to support) children's emerging reading.

1.9.1 Summary of the main thesis aims

1) To validate the concurrent DIET and DIER models of Kim (2015; 2016; 2017) for predicting listening comprehension longitudinally including a place for theory of mind
2) To examine whether it is the socially specific aspect of theory of mind which assists concurrent listening comprehension, or if instead it is the broad metacognitive nature of theory of mind.
3) To compare theory of mind to broad metacognition as a predictor of listening comprehension longitudinally.
4) To compare theory of mind to broad metacognition as a predictor of reading comprehension both concurrently and longitudinally looking at both direct and indirect models of reading comprehension.
5) To assess whether maternal mental state talk predicts early listening comprehension and if so if this is mediated by theory of mind.

These five aims form the five results chapters of this thesis and together they aim to give a better understanding of the influence ToM has on early LC and RC development.

2 Methodology

2.1 Overview and aims

This thesis had five aims explored in five results chapters. The first aim was to assess if Kim's cognitive DIET model of listening comprehension (LC) is valid longitudinally (Kim, 2015; Kim, 2016; Kim, 2017). This aim was novel as although Kim highlights a need for longitudinal work this has never been conducted. The second aim was to compare the contribution of socially specific metacognition (or theory of mind) to broad metacognition in predicting LC concurrently, and the third aim was to explore these relationships longitudinally. Aim four expanded the previous aims to reading comprehension (RC) to determine the contribution of metacognition for reading comprehension (RC) theory of mind to broad metacognition in predicting LC within DIET and DIER models (Kim, 2015; Kim, 2016; Kim, 2017) concurrently and then longitudinally. Aims two to four were novel as they explored the social component of theory of mind (ToM) in supporting LC and RC comparing it to broader metacognitive skills, to determine if the role of metacognition in LC and RC is socially specific. The final aim was to investigate the effects of maternal mental state talk on ToM and LC ability, which are three factors which have not previously been explored together. This methodology chapter will explain the methods used to address these five aims and justify their use. It will give descriptive information about the participants and clarify the measures and procedure in detail.

2.2 Study design

The design was longitudinal with three time points. Preschool children were recruited the year before they began primary school, when they were aged three to four years (Time 1). This age was crucial because here the children had not yet had any formal training in reading and so skills that precede reading could be measured. The children were assessed again a year later when they were four to five years old (Time 2), and again a further 10 months later when they were five to six years old (Time 3). The sample comprised two cohorts; Cohort 1 who were followed for all three time points, and Cohort 2 who were followed up to Time 2. The longitudinal approach enabled children's progress in a number of skills to be tracked over two years. The inclusion of three time points meant that consistent developmental patterns could be observed, and causal relationships could be suggested.

Two cohorts were recruited, one through schools and one directly via parents (see Section 2.3.1 for further details). A second cohort was needed because Cohort 1 (who were recruited through school) showed low uptake for parental measures (see Section 2.7 for information on these measures) and the inclusion of Cohort 2 resulted in an increased sample for these parental measures. All measures administered to the cohorts were the same except for an additional measure of LC, The Oral and Written Language Scales (OWLS: Carrow-Woolfolk, 2011) at Time 1 for Cohort 2. This inclusion of the OWLS was to remain consistent with recent published work (Kim & Phillips, 2014; Kim, 2015; Kim, 2016; Kim, 2017). Additionally, it strengthened the measurement of LC as an outcome measure as it included a broader range of items and so increased sensitivity.

2.3 Participants

During Time 1 children had not yet begun primary school and so had received no formal teaching in reading skills. Table 2.1 gives demographic information and attrition rates. After Time 1, ten participants were excluded: five due to very low functioning English ability (which caused difficulties with engagement in the measures), and five because they were unable to participate in both testing sessions due to illness. In these cases, their data for Time 1 was removed, and they were not assessed at future time points. These excluded participants are not included in Table 2.1.

Table 2.1

Participant demographics for all three time points and attrition rates

Characteristic		Time 1	Time 2	Time 3
N	Both cohorts	204	162(21%[a])	-
	Cohort 1	150	114 (24%[a])	107(29% [b])
	Cohort 2	54	48(12% [a])	-
Gender	Males Cohort 1	80	59	54
	Females Cohort 1	70	55	53
	Males Cohort 2	25	23	-
	Females Cohort 2	29	25	-
Mean age[c] (SD)	Both cohorts	4;1 (4.55)	5;1 (4.43)	-
	Cohort 1	4;3 (3.64)	5;2 (3.6)	6;1 (3.65)
	Cohort 2	3;9 (3.33)	4;9 (3.36)	-

[a]Attrition rate from Time 1 to Time 2. [b]Attrition rate from Time 1 to Time 3. [c] Mean age in years;months

For many of the children English was not their first language. This was case for 69 children of the Time 1 sample (33.82%). Of these, 64 came from Cohort 1 (42.66% of the cohort) and five from Cohort 2 (9.26% of the cohort). Across the children with English as an additional language (EAL) there were 24 languages represented and these included: Chinese, Polish, French, Kurdish, Bulgarian, Bengali, Swiss German, German, Portuguese, Hindi, Italian, Marathi, Gujarati, Spanish, Tamil, Urdu, Macedonian, Turkish, Tagalog, Oriya, Romanian, Korean, Albanian, and Mongolian. Information on languages spoken was gained by a parental questionnaire (Appendix 1). For those participants whose parents did not complete the questionnaire this information was gained from each child's class teacher.

Attrition rates for each time point and cohort are shown in Table 2.1. The attrition rates across all time points compares favourably with other similar longitudinal research, which have reported attrition rates of 26% (Hood, Conlon, & Andrews, 2008), 28% (Cain & Chiu, 2018; Sénéchal & LeFevre, 2002) and 29% (Lonigan et al., 2000). For the most part those children who left the project did not differ to those who remained for either Cohort 1 or Cohort 2. The age, gender and non-verbal ability (as measured at Time 1) was not significantly different in children who dropped out (at either Time 2 or Time 3) compared to those who remained in the project. For Cohort 1, mean vocabulary score (as measured by The British Picture Vocabulary Scale at Time 1; see Section 2.5.1) was not significantly different between those who remained in the project to those who left. However, in Cohort 1 more children with EAL left the project at Time 2 than English only speakers. In fact, 72.2% of the 36 children who left at Time 2 from Cohort 1 had EAL. For Cohort 2, there was no difference in first language of those who left to those who remained, but those who dropped out at Time 2 did have a significantly lower ($p < 0.05$) vocabulary score at Time 1 than those who remained. These differences were minimal but should be noted.

2.3.1 Participant recruitment

Two cohorts of participants were recruited. Cohort 1 was recruited through schools with a preschool class. Cohort 2 was recruited directly via parents through: poster (see Appendix 2) and flyer distribution (see Appendix 3), word of mouth, the social media site Facebook, and through the help of the charity group “Learn to Love to Read” based in Wandsworth, London, and their contacts.

2.3.1.1 Cohort 1

For Cohort 1, 12 schools with previous contact to the university and the researcher were telephoned and given information about the project. Three schools expressed an interest. After a meeting with the researcher to discuss the project further, each of the three headteachers consented to participation. One school was in Northwich, Cheshire in North West England (School 1) and the other two were in Sutton, Surrey in South East England (Schools 2 and 3). School 1 and 2 were primary schools whereas School 3 was a federation of an infant school and a nursery (taking children from preschool to Year 2). Ofsted reports suggest that the three schools were broadly similar with the exception of the number of EAL children in attendance. Table 2.2 highlights these characteristics.

Table 2.2

Characteristics of participating schools as taken from Ofsted reports

	School 1	**School 2**	**School 3**
Location	North West England	South East England	South East England
Ofsted date	March 2019	October 2017	September 2007
Ofsted grade	Good	Good	Outstanding
School size	Average size primary school (310 on roll)	Larger than the average primary school (628 on roll)	Larger than the average infant school (269 on roll)
Children receiving free school meals	Well below average	Broadly average	Broadly average
EAL children in attendance	Well below average	Well above the national average	Well above the national average

For School 1, the proportion of pupils from minority ethnic groups was below average, as is the proportion of EAL speakers. The two South East based schools were much more diverse. School 3 was named as having a much larger than average proportion of children from minority ethnic groups and who speak EAL. At this school 22 different languages were represented. School 2 was reported as having a majority of White British pupils, but other pupils represent a wide range of heritages. Although EAL was not the main focus of the thesis, due to this proportion of EAL children within the sample during analysis differences were explored between EAL and monolingual children.

During the preschool year in all three schools there was no formal literacy instruction given. However, the children took part in games to promote phonological awareness, and were read to daily. Formal literacy instruction began in all schools the following year when children entered Reception and began full time compulsory education.

2.3.1.2 Cohort 2

Cohort 2 was recruited directly via parents through poster (See Appendix 2) and flyer distribution (See Appendix 3), word-of-mouth, through the social media site Facebook, and with the help of the charity group "Learn to Love to Read" and their contacts. These children all lived within South West London or South East London. Many lived within close proximity to the University of Roehampton. Despite being recruited directly via parents the majority of these children attended preschool, with 98% of parents reporting their child attended some form of preschool for an average of 19.53 (SD = 7.67) hours a week. In this way children in Cohort 2 did not differ from those in Cohort 1.

Most of the children in Cohort 2 were from high earning middle class families, with 80% of parents at Time 2 reporting their family income was £50,000 or more a year, and of these 67% stating it was £70,000 or more. This is well above average for the UK based on the 2016/17 HBAI report which gave the mean yearly household income (2 adults) as £32,247 (HM Revenue and Customs, 2017). The household income for Cohort 2 is also more than that of Cohort 1, as 39% of parents in Cohort 1 reported their family income was £50,000 or more a year, and of these only 16%

stated it was £70,000 or more (also at Time 2). Additionally, mothers in Cohort 2 had generally reached a higher level of educational achievement with 83% of mothers in Cohort 2 stating they had at least an undergraduate degree, compared to 53% in Cohort 1. Although these differences between cohorts are acknowledged, socioeconomic status (SES) was not used in any analysis. This is because other similar research into reading and listening comprehension has not included SES in models (Atkinson et al., 2017; Kim, 2015; Kim, 2016; Kim, 2017; Strasser & Río, 2014).

2.4 Ethical consideration

The research for this project was submitted for ethics consideration under the reference PSYC 16/ 210 in the Department of Psychology and was approved under the procedures of the University of Roehampton's Ethics Committee on 07.04.16 and 06.10.16. For each cohort a different strategy to consent was taken.

2.4.1 Cohort 1 consent

Cohort 1 was recruited directly through three schools. Three levels of consent were obtained: headteacher consent, parental consent and child assent. The headteachers of the three recruited schools gave their consent for an opt-out method (see Appendix 4 for a copy of the headteacher consent form) in order to gain a large representative sample. There were no formal exclusion criteria and so information letters (see Appendix 5) were sent to all parents of the preschool classes at each of the three schools. Parents returned the signed slip to the researcher if they wished their child *not* to take part in the research, thus opting out. Across the three schools four parents chose to opt-out. These children were not tested. This letter also informed the parent about the subsequent home measures which they could opt-in to later.

For the school based tasks, opt-out consent was chosen to recruit the largest representative sample possible. Ethically, an opt-out method was appropriate as throughout the project measures were presented to children in the form of engaging tasks. These tasks were typical of the types of activities that the children already undertook during their school day. They were therefore not doing anything particularly unfamiliar or anxiety provoking, and so there were considered to be no major ethical issues raised by taking part. Thus, opt-out consent was requested and approved by the university ethics committee. Although it was made known to the parent that this was a long-term project, and that they could withdraw at any time by contracting the researchers, parents were only given an opt-out form once (at the beginning of Time 1). However, before subsequent time points they were sent a courtesy letter (see Appendix 6) informing them when the researcher would be in school and reminding them that they could withdraw by contacting the researcher.

In addition to headteacher and parental consent (opt-out), at the start of each testing session each child was verbally asked for their assent. At Time 1, three children did not give verbal assent, and so they did not take part in Time 1 or subsequent time points. At Times 2 and 3 all children gave their assent. Children were also monitored during the testing sessions, and in some cases if it appeared that the children were disengaged and reluctant to be taking part in a particular measure, it was discontinued. At Time 1 this occurred for 27 individual measures (1.4% of all measures administered) across 21 children. The subsequent 129 children participated in all measures. At Time 2 this occurred for five individual measures (0.3% of all measures administered) across three children. The remaining 111 children participated in all measures. At Time 3 this occurred for four individual measures across two children (0.2% of all measures administered). The remaining 105 children participated in all measures. For more information on missing data and how this was dealt with see individual results chapters.

The initial parental letter stated that home based measures would also take place. For these (the questionnaire and the book sharing activity) parents were asked to sign separate consent forms for each activity opting-in to participate (see Appendix 1 and 7). These letters were distributed via the class teachers once the school testing sessions had begun.

2.4.2 Cohort 2 consent

Cohort 2 was recruited directly through parent contact. As with Cohort 1 there were no formal exclusion criteria. Parents consented to all activities using only one consent form (see Appendix 8 for this consent form) including consent for: the one-to-one child measures with the researcher, the questionnaire, and the book sharing activity. As with Cohort 1, children also gave their assent verbally at the beginning of their sessions. All children from Cohort 2 gave their assent at all time points. However, as with Cohort 1 some children were reluctant to take part in some of the individual measures. At Time 1 this occurred for 18 individual measures (2.4% of all measures administered) across 11 children. The other 43 children engaged in all measures. At Time 2 this occurred for four individual measures (0.41% of all measures administered) across three children. All measures were the same as Cohort 1 with the exception of the addition of The Oral and Written Language Scales (Carrow-Woolfolk, 2011) at Time 1.

At the end of the project each of the participating schools received a £50 book voucher. For Cohort 2, at the end of each individual session each of the families received a £10 Amazon voucher.

2.4.3 Data confidentiality

Data were kept confidential, so that a child's name was never directly linked with their raw or processed data for any measure. Instead data was recorded and stored under an anonymous identification number. A password protected file linked the anonymous identification number to the child's name. This was necessary to link a child's data across time points.

Data was securely stored in confidential computer files and in locked filing cabinets at the University of Roehampton. These were accessible only to the study investigators. Additionally, audio recordings were stored securely on password protected storage and only listened to by the investigators. These files will be stored securely for at least ten years before being destroyed.

2.5 Measures administered to children

The following skills were assessed at all three time points: vocabulary, syntactic knowledge, listening comprehension, decoding, working memory, theory of mind, metacognition, and non-verbal ability. Inference making, comprehension monitoring and reading comprehension were assessed at Time 3. Table 2.3 shows the measures used to test each skill at each time point. A description of each and justification for its use follows. As shown in Table 2.3, in some cases, the same measure of a skill was not used across all timepoints because the measure was not suitable to use across a span of two or three years due to ceiling and floor effects. Instead, more advanced measures that tapped the same construct had to be used at progressive time points, or existing tasks had to be modified. This is consistent with similar longitudinal work (e.g. Atkinson et al., 2017; Gooch, Thompson, Nash, Snowling, & Hulme, 2016). Additional measures including inhibition, card sorting, and other decoding measures were taken at each time point as part of a wider study but were not used in any analysis for this thesis.

Table 2.3

Measures administered to children at each of the three time points

Skill	Measure	Time 1 (4;1)[a]	Time 2 (5;1)[a]	Time 3 (6)[a]
Vocabulary	BPVS-III	✓	✓	✓
Syntax	Sentence Structure (CELF-Preschool 2[uk])	✓	✓	-
	Sentence Structure (CELF-4[uk])	-	-	✓
Decoding precursors	Preschool Repetition Task	✓	✓	-
	Letter Sound Knowledge (YARC)	✓	✓	-
Decoding	Single word reading DTWRP	-	✓	✓
Listening comprehension	NARA-II	✓	✓	✓
	OWLS[b]	✓	✓	✓
Theory of mind	Unexpected contents	✓	-	-
	Unexpected locations	✓	-	-
	Belief desire reasoning	-	✓	-
	Unexpected locations second-order false belief	-	✓	-
	Strange Stories	-	-	✓

Broad metacognition	Source-monitoring tunnel task	✓	✓	-
	Source-monitoring event task	-	-	✓
	Metalinguistic knowledge synonym judgment task	✓	✓	-
	Metalinguistic knowledge homonym selection task	-	-	✓
	Metamemory task	✓	✓	✓
Working memory	Reverse word span	✓	✓	✓
Inference making	Inference oral stories	-	-	✓
Comprehension monitoring	Comprehension monitoring of stories	-	-	✓
Non-verbal ability	Block design (WPPSI – III)	✓	-	-
Reading comprehension	YARC	-	-	✓

[a] Average age in years;months [b] OWLs measure only administered for Cohort 2 at Time 1 and not Cohort 1

2.5.1 Vocabulary

At all three time points receptive vocabulary was measured using the long form of the British Picture Vocabulary Scale: Third Edition (BPVS-III, L. Dunn, Dunn, & National Foundation for Educational Research, 2009), a standardised assessment for preschool to secondary children (3-16 years). Participants listened to a word spoken by the researcher and were asked to indicate the meaning by pointing to one of four coloured pictures. For example, they heard the word "*spoon*", and were required to choose from pictures of: a spoon, a tropical fish, a decorated cake, and a drinking glass (see Figure 2.1). The assessment consisted of 14 sets of 12 words of increasing difficulty. It was discontinued after a full set was completed in which eight or more errors were made. Following standard procedure, at Time 1 participants began at Set 1, at Time 2 they began at Set 2 and at Time 3 they began at Set 4.

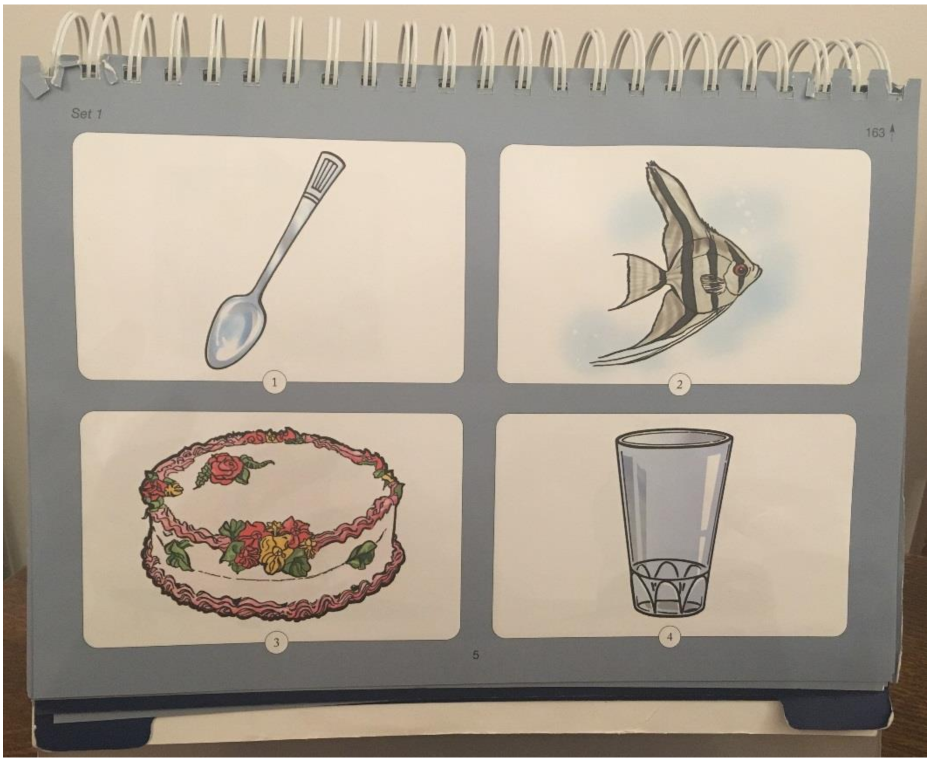


Figure 2.1: An example item from Set 1 of the BPVS in which participants heard the word "*spoon*" and were required to point to the corresponding picture. The correct response is picture 1 (top left).

To generate the raw score, the number of errors made across all sets was subtracted from the number of trials administered to the child, which gave the total number of correct responses. The maximum score was 168, which was given if a participant completed 14 sets with no errors.

This measure was selected because it is a well-used standard test of vocabulary (e.g. Atkinson et al., 2017; Babayiğit, 2014; Baron-Cohen, Leslie, & Frith, 1986; Bright-Paul et al., 2008; Cain & Bignell, 2014; Hughes & Dunn, 1997; Muter, Hulme, & Snowling, 1997; Nation, Clarke, Wright, & Williams, 2006; Slade & Ruffman, 2005). Its wide age range meant it could be administered at all three time points. In accordance with this, others have administered the BPVS to children the same age as the current sample, for example Bright-Paul, Jarrold and Wright (2008) used it with children aged three to six years, Hughes (1998a) with children aged three and four, and Atkinson et al., (2017) with children aged three to six years. It has also been used as a measure of vocabulary in similar work, such as that into reading comprehension (Atkinson et al., 2017; Cain, 2006; Nation et al., 2006), listening comprehension (Babayiğit, 2014; Cain & Bignell, 2014), theory of mind (Baron-Cohen et al., 1986; Slade & Ruffman, 2005) and maternal mental state talk (Hughes & Dunn, 1997; Meins et al., 2013).

A further justification for inclusion was the reliability of this measure. The BPVS-III was originally reported as having excellent reliability by Dunn et al., (2009), and by subsequent studies with reports of a Cronbach alpha coefficient (Cronbach, 1951) for internal reliability of $\alpha = .94$ in a group of typically developing British children aged five to eight years (Holliman, Wood, & Sheehy, 2010), and of α =.96 for three to four year olds, and α =.94 for six year olds (Atkinson et al., 2017). Here, reliability can be claimed to be excellent using the commonly used interpretation as follows: α larger or equal to .9 = Excellent; .between .8 and .9 = Good; between .7 and .8 = Acceptable; between .6 and .7 = Questionable; between .5 and .6 = Poor; and .5 or below = Unacceptable (George & Mallery, 2003). These parameters were used to assess the reliability of all measures.

2.5.2 Knowledge of syntax

During Times 1 and 2 knowledge of syntax was assessed through the sentence structure subtest of the Clinical Evaluation of Language Fundamentals-Preschool 2^{UK} (CELF-Preschool 2, Wiig et al., 2006). For age appropriateness, the same subset of the Clinical Evaluation of Language Fundamentals-4$^{\text{UK}}$ (CELF-4, Semel, Wiig, & Secord, 2006) was used at Time 3. From this point onwards, unless otherwise stated CELF-Preschool 2 and CELF-4 will be used to refer to the sentence structure subset of each of these measures.

During the CELF-Preschool 2, participants listened to a sentence spoken by the researcher, and were asked to point to one of four coloured pictures which depicted that sentence. For example, children heard the sentence, "*Mum showed the dog the cat*", and were required to choose the meaning from pictures of: a woman showing a dog a cat, a woman sat with two cats, a woman sat with a dog on her knee and a cat at her feet, and a woman showing a big dog a small dog (see Figure 2.2). The assessment consisted of 22 sentences, and in accordance with standard procedure was discontinued when a child made five consecutive errors.

Figure 2.2: An example item from the CELF-Preschool 2 in which participants heard the sentence, "*Mum showed the dog the cat*" and were required to choose the meaning from pictures. The correct response is the top left picture.

This measure was chosen as it is a well-used standardised test (e.g. Bishop, Adams, & Norbury, 2006; Cabell et al., 2011; Cabell, Justice, Logan, & Konold, 2013; Gooch et al., 2016; Justice, Kaderavek, Fan, Sofka, & Hunt, 2009; Nation et al., 2010; C. C. Peterson, Wellman, & Slaughter, 2012). It was age appropriate to use at Time 1 and 2 as it is normed at three to six years and has been used by others with similar aged children. For example, Gooch et al. (2016) used it with children aged four and five years old, and Cabell et al., (2011) used it with children aged three to five. It has also been used in similar research, such as that into reading comprehension (Poe, Burchinal, & Roberts, 2004; Snowling, Duff, Nash, & Hulme, 2016), listening comprehension (Alonzo et al., 2016; Piasta, Groom, Khan, Skibbe, & Bowles, 2018), and theory of mind (Diaz & Farrar, 2018; C. C. Peterson et al., 2012).

In addition, the CELF-Preschool was chosen for its reliability, with an original test-retest reliability reported as α = .78, as evaluated by a study with a group of 120 US children aged three to five years on two separate occasions (Wiig et al., 2006). Additionally, internal consistency was also originally estimated as α = .78 (Wiig et al., 2006). This internal consistency has been confirmed by others, with Gooch et al., (2016) stating α = .78–.83 for English speaking British children aged three and four.

At Time 3 the same subset of the CELF-4 (Semel et al., 2006) was used to assess knowledge of syntax. Presentation took the same form as the CELF-preschool 2, in that participants listened to a sentence spoken by the researcher and were required to point to one of four coloured pictures which depicted that sentence. For example, children heard the sentence, "*The boy is going down the ramp*", and were requested to choose the meaning from pictures of: a boy in a wheelchair at the bottom of a ramp, a boy in a wheelchair going up a ramp, a boy in a wheelchair going down a ramp, and a boy in a wheelchair at the top of a ramp (see Figure 2.3). The assessment consisted of 26 sentences and in accordance with standardised procedure all items were administered with no discontinuation rule.

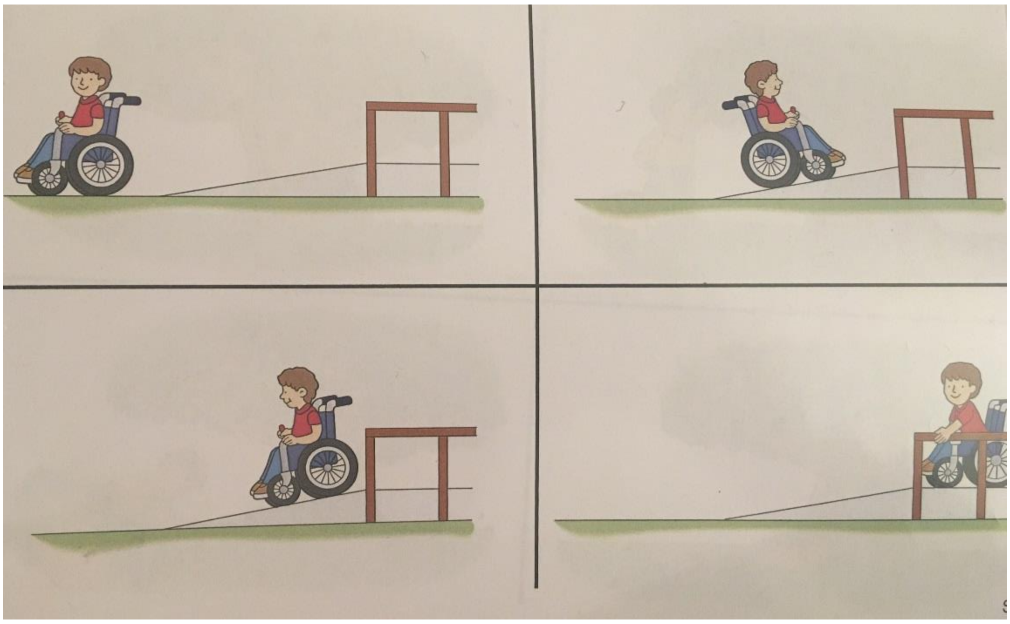


Figure 2.3: An example item from the CELF-4 in which participants heard the sentence, "*The boy is going down the ramp*" and were required to choose the meaning from pictures. The correct response is the bottom left picture.

The CELF-4 was used at the latter time point rather than the preschool edition because it is normed from 5-16 years and was therefore deemed more age appropriate. Importantly, both Wiig et al. (2006) and Paslawski (2005) state that there is a smooth transition between the CELF-Preschool 2 and the CELF-4, because of overlapping and correlated items. Owing to this, longitudinal studies have used the sentence structure subset of the CELF-Preschool 2 at their earlier time points, and the CELF-4 at latter time points (e.g. Gooch et al., 2016), to make direct age comparisons across time. It was therefore justifiable for this thesis to do the same.

This subset of the CELF-4 is a well-used standardised test (e.g. Bowyer-Crane et al., 2017; Foorman et al., 2015; Gooch et al., 2016). The sentence structure subset of the CELF-4 was chosen for its reliability, with an original Cronbach's alpha for children aged five to eight years reported as ranging between α =.62 and .77. This has been confirmed by subsequent studies, for example Bowyer-Crane et al., (2017) report α = .70 for a group of six years olds including English only speakers and EAL children, and Gooch et al., (2016) report coefficients ranging between .78-.83 for English speaking children aged five to eight years.

2.5.3 Listening comprehension

At all three time points LC was assessed using the Neale Analysis of Reading Ability-Second Revised British Edition (NARA-II) (Neale, 1999). This assessment was originally designed to measure RC for six to nine year olds, whereby children read passages aloud and answer comprehension questions. Here however it was administered as a measure of LC and stories were read aloud to the participants by the researcher (As used by Bowyer-Crane et al., 2008; Burgoyne, Whiteley, & Spooner, 2009; Cain & Bignell, 2014; Nation et al., 2010).

Form 1 stories of the NARA-II were used which included six levels of story of increasing length and complexity. After hearing each story, children were required to answer comprehension questions about the passage. For example, they heard a story about a girl called Kim, who was on her way to school when she saw two children on bikes crash into each other. Kim ran to help, only to find out from the children that they were taking part in a staged road safety lesson. The comprehension questions consisted of both literal questions (e.g. *"Where was Kim going?*"), and inferential questions (e.g. "*How do you think Kim felt?*").

At all time points children began on the Level 1 story. At Times 1 and 2 if the child correctly answered half or more of the comprehension questions, they proceeded to the next level story. This continued until the child was unable to correctly answer half (or more) of the comprehension questions for a given story, when the test was discontinued. At Time 3 they were administered the Level 2 story even if they failed to answer half the questions at Level 1. The test was then discontinued if they answered less than half at Level 2. There was a maximum score of 44. See Appendix 9 for all stories and questions.

The task was selected because previous UK studies have used the NARA as a measure of LC (e.g. Bowyer-Crane et al., 2008; Burgoyne et al., 2009; Cain & Bignell, 2014; Clarke, Snowling, Truelove, & Hulme, 2010; Nation et al., 2010) and a Norwegian translation has also been used as a measure of LC for Norwegian speaking children (Lervåg et al., 2018). These studies administered the NARA orally to similar aged children, for example Nation et al., (2010) to children aged six years old and Bowyer-Crane et al., (2008) to four year olds. Additionally, as the NARA stories are of increasing length and complexity, the measure could be used at all time points and so comparisons could be made across ages, as demonstrated by Lervåg et al., (2018). Reliability of administering the NARA as a measure of LC is reported as good. For a sample of children with poor attention and high hyperactivity and age matched controls, Cain and Bignell (2014) report a Cronbach's alpha of α =.84 and Lervåg et al., (2018) report α =.82 for a group of Norwegian children aged seven.

At all time points for Cohort 2 and at Times 2 and 3 for Cohort 1, the listening comprehension subset of the Oral and Written Language Scales (OWLS-II) (Carrow-Woolfolk, 2011) was administered. The OWLS is normed at ages three to 21 years. From this point onwards, unless otherwise stated OWLS will be used to refer to the listening comprehension subset of this measure. This subset of the OWLS consists of 130 items, arranged in increasing order of difficulty. Each item was presented to the child by reading the verbal stimulus aloud while the child looked at four coloured pictures numbered 1 to 4. The child was required to select the picture which best depicted the meaning of the verbal stimulus. An early example is: "*Show me the girl saying, "good-bye*", to which the child had to select the meaning from: a picture of a girl putting her coat on, a girl hurrying along a path waving to a woman, a boy getting into a car as he waves at a man, and a girl and woman planting flowers in a garden (see Figure 2.4). The second picture was the correct response.

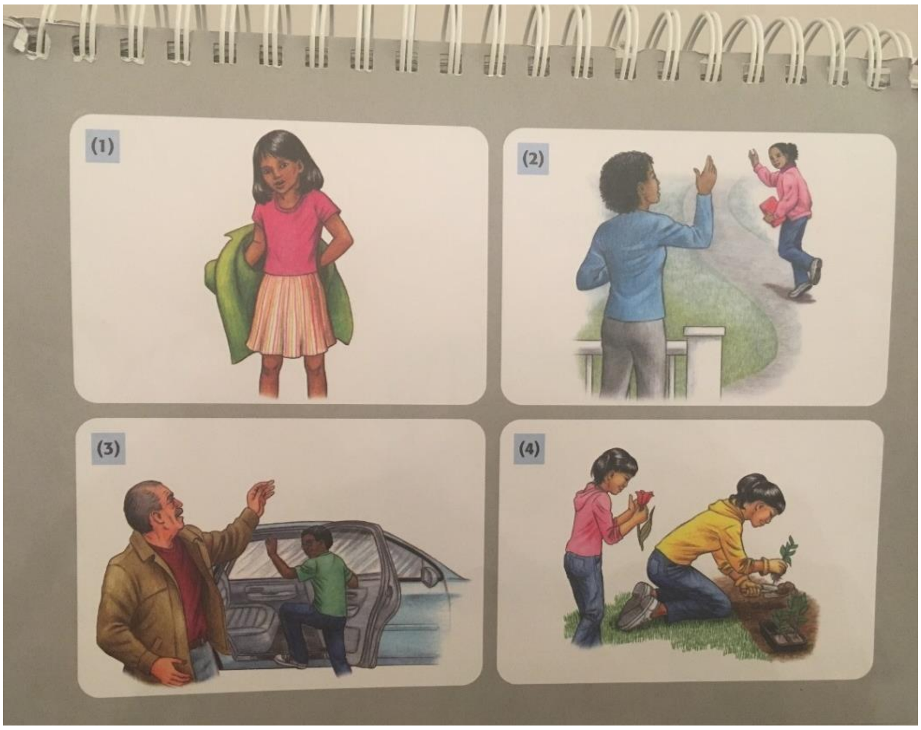


Figure 2.4: An early item from the OWLS in which the participant heard the sentence: "*Show me the girl saying, "good-bye*" and had to select the appropriate picture. The correct response is picture 2 (top right).

A later example is: "*The boy to whom the girl with the broken arm had given a football brought her a glass of water*", to which the child had to select the meaning from: a picture of a girl talking to a boy with a broken arm holding a ball, a picture of a boy giving a girl with a broken arm and a ball a glass of water, a picture of a boy with a ball and a girl with a broken arm talking, and a picture of a girl with a broken arm being given a glass of water by a boy with a ball (see Figure 2.5), of which the last was the correct response. Children could respond either nonverbally by pointing to the picture, or verbally with the number of the picture.

Figure 2.5: A later item from the OWLS in which the participant heard the sentence: "*The boy to whom the girl with the broken arm had given a football brought her a glass of water*" and had to select the appropriate picture. The correct response is picture 4 (bottom right).

In accordance with standard protocol the test was discontinued when a child made four errors in a row. Following the standard procedure, at Time 1 all children began at item 1, at Time 2 children began at item 15, and at item 30 at Time 3. For these later time points a basal score of seven consecutive items had to be established. The maximum total score was 130.

The scale consists primarily of lexical/semantic, syntactic, and supralinguistic items. Lexical/semantic items measure a range of linguistic structures such as nouns e.g. "*Show me the car*", and verbs and idioms. Syntactic items require comprehension of noun and verb modulations (such as number, tense, gender, voice, person, and case), and syntactic contractions (such as embedded sentences, coordination, direct/indirect objects, inflections, and functions). Early syntactic items measure the understanding of plurality e.g. "*Show me the pencils*". Whereas an example of a later syntactic item is *"The girl he waved to was sitting"*, with the four pictures depicting variations of boys and girls sitting and waving. The participant was required to understand the complex embedded sentences and the personal pronoun to answer correctly. Supralinguistic items require language analysis on a higher level than decoding literal lexical or syntactic structures, such as comprehension of figurative language (the understanding of similes or metaphors or humour) and derivation of meaning from context (including the use of logic, inference and other higher order thinking skills). For example, the item *"Sarah joined Jada in a cup of tea. Which pictures best shows what happened?"* requires the understanding of the double meaning of "join someone in a cup of tea". Likewise, in the later item *"The young moon's back rested on the arms of the tall trees"* understanding of figurative language was required.

This measure could be used at all time points because it is normed at ages three to 21 years. Due to this it has been used by others with similar aged samples, for example by Kim (2015) with children aged six, and by Kim and Phillips (2014) with children aged five to six years. Additionally, this scale was chosen was because it taps a broad range of LC understanding, making it a sensitive measure. Furthermore, unlike the NARA, responses on this scale can be nonverbal; this means that scores will not be reduced by a child's expressive language ability. Moreover, recent research (e.g. Kim & Phillips, 2014; Kim, 2015; Kim, 2017) has used this scale as an outcome measure of LC when using ToM and other high-level cognitive skills to predict it. The inclusion therefore was in line with these recent publications from the USA.

Further justification for inclusion was the excellent reported reliability of this scale. Reliability was originally reported by Carrow-Woolfolk (2011) as excellent α =.98 for internal consistency for three to four year olds (n = 321), and α =.98 for five to seven years (n = 320). This is supported by subsequent studies, such as Kim and Petscher (2016) who report α = .93 and Kim (2017) α = .94 for children aged five to seven years.

2.5.4 Precursors to decoding

A child's ability to repeat real words and non-words is the best predictor of later decoding skills (Chiat & Roy, 2007; Gathercole & Baddeley, 1989). At Time 1 and 2 this was assessed using the Preschool Repetition Task (Chiat & Roy, 2007). Likewise, letter sound knowledge is a very good predictor of later decoding (Chard, Pikulski, & Templeton, 2000; Gough & Tunmer, 1986; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012), and so the Letter Sound Knowledge subtest of the York Assessment of Reading for Comprehension: Early Reading (Hulme et al., 2009) was also administered at Times 1 and 2.

The Preschool Repetition Task (Chiat & Roy, 2007) includes both a word repetition and non-word repetition task. During the word repetition the researcher read aloud 18 familiar words. For example, "*police*" (pəˈliːs). Participants were required to repeat each word. The non-word repetition took the same format but children were instructed that they may not have heard the words before to ensure they were not discouraged by the unfamiliarity of the words. An example of a non-word is "*lopice*" (ləˈpis). For the full list of both words and non-words see Appendix 10. For both word and non-word items, there was an equal number of one, two, and three syllable words (6 of each). The non-words were phonologically matched to the real words, so that one syllable non-words were created by altering the vowel of the one syllable real words, e.g. *egg* (ɛg) → *oog* (ʊg), and two or three of the consonants were altered in the two and three syllable words, e.g. *machine* (məˈʃiːn) →*shameen* (ʃəˈmin), *dinosaur* (ˈdaɪnəˌsɔ) → *sinodaw* (ˈsaɪnəˌdɔ). There was no discontinuation rule with all items administered. The maximum score for the word repetition was 18, and for non-word repetition also 18.

This measure was specifically designed for preschool children, so was ideal for the age range of the current research. In this vein the measure is widely used as an early measure of decoding in preschool and young primary school children (Beattie & Manis, 2014; Clark, McRoberts, Van Dyke, Shankweiler, & Braze, 2012; Dispaldro, Deevy, Altoé, Benelli, & Leonard, 2011; Highman, Leitão, Hennessey, & Piek, 2012). For example, it has been used by Dispaldro et al., (2011) with three and four year olds and by Beattie and Manis (2014) with five year olds. This test was also selected for its excellent reliability. For their sample of 315 two to four year olds Chiat and Roy (2007) report α = .92 for internal reliability, and α = .93 for test-retest reliability (n = 41). This excellent reliability is supported by subsequent research which has reported a α = .93 for internal reliability for a US English speaking sample aged five years (Beattie & Manis, 2014).

The Letter Sound Knowledge subtest of the York Assessment of Reading for Comprehension Early Reading (YARC) (Hulme et al., 2009) was used at Time 1 and Time 2. Participants were asked to produce the sound made by each of the 26 letters. Unlike the original task, the six digraphs were not included at Time 1, as these were deemed too advanced for children of this age. The digraphs were however included at Time 2. Examples of digraphs include "*ee*" and "*th*". The printed letters were presented to the child in a random generated fixed order, to avoid bias from previous alphabet rote learning. There was no discontinuation rule which avoided possible biases based on children's greater familiarity with the letter shapes and associated sounds in their own name. The task was only discontinued midway through in situations of extreme boredom and fatigue. The maximum score was 26 at Time 1, and 32 at Time 2 (when digraphs were included).

This task was selected because it is well used in the literature (Bowyer-Crane et al., 2017; Burgoyne et al., 2012; Dilnot, Hamilton, Maughan, & Snowling, 2016; Duff, Mengoni, Bailey, & Snowling, 2015; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; Gooch, Hulme, Nash, & Snowling, 2014; Snowling, Duff, Petrou, Schiffeldrin, & Bailey, 2011). Notably, this assessment has been administered to children of a similar age to the current research, for example Fricke et al., (2013) administered it to four year olds. It is also used in similar work exploring early predictors (aged four) of reading comprehension aged six (Bowyer-Crane et al., 2017). Moreover, reliability has been reported to be excellent, with an internal consistency of $\alpha = .98$ reported by Snowling et al. (2011) for English speaking six year olds. Further, Duff et al. (2014), and Fricke et al. (2013) report a Cronbach alpha $\alpha = .95$ in samples of English speaking six and four year old respectively.

2.5.5 Decoding

At Times 2 and 3 decoding was measured using the Diagnostic Test of Word Reading Processes ((DTWRP: Forum for Research into Language and Literacy, 2012). The DTWRP is a test of word and non-word reading normed for five to 12 year olds. It comprises 90 items; 30 exception words which provide a measure of lexical-semantic processes, 30 non-words which provide a measure of phonological recoding processes, and 30 regular words which can be read by either process.

Children were told that they were going to read some alien words, and the first card (with pictures of aliens and their names) was laid in front of them (Appendix 11a). The children were asked to read the alien's names starting with the two practice words "*und*" and "*heg*". Feedback was given for these practices words only (e.g. "*Yes well done*" or "*Good try but that is not quite right, this says und*".) Children were then instructed to read the rest of the ten "alien words" of increasing difficulty one by one. For example, an early word was "*un*" and a later word was "*pertle*". The card was discontinued after five consecutive errors were made. If the child had not met the discontinuation criteria, the next non-words card (Appendix 11b) was placed in front of the child and they were encouraged to read these words. The card compromised 20 more non-words of increasing difficulty. For example, an early word was “*sus*” and a later word was “*experorium*”. As before, the card was discontinued after five consecutive errors were made. The participant was then told they would now read some real words, and the exception words card was placed in front of them (Appendix 11c). They were asked to read the words on this card. This card consisted of 30 words irregular words of increasing difficulty. For example, an early word was “*his*” and a later word was “*miscellaneous*”. As before, the card was discontinued after five consecutive errors were made. Lastly, the regular words card was placed in front of the child (Appendix 11d) and they were asked to read the words on this card. This card consisted of 30 regular words of increasing difficulty. For example, an early word was “*up*” and a later word was “*anecdote*”. As before, the card was discontinued after five consecutive errors were made. A point was given for each word read correctly so that there was a total score of 90 (30 from non-words, 30 from exception words, and 30 from regular words). To score children must have blended the word together and not just sounded out the individual letters.

The DTWRP was selected because it has been used with children aged five and six years olds (Bowyer-Crane et al., 2017; Cunningham, Witton, Talcott, Burgess, & Shapiro, 2015; Ricketts, Davies, Masterson, Stuart, & Duff, 2016). The measure was also age appropriate to use at both Time 2 and 3 as it is normed from five to 12 years and has been used by other reading comprehension longitudinal studies. For example, Bowyer-Crane et al., (2017) with four to six year olds and with five to nine year olds (Duff et al., 2015). The DTWRP also has excellent reliability, with reports of a Cronbach’s alpha of α =.97 for regular words, and α =.96 for irregular words in a sample of British six year olds (Cunningham et al., 2015), and α =.99 for the total of both also for a sample of British six year olds (Bowyer-Crane et al., 2017).

2.5.6 Theory of mind

Different ToM measures were used at different time points to be age appropriate. At Time 1, two false belief tasks were administered to assess ToM understanding: the unexpected contents task (Hogrefe, Wimmer, & Perner, 1986) and the unexpected locations task (Wimmer & Perner, 1983). At Time 2, the belief desire reasoning task (Harris, Johnson, Hutton, Andrews, & Cooke, 1989) and the unexpected locations second-order false belief task were used (Perner & Wimmer, 1985). At Time 3, the Strange Stories task (O'Hare, Bremner, Nash, Happé, & Pettigrew, 2009) was administered.

The unexpected contents task (Hogrefe et al., 1986), commonly known as the "Smarties® Task", involves a Smarties® box containing pencils, instead of Smarties®. Children must recall their own false belief, and attribute, and explain the false belief of a character. During this task, children were introduced to a Playmobil® doll called Jenny. They were told that Jenny would be put away, where she could not see or hear what was happening. The child was then required to predict the likely content of a standard Smarties® box. If the participant did not know, or gave the wrong answer (Smarties®, chocolate, sweeties, sweets, or other types of sweets e.g. Skittles were all accepted as correct answers), the test was abandoned. If the child correctly answered Smarties® (or one of the similar answers listed above), they were invited to have a look at what was inside the box. The researcher opened the box to reveal, unexpectedly, that there were pencil crayons inside the box, and not Smarties®. The box was then closed, and the children asked if they remembered what was inside. If they remembered correctly, they were asked the first test question: "*When I first showed you this box, all closed up like this, what did you first think was in there?*"

To answer this question correctly children had to recall their own false belief (that they had first thought there were Smarties® in the box). Jenny was retrieved, and the child was asked the second test question: "*When we first show Jenny this box, before she looks inside, what will she say is in there?"* If they gave no response they were given a force choice "*Will Jenny say there are Smarties® or pencils in the box*?" To correctly answer this question, children had to attribute a false belief to Jenny. For the third test question, they were asked to justify their answer: "*Why will Jenny say there are Smarties®/pencils in there?*" The children were awarded one point for each correct answer, and so could receive a maximum score of three. Correct answers for the justification question were statements such as "S*he didn't see there were pencils*" or "*It is a Smarties® box so she will say Smarties®*". Figure 2.6 depicts the stimuli used.

Figure 2.6: Stimuli used for the unexpected contents task. A Playmobil® doll called Jenny and a Smarties® box containing pencil crayons instead of Smarties®.

As the scoring of this measure required a judgment decision, 25% was scored by a second researcher and Cohen's Kappa was run to determine if there was consistency between scorers. Based on guidelines by Altman (1990) this agreement can be said to be excellent with 96% of total agreement (Kappa = .94, $p < .001$).

This measure was used because it is standard measure of ToM for use with preschool children (Wellman et al., 2001), and thus is widely used with this age group (e.g. Atkinson et al., 2017; Ebert, 2015; Gopnik & Astington, 1988; Guajardo & Cartwright, 2016; Happé, 1995; Lockl & Schneider, 2007; Meins et al., 2002; Strasser & Río, 2014). For example, Gopnik and Astington (1988) used it with three to five year olds from the US, and Atkinson et al., (2017) with a group of British three to four year olds. This ToM test has been used in longitudinal studies (Ebert, 2015; Lockl & Schneider, 2007; Meins et al., 2002), and in other similar research such as that comparing ToM to other forms of metacognition (Ebert, 2015; Lockl & Schneider, 2007), research into ToM and RC (Atkinson et al., 2017; Guajardo & Cartwright, 2016), and ToM and LC (Strasser & Río, 2014). Additionally, reliability for this test is also reported as good for children of varying abilities (Hughes et al., 2000).

The unexpected locations task (Wimmer & Perner, 1983), also known as the "Sally-Anne Task", requires children to predict and explain a character's false belief of a transferred object. Two small boxes, one blue and one red, with lids were placed equidistant from the child. The child was required to watch as the researcher played out a story using two Playmobil® dolls, a girl named Sally and a boy named Anthony (although the girl was called Anne in original versions of the task). Sally was shown to play with a ball before becoming tired, placing it in the blue box, and going away to sleep. The child was told that Sally could not hear or see them. Then, Anthony appeared looking for something to play with. He went into the blue box, found the ball and played with it. When Anthony finished playing, he placed the ball in the red box. Sally was then shown to wake up. The child was told that Sally wanted her ball, and they were asked: "*Where will Sally look first?*" To answer this question correctly, participants had to attribute to Sally the false belief that the ball would still be in the blue box, where she had placed it and seen it last. They were then asked to justify their answer ("*Why will Sally look there?*"). Next, they were asked two control questions to ensure they had followed the story ("*Where did Sally put the ball in the beginning?"* and "*Where is the ball now*"). See Figure 2.7 for the stimuli. A maximum score of two was given, one for the test question, and one for the justification question. Participants must have answered both the control questions to score at all.

Figure 2.7: Stimuli used for the unexpected locations task including Playmobil® dolls called Sally and Anthony. Anthony moved Sally's ball from the blue box to the red box when she was away from the scene.

As the scoring of this measure required a judgment decision, 25% was scored by a second researcher and Cohen's Kappa was run to determine if there was consistency between scorers. Based on guidelines by Altman (1990) agreement can be said to be excellent with 98% of total agreement (Kappa = .96, $p < .001$).

The ToM test was chosen because it is the standard in the literature for use with preschool children (Wellman et al., 2001), and is widely used in this age group (e.g. S. T. Baker, Leslie, Gallistel, & Hood, 2016; Baron-Cohen et al., 1985; Carlson, Moses, & Breton, 2002; Davies, Andrés-Roqueta, & Norbury, 2016; Happé, 1995; Jenkins & Astington, 1996; Stone, Baron-Cohen, & Knight, 1998; Wellman et al., 2001). For example, Atkinson et al., (2017) used it with a group of British three to four year olds, and Carlson et al., (2002) with a sample of US children aged three and four. This ToM test has been used in longitudinal studies (Atkinson et al., 2017; Ensor et al., 2014), and in other similar research such as that comparing ToM to other forms of metacognition (Bright-Paul et al., 2008; Doherty & Perner, 1998), research into ToM and RC (Atkinson et al., 2017), and ToM and LC (Strasser & Río, 2014). Additionally, reliability for this test is reported as good for children of varying abilities (Hughes et al., 2000).

The belief desire reasoning task (Harris et al., 1989) was used at Time 2. The task involves one character (Chris the crocodile) playing a “nasty surprise” on another character (Danny the dog). Children were informed that Chris the crocodile is naughty, and that he likes to play tricks on his friends. They were also told that Danny’s favourite drink was Coca-Cola®, and that he hates milk. Two emotion contingency questions were asked to ensure that children were following the story. Question 1 asked how Danny felt when he received a can of Coca-Cola® (correct answer: happy), and question 2 asked how he felt when he receives milk (correct answer: not happy). The researcher then told a story with soft toys and props, in which Chris replaced the content of Danny’s Coca-Cola® can with milk whilst Danny was away from the scene. On Danny’s return the children were informed that he was thirsty, and that he could see the Coca-Cola® can on the table, but could not see the contents. Children were then asked the belief-desire question, asking them to state whether Danny was happy or not happy when he saw the can on the table (question 3). The correct answer to question 3 was happy. They are also asked to justify their answer (question 4), to which answers such as "*he thinks there is coke in the can, so he is happy*" or "*because he has a can of coke*" were correct. The children were then asked what Danny thinks is in the can (question 5; correct answer: Coke) and what was really in the can (question 6; correct answer: milk). Danny was then shown to drink from the can, and the children were asked whether he is now happy or not happy (question 7; correct answer: not happy), and asked to justify their answer (question 8), to which answers such as "*because he just drank milk and he hates milk*" or "*because the crocodile swapped his coke for milk*" were correct. See Figure 2.8 for the stimuli used.

Figure 2.8: The stimuli used in the belief desire reasoning task. A stuffed dog called Danny and his naughty friend Chris the Crocodile.

This task was scored out of three. To score the first point children must have identified that on his return Danny would hold the false belief that Coke was in the can (question 5). To score this point they must have also stated that the can really contained milk (question 6). To score the second point children must have stated that on his return he would have felt happy (question 3). To score this point they must have also answered correctly questions 1, 2, 6, 7 and 8. To score the final point children must have given a correct justification for why Danny felt happy on his return (question 4). To score this point they must have also answered question 3 correctly.

As the scoring of this measure required a judgment decision, 30% was scored by a second researcher and Cohen's Kappa was run to determine if there was consistency between scorers. Based on guidelines by Altman (1990) agreement can be said to be excellent with 92% of total agreement (Kappa = .89, $p < .001$).

This task was chosen because it was age appropriate as others have successfully administered it to four year olds (Avis & Harris, 1991; Devine, White, Ensor, & Hughes, 2016; J. Dunn, Cutting, & Fisher, 2002; Hughes & Dunn, 1998; Lecce et al., 2014; Lundy, 2013; Vinden, 1999) and five year olds (Avis & Harris, 1991; Devine et al., 2016; Hughes et al., 2005; Lecce et al., 2014; Vinden, 1999). This ToM test has been used in longitudinal studies (Ensor et al., 2014), and in other similar research such as that comparing ToM to other forms of metacognition (Lecce et al., 2014), research into ToM and RC (Atkinson, 2014), and ToM and LC

(Strasser & Río, 2014). Additionally, reliability for this test is reported as good for children of varying abilities (Hughes et al., 2000).

The unexpected locations second-order false belief task was also used at Time 2 (Perner & Wimmer, 1985). Here, children were told a story about two siblings, Mary and Simon, using a series of four picture cards (See Figure 2.9). The first card was presented, and children were told that grandad had given Mary and Simon some chocolate, but told them to put it away until mum said they could eat it. The second picture card showed Mary and Simon putting the chocolate away in the fridge, and the participants were then told the siblings went out to play. With the third picture card, participants were told that Simon came back inside for a glass of water and decided that he wants to keep the chocolate all to himself, so he took the chocolate out of the fridge and put it in his bag. At this point, the participants were asked two control questions to check they were following the story (*"Where does Mary think the chocolate is?"* and *"Where has Simon put the chocolate really").* If a child failed either of these questions the story was repeated. If they still answered incorrectly the task was abandoned.

The fourth picture showed Mary looking through the window as Simon put the chocolate in his bag, and participants were told that Mary was playing by the window and had seen everything that Simon was doing. They were also told that Simon was so busy hiding the chocolate he did not notice Mary watching him. Participants were then told that later mum called Mary and Simon in for tea and told them that they could have some chocolate. Participants were asked the test question

(*"Where does Simon think Mary will look for the chocolate?*") followed by a justification ("*Why does Simon think that?"*). Finally, participants were asked a reality control question ("*Where is the chocolate really"*) and a memory control question ("*Where was the chocolate first of all?*"). One point was awarded for the second-order false belief test question, if both the reality and memory control questions were also answered correctly. A further point was awarded for the justification test question if the test question was correct. This gave a maximum score of two.

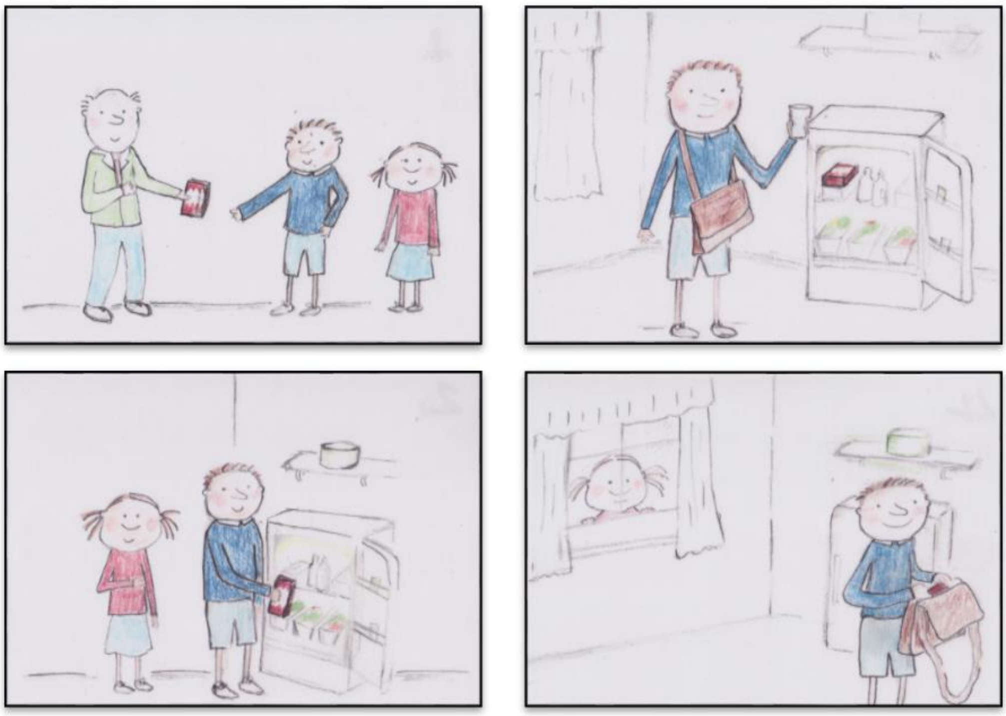


Figure 2.9: The four picture cards used to tell the story during the unexpected locations second-order false belief task.

As the scoring of this measure required a judgment decision, 30% was scored by a second researcher and Cohen's Kappa was run to determine if there was consistency between scorers. Based on guidelines by Altman (1990) agreement can be said to be excellent with 92% of total agreement (Kappa = .85, $p < .001$).

This task was chosen because it was age appropriate to use at Time 2, with other research administering it to both four year olds (Bosco & Gabbatore, 2017; Guajardo & Cartwright, 2016; Hayashi, 2007; Lavoie, Leduc, Arruda, Crossman, & Talwar, 2017; Lockl & Schneider, 2007) and five year olds (Arslan, Verbrugge, Taatgen, & Hollebrandse, 2015; Bosco & Gabbatore, 2017; Guajardo & Cartwright, 2016; Hayashi, 2007; Lavoie et al., 2017; Lockl & Schneider, 2007). This ToM test has been used in longitudinal studies (Ensor et al., 2014), and in other similar research such as that comparing ToM to other forms of metacognition (Lockl & Schneider, 2007), research into ToM and RC (Atkinson, 2014; Guajardo & Cartwright, 2016), and ToM and LC (Strasser & Río, 2014). Additionally, reliability for this test is reported as good for children of varying abilities (Hughes et al., 2000).

At Time 3 the strange stories as developed by White, Hill, Happé, and Frith (2009) was used to assess ToM. Here five of the eight mental state stories by White et al., (2009), based on the initial work by Happé (1994), were used. These stories were developed as an advanced test of ToM for older children (White et al., 2009). The aim was to tap into mentalising concepts through a selection of simple stories where, in each, the protagonist either makes a belief based misunderstanding or is motivated to tell an untruth. Successful performance required attribution of mental states such as desires, beliefs, intentions or sometimes higher order mental states such as one character's belief about what another character knows. For example, in a belief based misunderstanding story a burglar is running away after having stolen from a shop and, as he runs past a policeman, he drops his glove. The policeman retrieves the glove and calls for the burglar to stop so that he could return the glove, but the burglar puts up his hands and gives himself up admitting he stole from the shop. Children were required to understand the burglar's false belief of the policeman's intentions. In an untruth story, Brian is said to be a greedy boy. At school it is sausages and beans for lunch, which is Brian's favourite, and so to gain extra sausages he tells the server that he will not be having any dinner when he gets home, even though the truth is that his mother will be making him a lovely meal. Children were required to understand that Brian tells a lie to provoke sympathy because of his desire to receive extra sausages.

The five stories were presented on a tablet computer using a recorded PowerPoint presentation in one of three orders. The researcher introduced the task to the child by explaining that they were going to watch and listen to some stories, and that they should listen carefully as they would be asked some questions to see what they thought of the stories. The presentation of each story included a sequence of three coloured cartoon pictures that appeared on the screen as each story was recounted using an audio recording. Figure 2.10 shows the sequence of three pictures for the untruth story about Brian described above.

Figure 2.10: The sequence of three pictures for the untruth story about Brian described above. The pictures were shown on a tablet computer in PowerPoint with a voice over telling the story. Each picture was flashed onto the screen at the appropriate point in the story.

After each story the researcher asked a corresponding question to assess whether the child had understood the misunderstanding or the untruth told. In line with the scoring of White et al., (2009) for each story participants scored no points if they gave an irrelevant or factually incorrect response, one point if they gave some factually correct information but were not fully able to understand that an untruth/misunderstanding had taken place and the reason or consequences of it, and two points if they gave an answer which showed an advanced understanding of why the character believed what they did or why they told an untruth. For example, in the burglar story described above they were asked "*Why did the burglar do that*". Children scored no points if they gave a factually incorrect or irrelevant answer. They gained one point if they referred to something factually correct in the story, and two points if the referred to the belief that the policeman did not know that the burglar had burgled the shop. In the story of Brian and the sausages they were asked *"Why does Brain say that?"* Children scored no points if they referred to a motivation that missed the point of sympathy elicitation/deception, or if they were factually incorrect. They scored one point if they made a reference to his state (greedy) and the outcome (to get more sausages), and they scored two points if they referred to the fact that he was trying to elicit sympathy and be deceptive to gain the extra sausages. During administration the researcher provided positive engagement but gave no direct feedback on the accuracy of child responses. The total possible score was 10. See Appendix 12a for full script with scoring criteria and Appendix 12b for all story pictures.

As the scoring of this measure required a judgment decision, 25% was scored by a second researcher and Cohen's Kappa was run to determine if there was consistency between scorers. Based on guidelines by Altman (1990) agreement can be said to be excellent with 87% of total agreement (Kappa = .83, $p < .001$).

This measure was chosen as it is well used (e.g. Atkinson, 2014; F. Bianco & Lecce, 2016; Devine et al., 2016; Ensor et al., 2014; Kirk et al., 2015; Lecce, Bianco, Devine, Hughes, & Banerjee, 2014; Lecce et al., 2014; Lecce, Caputi, & Pagnin, 2014; O'Hare et al., 2009; Wang, Devine, Wong, & Hughes, 2016). Additionally, it has been used by others with similar aged children. For example, Leece et al., (2014) administered the stories to four and five year olds, Atkinson (2014) administered them to six year olds, Kirk et al., (2015) to five and six year olds, and O'Hare et al.,

(2009) to five to 12 year olds. This ToM test has been used in longitudinal studies (Ensor et al., 2014), and in other similar research such as that comparing ToM to other forms of metacognition (Lecce et al., 2014; Lecce et al., 2014), and research into ToM and RC (Atkinson, 2014).

An additional advantage of the strange stories is that they have been reported as being reliable with good test-retest reliability (Devine & Hughes, 2016b; Hughes et al., 2000; Hutchins, Prelock, & Chace, 2008). Further, White et al., (2009) emphasise that a strong correlation between a ToM battery (battery including unexpected locations tasks and unexpected contents tasks) and the mental state stories which provides a validity check that both measures are tapping the same underlying ability.

2.5.7 Broad metacognition

At all time points, three types of metacognitive skills were assessed: source monitoring, metalinguistic knowledge and metamemory.

2.5.7.1 Source monitoring

Source monitoring, the understanding of the source of one’s knowledge, was assessed at Times 1 and 2 using the tunnel task as used by O'Neill and Gopnik (1991). Here children had to identify an object within a closed tunnel and recognise the source of their knowledge of that object. For this task a red “tunnel” (approximately 25 cm x 17 cm x 15 cm) was used. The tunnel was made from thick cardboard lined inside with cream fabric, and covered with red felt. At either end the openings were concealed with a red felt flap, so that a child could not see inside when the flaps were. Figures 2.11 and 2.12 show the tunnel and a selection of objects used during the task. The tunnel was placed on the table or floor in front of the child, with one opening facing them and the other facing the researcher. The child was told that the researcher would “*put different objects inside the tunnel*”. For each trial, they would either be allowed to LOOK inside, or the researcher would TELL that what was inside, or they would be allowed to put their hand inside to FEEL the object. The researcher would then ask them “*what is inside, and how you knew that was what was inside?*”.

*Figure 2.11:*The big red tunnel from above and a selection of objects used within the task.

Figure 2.12: Side view of the red tunnel with the felt flat up so that the object inside can be seen by the child.

The training trials then began and consisting of three trials, one for each of the three types of source information (SEE, TELL, FEEL). The training trials were very similar to the experimental trials except that the three types of source were *explicitly* identified, and the child received feedback about their responses. The objects used in the training trials were: a helicopter, a toothbrush, and a cup. On the SEE trial the toy helicopter was placed in the tunnel, and the researcher said, "*lift the tunnel up, can you see what's inside*". On the TELL trial the toothbrush was placed in the tunnel, and the researcher said, "*This time you can't look inside, but I am going to tell you what's inside, there is a toothbrush inside*". On the FEEL trial the cup was placed inside the tunnel, and the researcher said, "*this time you can't look inside, but you can put your hand in and feel what is inside.*" The order of presentation was fixed. In each case the child was first asked to identify the object inside the tunnel. The identification question explicitly mentioned the source in each trial (e.g. "*what did* *you see inside the tunnel?*”). After correctly identifying the object they were asked the source question (“*How did you know that’s what was inside?*”). If they gave no response or a general response (e.g. “*because*”), they were given a forced choice question (“*Did you see it, did I tell you, or did you feel it?*”). For the training trials only, they were given explicit feedback: “*That is right, you saw the helicopter inside the tunnel*” or “*no you saw the helicopter inside the tunnel*”.

Directly after the training trials there were six experimental trials. These trials used: a toy horse and a box of crayons for the SEE trials, a plastic spoon and a ball for the FEEL trials, and a toy car and a pair of plastic scissors for the TELL. The experimental trials continued in the same way as the training trials except that the source of knowledge was not explicitly mentioned or referred to in the identification question. Instead the researcher simply asked, “*What’s inside?”,* and no feedback was given. The order of the six experimental trials was counterbalanced, and no two successive trials were of the same source type. For each trial, the child had to correctly recognise the object and correctly identify the source to receive one point. For example, they had to identify the horse, and then say that they saw the horse inside the tunnel. The maximum score was therefore six.

This task was chosen because it was age appropriate and appealing. It has been used by others with similar aged samples, for example Carlson, Claxton, and Moses (2015) with children aged three and four years, by O'Neill and Gopnik (1991) with children aged three to five years old, and by Bright-Paul et al. (2008) with those aged three to six years. It has also been used in similar research comparing ToM ability to source monitoring (Bright-Paul et al., 2008).

At Time 3 the source monitoring measure was modified to remain age appropriate. Here, instead of assessing children's understanding of the source of their knowledge of an object hidden inside a tunnel instead children's understanding of the source of their knowledge about an event was assessed. This modification was based on the measures used by Ozturk and Papafragou (2016), and Gopnik and Graf (1988). Stimuli were presented to the child on a tablet computer. As with the tunnel task there were three types of trials; SEE trials, HEAR trials and INFER trials, each responding to the source of knowledge type. In the SEE trials children saw an event happening on the screen (e.g. an animation of a frog jumping). In the HEAR trials children heard a character utter a sentence describing an event (e.g. a male voice said, *"I played basketball"*). In the INFER trials children were given a clue about what happened and had to infer (or "*work out*") what had happened from the clue (e.g. they saw a still picture of a living room strewn with Christmas wrapping paper and had to make the inference that a family had opened Christmas presents). For each trial the child was asked *"What happened?"* and then *"How do you know that is what happened?"* As with the tunnel task, if the child gave no response or an incoherent response to this last question they were given forced choice "*Did you SEE it happen, did you HEAR about it, or did you work it out from a CLUE?"* A correct response to both questions was needed to score one point.

Three training trials preceded the experimental trials, one for each type of source information (SEE, HEAR, INFER). The training trials were very similar to the experimental trials except that the three types of source were *explicitly* identified in the question, and the child received feedback about their response. On the SEE training trial children were told *"Right let's see what happened*" and were then shown an animation of a car driving down a road. They were then asked, "*What did you see happen?"* and *"How did know that is what happened?"* On the HEAR training trial children were told *"This time I am not going to let you see what happened but, I will let you hear what happened"* and were then played an audio clip of a girl stating, *"I read a book*". They were then asked, *"What did you hear happened?"* and *"How did know that is what happened?"* On the INFER training trial children were told *"This time I am not going to let you hear what happened but, I will give you a clue about what happened so you can work out what happened*". They were then shown a still picture of a girl crying with a wall full of red scribbles in the background and were asked "*Here is the clue, can you work out what happened?"* and *"How did know that is what happened?"* For the training trials feedback (such as: *"That is right you knew the girl drew on the wall because you worked it out from a clue"*) was given. Training trials were presented in this fixed order. The six experiment trials (two of each of the three types of trial) followed directly after the training trials presented in one of three orders. The total score was six.

As previously stated this measure was chosen for Time 3 to ensure the task was age appropriate. It was felt that if the tunnel task, as used to measure source monitoring at previous time points, was administered, scores would be at ceiling. For example, although Bright-Paul et al. (2008) did administer the tunnel task up to the age of six, for children aged five to six years the mean score was 5.66 (maximum score = 6). Given this, it was felt that this modification (which still taps the same underlying ability with a very similar scoring procedure) was appropriate. Moreover, Ozturk and Papafragou (2016) have administered this version successfully to five to seven year olds.

2.5.7.2 Metalinguistic Knowledge

Metalinguistic knowledge, the ability to reflect on the use of language (Doherty & Perner, 1998; Doherty, 2000), was assessed through a synonym judgment task at Times 1 and 2 (Doherty & Perner, 1998; Doherty, 2000). The task assessed children's knowledge of synonyms, i.e. words that have the same or nearly the same meaning as other word. The measure assessed the understanding that some objects have two or more names, for example, a sofa can also be called a settee or a couch. At Time 1 the same task used by Doherty and Perner (1998) and Doherty (2000) was used. At Time 2 this task was adapted to be age appropriate.

The Time 1 task consisted of three parts: the vocabulary check, the modelling phase, and the testing phase. The aim of the vocabulary check was to examine knowledge of the synonyms used later in the actual test, and to alert the child to the distinctions made later in the test. Four laminated sheets, each with four coloured pictures on were used (see Appendix 13a). Each of the sheets had on them two pictures which were used as experimental items later (truck/lorry and woman/lady on two of the sheets; TV/television and coat/jacket on the other two). The other two pictures on the sheet were either: a rabbit, a cat, an apple, a bird or a daisy. Participants listened to a word spoken by the researcher and were required to indicate what it meant by pointing to the corresponding picture. For the experimental words, both synonyms were used, e.g. children were asked to point to both a "*lorry*" and then later a "*truck*", which was represented by the same picture. If they hesitated they were given encouragement, and the question was repeated. If they responded incorrectly to one of the experimental words they were told that the object has two names.

The objective of the modelling procedure was to model the testing procedure to the child. A white bear glove puppet called Timmy was used alongside two laminated sheets, one with a picture of a rabbit/bunny on it, and the other with a picture of a cup/mug (see Appendix 13b). First the child was shown the picture of a rabbit, and it was explained that the picture could be called either a rabbit or a bunny. The child was asked to choose one name to call it. Then Timmy was introduced, and the child was told that it was Timmy's job "*to say the other name NOT the one that you said*". Addressing the puppet, the researcher asked Timmy to say the other name, not the one that the child had said. The puppet then gave the wrong answer (e.g. if the child had said bunny, Timmy also said bunny, and if the child had said rabbit, Timmy also said rabbit). The researcher then asked the child "*is that what he should have said*?" paused slightly for the child's response, before saying "*No, because you said rabbit/bunny and it is his job to say the other name not the one that you said*". This phrase was said irrespective of the child's response, because if they gave the correct response it acted as a confirmation of what they had just said, and if they gave the incorrect response it corrected their response. The researcher then encouraged Timmy to have another go and this time Timmy gave an unrelated word

("*Elephant*"). The researcher once again asked the child if this is what he should have said, paused slightly for the child's response, before saying "*No because you said rabbit/bunny and it is his job to say the other name not the one that you said*". Timmy was given one last go, to which he gave the correct answer (bunny/rabbit dependent on what the child had said), and the child was asked if this is what he should have said. The researcher gave feedback by saying: "*Yes! Because you said rabbit/bunny so he said rabbit/bunny the one that you didn't say*". The second modelling trial continued using the same procedure, except that a picture of a cup/mug was used. As with the first modelling trial, Timmy was given three chances, until he eventually gave the correct answer.

For the testing phase the modelling phase continued but with four different pictures (woman/lady, truck/lorry, TV/television, coat/jacket; see Appendix 13c). During this phase Timmy was only given one chance to answer, and the child was given no feedback. The four items were presented in a fixed order. It was also fixed so that in the first trial Timmy always gave the correct answer (woman/lady). For the second trial (truck/lorry) he was incorrect in that he said the same as the child. For the third trial (TV/television) he was also incorrect as he gave an unrelated answer (woman). For the final trial (coat/jacket) he was correct. See Figure 2.13 for the stimuli used including Timmy the puppet and the woman/lady card used in Trial one. For clarification of the presentation of the fixed trials and the researcher's responses based on the responses of the child see Figure 2.14. For each trial children scored one point for correctly identifying whether that was what Timmy should have said. There was a maximum score of four.

Figure 2.13: Stimuli used at Time 1 for the synonym judgment metalinguistic knowledge measure. Including Timmy the puppet and the woman/lady card used in Trial one.

This task was chosen because it was age appropriate to use at Time 1 as it has been used by Doherty and Perner (1998) and Doherty (2000) with British children aged three to five years old, and Japanese children aged three to four years (Doherty & Itakura, 1995). It has also been used in similar research comparing metalinguistic knowledge to ToM (Doherty & Itakura, 1995; Doherty & Perner, 1998; Doherty, 2000).

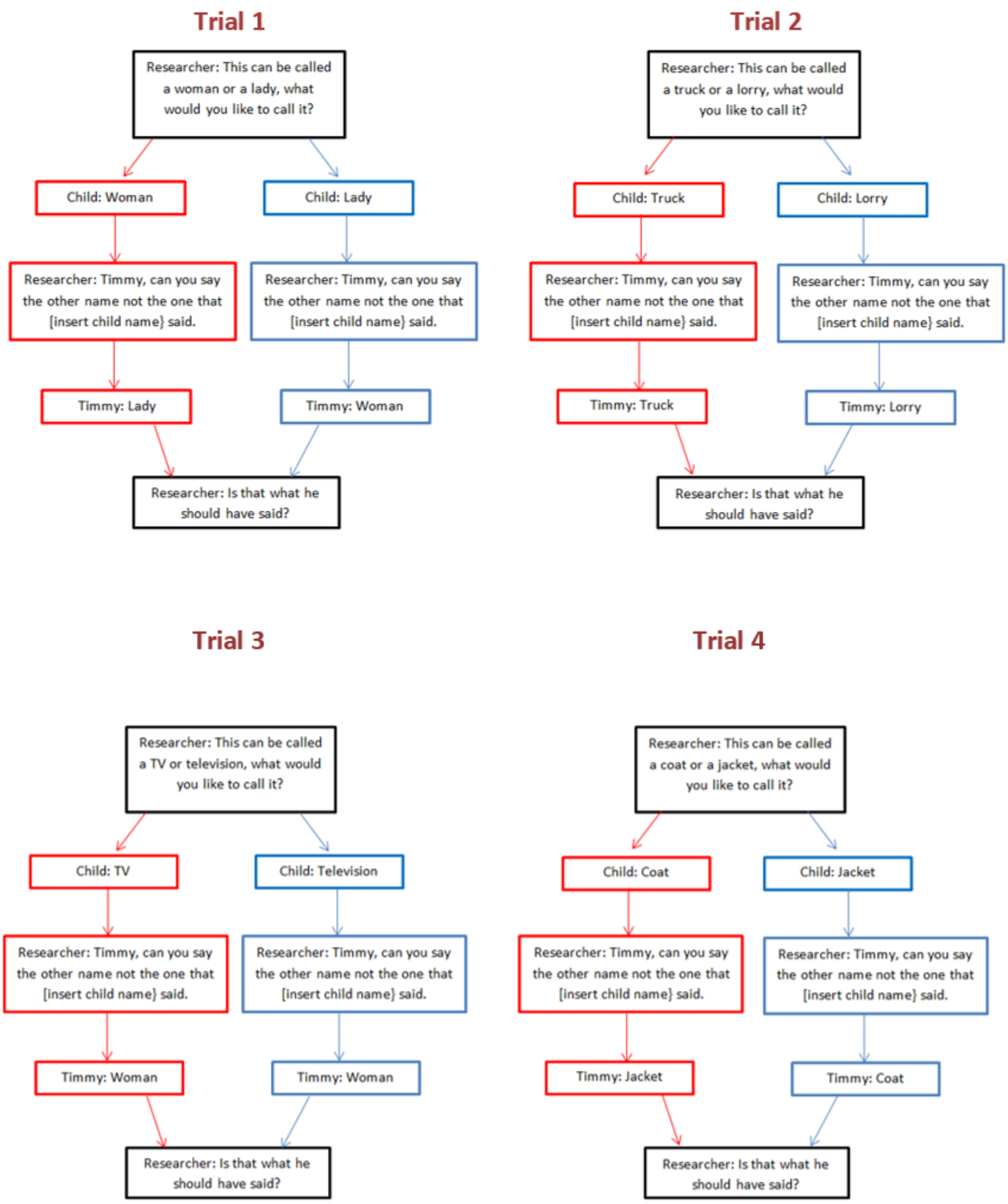


Figure 2.14: Flowcharts clarifying the fixed trial orders and responses made by the child and researcher for the synonym judgment task. The red and blue routes on the flowcharts show the responses made by the researcher dependant on the responses made by the child. Correct answers for this task are yes for Trial 1, no for Trial 2 and Trial 3, and yes for Trial 4. Each correct answer received one point.

At Time 2 the measure was modified to be more age appropriate, this included more test trials and fewer training trials. This modified version closely resembled tasks used by Nation and colleagues with older children (Nation & Snowling, 1998; Nation & Snowling, 2004). Fourteen cards (see Appendix 14a) with pictures of objects which also have another name (or synonym), for example jumper/sweater, spaghetti/pasta and Hoover/vacuum cleaner were used. Children were told they were going to play a picture game and the first card (a picture of a bunny/rabbit) was presented. They were told that some objects have two names, and so this picture can be called a rabbit, but also a bunny. The next card was presented (a picture of a mug/cup), and children were told *"This is a mug, can I also call this a bike?"* After waiting a few seconds for the child to respond, the researcher explained that a mug and a bike are not the same thing and that the picture cannot be called a bike. Children were then told that in the game they would be show some more pictures and that the researcher would tell them what the pictures were, and that they would be asked if they could also call the picture another name. The 12 test trials then began. Half the trials were synonym matches (for example, a boat could also be called a ship) and half were non-matches (for example, a lady could not also be called a jacket). For each trial the child was presented with the picture card, for example a picture of a jumper/sweater, and the researcher said, "*This is a jumper, can I also call this a sweater*", and children were required to answer yes or no (correct answer: yes). In one of the non-match trial a picture of a Television/TV was presented and children were asked, "*This is a television, can I also call this pasta"* (correct answer: no). For a full list of the synonyms (and non-matched synonyms) see Appendix 14b. The trials were presented in a fixed random order and children were awarded one point for each test trial, and so the maximum score was 12.

At Time 3 children's understanding of the use of homonyms (two words with the same spelling or pronunciation but with different meanings) was used to assess metalinguistic knowledge. This homonym selection task was based on the Diaz and Farrar (2017) modification of Doherty (2000). The measure consisted of two parts: the vocabulary check and the testing phase. The aim of the vocabulary check was to ensure that children were familiar with both uses of the four target words. Two laminated sheets, each with four coloured pictures on were used (see Appendix 15a). The first vocabulary card had one version of the test words and the second vocabulary card had the other. The test words were bat, glasses, (k)night, and letter. The pictures on the first vocabulary card were: a bat (nocturnal flying mammal), a pair of glasses used to improve sight, the night (sky with moon and stars) and a letter (the letter "A"). Pictures on the second card were: a bat (wooden sports equipment used to hit a ball), glasses (used to drink beverages from), a knight (a man who serves his sovereign or lord as a mounted soldier in armour), and a letter (a written, typed, or printed communication, sent in an envelope by post). The researcher assessed the child's understanding of each of the words by asking them to point to each, e.g. "*Can you point the bat, and now to the glasses..."* Any incorrect responses were recorded.

Children were then told "*Great! Now let's look at some more*" and the testing phase began. For the test trials four laminated sheets were used each containing both versions of one of the target words and two unrelated pictures. For example, the first card contained a picture of a mammal bat, a wooden bat, a man's shirt and a chocolate cake. The researcher asked the child "*Which two have the same name?"* and then a further justification question *"Why did you point to those two".* Children were awarded one mark for pointing to the correct two pictures, and a further mark for giving a correct justification. Correct justifications included: "*because they have the same name"* or a variation on "*because that is a bat which flies, and that is a bat which you hit a ball with*". "*Because they begin with the same letter*" or similar was not accepted as this did not demonstrate the understanding that the two objects had the same name. All cards were administered in the same order for all participants. The maximum total score was eight. See Appendix 15b for the test picture cards.

This measure was chosen as it has been used with similar aged children aged two to six years by Diaz and Farrar (2017) to compare ToM and metalinguistic knowledge development. As the scoring of this measure required a judgment decision, 25% was scored by a second researcher and Cohen's Kappa was run to determine if there was consistency between scorers. Based on guidelines by Altman (1990) agreement can be said to be excellent with 90% of total agreement (Kappa = .88, $p < .001$).

2.5.7.3 Metamemory

Metamemory is the thinking about one's own, and other's memory capabilities, and the understanding of how best human memory works (Ebert, 2015). To assess this, a task similar to that of Ebert (2015) was used. Ebert's task was a modified version of an assessment originally used by Wellman (1977), and similar tasks have been used by others (e.g. Lockl & Schneider, 2007). For each trial, the child was shown a pair of laminated sheets with pictures of objects on, while the researcher described the associated memory learning circumstance. The child was then asked to decide which card was easier to remember, or whether there was no difference in difficulty between the two cards. Playmobil© dolls were used to illustrate the memory circumstances to the children.

The memory circumstance used in the first trial was "study time". For this trial, children had to understand that the more time an individual is given to learn a group of pictures, the more likely they are to remember. Two dolls, one male (named Tom) and one female (named Polly), sat on the table facing the child. Two laminated sheets with the same pictures of six objects (Appendix 16a) were laid out on the table; one in front of Tom and one in front of Polly. It was explained to the child that Tom and Polly had the same pictures to learn, but that Polly "*only has a short time to look at her pictures*". As this was said a piece of thick card was used to cover Polly's pictures so that it appeared as if she could no longer see them. Tom was given longer to look at his pictures, and it was a few seconds until his pictures were covered up with another piece of card. The child was then asked, "*It is easier for Tom or for Polly to remember their pictures or will they remember the same*?" The child was then asked a justification question "*Why is it easier for….to remember his/her pictures*". For this first trial only, the child was also asked a control question: "*Which one had more time to remember the picture*?" The child was given one point if they correctly answered that it would be easier for Tom to remember his pictures, as he had had longer to look at them (correctly answering both the test question and the justification question). To score this point, they must have also correctly answered the control question.

The memory circumstance used in the second trial was “number of items”. For this trial, children had to possess the knowledge that a shorter list is easier to remember than a longer list. It was explained to the child that Tom and Polly had “*some more pictures they had to remember*”. This time two different laminated sheets were laid out in front of Tom and Polly (see Appendix 16b). Tom’s sheet had six pictures on it, whereas Polly’s only had three. As with the first trial the child was asked who would find it easier to remember their pictures and then asked to justify their answer. They were given one point if they correctly recognised that it would be easier for Polly to remember her pictures as she only had three to remember, whereas Tom had six. Figure 2.15 shows how this trial was presented to the children.

The memory circumstance used in the third trial was “random versus categorised order”. For this trial, children were required to understand that categorised items are easier to remember than items presented in a random order. This time a sheet with nine pictures on it was laid out in front of Tom (Appendix 16c). Tom’s pictures were ordered into categories so that the three pictures which were food items were all in a row, the three pictures which were furniture items were all in another row, and the three pictures which were clothing were all in a row. Polly’s sheet (Appendix 16d) had the same pictures on it, but her pictures were not sorted into categories, instead they were mixed up randomly on the page. Children were given one point if they correctly stated that it would be easier for Tom to remember his pictures, as they were sorted into categories on the page.

The memory circumstance used in the fourth trial was “hair colour”. For this trial, children had to possess the knowledge that hair colour does not affect memory capability. Two more dolls were used who were identical except that one had black hair, and the other had blonde hair. It was explained to the child that two more people had some pictures to learn, and that they had the same pictures to learn and the same about of time to look at their pictures, but that one doll had blonde hair and the other had black hair. The two dolls were placed sitting where Tom and Polly had been. The same laminated sheets from the first trial (Appendix 16a) were used. One was laid in front of the blonde-haired doll, and the other in front of the black-haired doll. As with the other trials the child was asked: “*Is it easier for the blonde haired one, or for the black haired one to remember their pictures, or will they remember the same?*” Children scored one point for correctly recognising that the dolls would remember their pictures the same, as they had the same pictures to remember and hair colour does not affect memory ability. For the whole task a maximum score of four could be given.

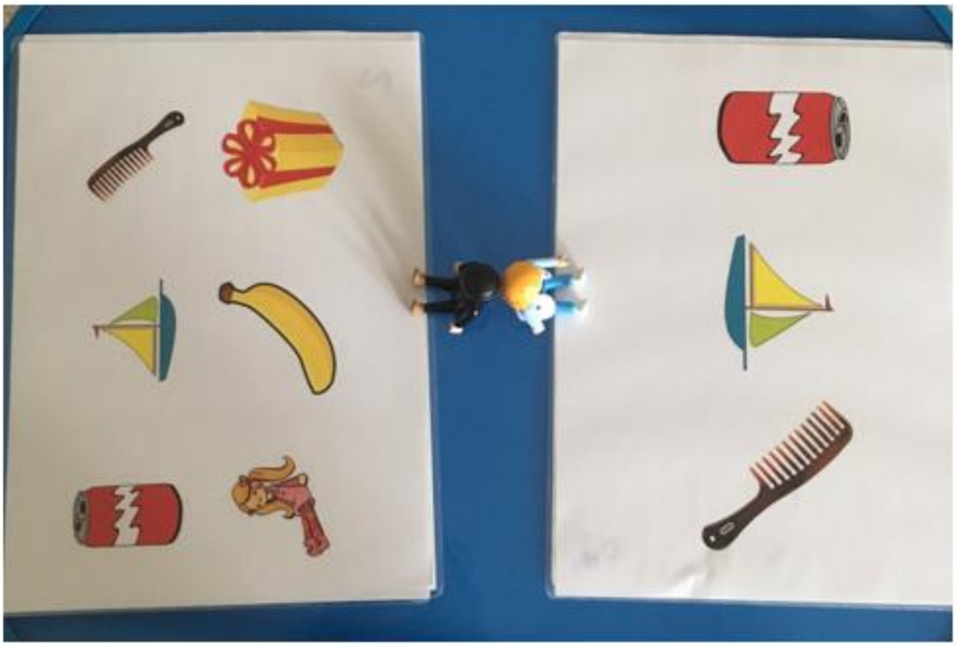


Figure 2.15: Trial two where the memory circumstance was “number of items”. Polly (right) has three pictures to remember and Tom(left) has six to remember. The correct response is that Polly would find it easier to remember her pictures as she has fewer items.

As the scoring of this measure required a judgment decision, 25% at Time 1, 30% at Time 2 and 46% at Time 3 was scored by a second researcher and Cohen’s Kappa was run to determine if there was consistency between scorers. Based on guidelines by Altman (1990) agreement for all time points can be said to be excellent with 96% of total agreement (Kappa = .94, $p < .001$) at Time 1, 90% of total agreement (Kappa = .87, $p < .001$) at Time 2, and 92% of total agreement (Kappa = .87, $p < .001$) at Time 3.

The task was selected because it was age appropriate to use at all three time points as it was used by Ebert (2015) with children aged three to six years old, and a similar task was used by Lockl and Schneider (2007) with a sample aged three to six years old. This research was also similar to the current as it compared metamemory to ToM (Lockl & Schneider, 2007). Additionally, internal reliability for the similar task used by Lockl and Schneider (2007) was reported to be α =.79.

2.5.8 Working memory

At all time points the reverse word span task (Slade & Ruffman, 2005) was used to measure working memory. This is a modified version of the Backwards Digit Span task by Davis and Pratt (1995). The task required children to orally reverse short sets of words spoken by the researcher. A teddy bear (named Eddie) was used to help the children learn the "*backwards game*". The children were told that in the game the researcher would say some words, e.g. "*horse – sheep*", and that Eddie would say them in a backwards order e.g. "*sheep – horse*". As the researcher spoke the set of words, Eddie was used as a prop, moving him from side-to-side, so that when they said the words "horse – sheep" (in their forwards order) Eddie was moved from right to left, and when they said the words "*sheep – horse*" (in their backwards order) Eddie was moved from left to right. This helped in distinguishing forwards and backwards to the children. There was one further training trial followed by two practice trials in which children were instructed that they should "*try the game now*", and "*Eddie will help you*". Here they were asked to reverse the words, but if they could not do so successfully Eddie gave the answer for them.

The test trials then began, with three sets of two words e.g. "*scarf – coat*", and three sets of three words e.g. "*plane – dog – pear*". For the test trials, no feedback was given but the task was discontinued if the child could not reverse the words in the first two trials. At Time 3 one extra trial ("*Ball – Cat – Tree"*) was added. One point was awarded for correctly reversing two words, and two points for correctly reversing three words, half points were awarded for reversing two words that were not adjacent. The maximum score was nine at the first two time points and 11 at Time 3. This task was chosen because it has been used in the past with a sample of similar age children (Slade & Ruffman, 2005), and by Atkinson et al., (2017) with children aged three to four year old also looking at RC and ToM.

2.5.9 Inference making

At Time 3 only inference making was measured using an oral story which required the ability to make both local coherence inferences and global coherence inferences. The story used was entitled "A new pet" and was developed by Language and Reading Research Consortium and Muijselaar (2018) based on research by Cain and Oakhill (1999; 2014) with older children.

Children were read aloud the story by the researcher and then required to answer the eight inference questions. Four questions involved local inferences and four required children to make global inferences. For local inference questions children needed to integrate information from different parts of the passage. For example, to answer the question "*What did Tim buy at the shop?*" it was necessary to connect the information that Tim went to the shop and that he needed wood and nails from two different sentences to give the answer "*he bought wood and nails*". On the other hand, global inference questions required children to fill in details not explicitly stated within the passage to construct a globally coherent representation of text meaning. For example, to answer the question "*What sort of animal was Sparky?*" children needed to infer that Sparky was a dog because he was soft, furry and playful and had a kennel. The maximum total score was eight. See Appendix 17 for the story and inference questions and answers.

This measure was chosen as it is based on a wealth of research by Cain, Oakhill and colleagues (e.g. Cain & Oakhill, 1999; Cain et al., 2001; Currie & Cain, 2015; Freed & Cain, 2017; Oakhill & Cain, 2012; Silva & Cain, 2015). Additionally, Language and Reading Research Consortium and Muijselaar (2018) administered this story to similar age children aged five to six years. They also report the measure to be reliable with a Cronbach's alpha of $\alpha = .71$ for six year olds (Language and Reading Research Consortium & Muijselaar, 2018).

2.5.10 Comprehension monitoring

At Time 3 only comprehension monitoring was measured by children's ability to monitor their comprehension of short stories to determine if the stories were sensical. The stories were that used by Cain and Yeomans-Maldonado (2017) and were based on previous work with older children where children have read the stories aloud to themselves (Cain & Oakhill, 2006; Florit & Cain, 2011; Oakhill & Cain, 2012). Here, children listened to five quick stories of five short lines read orally to them by the researcher. Directly after each story they were required to state whether the story "*made sense*" or "*did not make sense*". An example of not making sense was a story which stated that a pet rabbit never goes outside, but later on it asserted that the rabbit plays in the garden every day. Children were given a practice story first in which the researcher gave them explicit feedback, e.g. *"that's right this story does not make sense as firstly it says it is Katie's sixth birthday today, and then later it says she is ten years old".* This practice was to alert children to what was meant by not making sense. They scored one point per story and so the maximum possible score was five. See Appendix 18 for the stories.

This measure was chosen as it is based on a wealth of research by Cain, Oakhill and colleagues (Cain & Oakhill, 2006; Florit & Cain, 2011; Oakhill, Hartt, & Samols, 2005; Oakhill & Cain, 2012). Additionally, Cain and Yeomans-Maldonado (2017) administered the measure to similar age children aged five to seven years. They also report the measure to be reliable with a Cronbach's alpha of α = .73-.84 for internal consistency across children aged five to seven (Cain & Yeomans-Maldonado, 2017).

2.5.11 Non-verbal ability

To measure non-verbal intelligence, the block design subset of the Wechsler Preschool and Primary Scale of Intelligence - Third Edition (WPPSI-III; Wechsler, 2002) was used at Time 1 only. This required children to recreate several geometric patterns of increasing difficulty, using coloured blocks. For example, recreating a checked square pattern in which two white and two red blocks fit together. In line with standard procedure, the first ten trials used red and white blocks, and the last ten trials used two-toned blocks. For the first 12 trials the researcher built the pattern with oral commentary, before encouraging the child to make one that looked the same using a different set of blocks. For trials 13 to 20 the target pattern for each trial was shown to the child in picture form from the stimulus book, and they were encouraged to make one that looked the same as the picture. For trial 12 only the child saw both the picture of the pattern in the stimulus book, and the researcher creating the pattern. For the first six trials children were given 30 seconds to recreate the pattern, but if they could not do so in this time, they were shown how to create the pattern again, and given a further attempt. They scored two points if they created it in the first 30 seconds, 1 point if they created it in the additional 30 seconds, and 0 points if they failed to recreate it. In line with standard procedure from trial 7 onwards they were only given one attempt to recreate the pattern, with a time limit of 30, 60 or 90 seconds. Here they scored two points for successfully creating the pattern in the time limit. The task was discontinued after three consecutive scores of zero. The maximum score was 40.

This task was chosen because is a well-used standardised measure, often used for controlling for non-verbal intelligence in reading research (e.g. Cain & Oakhill, 2011; Hulme et al., 2012; Marshall, Snowling, & Bailey, 2001; Nation, Clarke, & Snowling, 2002). Additionally, it is also well-used with children of this age, for example, Guo, Piasta, and Bowles (2015) administered it to children aged three to five years and Hulme et al., (2012) in children aged four to five. The measure has good reliability, with split-half reliability α= .84, and .76 for test-retest reliability as originally reported by Wechsler (2002). Additionally, subsequent studies have reported α = .79 for internal consistency for children aged three to five (Guo et al., 2015).

2.5.12 Reading comprehension

At Time 3 only children's RC was assessed using Form A of the Primary passages of the York Assessment of Reading for Comprehension (YARC: Snowling et al., 2011). This standardised test comprises graded fiction and non-fiction passages and is normed from ages five to 11 years. Children read aloud two passages and were then asked eight comprehension questions per passage.

Consistent with standardised procedure to prepare for the YARC children were administered the Single Word Reading Test (SWRT). The SWRT consists of six blocks of ten single words of increasing difficulty which children were required to read. A child's raw score on the SWRT determined their passage YARC start level. For example, if a child scored below 19 on the SWRT they started at the beginner level passage, and if they scored between 19 to 24 they started at Level 1. In line with standardised procedure, with the exception of the beginner's passage, children were timed as they read the passages to calculate their reading rate. The accuracy of their reading was recorded by noting all reading errors and the type of error. These errors included, mispronunciations (when words were wrongly pronounced or were only partially decoded), substitutions (when an incorrect real word was given instead of the word in the passage), reversals (a type of substitution error when letters of a word were reversed e.g. was $\rightarrow$ saw, on $\rightarrow$ no), refusals (when children were unable to attempt the word), additions (when children inserted a word or part word in the text) and omissions (when children omitted a word). For the beginner's passage only if a child made more than 16 reading errors the measure was discontinued, and no comprehension questions asked; this occurred for three children.

Directly after the child had read the passage, the researcher asked the eight comprehension questions. If a child scored more than four on the comprehension questions of their first passage they proceeded to the next passage, but if they scored less than four (with the exception of the beginners passage where they still preceded to the Level 1 passage) they dropped down to the previous level for their second passage. A reading comprehension ability score was computed and used in all analysis, as different children read different level passages the ability score reflects both the raw score and the difficulty of the passages they have read.

The YARC was selected as it is well used standardized measure of RC in the UK (Atkinson et al., 2017; Babayiğit, 2015; Cunningham & Carroll, 2015; Duff et al., 2015; Fricke et al., 2013). It was also age appropriate as it is normed from Year 1 onwards and thus has been used by others with similar aged children. For example, Atkinson et al. (2017), Fricke et al. (2013), and Duff et al., (2015) all administered it to five to six year olds who were in Year 1. It has also been used by a recent longitudinal study into RC and ToM (Atkinson et al., 2017). Additionally, the YARC is found to be a reliable measure of RC with original internal reliability reported as α = .71–.84 depending on the passages administered (Snowling et al., 2011). Subsequently, Fricke et al., (2013) report α =.77 and Atkinson et al, (2017) of α =.64 both for children aged six years.

2.6 Recordings and scoring of measures

During the sessions many of the measures were recorded on a digital recorder to score from the recordings later. When this was the case answers were still written and scored live. For the measures which required children to give long oral answers (NARA at all time points, inference making at Time 3, metalinguistic knowledge at Time 3, strange stories at Time 3 and YARC at Time 3) recordings were made to ensure that the researcher had written the child's answer accurately. For the decoding measures in which children were often required to give phonetically specific responses in which the researcher could have easily misheard live and scored inaccurately (word repetition and non-word repetition at Times 1 and 2, and single word reading at Times 2 and 3) responses were scored both live and from the recordings, and the scores compared. In a small number of cases, either due to researcher error or uncontrolled circumstances, recordings were not made. Therefore, as the scoring made from recordings was highly similar to the scoring performed live, live scoring only was used in all final analysis.

2.7 Home based measures

There were two home based measures: the parental questionnaire, and the book sharing activity. These were administered at Time 1 and Time 2.

2.7.1 The parental questionnaire

Parents/caregivers were asked to complete a questionnaire to collect information about family demographics and their child's use of mental state words. The questionnaire was available online via Qualtrics and in paper form (see Appendix 1). This questionnaire also asked for information about the home literacy environment including family reading habits but this data was taken as part of a wider project and was not used in this thesis.

To measure the mental state vocabulary used by the child, a vocabulary checklist was included in the questionnaire. So that parents did not guess the aim of the checklist it included 40 general words typical in children's vocabularies, e.g. "Big" "Empty" "Red", as well as 43 mental state words such as cognitive terms, for example "Think", mental state expressions referring to desires, for example "Hope", and emotions, for example "Angry". The general words were taken from the descriptive words section of the Oxford Communicative Development Inventory (A. Hamilton, Plunkett, & Schafer, 2000), which is a UK adaptation of the MacArthur-Bates Communicative Development Inventory. The mental state words came from a list of mental state words and phrases used to code a mental state talk activity by Ruffman, Slade, Devitt and Crowe (2006). This list was based on the criteria of Bartsch and Wellman (1995). General words were chosen from the descriptive words section of the Oxford Communicative Development Inventory as this section of words from the inventory was most similar to the mental state words, and so parents would be less likely to distinguish between the two types of words and guess the aim of the checklist. In the case of word overlap (that both lists included the same words, for example this occurred for the words "Happy" and "Scared") the word only appeared once in the checklist and was scored as a mental state word. The two types of words were presented in a fixed mixed order. Parents were required to indicate whether their chid understood but did not say the word yet, understood and said the word, or did not know the word at all.

The last section of the questionnaire required the parent to give demographic information such as: information about their marital status, their occupation, highest educational achievement, details of other children in the household, annual household income, and information about the languages spoken within the home.

2.7.2 The book sharing activity

The book sharing activity was used to measure the frequency of mental state utterance used by mothers and children during shared book reading. This activity was similar to that used by Ruffman et al. (2006) and more recently Carr, Slade, Yuill, Sullivan, and Ruffman (2018). Mothers were instructed to look at ten pictures with their child either on a phone/tablet screen or paper copies of the pictures. These pictures were taken from The Thorpe Interaction Measure (Thorpe, 1996). Mothers were asked to talk about the pictures as they would a storybook at bedtime. For those who viewed the pictures on their device (Cohort 1), they were asked to record the conversation also on the device. For Cohort 2 the pictures were viewed in paper form, and the conversation was recorded by the researcher using a digital voice recorder.

The ten pictures depicted children and their families taking part in normal everyday activities. For example, two children beneath a Christmas tree playing, a girl and a boy playing chess, and a girl riding a tractor. The book sharing activity was the same at both Time 1 and Time 2, except that different pictures were used (see Appendix 19a and 19b). The mothers were told that there was no time limit for the task and were encouraged to talk about each of the pictures for as long as they, and their child, wanted to. The conversation between child and mother was transcribed and coded later.

2.7.2.1 Coding of book sharing activity

The audio recordings were first transcribed by a transcriber and then coded by the researcher. Although others have used a slightly different mental state coding system (e.g. Devine & Hughes, 2017; J. Dunn, Brown, & Beardsall, 1991; Ensor et al., 2014; Symons et al., 2005) coding of the book sharing activity was based on that of Ruffman, Slade and Crowe (2002) and therefore driven by the broad mental state categories used by Bartsch and Wellman (1995). This coding has also been used by the more recent study of Carr., et al. (2018). This coding system was chosen as this thesis used the same pictures as Ruffman et al., (2002) and the study also had very similar aged participants, so terms used by child and mothers were expected to be very similar.

An utterance refers to a string of words identified by a grammatical mark of completeness or a pause (Golinkoff & Ames, 1979). All utterances were coded, into one of four types of mental state category (see Tables 2.4) or one of ten non-mental state categories (See Tables 2.5). These categories were in line with Ruffman et al. (2002), however an additional four categories (prompt questions, unrelated utterances, short responses and nonsensical utterances) were added for reoccurring utterances which did not fit within the categories of Ruffman et al. (2002). In some cases, utterances were coded under several categories. For example, the utterance "*Why is she crying?*", was coded as a physical utterance as it included the physical state of crying, but also as a causal utterance as it asked a question about a cause. Utterances could also be coded into one category several times, for example the utterance "*I don't think we know these babies*" was coded as cognitive twice as it made two references to cognitive states (think and then know). Thirteen percent of transcripts were coded by a second researcher. Cohen’s kappa inter-rater reliability was calculated for each category. Based on guidelines by Altman (1990), all categories were found to have moderate to very good reliability with kappa’s ranging from .82 to .69

The total number of utterances made by both the child and the mother and the mean length of utterances (MLU) was also recorded i.e. the mean number of words in an utterance. This was in line with Ruffman et al. (2002) and was important because it could be the amount of language a child is exposed to which is important rather than the content of the language exchange.

Table 2.4

The mental state coding categories with examples.

Category	Example	Notes
Cognitive terms	"I was thinking it could be grandad" "I think those are the Christmas presents" "I think he is trying to pick him up" "They're thinking about what to do." "I know what it is" "How do you know he's grumpy?" "Do you know which one is going to win?"	Cognitive terms referred to either thinking or knowing something. This included the cognition of either the speakers or others.
Desire terms	"I hope he makes it better" "Coz babies like that" "He doesn't want to cuddle"	These were terms referring to the desire of the speaker or others.

Modulations of assertion	"He could be hiding or he could be hot" "What is that they are doing I wonder" "They are breaking it or maybe they are fixing it" "So maybe he's a bit younger"	These were utterance gaging certainty. Context was used to determine if an utterance was truly used to modify certainty.
Emotion terms	**"**Does it look happy or sad?" "She doesn't look very happy." "He's a bit grumpy."	These were utterances about emotions of self or others. Emotion terms were distinguished from desires.
Other mental states	**"**Do you remember that old lady who lived in the shoe?" "She is pretending to be the mummy" "Imagine there is a switch in there"	This category described other mental activities not captured by the other categorises for example, remembering, realising, considering and having an idea. Wonder and expect could be coded as modulations of assertion and so context was used. e.g. "I expect so" was coded as a modulation of assertion. Whereas, "He expects her to cry" would be coded as other mental state term.

Table 2.5

The non-mental state coding categories with examples.

Category	Example	Notes
Simple description	"She's getting washed" "They is just shearing the sheep" "Mummy giving milk to her baby"	These comments added nothing more to the picture.
Physical state term	"The baby's crying" "He's got the giggles" "She is laughing at him"	These were utterances about physical states. These were not referred to as mental state words because they did not refer to internal experiences. Although some words such as "cry" and "laugh" have strong emotion links they were coded as physical because they described physical manifestations whereas an emotion term such as "happy" also refer to an internal experience.
Causal utterance	"Why is the cat sitting on his head?" "Why does it fall down?" "Because that girl messed it up."	These utterances occurred when mother and child talked or asked questions about causes.

Elaborations of Theme	"They could catch fish" [no fish present in picture]. "He's going to throw it at the baby" [nobody throwing anything in the picture]. "The daddy's saying 'no shoes on the chair'"	These occurred when mother and child expanded on the content of the picture, referring to something that was not actually presented in the picture.
Links to the child's life	"You have got one like this." "We used to play that" "It looks a bit like uncle ****"	This occurred when mother and child connected the pictorial content to their own life.
Factual utterance	"A baby sheep is called a lamb" "A tortoise lives on the land" "Babies have bibs"	These utterances were attempt to teach general principles that did not include causal information.
Orientating responses	"Mum look!" "Look closely." "Let's look at this one."	These were attempts to focus the others attention on a picture or element of a picture. Attempts to move onto the next picture was also coded as orientating responses.

Self-repetition		Utterances where children or parents repeated themselves
Reputation of others		Utterance where they repeat the other (mother repeats child or child repeats mother).
Unrelated utterances	"That's the postman" "Careful of mummy's tea" "No don't press that darling"	These were utterance that took place during the recording but were off task. These often occurred when child got distracted
Prompt questions	"What do you mean?" "What is this?" "Anything else?"	This occurred where questions were asked or prompts were given to encourage the other to add detail to what they were speaking about.
Short responses with no elaboration	"Yeah" "No"	These occurred when a speaker answered a question and they answered briefly with no elaboration.
Nonsensical	"Waw woo" "Rarrrr!"	This was an utterance that was gobbledegook or did not make sense.

2.8 Cohort 1 procedure

Children in Cohort 1 were tested individually in the school setting in the summer term of their preschool year (Time 1), again a year later (*M* = 366.31 days later, *SD* = 4.75) in the summer term of their Reception year (Time 2), and for a final time a further 10 months later in the spring or summer term of Year 1 (*M* = 309.61 days later or 10 months 5 days, *SD* = 12.95) (Time 3). At each of the three time points they took part in two sessions of approximately 20 minutes each. The two sessions were administered within two weeks of each other at Time 1 and within 4 days of each other at the latter two points. At Time 1 the mean time difference between the two sessions was 4.42 days (*SD* = 2.62), at Time 2 it was 2.06 days (*SD* = 1.38), and at Time 3 it was 1.14 days (*SD* = 0.97).

At all time points measures were divided between the two sessions and to avoid order effects the order was counter-balanced both within the session itself, and in the order of the session in which the child took part in first. At each time point there were eight orders in which the child could take part. The sessions took place in a quiet area either within the main classroom or just outside it. The child and the researcher sat at a table facing one another and the tasks were presented to the child on the table.

At each time point, after the child had taken part in their first session the questionnaire was sent home to parents by the class teacher. Parents were asked to either complete this online via Qualtrics or return the hard copy to school. At Time 1, 58 returned the questionnaire (39% of Time 1 sample), at Time 2, 46 returned the questionnaire (40% of Time 2 sample), and at Time 3, 44 returned the questionnaire (41% of Time 3 sample). The last page of the questionnaire gave details of the book sharing activity. Parents were asked to consent to take part in the activity and asked to give their email address so that the researcher could contact them. The researcher emailed them with instructions for the activity for both Apple and Android phones (see Appendix 20a and 20b), and the pictures for the activity (Appendix 19a and 19b). After they had made their recordings parents were asked to return them to the researcher via email. At Time 1, of the 58 who returned the questionnaire eight of them also returned recordings, at Time 2 three completed the book sharing activity.

2.9 Cohort 2 procedure

Due to low numbers on the questionnaire and the book sharing activity for Cohort 1, Cohort 2 was recruited. Children in Cohort 2 were tested in the autumn or spring term of their preschool year (Time 1) and again a year later (*M* = 365.02 days later, *SD* = 10.66) in the autumn or spring term of their Reception year (Time 2). This cohort took part in the same measures as Cohort 1 but they were administered in their home. This occurred in a quiet room in their house with the child and researcher sitting on the floor or a table facing each other. Unlike Cohort 1, the children took part in all measures in one single session. The order of the measures within the session was counterbalanced in the same way. During this session parents also filled in the questionnaire, and children and parents took part in the book sharing activity together whereby the researcher recorded them talking about the pictures using a digital voice recorder. The only measure which was not counterbalanced was the book sharing activity as this was always conducted first to act as a “warm up” for the child, and to prevent the parent from guessing the true nature of the task after watching the administration of the ToM measures. The single session took approximately an hour and parents were present for the duration. Children from this cohort only participated in Times 1 and 2 but there are plans to follow these children to Time 3 outside of this thesis as part of a wider project.

2.10 Analysis

The primary data analytic strategy across the first four chapters was structural equation modelling (SEM) using AMOS Version 25 (Arbuckle, 2016). SEM is a multivariate statistical method which takes a confirmatory (hypothesis testing) approach to test the hypothesized interrelationships among variables (Morrison, Morrison, & McCutcheon, 2017). The advantage of SEM over regression is that it allows for many interrelationships to be tested simultaneously (Von der Embse, 2016). In essence, SEM tests the fit of a proposed model which includes relationships between observed variables and latent variables (Byrne, 2016). Observed variables are those that are directly measured whereas latent variables are those which are not directly observed but instead are inferred from other variables which are observed (Byrne, 2016). As listening comprehension and reading comprehension are complex skills with many sub-skills, the use of SEM is of great value here. Moreover, SEM allows for indirect and direct effects to be studied which could be masked if a multiple regression approach examining unique contribution was used (Kim & Pilcher, 2016, p.10).

Selecting appropriate indices to evaluate hypothesized models is pertinent when using SEM analysis. Here, a number of commonly accepted model fit indices were adopted with criteria suggested by Lei and Wi (2007) and Hu and Bentler (1999). These included; a non-significant chi-square statistic (χ^2) (p >.05), comparative fit index (CFI) of $\geq .95$, Tucker-Lewis Index (TLI) of $\geq .95$, and a standardized root mean square residual (RMSEA) of $\leq .06$. Excellent fitting models would reach or exceed all of these indices, good fitting models would reach or exceed most of these indices and approach the others. These were also the desirable cut off values used by Kim (2017) and other reading research using SEM (Catts et al., 2015; Foorman et al., 2015; Oakhill & Cain, 2012; Puglisi, Hulme, Hamilton, & Snowling, 2017).

The sample size required for SEM also needs to be considered. SEM analysis is sensitive to sample size but there is much dispute over the minimum number of participants needed (E. J. Wolf, Harrington, Clark, & Miller, 2013). For example, Boomsma (1982; 1985) suggests a minimum sample size of 100, whereas Nunnally and Bernstein (1967) advocate ten participants per variable. In order to stay mindful of this, throughout results chapters and models power analysis and analysis of sample size required was run. Due to a smaller sample population for Chapter 7 (under 50 participants) mediation analysis rather than SEM was planned for this chapter. Mediation looks at three variables to assess if one variable facilitates the relationship between the other two variables (Hayes, 2009).

2.11 Summary

To summarise, this thesis investigated the metacognitive skills, language skills and social environments that support and predict early LC and RC. To do so, it employed a longitudinal design whereby children's patterns of development in a number of skills was tracked across three time points from preschool into Year 1. Children were recruited in two cohorts; Cohort 1 who were recruited directly through primary schools with a preschool class, and Cohort 2 who were recruited directly via parents. The recruitment of these two cohorts resulted in a large representative sample. At the three time points children participated in sessions which included a number of standardised measures and cognitive tasks. In some cases their parents also participated in mental state talk measures. The key skills under investigation were: vocabulary, syntactic knowledge, listening comprehension, decoding, theory of mind, broad metacognition, working memory, inference making, comprehension monitoring and reading comprehension. Additionally, mental state talk was assessed in the home environment. Measures were chosen based on their age appropriateness, wide and standard use in the relevant literature, and their reliability. The following five chapters used the methods described here in order to pursue the five main aims.

3 A longitudinal replication of the Direct and Indirect Effects Model of Text Comprehension (DIET)

3.1 Introduction

Listening comprehension is the processing and understanding of language received orally (Gough & Tunmer, 1986). During listening comprehension (LC) lexical information is used to achieve sentence and discourse interpretation (Hoover & Gough, 1990). LC is a complex skill and research suggests that proficient LC requires many sub-components (Lervåg et al., 2018). Recent work by Kim (2017) proposes the Direct and Indirect Effects Model of Text Comprehension (DIER) which contains sub-components of LC arranged in a hierarchical model. Kim (2017) only provides evidence for this model cross-sectionally at the age of seven years and so the main aim of this chapter was to replicate the model longitudinally for younger children.

3.1.1 Sub-components of listening comprehension

Research has proposed many sub-components of LC but the skills most often cited are: vocabulary, syntactic knowledge and working memory. Vocabulary is important for LC because most individual words of a sentence must be understood before meaning can be taken from the whole sentence. Owing to this, the relationship between preschool vocabulary and concurrent LC is strong (Florit et al., 2011; Florit & Levorato, 2012; Lepola et al., 2012), and longitudinal studies show that earlier vocabulary predicts later LC (Florit et al., 2009; Kendeou et al., 2008; Sénéchal et al., 2008).

Syntactic knowledge is important for LC because the order of words and grammatical rules can change the meaning of a sentence, and so understanding of these is essential for the correct meaning to be achieved. Indeed, research shows that for children aged six, syntactic knowledge directly predicts concurrent LC after controlling for vocabulary and working memory (Kim, 2015), and Potocki, Ecalle and Magnan (2013) found that in five year olds syntactic knowledge explained 3% of unique variance in LC. Longitudinal work has found that syntax at the age of five predicts LC aged eight (Alonzo et al., 2016).

Working memory assists LC as information must be held while attention is given to making inferences about what has been heard, and information is connected and integrated (Florit et al., 2009), and so research shows a strong relationship between LC and working memory. Working memory at the age of four and five years old has been shown to predict concurrent LC over and above vocabulary and verbal IQ (Florit et al., 2009; Florit, Roch, & Levorato, 2013). Longitudinal work has found that working memory at the age of five predicts LC aged eight (Alonzo et al., 2016). Another skill recently suggested as a sub-skill of LC is theory of mind.

3.1.2 Theory of mind and listening comprehension

There has been limited research exploring the role of theory of mind (ToM) in LC. ToM may be important for LC as its social component could aid the listener in understanding the viewpoint, desires and intentions of the speaker and could also lead to better awareness of social information and details within the spoken passage. In support of this, research in the early 2000s showed that children who were better at making sense of a speakers' meaning and intentions (as measured using referential communication games) also performed better on false belief tasks (Astington, 2004; Nilsen & Fecica, 2011; Resches & Pereira, 2007). These referential communication games traditionally involve two identical sets of objects, one for the child and one for the researcher, and the researcher asks the child to identify specific objects e.g. "*find the small brick*" (Lloyd, Boada, & Forns, 1992). To perform well the child must identify the correct referent.

In a review of the sub-skills needed for LC, Kim and Pilcher (2016) hypothesize that ToM may be involved in cross-checking the meaning taken from passages and filling in missing information. Kim and Pilcher (2016) use the situation model to explain this. They suggest that successful comprehension ultimately requires construction of the "situation model" or the "mental model" (Graesser, Singer, & Trabasso, 1994; Zwaan et al., 1995), that is a mental representation of what the passage is about (Kintsch, 1988). This mental representation may include information about characters, intentionality (or goals) and causation (Graesser et al., 1994). In a similar way Perner and colleagues describe the possession of a ToM as mental model building (Johnson-Laird, 1983; Perner, 1991) in that a belief is a mental model of the world. It has been suggested that children fail a false belief test because they do not understand that mental models or representation can differ from true reality (Lillard & Flavell, 1992; Perner, 1991). If ToM is concerned with the constructing and understanding of mental models, then this ability could assist with the creation of mental models during LC.

Despite this explanation, there have been a limited number of studies concerned with the role of ToM in LC. These studies are presented in Table 3.1. The first was conducted by Strasser and Rio (2014) who, against their hypothesis, found that in five year olds ToM (although correlating) did not make a significant contribution to LC over and above vocabulary and working memory. However, subsequent research has contradicted this, instead finding that ToM directly predicts LC for four to seven year olds (Kim, 2015; Kim, 2016; Kim, 2017; Pelletier & Beatty, 2015).

Table 3.1

Summary of pre-existing research into the role of theory of mind in listening comprehension

Author	N	Population	Age[a]	Design	ToM measure	LC measure	Analysis	Findings
Strasser & Rio (2014)	257	TD Spanish speakers	5;7	Cross-sectional	• Unexpected contents (Hogrefe et al., 1986) • Unexpected locations (Wimmer & Perner, 1983) • Belief desire reasoning (Harris et al., 1989) • Real–apparent emotion (Wellman & Liu, 2004) • Unexpected locations second- order (Perner & Wimmer, 1985)	• Recall of a wordless picture book read to them	Regression	ToM did not make a significant contribution to LC
Kim (2015)	148	TD Korean speakers	6;1	Cross-sectional	• First-order false belief (Gwon & Lee, 2012) • Second-order false belief (Caillies, Hody, & Calmus, 2012)	• Subset of OWLS[b] (Carrow-Woolfolk, 2011) • Subtest of CASL[c] (Carrow-Woolfolk, 1999)	Structural equation modelling	ToM directly predicted LC

Pelletier & Beatty (2015)	186	TD English speakers	4;6	Cross-sectional	• Two batteries consisting of a first order and a second order false belief tasks	• Fable comprehension. Read fables and asked recall questions	Regression	ToM directly predicted LC
Kim (2016)	201	TD Korean speakers	6;7	Cross-sectional al	• First-order false belief (Gwon & Lee, 2012) • Second-order false belief (Caillies et al., 2012)	• Story recall as devised by the author	Structural equation modelling	ToM directly predicted LC
Kim (2017)	350	TD English speakers	7;6	Cross-sectional	• Second-order false belief task (Kim & Phillips, 2014)	• Subset of OWLS[b] (Carrow-Woolfolk, 2011) • Subtest of TNL[d] (Gillam & Pearson, 2004) • An experimental expository comprehension task	Structural equation modelling	ToM directly predicted LC

[a] Mean age in years;months [b] OWLS = Oral and Written Language Scales [c] CASL = Comprehensive Assessment of Spoken Language [d] TNL = Test of Narrative Language

3.1.2.1 Kim's models of listening comprehension

As illustrated in Table 3.1, Kim has dominated research into ToM and LC. Using structural equation modelling, based on cross-sectional data, she proposes cross-sectional hierarchical models of LC, in which low-level skills predict high-level skills which then predict LC. Kim often refers to LC as text comprehension, stating that it should not be differentiated from reading comprehension (RC) as oral language comprehension and RC tap into the same processes (Kim, 2015, p. 102). Across three studies (Kim, 2015; Kim, 2016; Kim, 2017) this model has remained largely the same (Figures 3.1-3.3 show the three models). As it reflects current thinking the predominant focus of this chapter is the most recent model (Kim, 2017).

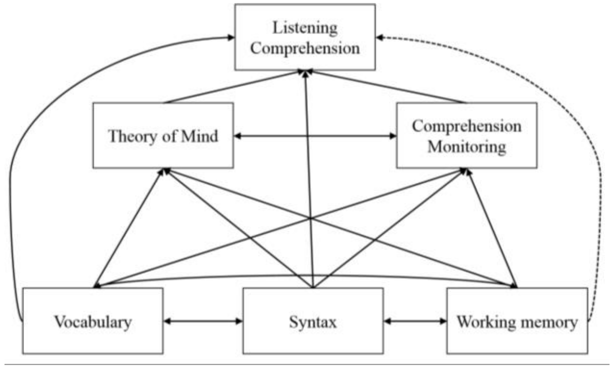


Figure 3.1: Best fitting model of listening comprehension as proposed by Kim (2015). *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

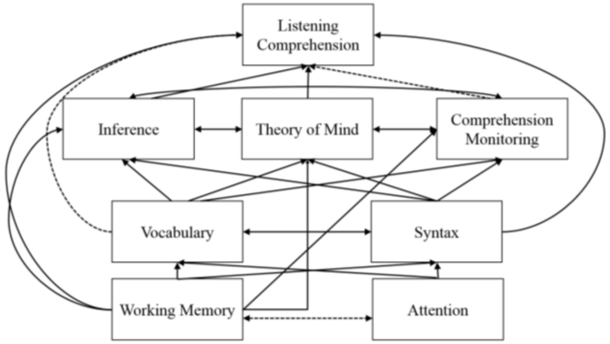


Figure 3.2: Best fitting model of listening comprehension as proposed by Kim (2016) . *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

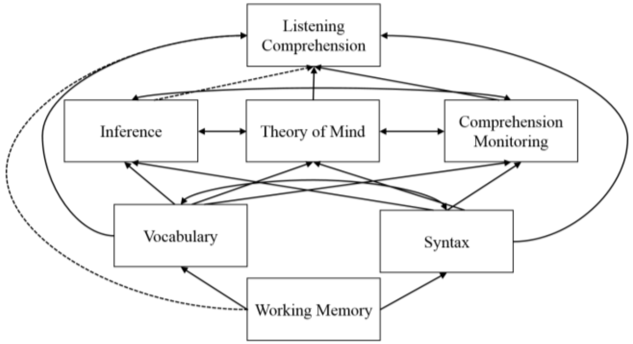


Figure 3.3: Best fitting model of listening comprehension as proposed by Kim (2017). *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

In addition to the sub-components of vocabulary, syntactic knowledge and working memory already discussed (in Section 3.1.1), Kim's models suggest a place for the high-order skills of comprehension monitoring and inference making. This is consistent with past research which shows that inference making is important for LC to establish global coherence (Kim & Pilcher, 2016). Indeed, for preschool children total number of inferences is related to LC (Tompkins et al., 2013), and longitudinal work shows that inference aged five predicts LC aged eight (Alonzo et al., 2016). Research also shows comprehension monitoring to be important for LC because in order to ensure that the correct meaning has been taken from the oral passage children may be required to monitor their comprehension (Kim & Phillips, 2014), for example, reflection strategies and playing back phrases and sentences in their mind (Carretti, Caldarola, Tencati, & Cornoldi, 2014). Concurrent comprehension monitoring has been found to correlate with LC for children aged five to eight years (Kim & Phillips, 2014), and longitudinal work shows that comprehension monitoring aged five predicts LC aged eight (Alonzo et al., 2016).

Theoretically, these models are based on a multi-level representation framework based on the situation model (Graesser et al., 1994) which states that successful comprehension is only achieved when an accurate, rich and elaborate mental picture of the situation portrayed within the oral passage is obtained (Kim, 2016; Kim, 2017; Kintsch, 1988). During LC a child must build a mental representation of the message in the oral discourse in order to take meaning from it (Kim, 2016; Kim, 2017; Paris & Stahl, 2005; Van Dijk et al., 1983; Zwaan, 2016). The framework suggested by Kim and Pilcher (2016) is based on a combination of the construction-integration model of Kintsch and colleagues (Kintsch, 1994; Van Dijk et al., 1983), the constructionist model (Graesser et al., 1994) and the landscape model (Van den Broek, Rapp, & Kendeou, 2005). The framework has three levels: surface code (where the listener extracts key words and phrases from the passage), text-base (where literal meaning is taken from the passage) and situation model (where the listener integrates this literal meaning with their prior knowledge to create a mental picture of the passage). The framework hypothesizes that different levels require different language and cognitive skills. Figure 3.4 conceptualises this and shows how foundational cognitive skills (e.g. working memory), foundational language skills (e.g. vocabulary and syntactic knowledge), and higher-order cognitive skills (inference making, ToM and comprehension monitoring) map onto the surface code, text-base and situation model. Kim and Pilcher (2016) suggest that comprehension monitoring is involved in evaluating initial local propositions, whilst inference making and ToM are involved in validating propositions and filling in missing information. Kim (2016) suggests that within a model ToM relates to inference making because both require some level of reasoning, but the author also suggests that ToM is independently related to LC, after controlling for inference making, because thoughts, beliefs and intentions of storytellers and characters are vital to plots of narrative and therefore the ability to think about one’s own and others’ thinking and mental status would be critical to the understanding of these elements within narratives.

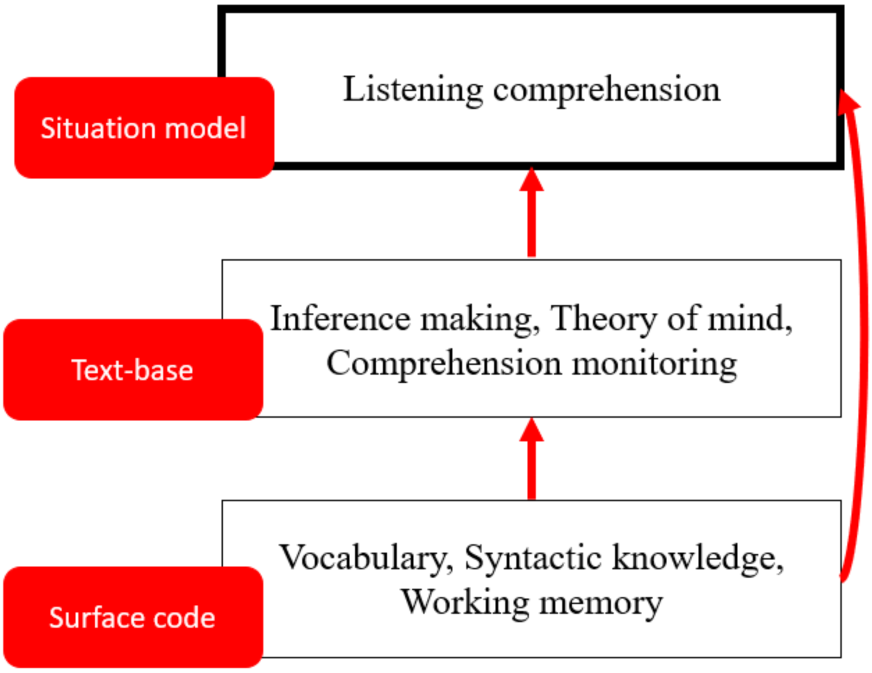


Figure 3.4: The theoretical framework for the DIET model (Kim & Pilcher, 2016; Kim, 2017).

3.1.3 Limitations of past studies

Research into ToM in LC has been, so far, cross-sectional only (see Table 3.1). Although Kim proposes a DIET model of LC (Figures 3.1-3.3), the concurrent and correlational nature of the studies means that the direction of relation in the model is based on theory. Kim acknowledges that future work should track the development of LC and its sub-components across time to validate the model (Kim, 2015, p.30). Kim also recognises that future research should assess if the model can be generalised across different developmental phases (i.e. younger and older children) and with children of different languages and orthographies beyond the Korean and US samples used by her studies (Kim, 2015, p.30; Kim, 2017, p.328). It is possible that the model will not be generalisable across developmental phases as it is recognised that skills such as RC and LC rely more heavily on certain sub-components at different developmental stages, e.g. the changing role of decoding and oral skills in RC (Storch & Whitehurst, 2002).

Historically there has been a lack of LC longitudinal work (Lepola et al., 2012) and this continues to be the case (see review in Table 3.1). Longitudinal studies have advantages over cross-sectional studies such as the ability to document the development trajectory of a specific skill, identify precursors of an ability, and examine how the relationship between two (or more) related skills interact and progress over time (Grammer, Coffman, Ornstein, & Morrison, 2013; Kendeou et al., 2008). Therefore, longitudinal studies can determine the direction and magnitude of causal relationships (Ebert, 2015). The applications of longitudinal work are of importance, because if earlier predictors of LC are known this has potential to shape the focus of the early years classroom and could form interventions for those with weaker oral language skills.

3.1.4 The present study

Recent research has begun to explore the contribution of ToM to LC and has proposed a model of concurrent LC which includes a direct role for ToM (Kim, 2015; Kim, 2016; Kim, 2017). However, no study to date has been longitudinal and so there is a need to validate this model for use longitudinally to determine the directionality of relationships, particularly to assess if preschool skills predict later

LC in the same way as suggested by Kim's DIET model. This thesis aimed to address this by carrying out a longitudinal study measuring LC and its sub-components across 22 months with children aged four years to six years old. This age group is important for longitudinal study because although children can, to a certain extent, comprehend oral passages before the age of three, it is not until this age (when they are in their preschool year) that LC and its sub-skills can be reliably measured (Carrow-Woolfolk, 2011; L. M. Dunn, Dunn, Whetton, & Burley, 1997). Therefore, unlike previous research, in this research the earliest pre-cursors of later LC could be determined.

This chapter sought to do three things:

1) Firstly, it aimed to validate the DIET model of LC cross-sectionally (Kim, 2017: see Figure 3.3.) with a UK sample of broadly similarly aged children (aged six years). This was of importance because only the one research group has explored models of LC that include a direct role for ToM, so validation is needed. Additionally, it was acknowledged by the author that future work is needed to extend the model to different types of sample i.e. a UK sample (Kim, 2015, p.30). The DIET model (Kim, 2017) was used for the validation rather than previous models as it represents most current thinking.
2) Secondly, this chapter aimed to validate the DIET model longitudinally[3] using skills at the age of four years to predict LC 12 months later (aged five). This aim was important and novel because there is a lack of longitudinal LC

[3] It should be noted that this model was not a direct longitudinal replication of the DIET model as it did not include inference making and comprehension monitoring. This is because it is not possible to measure these skills before Time 3, as to date, no reliable UK based measures exist for these skills for administration before the age of six. See Section 1.5.1 and 1.5.2 in Chapter 1 for more details.

research (Lepola et al., 2012, p.260) and none exploring the role of ToM (Kim, 2015, p.30). More specifically it was beneficial to explore this age group because; firstly the most early pre-cursors of later LC could be determined, and secondly LC is an important skill needed in the first year of primary school for children aged five (Hogan et al., 2014).

3) Lastly, this chapter aimed to assess whether the DIET model could predict LC even further across time (22 months later)[1], with skills at the age of four used to predict LC aged six. Longer term longitudinal studies have value because they can track change over time and because they are more valid for examining cause-and-effect relationships than cross-sectional study or shorter-term longitudinal study (Caruana, Roman, Hernandez-Sanchez, & Solli, 2015; Sontag, 1971).

Together these aims gave a deeper understanding of the influence ToM has to early LC development. Moreover, whilst addressing these three aims non-verbal ability (as measured at aged four) and age were controlled for, thus this research is more tightly controlled than that of Kim (2017). In developmental studies controlling for these variables allows for clarity and more certainty that the skills under direct investigation are the cause of the outcome, and that it is not driven by other highly correlated factors e.g. age, non-verbal ability or gender. Age and non-verbal ability are commonly controlled for in reading and social cognition studies (e.g. Atkinson et al., 2017; Cain et al., 2004; Devine et al., 2016; Ricketts et al., 2013) because of the strong relationship they have to cognitive and language skills.

It was hypothesized that:

1) The DIET model would be validated in a UK sample of children aged six (Time 3; Model 1). Specifically, after controlling for non-verbal ability and age, and other low-level skills in the model, ToM would make a direct significant contribution to LC.
2) The DIET model would hold longitudinally across 12 months, with skills at the age of four (Time 1) predicting LC aged five (Time 2; Model 2). In particular, after controlling for non-verbal ability and age, and other low-level skills in the model, ToM aged four would make a significant direct contribution to LC aged five.
3) The DIET model would hold even further across time, with skills at the age of four (Time 1) predicting LC 22 months later when children were six (Time 3; Model 3). Specifically, after controlling for non-verbal ability and age, and other low-level skills in the model, ToM aged four would make a significant direct contribution to LC aged six.

3.2 Method

3.2.1 Participants

The three models tested here included different sets of participants because Cohort 2 were not followed to Time 3 (see method chapter for explanation of this) and could therefore not appear in analysis for models predicting Time 3 LC. For more information on the participants including recruitment see Section 2.3 in Chapter 2.

Model 1 which was based on the DIET model from Kim (2017), used skills measured at Time 3 (when children were six) and so only included participants from Cohort 1 (n = 107). Model 2 which predicted LC at the age of five (Time 2) using skills aged four (Time 1), included both cohorts (n = 162). Model 3 which predicted aged six LC (Time 3) using aged four skills (Time 1), only included participants from Cohort 1 (n = 107). Descriptive statistics of the participants included in each of the models is shown in Table 3.2.

Table 3.2

Descriptive statistics of participants in each of the three models

	N	N males	N females	Mean age[a]
Model 1	107	54	53	Time 3: 6;1 (3.65)
Model 2	162	82	80	Time 1: 4;1 (4.55) Time 2: 5;1 (4.43)
Model 3	107	54	53	Time 1: 4;3 (3.64) Time 3: 6;1 (3.65)

Note. [a] Mean age in Years; Months

3.2.2 Materials and measures

Table 3.3 shows the measures administered at each time point. For comprehensive details on each of these measures refer to Section 2.5 in Chapter 2.

3.2.3 Procedure

Participants in Cohort 1 were administered the measures in Table 3.3 across the three time points within two 20 minute sessions in a quiet place in or just outside their school classroom. Participants in Cohort 2 were administered the measures in Table 3.3 at the first two time points in a single session in a quiet place in their own home. For complete procedural details refer to Chapter 2 Sections 2.7 and 2.8.

Table 3.3

Measures administered to children at each of the three time points

Skill	Measure	Time 1	Time 2	Time 3
Listening comprehension	NARA	-	✓	✓
	OWLS	-	✓	✓
Vocabulary	BPVS-III	✓	-	✓
Syntactic Knowledge	Sentence Structure (CELF-Preschool 2uk)	✓	-	-
	Sentence Structure (CELF-4)	-	-	✓
Theory of mind	Unexpected contents task	✓	-	-
	Unexpected locations task	✓	-	-
	Strange Stories	-	-	✓
Working memory	Reverse word span task	✓	-	✓
Comprehension monitoring	Comprehension monitoring stories	-	-	✓
Inference making	Inference oral story	-	-	✓
Non-verbal ability	Block design subset of WPPSI-II	✓	-	-

Note: Many of these measures were administered at Time 2 or other time points but only the specific measures used in the analysis for this chapter are listed here. For further details see Chapter 2.

3.2.4 Analysis

The primary data analytic strategy was structural equation modelling (SEM) using AMOS Version 25 (Arbuckle, 2016). Details on the use of SEM and justification of its use can be found in Section 2.10 of Chapter 2. For each of the three models tested, first descriptive statistics were computed and then initial correlation and regression analysis were carried out. The fit of the hypothesized models was then assessed using SEM. During the SEM analysis, a latent variable was created for LC using two measures (NARA and OWLS). All other language and cognitive skills were assessed by a single measure for each construct and so therefore observed variables were used.

3.3 Results

Results for the three models tested are outlined separately. Model 1 was a cross-sectional model using data from Time 3 only (aged six). Model 2 was a longitudinal model with Time 1 (aged four) skills predicting Time 2 LC (aged five). Model 3 was a longitudinal model with Time 1 (aged four) skills predicting Time 3 LC (aged six).

3.3.1 Model 1

Model 1 aimed to validate the cross-sectional DIET model of Kim (2017) shown in Figure 3.3 but controlling for concurrent age and non-verbal ability (as measured at Time 1) and in a slightly younger sample (on average 17 months younger) from the UK.

3.3.1.1 Descriptive statistics and preliminary analysis

Table 3.4 shows descriptive statistics for the variables included in Model 1.

Table 3.4

Descriptive statistics for variables included in Model 1[a]

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Listening comprehension	NARA	106	44	1-16	6.58	3.38	.69	-.31
	OWLS	107	130	36-94	63.66	12.32	.38	-55
Vocabulary	BPVS-II	107	168	48-130	87.03	13.28	.14	1.29
Syntactic knowledge	CELF	107	26	7-26	20.24	3.74	-.99	1.06
Theory of mind	Strange Stories	106	10	0-8	2.33	1.57	.92	1.43
Inference making	Inference stories	106	8	1-8	4.21	1.89	-.20	-.53
Comprehension monitoring	Stories	107	5	1-5	2.67	.98	.09	-.37
Working Memory	Reverse word span	107	11	0-11	6.69	3.78	-.25	-1.42
Non-verbal ability	Block design	107	40	0-32	15.34	6.50	-.12	-.13
Time 3 age[b]		107		66-80	72.95	3.68	-.11	-1.10

[a] With the exception of non-verbal ability (which was administered at Time 1, when children were aged four) all measures were administered at Time 3 [b] Age in months

For each of these variables univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995), Kline (2005) and Weston and Gore Jr (2006) (skewness < +3/-3 and kurtosis < 10/-10). These skewness and kurtosis ranges have also been used by similar SEM work into LC (e.g. Lepola et al., 2012). Skewness and kurtosis values are shown in Table 3.4 and are all well within these proposed values, they are also within the more traditional and stringent ranges of: skewness < +1/-1 and kurtosis < +2/-2 (Gravetter & Wallnau, 2010). Multivariate normality assumptions were also checked and were not violated as confirmed by a Mardia's coefficient of multivariate critical ratio of -.57 (a value < 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). Correlations between measures are displayed in Table 3.5. All variables were weakly to moderately related to the two measures of LC.

Table 3.5

Correlation matrix of measures included in Model 1

	1	2	3	4	5	6	7	8	9	10
1. Time 3 NARA	-									
2. Time 3 OWLS	.62**	-								
3. Time 3 BPVS	.54**	.50**	-							
4. Time 3 CELF	.58**	.37**	.49**	-						
5. Time 3 Strange Stories	.38**	.32**	.35**	.25*	-					
6. Time 3 inference stories	.45**	.51**	.52**	.49**	.29**	-				
7. Time comprehension monitoring stories	.21*	.23*	.19*	.21*	.16	.10	-			
8. Time 3 reverse word	.36**	.33**	.39**	.32**	.30**	.31**	.10	-		
9. Time 1 block design	.46**	.32**	.45**	.47**	.28**	.33**	.18	.28**	-	
10. Time 3 age	.22*	.13	.13	.14	.09	.11	.18	-.01	.41**	-

Note. ** = p < .001, * = p < .05

Preliminary regression analyses were then performed to assess whether the variables in the proposed model had predictive power in calculating LC and therefore whether it was logical to proceed with the model. Separate linear regressions were carried out for each individual variable in the model (working memory, vocabulary, syntactic knowledge, inference making, ToM and comprehension monitoring) for its ability to predict each of the two outcomes measures of LC (NARA and OWLS). All variables except comprehension monitoring significantly predicted NARA scores ($p < .001$) explaining between 13.7% and 33.6 % of unique variance, whereas comprehension monitoring significantly predicted NARA scores ($p < .05$) explaining 5% of unique variance. Most of the variables also significantly predicted OWLS scores ($p < .001$) explaining between 10.9% and 26 % of unique variance, and ToM significantly explained 8% and comprehension monitoring 5% of unique variance (both $p < .05$). Given this full modelling with LC as an outcome measure was carried out.

3.3.1.2 SEM analysis

Prior to SEM analysis, missing data (see Table 3.4 for details of which measures) was imputed using expectation maximization (EM) in SPSS. This method of data imputation was used because missing data was minimal (three cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage for data sets with less than 250 participants. It is also the method used by other similar work within the field (Cain & Chiu, 2018; Catts et al., 2015; Foorman et al., 2015; L. Hamilton, Hayiou-Thomas, Hulme, & Snowling, 2016; Puglisi et al., 2017). Furthermore, Little’s MCAR test reported the data to be missing at random ($p =.25$). A model

(Figure 3.5) based on the DIET model was then fitted to the data for LC using AMOS. The model had a good fit as the desirable indices were either met or approached, χ^2 (10) = 16.50, *p* = .09; CFI = .98, TLI = .89, and RMSEA = .08. Seven multivariate outliers were identified using Mahalanobis d-squared (those with a p < .05). When these seven participants were removed and the model re-run, the model fit indices changed only slightly to χ^2 (10) = 16.72, *p* = .10; CFI = .98, TLI = .89, and RMSEA = .08. Therefore, the model with all 107 participants included is presented as the final model.

Standardised path coefficients of the model are shown in Figure 3.6. In Figure 3.6, and in all subsequent models, observed variables are depicted with a square and latent variables are depicted with a circle. For this first model, see Appendix 21 for the raw model including the standardised path coefficients, the measurement model and the error variables. As presented in Figure 3.6, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .28, p =.001), grammatical knowledge (β = .20, p = .03) but not to LC (β = .14, p = .21). Grammatical knowledge was not significantly related to ToM (β = .05, p = .66), or comprehension monitoring (β = .13, p = .23), but was to inference making (β = .30, p = .002). Vocabulary was significantly related to ToM (β = .27, p = .01) and inference (β = .36, p <.001), but not to comprehension monitoring (β = .11, p = .35). Both vocabulary (β = .33, p = .01) and grammatical knowledge (β = .35, p =. 01) were directly related to LC. ToM was significantly independently related to LC (β = .19, p = .04) but inference (β = .23, p = .09) and comprehension monitoring were not (β = .11, p = .31). A total of 40% of variance in Time 3 LC was explained by the concurrent skills in the model.

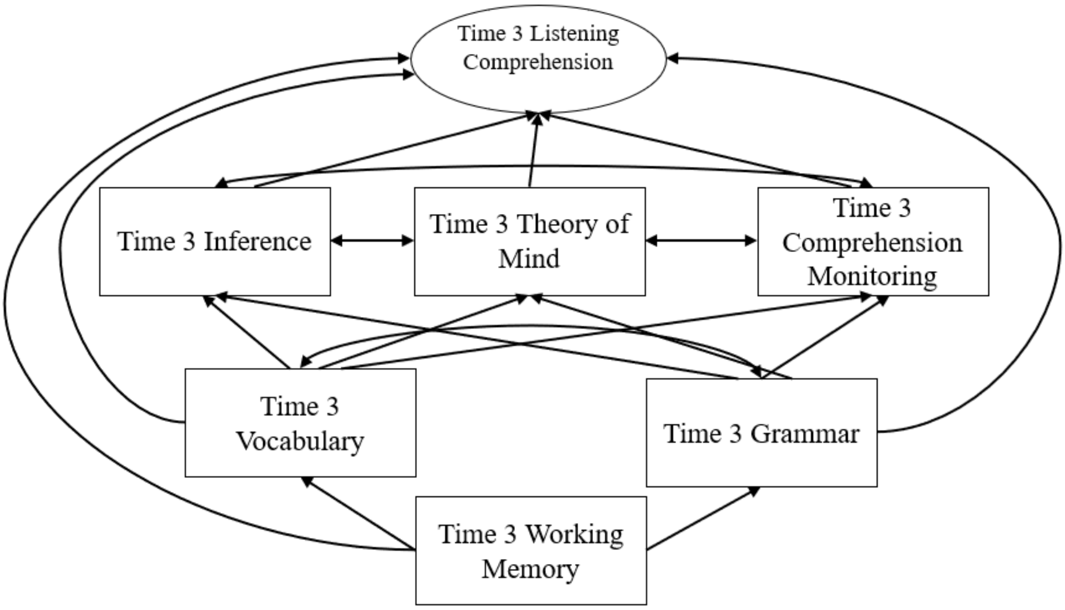


Figure 3.5: Hypothesized Model 1 based on the DIET model (Kim, 2017) after controlling for age and earlier non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

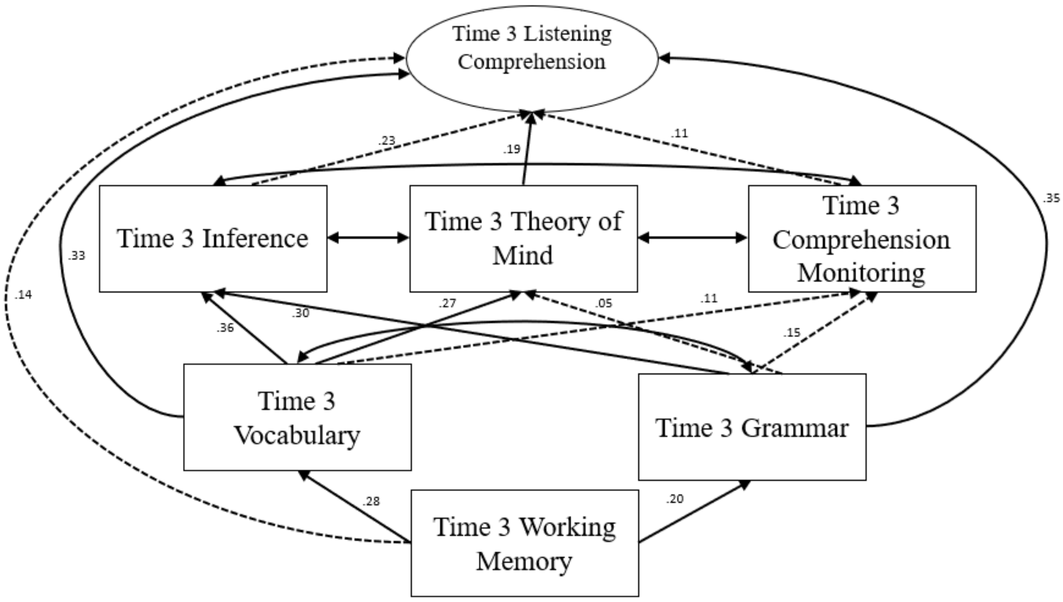


Figure 3.6: Model 1 with standardised path coefficient weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

3.3.2 Model 2

Model 2 hypothesized that the DIET model could be used longitudinally to predict LC across a year from the age of four (Time 1) to the age of five (Time 2)[4].

3.3.2.1 Descriptive statistics and preliminary analysis

First, a ToM composite consisting of the two ToM measures was computed by summing scores from the ToM measures from Time 1 (unexpected contents and unexpected locations). The composite was used to give a more sensitive measure of ToM and was justified as the individual measures correlated significantly ($r = .29$, $p < .001$). This composite was used in all further analysis. Table 3.6 shows descriptive statistics for the variables used in Model 2. All variables in the model met the univariate normality assumptions for SEM as assessed by skewness and kurtosis values (as stated by Chou & Bentler, 1995; Kline, 2005). Skewness and kurtosis values are also shown in Table 3.6. Multivariate normality assumptions were also checked and were not violated as demonstrated by a Mardia's coefficient of multivariate critical ratio of .01. Correlations between measures are displayed in Table 3.7. All variables were weakly to moderately related to the two measures of LC.

[4] Note that in contrast to the DIET model and Model 1, inference making and comprehension monitoring were not included in Model 2 as these were not measured until Time 3 (see footnote 1 in Section 3.1.3 for a justification of this).

Table 3.6

Descriptive statistics for variables included in Model 2 measured at Time 1 (aged four) and Time 2 (aged five)

Skill	**Measure**	**N**	**Max**	**Range**	**Mean**	**SD**	**Skewness**	**Kurtosis**
Time 1								
Vocabulary	BPVS-II	162	168	10-102	54.61	17.15	-.14	-.49
Syntactic knowledge	CELF-Preschool	161	22	0-21	12.14	4.33	-.69	.02
Theory of mind	ToM composite	158	5	0-5	2.03	1.72	.25	-1.04
Working memory	Reverse word span	160	9	0-9	1.07	2.21	1.95	2.57
Non-verbal ability	Block design	162	40	0-32	16.56	5.86	-.47	.40
Time 2								
Listening comprehension	NARA	161	44	0-13	3.86	3.62	.57	-.79
	OWLS	162	130	24-98	59.98	15.06	.49	.26
Age [a]	-	162	-	50-69	61.18	4.43	-.17	-.69

[a]Age in months

Table 3.7

Correlation matrix of variables used in Model 2

	1	2	3	4	5	6	7	8
1. Time 2 NARA	-							
2. Time 2 OWLS	.56**	-						
3. Time 1 BPVS	.35**	.38**	-					
4. Time 1 CELF	.38**	.37**	.64**	-				
5. Time 1 ToM composite	.27**	.30**	.39**	.48**	-			
6. Time 1 Reverse word	.20*	.21**	.42**	.40**	.38**	-		
7. Time 1 Block design	.17*	.29**	.61**	.48**	.30**	.37**	-	
8. Time 2 Age	.01	.06	.09	.15	.15	.19*	.10	-

Note. ** = $p < .001$, * = $p < .05$

Preliminary regression analyses were then performed to assess whether the variables in the proposed model had predictive power in calculating Time 2 LC and therefore whether it was logical to proceed with the model. Separate linear regressions were carried out for each individual variable in the model (vocabulary, working memory, syntactic knowledge and ToM) for its ability to predict each of the two outcome measures of LC (NARA and OWLS). All variables significantly predicted NARA scores ($p < .05$) explaining between 6% and 14.7% of unique variance. All of the variables also significantly predicted OWLS scores ($p < .05$) explaining between 4.5% and 14.1 % of unique variance. Given this full modelling using these variables and LC as an outcome measure was carried out.

3.3.2.2 SEM analysis

As with Model 1 prior to SEM analysis missing data (See Table 3.6 for details of which measures) was imputed using EM, again this method of data imputation was used because missing data was minimal (eight cases) and Little's MCAR test reported the data to be missing at random ($p =.34$). A longitudinal model (Figure 3.7) of Time 1 skills was fitted to the data for Time 2 LC. The model had a good to excellent fit, $\chi^2 (6) = 10.42$, $p = .11$; CFI = .99, TLI = .94, and RMSEA = .07. Seven multivariate outliers were identified using Mahalanobis d-squared (those with a $p < .05$), when these seven participants were removed and the model re-run, the model fit indices changed only slightly to $\chi^2 (6) = 8.38$, $p = .21$; CFI = .99, TLI = .96, and RMSEA = .05. Therefore, the model with all 162 participants included is presented as the final model.

Standardised path coefficients of the model are shown in Figure 3.8. As presented in Figures 3.8, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .22, $p < .001$) and grammatical knowledge (β = .25, $p < .001$), but not directly to later LC (β = .02, p = .90). Vocabulary was not significantly related to ToM (β = .12, p = .24). Grammatical knowledge was related to ToM (β = .40, $p < .001$). Vocabulary (β = .47, p = .03) and grammatical knowledge (β = .48, p = .02) were also directly related to later LC. ToM was significantly independently related to LC (β = .30, p = .04). A total of 47% of variance in Time 2 LC was explained by the Time 1 skills in the model.

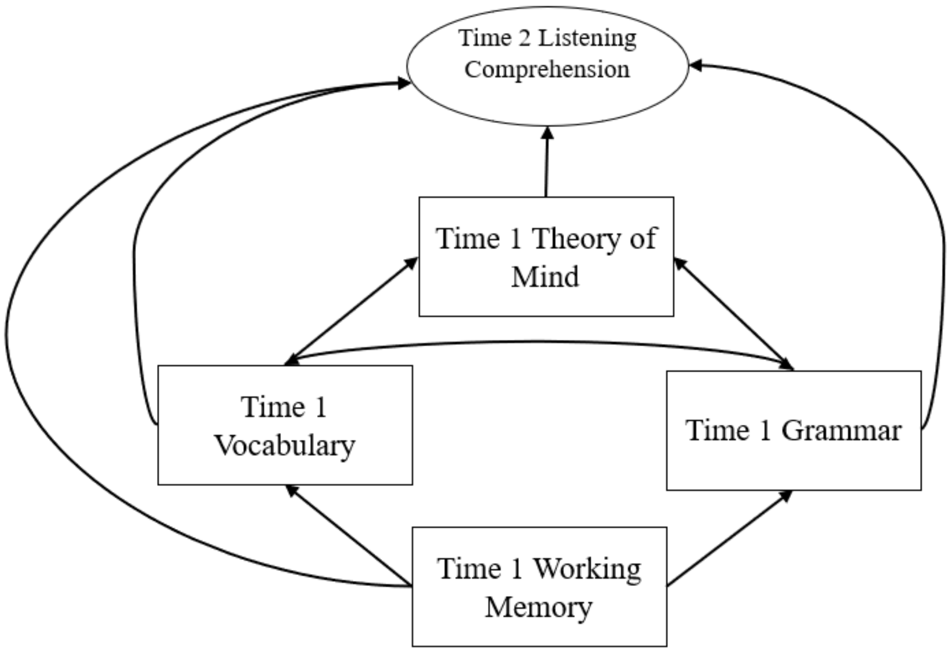


Figure 3.7: Hypothesized Model 2 after controlling for age and earlier non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

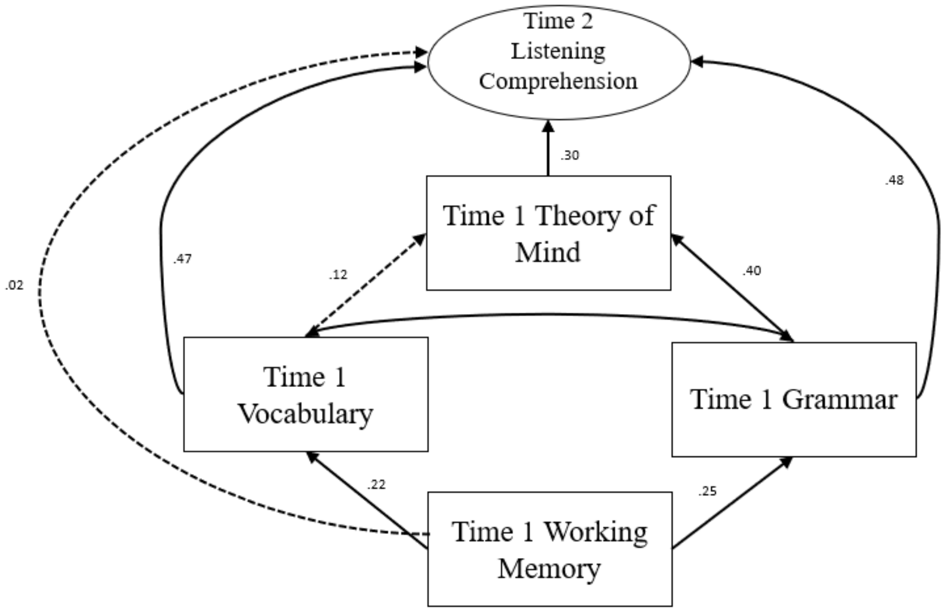


Figure 3.8: Model 2 with standardised path coefficient weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

3.3.3 Model 3

Model 3 proposed that the DIET model would predict LC over 22 months from the age of four (Time 1) to the age of six (Time 3)[5].

3.3.3.1 Descriptive statistics and preliminary analysis

Table 3.8 shows descriptive statistics for the variables used in Model 3. As with the previous model a composite of the two ToM measures was used. The individual ToM measures correlated significantly ($r = .24$, $p < .05$). All variables in the model met the univariate normality assumptions for SEM as assessed by skewness and

[5] Note that as with Model 2 inference making and comprehension monitoring were not included in this model as these were not measured until Time 3 (see Footnote 1 in Section 3.1.3 for a justification of this).

kurtosis (values stated by Chou & Bentler, 1995; Kline, 2005). Multivariate normality assumptions were also checked and were not violated as demonstrated by a Mardia's coefficient of multivariate critical ratio of -.43. Correlations between measures are displayed in Table 3.9. All variables were weakly to moderately related to the two measures of LC.

Table 3.8

Descriptive statistics for variables included in Model 3 measured at Time 1 (aged four) and Time 3 (aged six)

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 1								
Vocabulary	BPVS-II	107	168	10-102	51.59	18.12	.13	-.43
Syntactic knowledge	CELF-Preschool	107	22	0-21	11.64	4.61	-.55	-.29
Theory of mind	ToM composite	104	5	0-5	1.94	1.70	.38	-1.02
Working memory	Reverse word span	106	9	0-8	1.00	2.12	2.01	2.71
Non-verbal ability	Block design	107	40	0-32	15.34	6.50	-.12	-.13
Time 3								
Listening comprehension	NARA	106	44	1-16	6.58	3.38	.69	-.30
	OWLS	107	130	36-94	63.66	12.32	.38	-.55
Age[a]	-	107	-	66-88	72.95	3.68	-.11	-1.10

[a] Age in months

Table 3.9

Correlation matrix of variables used in Model 3

	1	2	3	4	5	6	7	8
1. Time 3 NARA	-							
2. Time 3 OWLS	.62**	-						
3. Time 1 BPVS	.51**	.39**	-					
4. Time 1 CELF	.41**	.47**	.67**	-				
5. Time 1 ToM composite	.36**	.37**	.46**	.54**	-			
6. Time 1 Reverse word	.49**	.35**	.51**	.41**	.27**	-		
7. Time 1 Block design	.46**	.32**	.62**	.50**	.34**	.39**	-	
8. Time 3 Age	.22*	.13	.36**	.29**	.10	.29**	.41**	-

Note. ** = p < .001, * = p < .05

Preliminary regression analyses were then performed to assess whether the variables in the proposed model had predictive power in calculating Time 3 LC and therefore whether it was logical to proceed with the model. Separate linear regressions were performed for each individual variable in the model (vocabulary, working memory, syntactic knowledge and ToM) for its ability to predict each of the two outcome measures of LC (NARA and OWLS). All variables significantly predicted NARA scores ($p < .01$) explaining between 10.9% and 25.6% of unique variance. All of the variables also significantly predicted OWLS scores ($p < .01$) explaining between 11.4% and 23.7 % of unique variance. Given this, full modelling using these variables and LC as an outcome measure was carried out.

3.3.3.2 SEM analysis

As with the previous models prior to SEM analysis missing data (See Table 3.8 for details of which measures) was imputed using EM. Again, this method of data imputation was used because missing data was minimal (five cases) and Little's MCAR test reported the data to be missing at random ($p =.28$). A longitudinal model (Figure 3.9) of Time 1 skills was fitted to the data for Time 3 LC. The model had an excellent fit, $\chi^2 (6) = 8.51$, $p = .20$; CFI = .99, TLI = .96, and RMSEA = .06. Six multivariate outliers were identified using Mahalanobis d-squared (those with a $p < .05$), when these six participants were removed and the model re-run, the model fit indices changed only slightly to $\chi^2 (6) = 6.95$, $p = .33$; CFI = 1.00, TLI = .98, and RMSEA = .04. Therefore, the model with all 107 participants included is presented as the final model.

Standardised path coefficients of the model are shown in Figure 3.10. As presented in Figures 3.10, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .30, $p < .001$) and grammatical knowledge (β = .24, $p < .01$) as well as directly to later LC (β = .45, p = .003). Vocabulary was not significantly related to ToM (β = .17, p = .16). Grammatical knowledge was related to ToM (β = .43, $p < .001$). Neither vocabulary (β = .25, p = .20) or grammatical knowledge (β = .14, p = .48) were directly related to LC. Finally, ToM was not significantly independently related to LC (β = .25, p = .11). A total of 30% of variance in Time 3 LC was explained by the Time 1 skills in the model.

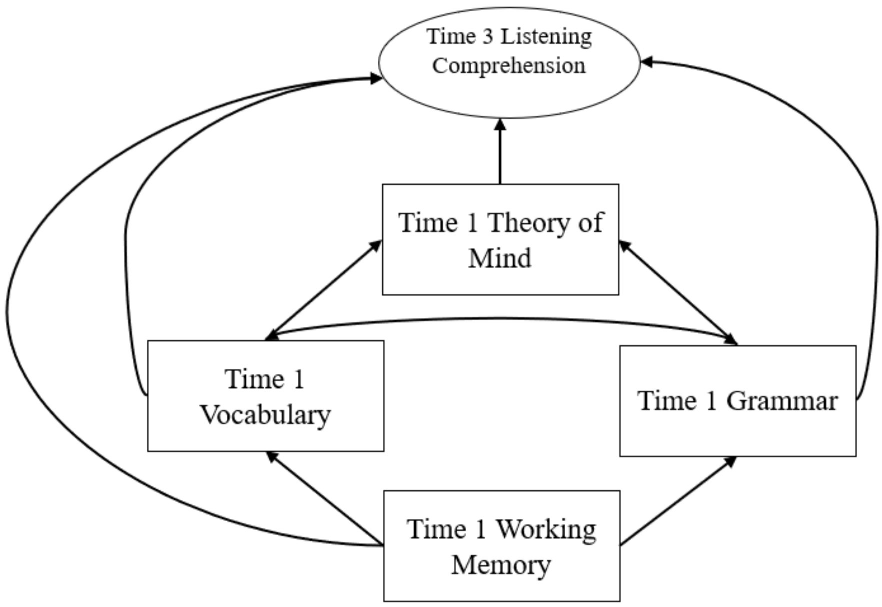


Figure 3.9: Hypothesized Model 3 after controlling for age and non-verbal ability.
Note. Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

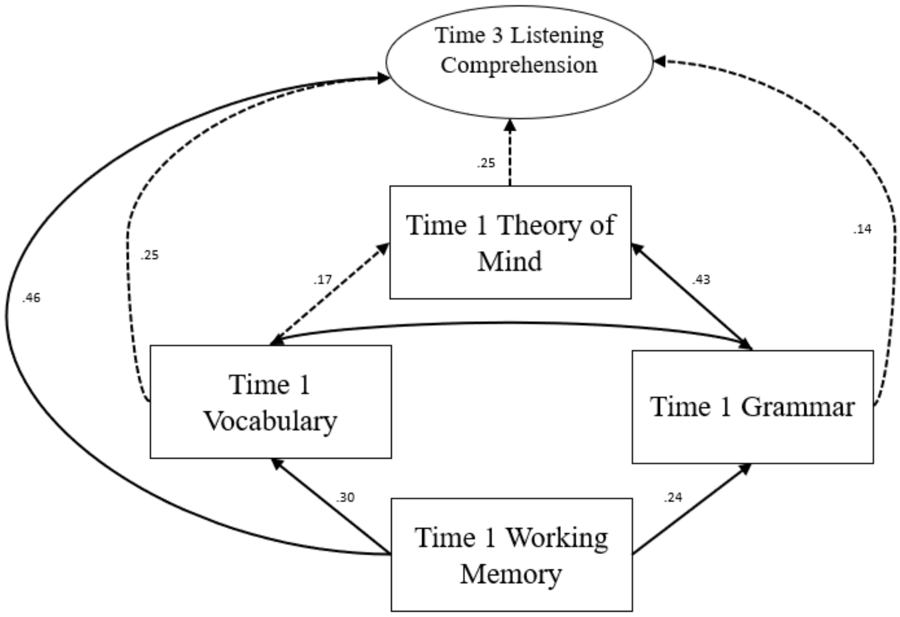


Figure 3.10: Model 3 with standardised path coefficient weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

3.3.4 Further analysis

3.3.4.1 Post-hoc statistical power analysis

Due to potential issues with sample size, power analysis of each model was calculated using an online power calculator (Soper, 2017). This analysis showed each of the three models to have sufficient power; observed statistical power = .99 for Model 1, .999 for Model 2 and .995 Model 3. SEM models are considered to have sufficient power if the observed statistical power reaches .80 (J. Cohen, 1988; P. Cohen, West, & Aiken, 2014). Furthermore, a calculation of sample size required, computed using an online calculator (Soper, 2015), suggested that a minimum of 87 participants would be needed for each of the three models in order to detect effects (actual sample sizes: Model 1 = 107, Model 2 = 162 and Model 3 = 107).

3.3.4.2 Comprehension monitoring and inference in Model 3

Neither Model 2 nor Model 3 included comprehension monitoring or inference making so they were not full longitudinal replications of Kim (2017), or indeed of Model 1. Model 2 and Model 3 did not include these skills as they were not measured until Time 3 because it is not possible to measure before this age as to date no reliable UK based measures exist for administration before the age of six. Given that Model 3 was predicting Time 3 LC and that comprehension monitoring and inference making were measured at Time 3, to assess if the inclusion of comprehension monitoring and inference making had an influence on Model 3, these skills as measured at Time 3 were included and the model re-run. When this was done the fit of Model 3 remained very similar: χ^2 (11) = 13.63, p = .25; CFI = .99, TLI = .97, and RMSEA = .05. In this new model comprehension monitoring was not significantly related to LC (β = .14, p=.21) but inference making was (β = .51, p<.001). In this new model ToM was still not significantly related to LC (β = .22, p=.08) but the p value was closer to approaching significance.

3.3.4.3 Children with English as an additional language

It is important to note that a proportion of the sample had English as a second language (30% of the sample for Models 1 and 3 and 22% for Model 2) which could have affected results. For example, children with English as an additional language (EAL) are known to have weaker oral language skills, particularly vocabulary skills, in comparison to their monolingual peers (Burgoyne et al., 2009) (See Section 1.10 in Chapter 1 for a more detailed description of the profiles of EAL children). This could have meant that, in the absence of a proficient vocabulary, the EAL children in the sample relied more heavily on other skills (e.g. working memory or even ToM) to assist with their LC, and thus the DIET model may not have fitted the sample in the same way it did with the monolingual children. Therefore, multigroup analysis was used in AMOS to compare the fit of all models for EAL participants to English only speaking participants. This was done using a chi-square difference test whereby a non-significant chi-square shows that the model fit is the same for both groups (Dimitrov, 2006). For each of the three models a non-significant chi-square demonstrated that the fit was no different for EAL and English only speaking participants. Model 1: $\chi^2(1) = .42$, $p = .52$, Model 2: $\chi^2(1) = .22$, $p = .64$, Model 3: $\chi^2(1) = .23$, $p = .64$.

3.4 Discussion

The goal of this chapter was to validate and expand work by Kim (2015; 2016; 2017) which has proposed the DIET model of LC comprising roles for cognitive and language skills including ToM. The first aim was to validate the cross-sectional model of Kim (2017) using broadly similarly aged children (current study mean age = 6 years and 1 month, Kim, 2017 mean age = 7 years and 6 months) in a UK population whilst controlling for age and non-verbal ability. The second aim was to extend the model longitudinally to assess whether it was capable of predicting LC over a period of a year. The final aim was to extent the model even further to evaluate whether it could predict LC across a greater period of time (22 month) from 4 years 3 months to 6 years 1 month.

Together these aims gave a better understanding of the role played by ToM in LC, a cognitive skill which had previously not been explored as a longitudinal predictor of LC. Extending the DIET model longitudinally is important because longitudinal studies enable precursors of an ability to be determined. Indeed Kim (2015) acknowledges the need for longitudinal research into the DIET model (p.116). The use of longitudinal deign aided in determining precursory skills (those in preschool) for later LC. This age timeline was important because preschool is arguably the first age in which LC and the skills within the model can be reliably measured (Carrow-Woolfolk, 2011; L. M. Dunn et al., 1997), so this study was able to determine the earliest precursors of LC. This has potential impactful implications for the early years classroom.

Findings showed that models including direct and indirect effects of the language skills of knowledge of grammar (syntax) and vocabulary, and the cognitive skills of working memory and ToM fitted well for LC; concurrently, 12 months later and 22 months later. These findings are important as longitudinal work into LC is rare (Lepola et al., 2012). This research is also the first to show some evidence that earlier ToM plays a role in later LC as findings showed this to be the case across 12 months (ToM measured at four years old predicting LC aged five).

3.4.1 A concurrent model of listening comprehension

Model 1 (Figures 3.5 and 3.6) showed that the cross-sectional DIET model (Kim, 2017) could be validated in a UK population of slightly younger children (on average 17 months younger) after controlling for age and non-verbal ability, and that this model had a good fit across all indices. Importantly for this thesis, the model included a direct significant relation between ToM and LC. This is consistent with other research which has also suggested a robust relation of ToM to LC (Kim, 2015; Kim, 2016; Pelletier & Beatty, 2015).

However, this contrasts with Strasser and Rio (2014) who did not find a direct relationship between ToM and LC. Reasons for this difference in findings could be the measures used to assess LC. Measures of LC often vary in their content (Cain, 2017), for example, in the wordless book task used by Strasser and Rio (2014) children were read a wordless book and then asked questions to measure their comprehension of the story. By contrast, in the LC subset of the Oral and Written Language Scales (OWLS; Carrow-Woolfolk, 2011) used by Kim (2015; 2017) and in the Paragraph Comprehension subtest Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999) used by Kim (2015), children's comprehension of single words, phrases and sentences was assessed using a picture pointing task. The difference between these measures demonstrates the differing nature of the task demands of measures which are used to contribute to the same construct of LC and could explain the contrasting findings of Strasser and Rio (2014) to other studies. To address this, the current study used two standardised assessments of LC, one tapping comprehension of sentences and requiring non-verbal responses (OWLS also used by Kim, 2017) and the other comprehension of passages requiring verbal responses (NARA). The use of these two measures was consistent with Kim (2017) and aimed to capture the full nature of LC something which the study by Strasser and Rio (2014) was perhaps not able to do.

Another possible explanation for the contrasting findings of Strasser and Rio (2014) is differences in the nature of the text used by the current research. Here, the LC NARA stories included more than one character interacting with one another (e.g. a child helping a bird, or a girl running to aid of children who have appeared to have crashed their bikes) and the OWLs measure items also often included differing characters and social information. In contrast, Strasser and Rio's LC measure was a wordless story book involving a capybara's effort to reach bananas, a story which appears less rich in social information.

Model 1 cannot claim to be a direct replication of the DIET model because not all significant paths were reproduced. The current model did not replicate the finding that grammar is significantly related to ToM. Yet, further inspection of Kim (2017) shows that this relationship was only marginal. Additionally, the current model showed a different relationship with comprehension monitoring to that suggested by Kim (2017), in that neither vocabulary nor grammar were significantly directly related to comprehension monitoring and comprehension monitoring was not significantly related to LC. This contrasting finding could be due to the age of the participants, as although the children in the current model were of a broadly similar age they were on average 17 months younger than the sample of Kim (2017). Firstly, children at this younger age will have less experience in encountering complex oral discourse which require them to monitor their comprehension, and secondly, they may not yet have acquired the skills to act on this to put monitoring into action whilst comprehending language. This is supported by the earlier model by the same author (Kim, 2016, see Figure 3.2) with children the same age as the current study (six years) which also found that comprehension monitoring did not to make a significant contribution to LC.

Although not the key concern of this chapter, it is also interesting to note that Model 1 did replicate the finding that working memory is not directly significantly related to LC. This is also consistent with very recent work which suggests a smaller role for working memory in LC than perhaps previously thought (Jiang & Farquharson, 2018).

3.4.2 Longitudinal models of listening comprehension

Models 2 and 3 extended cross-sectional findings longitudinally, demonstrating that LC can be predicted by cognitive and language skills both 12 months and 22 months previously[6] after controlling for age and non-verbal ability. Most importantly for this thesis, Model 2, which used skills measured aged four (Time 1) to predict LC aged five (Time 2), found that as with cross-sectional studies ToM made a direct

[6] It should be acknowledged that these was not full replications of the DIET model as they did not include comprehension monitoring or inference making.

contribution to LC. This is notable as this is the first study to show this extension from concurrent to longitudinal study for ToM.

In contrast, Model 3, which extended this further using skills aged four (Time 1) to predict LC aged six (Time 3), did not fully support this. Although the same model was a good fit, ToM was not shown to make a direct contribution to LC. However, sample size may have affected the power of the model to detect the contribution of ToM to LC. If a larger sample had been used a significant relationship could have perhaps been reached. Due to the use of two cohorts Model 3 had 34% fewer participants than Model 2. SEM is very sensitive to sample size (E. J. Wolf et al., 2013) and so if a sample size closer to that included in Model 2 had been used the relationship between ToM and LC may have reached significance. This said, when a post-hoc statistical power test was carried out on all three models (Soper, 2017), Model 3 was found to have sufficient power, and the power was only slightly weaker than the other two models (see further analysis Section 3.3.4 for details). Future research should endeavour to have a larger sample to explore if this was an issue.

As previously raised, these longitudinal models were not direct replicas of Model 1 or the DIET model as they do not include a role for inference or comprehension monitoring. This is because it is difficult to measure these skills before the age of six years old and to the knowledge of the author no UK based measures exist for children this young. However, further analysis (see Section 3.3.4.2) added Time 3 comprehension monitoring and inference making to Model 2 and Model 3 to consider how this could affect model fit. In both cases model fit remained excellent and inference making but not comprehension monitoring were significantly related to LC. This highlights further that aged five and six, inference making is important for LC whereas comprehension monitoring is not. Future research should endeavour to develop simpler measures of inference making and comprehension monitoring which can be administered to children before the age of six years.

3.4.3 Strengths, limitations and further directions

A strength of this study is that it was longitudinal and that in particular skills were first assessed very early on (age four years; Time 1). Research with young children has inherent problems regarding the reliability of measures. Selecting age appropriate measures is always a challenge with this type of research especially a pre-planned longitudinal study. However, here with the exception of working memory measured at Time 1, all measures were well distributed and were not at ceiling or floor for any time point. Even for this measure of working memory, although skewness and kurtosis values did not quite meet the traditional and stringent values of skewness (< +1/-1) and kurtosis (< +2/-2; Gravetter & Wallnau, 2010), they did meet values suggested for SEM (skewness < +3/-3 and kurtosis < 10/-10; Chou & Bentler, 1995; Kline, 2005) and so there were no prominent issues with distribution.

Future research should also explore the DIET model in older children. This is acknowledged by Kim, who states that the model should be validated across different developmental phases (Kim, 2015, p.30; Kim, 2017, p.328). Moreover, given that the model was able to predict LC a year later and ToM was shown to make a direct contribution, yet this was not the case for ToM predicting LC a further 10 months later, it would be interesting to see how this relationship holds up for predicting LC even further in the future. For example, when children are aged eight or nine and encountering even more complex language which could include more complex social details.

Although highlighting that ToM is important for LC, these findings provide no clarification why this might be. On one hand it was theorised that ToM assists LC because it is involved in the process of ensuring that the correct meaning has been taken from a spoken passage regarding the social information, such as characters' intentions, thoughts, and emotions, which are often critical aspects in understanding the key elements of a story (Kim & Pilcher, 2016). In this way ToM could aid the listener in building a mental model of the passage and its social content (Kim & Pilcher, 2016). On the other hand, ToM could assist with building a mental model of the passage not just in relation to social information, but also to build a more general representation of the passage beyond its social content. This links to the argument concerned with whether ToM is domain general or socially specified. Authors such as Perner believe ToM to be domain general as it is concerned with understanding all representations and this is not restricted to the representation of mental states (Iao et al., 2011; Perner, 1991), whereas others believe ToM to be socially specialised (He, Bolz, & Baillargeon, 2011). These arguments will be crucial in determining the nature of ToM which is important for LC and will be explored in the next chapter.

3.4.4 Conclusions

The findings of this chapter extend the work of Kim (2015; 2016; 2017) to show that a DIET model including language and cognitive skills cannot only predict LC concurrently, but also longitudinally (across 12 months) and from an early stage of development. Moreover, the work presented here is the first to show a longitudinal role for ToM in LC 12 months later. This adds to the growing evidence that ToM is important for LC and begins to suggest that it may be a precursor of later LC. Yet, this research was not able to show a direct relationship between ToM and LC 22 months later. This may have been because of a reduced sample size and so further work is needed to investigate this. Moreover, the findings presented here do not explain what it is about the nature of ToM which assists LC; the next chapter of this thesis endeavoured to do this.

4 Theory of mind versus broad metacognition in concurrent listening comprehension

4.1 Introduction

The previous chapter of this thesis (Chapter 3) provides evidence that theory of mind (ToM) is predictive of listening comprehension (LC), both concurrently (at age six years) and longitudinally across a 12 month period (ToM aged four predicting LC aged five). Yet, these findings do not explain why ToM is important for LC, specifically what it is about the nature of ToM which promotes and assists LC.

4.1.1 Why does theory of mind support listening comprehension?

A logical explanation for why ToM assists LC is that it is ToM's social nature. ToM is often defined as the ability to infer the mental states of others and predict and explain behaviour (Doherty, 2008; Premack & Woodruff, 1978). Therefore, if a child has a better understanding of the desires, intentions and perspectives of a speaker, or the mental states of a protagonist in an oral passage, then they may be able to comprehend the passage more fully. ToM may be involved in monitoring the meaning taken from passages and filling-in missing information regarding social information such as intentions, thoughts and emotions, which can be critical aspects of a spoken passage (Kim & Pilcher, 2016). This view is supported by research using referential communication which shows that children who are better at making sense of a speakers' meaning and intentions (e.g. "*find the small brick*") also perform better on false belief tasks (Astington, 2004; Nilsen & Fecica, 2011; Resches &

Pereira, 2007). Moreover, Pelletier and Astington (2004) demonstrated that children (aged three to five years) with more advanced ToM were better at re-telling wordless story books as their retelling included reference to characters' thoughts as well as their actions.

Measures of LC require children to understand social information in order to perform well. For example, in the NARA (Neale, 1999) administered orally to measure LC (as used in this thesis), children are read a story about a character called Kim who witnesses two children on bikes crash into each other and then runs to help them. An advanced ToM would give the listener the ability to infer that Kim feels scared or worried for the hurt children and this is why she ran to help. Moreover, in this story the children are actually recording a road safety video, and so Kim holds a false belief that the accident she witnessed is a real accident. An understanding of Kim's thoughts, feelings and her false belief will help the listener to comprehend this passage. In the same way, plots and narratives of children's stories often revolve around mental states and misunderstandings (Zunshine, 2019), and so the awareness of these mental states and false beliefs will lead to better comprehension when these stories are read to children.

It is hypothesised that the understanding of these social details may aid the listener in constructing a mental model of the passage which is crucial for proficient comprehension (Graesser et al., 1994; Kim & Pilcher, 2016; Kintsch, 1988). Indeed, ToM may directly contribute to model building itself. Perner and colleagues describe the possession of a ToM as mental model building (Johnson-Laird, 1983; Perner, 1991), in that a belief is a mental model of the world. It has been suggested that young children fail false belief tests because they do not understand that mental models or representations can differ from true reality (Lillard & Flavell, 1992; Perner, 1991). In other words, though young children are able to have a model (representation) of the world and also a model (representation) of what someone thinks, they are not yet able to build a higher-order meta-model (or meta-representation) that holds and connects the two. False belief understanding involves being able to represent (model) how someone is representing (modelling) something. So, for example, a child failing the Sally-Anne false belief task is not yet able to represent that Sally represents the ball being in the basket rather than as where it is in reality (the box). As such developments in ToM ability may directly underpin building mental models of the text and passages during LC.

Until now the social explanation has been the view taken by the majority of the literature (Dore et al., 2018; Kim & Pilcher, 2016). Yet, there is an alternative account. Instead it could be the general metacognitive nature of ToM which is important. Metacognition is defined as knowing about knowledge or thinking about thinking (Flavell, 1976; Flavell, 1979). Under this definition ToM is clearly a metacognitive skill as it is concerned with thinking about the mental states of others (Premack & Woodruff, 1978). Therefore, it may be that it is the broad metacognitive nature of ToM which facilitates LC, because it informs knowledge, actions, and understanding not necessarily related with mental states or social information

(Atkinson et al., 2017). This draws on the argument concerned with whether ToM is domain general or socially specified. Authors such as Perner argue ToM is domain general as it involves understanding representation in general and is not restricted to the representation of mental states (Iao et al., 2011; Perner, 1991), whereas others believe ToM to be socially specialised (He et al., 2011). If ToM is domain general then it is possible that during LC, ToM could aid with the construction of a mental model not just concerning social information but also representations of non-social aspects, such as space, time and objects (Graesser et al., 1994). If this is the case, then other forms of metacognition, not social in nature but drawing on comparable domain general ability, could also aid in creating this mental model during LC.

4.1.2 Theory of mind and other types of metacognition

As well as being a metacognitive skill itself, ToM has been linked to other (non-social) metacognitive processes in preschool and early years. These include: metamemory which is defined as knowledge about memory and how best the memory process works (Lockl & Schneider, 2007), metalinguistic awareness which is the ability to reflect on language as a carrier of meaning (Doherty, 2000), and source monitoring which is the understanding of the source of one's own knowledge (Bright-Paul et al., 2008).

Several studies have demonstrated a relationship between early ToM and metamemory. Lockl and Schneider (2007) found that ToM at the age of three significantly predicted metamemory aged five even after controlling for language competencies, when metamemory was measured using an interview with items adapted from Wellman (1977) which included activities such as brainstorming for strategies which children could use to remember to take their lunch to preschool. This finding was confirmed by the later work of Ebert (2015) using structural equation modelling and controlling for earlier metamemory ability. Moreover, a very recent training study has shown that ToM training for four and five year olds improves not only false belief understanding but also metamemory (Lecce & Bianco, 2018). Likewise, metalinguistic awareness has been linked to ToM. Doherty and Perner (1998) and Doherty (2000) found that for three and four year olds metalinguistic awareness (as measured by the understanding of synonyms and homonyms) was significantly associated with the ability to pass a false belief test. This is supported by more recent research with bilingual children which found that metalinguistic awareness (as measured by knowledge of synonyms) at the age of four predicted ToM a year later (Diaz & Farrar, 2017). ToM and source monitoring are also shown to be related. Bright-Paul et al., (2008) found that in three to six year olds ToM predicted children's ability to understand the source of their knowledge about a hidden object (e.g. they knew the object was a ball because they felt it, or they knew it was a toy horse because they saw it). With similar findings from Evans (2005) with a correlation found between false belief scores and scores on the same source monitoring task used by Bright-Paul et al., (2008) for children aged four years.

Research which shows that these non-social metacognitive skills are related to ToM in the early years suggests that these skills share the same underlying metacognitive nature as ToM (Atkinson et al., 2017). It is plausible that the understanding of memory, language and the source of one's knowledge depend on the same representational understanding (metarepresentation) as ToM does. For example, to pass a ToM false belief test children must think or represent how someone else thinks (or represents) to understand their perspective. Likewise, in metalinguistic awareness children must be able to understand or represent that an object can be both a "bunny" and a "rabbit". In source monitoring they must represent an object as being seen or being heard. In metamemory they must understand that that the same objects can be understood and remembered differently. Given this, to investigate if it is the social nature of ToM which is important for facilitating LC, or instead if it is the general metacognitive nature, a direct comparison of the contribution of ToM and these other types of non-social metacognition for LC is needed. If these non-social metacognitive skills predict LC, in the same way that ToM does, then this suggests it is because of the broad-metacognitive nature which these skills share. On the other hand, if these other types of metacognition do not predict LC, but ToM does, then this supports the idea that it is the social component of ToM which is important for LC.

4.1.3 The present study

To determine if it is the socially-specific element of ToM or the broad metacognitive nature which is important for assisting LC, this chapter compared the ability of ToM to predict LC, to the ability of other types of broad non-social metacognition (metamemory, metalinguistic awareness and source monitoring) to predict LC. The aim was to provide a deeper understanding of the role played by ToM in LC by explaining what it is about the nature of ToM which promotes and assists LC.

Three types of metacognitive tasks were used as an index of broad metacognitive ability and compared to ToM. Metamemory, metalinguistic awareness and source monitoring were chosen because, as noted above, they are robustly linked to ToM development in the early years and, arguably draw on the same underlying representational (model building) abilities. Moreover, these metacognitive skills can be reliably measured in young children (i.e. those aged four to six years) using hands-on tasks which can be administered in methodologically similar ways to false belief ToM tests and with similar language demands (Bright-Paul et al., 2008; Doherty, 2000; Ebert, 2015).

The fit of concurrent models of LC including (a) just ToM and (b) the other general types of metacognition were compared. Concurrent models were evaluated at the ages of four (Time 1), five (Time 2) and six years old (Time 3), to assess if the same fit could be observed at different ages as children begin to encounter more complex oral language. The period of four to six years is very important for both the development of ToM (Flavell, 1988; Wimmer & Perner, 1983) and other metacognitive skills (Lockl & Schneider, 2007) and also LC, and so this is a crucial age range to study. Although longitudinal studies have great value, it is important to first look at these ages separately to determine separate developmental changes e.g. determining if the pattern in relationships is different at younger ages (three and four years) and older ages (five and six years).

The models tested in this chapter were based on those from the previous chapter (See Figure 3.3 in Chapter 3) which were originally based on the DIET model of Kim (2017). As with the analysis in the previous chapter age and non-verbal ability were also controlled for (see Section 3.1.3 in Chapter 3). The aim was to examine whether it is the specifically social aspect of ToM that is important for LC. It was hypothesized that, if this is the case, then:

1) For all time points, models which included ToM would have a better fit than those which included a latent variable of non-social types of metacognition (metamemory, source monitoring and metalinguistic awareness).
2) For all time points, ToM would directly predict LC after accounting for all the other skills in the model including age and non-verbal ability, whereas the latent variable of non-social types of metacognition (metamemory, source monitoring and metalinguistic awareness) would not make a direct contribution to LC.

4.2 Method

4.2.1 Participants

For full information on participants including recruitment see Section 2.3 in Chapter 2. Table 4.1 shows the descriptive details of participants at each of the three time points in the current study.

Table 4.1

Descriptive details of participants at each of the three time points

	N	N males	N females	Mean age[a]
Time 1	204	105	99	4;1 (SD = 4.37)
Time 2	162	82	80	5;1 (SD = 4.43)
Time 3	107[b]	54	53	6;1 (SD = 3.68)

Note. [a] Mean age in Years; Months [b] At Time 3 only Cohort 1 were tested

4.2.2 Materials and measures

Table 4.2 shows the measures used at each time point. For comprehensive details on each of these measures refer to Section 2.5 in Chapter 2.

Table 4.2

Measures administered at the three time points

Skill	Measure	Time 1	Time 2	Time 3
Listening comprehension	NARA	✓	✓	✓
	OWLS	-	✓	✓
Vocabulary	BPVS-III	✓	✓	✓
Syntactic Knowledge	Sentence Structure (CELF-Preschool 2^{uk})	✓	✓	-
	Sentence Structure (CELF-4)	-	-	✓
Theory of mind	Unexpected contents task	✓	-	-
	Unexpected locations task	✓	-	-
	Belief desire reasoning	-	✓	-
	Unexpected locations second-order false belief	-	✓	-
	Strange Stories	-	-	✓
Metamemory	Metamemory task	✓	✓	✓

Metalinguistic awareness	Synonym judgment task	√	√	-
	Homonym selection task	-	-	√
Source monitoring	Tunnel task	√	√	-
	Events task	-	-	√
Working memory	Reverse word span task	√	√	√
Comprehension monitoring	Comprehension monitoring stories	-	-	√
Inference making	Inference oral story	-	-	√
Non-verbal ability	Block design	√	-	-

4.2.3 Procedure

Participants in Cohort 1 were administered the measures in Table 4.2 across the three time points within two 20-minute sessions in a quiet place in or just outside their school classroom. Participants in Cohort 2 were administered the measures in Table 4.2 at the first two time points in a single session in a quiet place in their own home. For complete procedural details refer to Chapter 2 Sections 2.7 and 2.8.

4.2.4 Analysis

The primary data analytical strategy was structural equation modelling (SEM) using AMOS Version 25 (Arbuckle, 2016). For each of the models tested, first descriptive statistics were computed and then initial correlational and regression analysis were carried out. During SEM analysis, a latent variable was created for LC using two measures (NARA and OWLS) at Time 2 and Time 3. For the broad metacognitive models, a latent variable for broad metacognition was also created which included the three types of non-social metacognition (metamemory, source monitoring and metalinguistic awareness). All other language and cognitive skills and LC at Time 1 (just NARA) were assessed by a single measure for each construct, and so therefore observed variables were used. For justification of this analysis and further details of its use see Section 3.2.4 in Chapter 3.

4.3 Results

At each time point a model which included ToM was compared to a model which included a latent variable of broad non-social metacognition (including metamemory source monitoring and metalinguistic awareness).

4.3.1 Time 1

4.3.1.1 Descriptive statistics and preliminary analysis

Table 4.3 shows descriptive statistics for all measures. It should be noted that a ToM composite consisting of the two ToM measures was computed by summing scores from the two ToM measures (unexpected contents and unexpected locations). This composite was used to give a richer measure of ToM and was justified as the individual measures correlated significantly ($r = .39$, $p <.001$) and because this type of composite is often used (Atkinson et al., 2017; Guajardo & Cartwright, 2016; Ruffman et al., 2002). This composite was used in all further analysis.

Correlations between measures are displayed in Table 4.4. All main measures (ToM, metamemory, metalinguistic awareness and source monitoring) were weakly to moderately related to the LC measure (NARA). Preliminary regression analyses were then performed to assess whether the metacognitive measures had predictive power in calculating LC and therefore whether it was logical to proceed with the models. Separate linear regressions were carried out for each individual metacognitive measure for its ability to predict NARA scores (the LC outcome measure). All variables significantly predicted NARA scores, with ToM shown to predict 6% ($p = .001$) of unique variance, metamemory 12% ($p < .001$), metalinguistic awareness 4% ($p = .006$), and source monitoring 9% ($p < .001$). Given this, full modelling using these variables and LC as an outcome measure was carried out.

Table 4.3

Descriptive statistics for each measure at Time 1

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Listening comprehension	NARA	204	44	0-12	1.60	2.35	1.89	3.00
Vocabulary	BPVS-III	204	168	10-102	53.58	16.85	-.06	-.49
Syntactic knowledge	CELF-Preschool	203	22	0-21	11.96	4.24	-.58	-.03
Theory of mind	Unexpected contents	200	3	0-3	1.09	1.16	.61	-1.10
	Unexpected locations	204	2	0-2	.95	.90	.40	-1.72
Metamemory	Metamemory task	203	4	0-4	.79	.97	.92	-.22
Metalinguistic awareness	Synonym judgment	201	4	1-4	3.30	.95	-.90	-.62
Source monitoring	Source monitoring tunnel	203	6	0-6	3.61	1.77	-.31	-.78
Working memory	Reverse word span	201	9	0-9	1.00	2.13	2.01	2.77
Non-verbal ability	Block design	204	40	0-32	16.80	5.73	-.56	.58
Age[a]		204	-	38-57	49.56	4.39	-.20	-.79

[a] Age in months

Table 4.4

Correlation matrix of all Time 1 measures

	1	2	3	4	5	6	7	8	9	10
1. Listening comprehension	-									
2. Vocabulary	.40**	-								
3. Grammar	.34**	.63**	-							
4. Theory of mind	.26**	.37**	.44**	-						
5. Metamemory	.35**	.40**	.46**	.34**	-					
6. Metalinguistic	.20**	.28**	.41**	.28**	.28**	-				
7. Source monitoring	.29**	.43**	.46**	.43**	.46**	.42**	-			
8. Working memory	.36**	.41**	.40**	.37**	.39**	.21**	.35**	-		
9. Non-verbal ability	.34**	.55**	.44**	.28**	.35**	.37**	.42**	.37**	-	
10. Age	.10	.12	.18*	.06	.21**	.13	.27**	.25**	.16*	-

Note. ** = p < .001, * = p < .05

4.3.1.2 SEM analysis

The fit of two models for LC was compared; one model included ToM (Model 1a see Figure 4.1) and the other included the broad (non-social) metacognitive skills (Model 1b see Figure 4.3). First, missing data was imputed using EM. This method of data imputation was used because missing data was minimal (13 cases) and Little's MCAR test reported the data to be missing at random (p =.17).

Variables in the models met univariate normality assumptions for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (Skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked but were not violated as confirmed by a Mardia's coefficient of multivariate critical ratio of 1.24 for Model 1a, and 1.54 for Model 1b (a value < 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008).

The same model indices were used to assess the fit as those used in the previous chapter (see Section 3.2.4 of Chapter 3 for the justification of their use). Model 1a did not have a particularly good fit with only the CFI indices reaching the desirable value; Model 1a; χ^2 (1) = 9.43, p = .002; CFI = .98, TLI = .47, and RMSEA = .20. Neither did Model 1b; χ^2 (13) = 17.74, p =.03; CFI = .99, TLI = .87, and RMSEA = .06.

For Model 1a, 10 multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05), when these 10 participants were removed and the model re-run, the model fit indices changed only slightly (to χ^2 (1) = 5.60, *p* = .018; CFI = .98, TLI = .66, and RMSEA = .15). For Model 1b, 10 multivariate outliers were identified using Mahalanobis d-squared, when these 10 participants were removed and the model re-run, the model fit indices also changed only slightly (to χ^2 (13) = 18.86, *p* = .04; CFI = .99, TLI = .87, and RMSEA = .05). Due to these only slight changes the full models with all 204 participants are presented here.

Standardised path coefficients of the models are shown in Figure 4.2 (Model 1a) and Figures 4.4 (Model 1b). In Model 1a after controlling for both age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .24, p < .001), grammatical knowledge (β = .27, p < .001) and LC (β = .21, p = .005). Grammatical knowledge was significantly related to ToM (β = .33, p < .001), but vocabulary was not significantly related to ToM (β = .13, p = .15). Vocabulary was directly related to LC (β = .19, p = .03) but grammatical knowledge was not (β = .06, p = .45). ToM was not significantly independently related to LC (β = .05, p = .34). A total of 22% of variance in Time 1 LC was explained by the concurrent skills in the model.

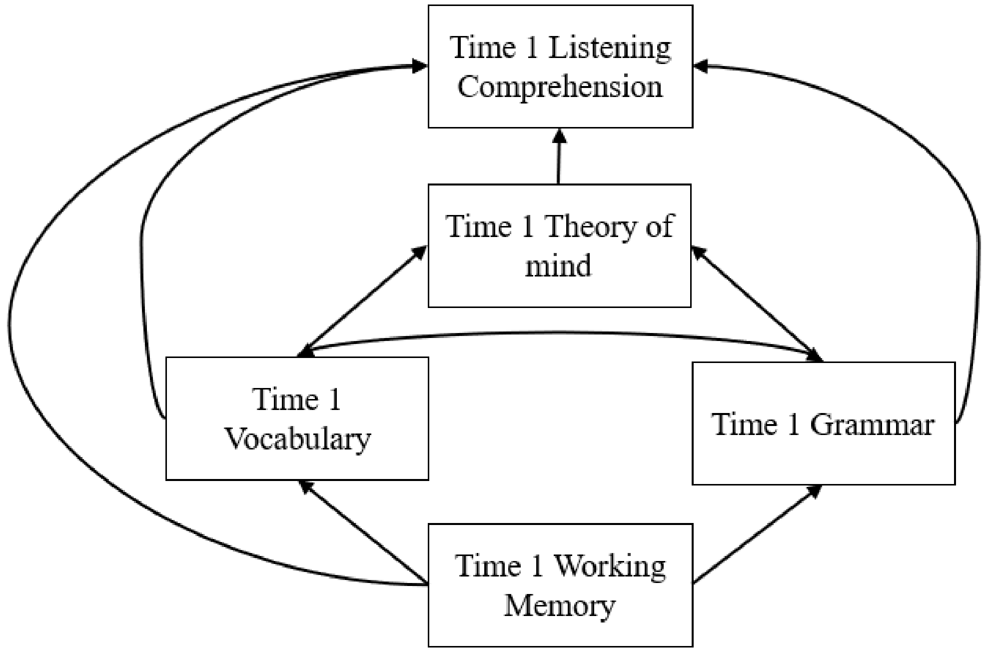


Figure 4.1: Hypothesised Model 1a after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

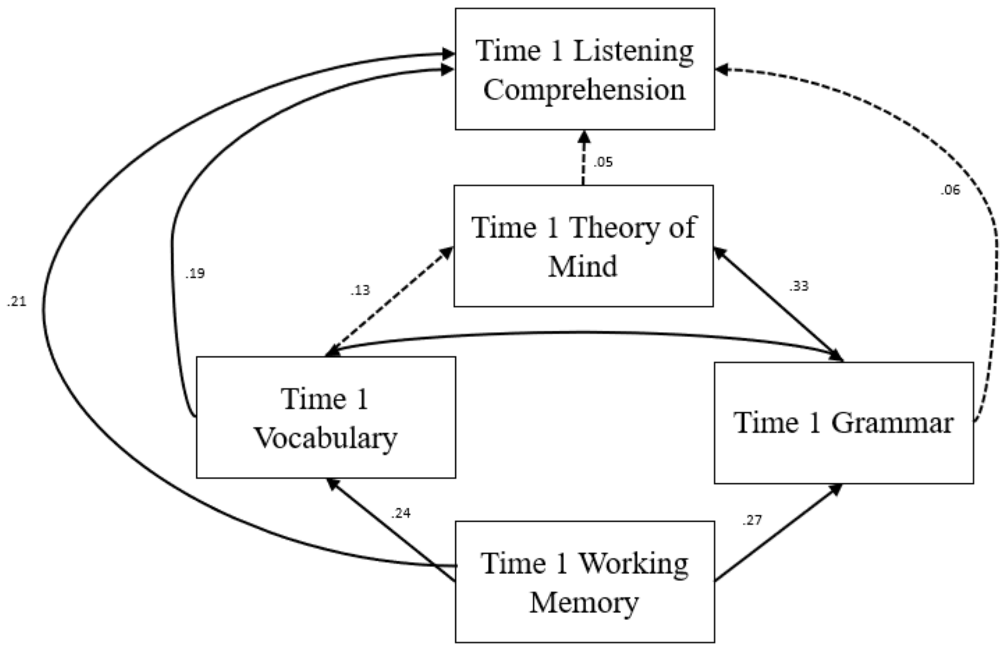


Figure 4.2: Model 1a with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

For Model 1b after controlling for age and non-verbal ability in all paths, working memory was significantly related to vocabulary (β = .24, p < .001) and grammatical knowledge (β = .27, p < .001). Grammatical knowledge was significantly related to metacognition (β = .57, p < .001) but vocabulary was not significantly related to metacognition (β = .07, p = .45). Vocabulary was directly related to LC (β = .18, p = .04) as was working memory (β = .20, p = < .01), but grammatical knowledge was not (β = -.06, p = .63). Metacognition was not significantly independently related to LC (β = .26, p = .14). A total of 24% of variance in Time 1 LC was explained by the Time 1 skills in the model.

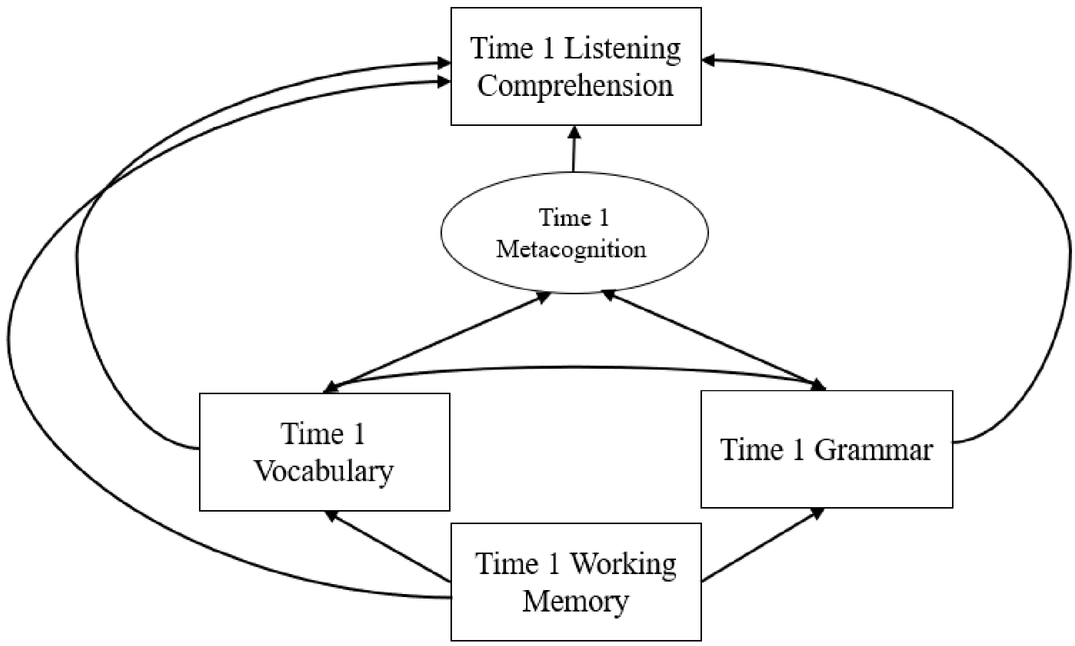


Figure 4.3: Hypothesised Model 1b after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

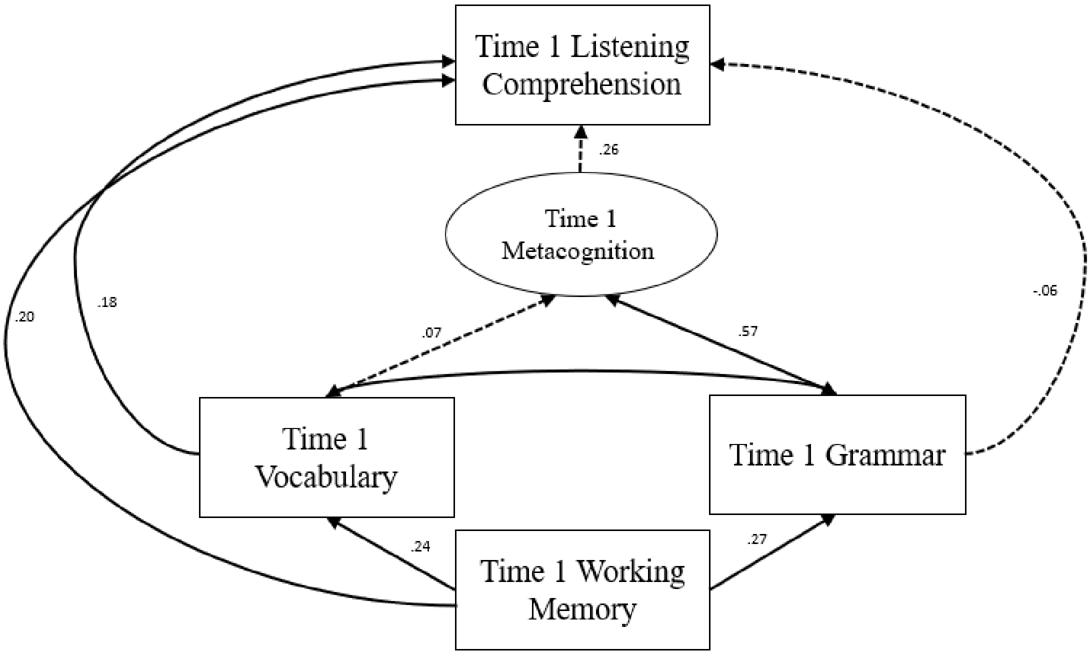


Figure 4.4: Model 1b with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent no-significant relations.

4.3.2 Time 2

4.3.2.1 Descriptive statistics and preliminary analysis

Table 4.5 shows descriptive statistics for all measures at Time 2. It should be noted that a ToM composite consisting of the two ToM measures was computed by summing the scores (belief desire reasoning and unexpected locations second-order**).** This composite was used to give a richer measure of ToM and was justified as the individual measures correlated significantly ($r = .32$, $p < .001$). This composite was used in all further analysis.

Table 4.5

Descriptive statistics for all Time 2 measures

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Listening comprehension	NARA	161	44	0-13	3.86	3.62	.57	-.80
	OWLS	162	130	24-98	52.97	15.06	.49	.26
Vocabulary	BPVS-III	162	168	24-105	78.61	13.19	-.90	1.81
Syntactic knowledge	CELF-Preschool	162	22	5-22	16.53	3.40	-.99	1.23
Theory of mind	Belief desire reasoning	162	3	0-3	1.47	1.21	.14	-1.55
	Unexpected locations second-order false belief	162	2	0-2	.60	.78	.82	-.85
Metamemory	Metamemory task	162	4	0-4	1.64	1.22	.26	-1.00
Metalinguistic awareness	Synonym judgment task	162	12	2-12	10.63	1.98	-2.18	4.94
Source monitoring	Source monitoring tunnel task	160	6	0-6	4.31	1.52	-.61	-.36
Working memory	Reverse word span task	162	9	0-9	3.51	2.90	.34	-.937
Time 1 non-verbal ability	Block design	162	40	0-32	16.56	5.86	-.47	.37
Age	-	162	-	50-69	61.18	4.43	-.17	-.69

[a]Age in months

Correlations between measures are displayed in Table 4.6. All metacognitive measures (ToM, metamemory, metalinguistic awareness and source monitoring) were weakly to moderately related to the two LC measures (NARA and OWLS). Preliminary regression analysis was then performed to assess whether the metacognitive measures had predictive power in calculating the two LC measures and therefore whether it was logical to precede with the models. Separate linear regressions were carried out for each individual metacognitive measure for its ability to predict the two outcome measures of LC (NARA and OWLS). All variables significantly predicted NARA scores, with ToM shown to predict 6% ($p = .002$) of unique variance, metamemory 12% ($p < .001$), metalinguistic awareness 5% ($p = .005$), and source monitoring 12% ($p < .001$). All variables also significantly predicted OWLS scores, with ToM shown to predict 9% ($p < .001$) of unique variance, metamemory 14% ($p < .001$), metalinguistic awareness 11% ($p < .001$), and source monitoring 13% ($p < .001$). Given this, full modelling using these variables and LC as an outcome measure was carried out.

Table 4.6

Correlation matrix of all Time 2 measures

	1	2	3	4	5	6	7	8	9	10	11
1. NARA	-										
2. OWLS	.56**	-									
3. Vocabulary	.44**	.41**	-								
4. Grammar	.28**	.50**	.48**	-							
5. Theory of mind	.25**	.30**	.37**	.29**	-						
6. Metamemory	.34**	.37**	.48**	.26**	.40**	-					
7. Metalinguistic	.22**	.32**	.41**	.34**	.31**	.41**	-				
8. Source monitoring	.35**	.36**	.44**	.28**	.35**	.58**	.42**	-			
9. Working memory	.28**	.40**	.49**	.34**	.34**	.43**	.34**	.51**	-		
10. Time 1 non-verbal	.18*	.29**	.48**	.34**	.36**	.33**	.36**	.38**	.41**	-	
11. Age	-.001	.06	.13	.17*	.13	.19*	.04	-.12	.10	.10	-

Note. ** = p < .001, * = p < .05

4.3.2.2 SEM analysis

As with Time 1, the fit of two models for LC was compared; one model included ToM (Model 2a see Figure 4.5) and the other included the other (non-social) metacognitive skills (Model 2b see Figure 4.7). First, missing data was imputed using EM. This method of data imputation was used because although the MCAR (data missing completely at random) assumption was violated (as shown by a significant Little's MCAR test, $p = .04$), MAR (data missing at random) assumptions were not violated because missing data was minimal with only 3 missing cases, i.e. less than 5% of cases (Schafer, 1999).

All variables in the models met univariate normality assumption for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (Skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were not violated as confirmed by a Mardia's coefficient of multivariate critical ratio of .694 for Model 2a. However, multivariate normality assumptions were violated for Model 2b as confirmed by a Mardia's coefficient of multivariate critical ratio of 3.272 (a value < 1.96 demonstrates normality; Gao, et al., 2008). As the univariate normality of each variable did not exceed +3/-3 for skewness or kurtosis, Gao, et al., (2008) recommend that the models should be still run but that caution should be taken when interpreting the chi-squared statistic as it could be inflated somewhat by the slight non-normality.

Neither Model 2a nor 2b were shown to have a good fit, with only the CFI indices reaching the desirable values. Model 2a; χ^2 (6) = 19.89, *p* = .003; CFI = .95, TLI = .78, and RMSEA = .12. Model 2b; χ^2 (20) = 69.02, *p* < .000; CFI = .90, TLI = .77, and RMSEA = .12.

For Model 2a, nine multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05), when these nine participants were removed and the model re-run, the model fit indices changed only slightly (to χ^2 (4) = 17.73, *p* = .001; CFI = .94, TLI = .80, and RMSEA = .15). For Model 2b, eight multivariate outliers were identified using Mahalanobis d-squared, when these eight participants were removed and the model re-run, the model fit indices changed only slightly (to χ^2 (20) = 75.42, *p* < .001; CFI = .88, TLI = .72, and RMSEA = .15). Due to these only slight changes the full models with all 162 participants are presented here.

Standardised path coefficients of the models are shown in Figure 4.6 (Model 2a) and Figure 4.8 (Model 2b). For Model 2a, after controlling for age and non-verbal ability working in all paths working memory was significantly related to vocabulary (β = .35, p < .001), grammatical knowledge (β = .27, p < .001), and LC (β = .29, p = .036). Grammatical knowledge was not significantly related to ToM (β = .15, p = .11), but vocabulary was significantly related to ToM (β = .20, p = .02). Both vocabulary (β = .35, p = .03) and grammatical knowledge (β = .55, p <.001) were directly related to LC. ToM was not significantly independently related to LC (β =

.19, p = .13). A total of 32% of variance in Time 2 LC was explained by the concurrent skills in the model.

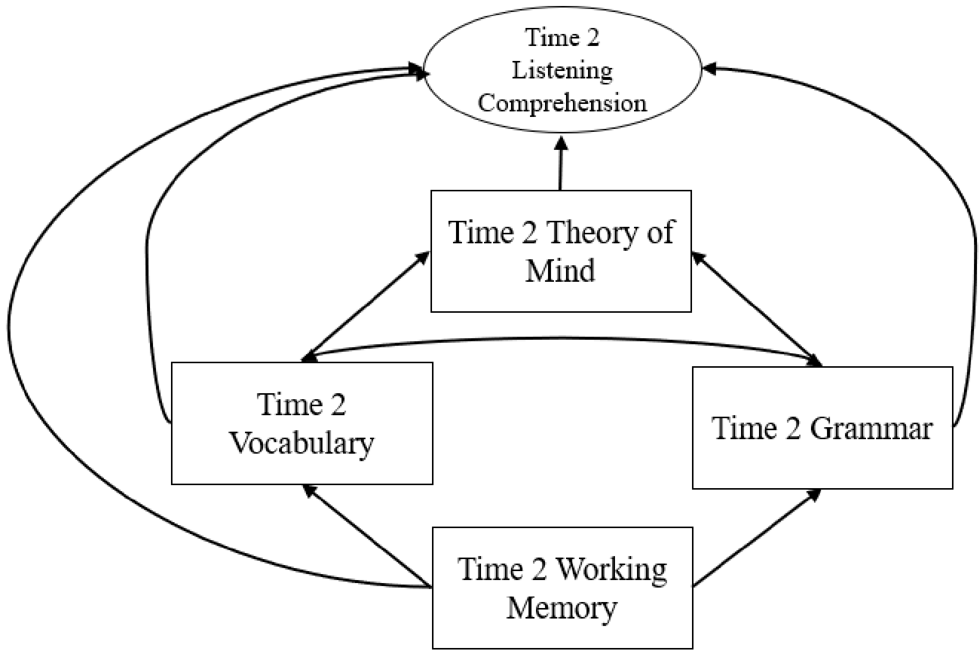


Figure 4.5: Hypothesised Model 2a after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

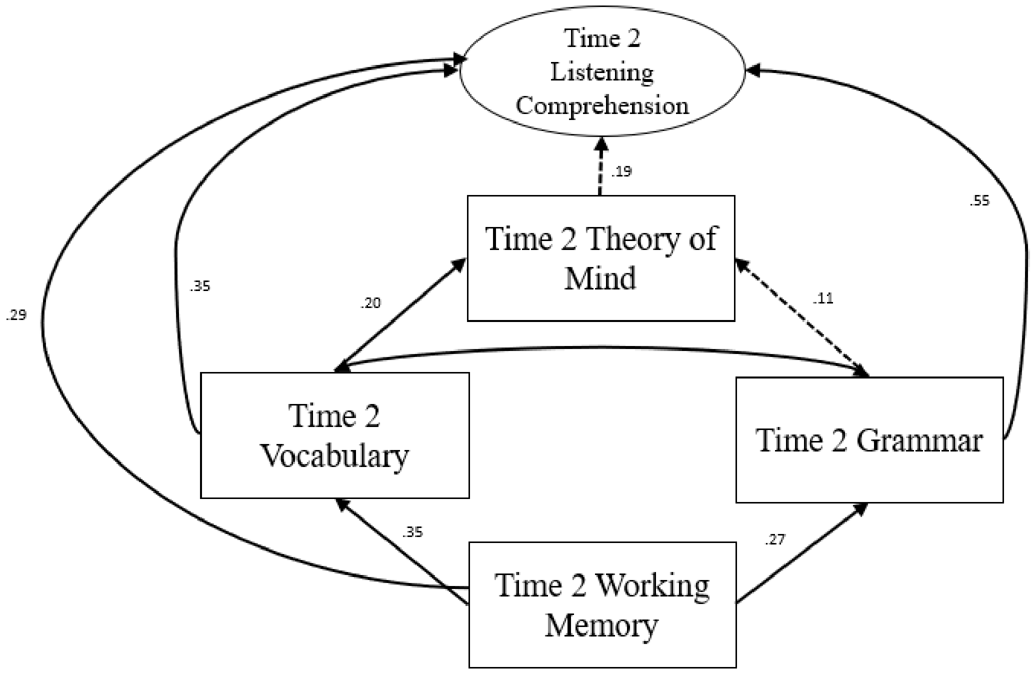


Figure 4.6: Model 2a with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

For Model 2b, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .35, p < .001) and grammatical knowledge (β = .27, p < .001) but not directly to LC (β = .17, p = .19). Grammatical knowledge was not significantly related to metacognition (β = .09, p = .29). Vocabulary was significantly related to metacognition (β = .48, p < .001). Vocabulary was not directly related to LC (β = .17, p = .32) but grammatical knowledge (β = .50, p <.001) was directly related to LC. Metacognition was significantly independently related to LC (β = .52, p < .01). A total of 26.5% of variance in Time 2 LC was explained by the concurrent skills in the model.

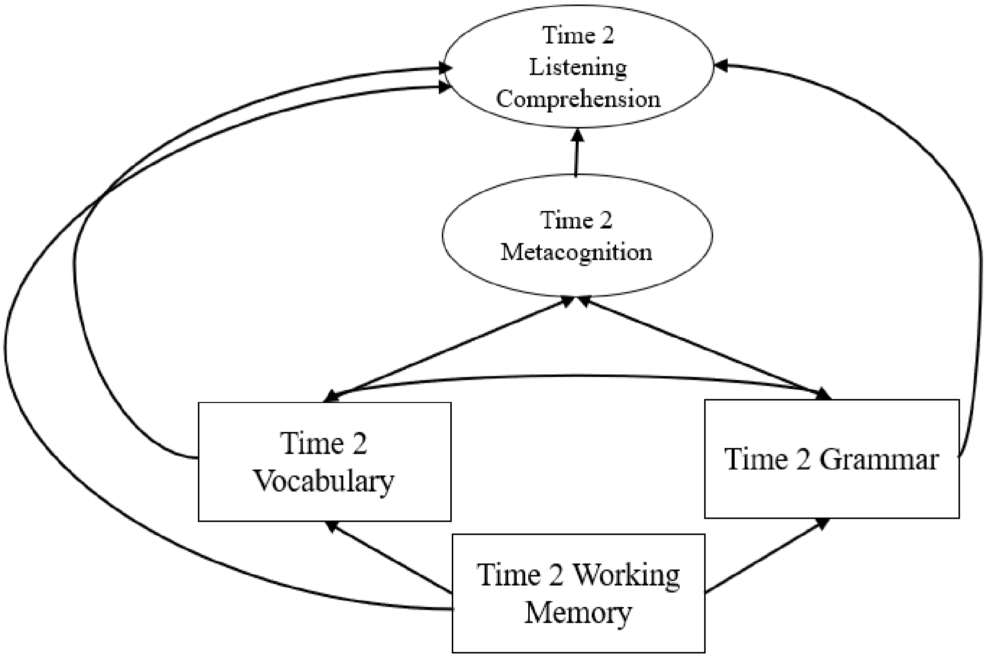


Figure 4.7: Hypothesised Model 2b after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

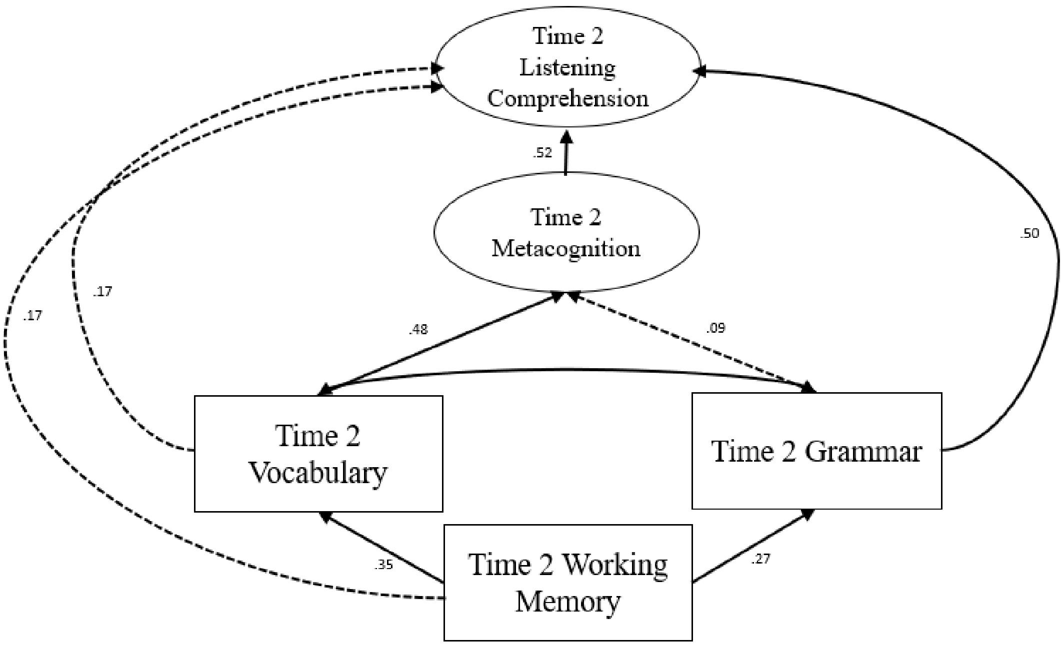


Figure 4.8: Model 2b with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

4.3.3 Time 3

4.3.3.1 Descriptive statistics and preliminary analysis

Table 4.7 shows descriptive statistics for all measures at Time 3. Correlations between measures are displayed in Table 4.8. All main measures (ToM, metamemory, metalinguistic awareness and source monitoring) were weakly to moderately related to the two LC measures (NARA and OWLS). At this timepoint, in contrast to the first two timepoints, comprehension monitoring and inference making were measured and included in the model. These were also weakly to moderately related to the two LC measures.

Preliminary regression analysis was then performed to assess whether the metacognitive measures had predictive power in calculating the two LC measures and therefore whether it was logical to precede with the models. Separate linear regressions were carried out for each individual metacognitive measure for its ability to predict the two outcome measures of LC (NARA and OWLS). All variables significantly predicted NARA scores, with ToM shown to predict 15% ($p <.001$) of unique variance, metamemory 11% ($p <.001$), metalinguistic awareness 4% ($p = .03$), and source monitoring 13% ($p < .001$). All variables also significantly predicted OWLS scores, with ToM shown to predict 10% ($p = .001$) of unique variance, metamemory 25% ($p < .001$), metalinguistic awareness 14% ($p < .001$), and source monitoring 12% ($p < .001$). Given this full modelling using these variables and LC as an outcome measure was carried out.

Table 4.7

Descriptive statistics of all Time 3 measures

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Listening comprehension	NARA	106	44	1-16	6.58	3.37	.69	-.29
	OWLS	107	130	36-94	63.66	12.31	.38	-.55
Vocabulary	BPVS-III	107	168	48-130	87.03	13.28	.14	1.29
Syntactic knowledge	CELF-4	107	26	7-26	20.24	3.74	-.99	1.06
Theory of mind	Strange Stories	106	10	0-8	2.33	1.57	.93	1.46
Metamemory	Metamemory task	106	4	0-4	2.70	1.22	-.63	-51
Metalinguistic awareness	Homonym selection task	107	8	0-8	5.28	2.17	-.34	-.69
Source monitoring	Source monitoring events task	105	6	2-6	4.46	1.13	-.18	-.67
Working memory	Reverse word span task	107	11	0-11	6.69	3.38	-.25	-1.42
Comprehension monitoring	Comprehension monitoring stories	107	5	0-5	2.67	.98	.09	-.37
Inference making	Inference oral story	106	8	0-8	4.21	1.89	-.20	-.51
Time 1 non-verbal ability	Time 1 block design	107	40	0-32	15.34	6.50	-.12	-.13
Age[a]	-	107	-	66-80	72.95	3.68	-.11	-1.10

[a] Age in months

Table 4.8

Correlation matrix of all Time 3 measures

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. NARA	-												
2. OWLS	.62**	-											
3. Vocabulary	.54**	.50**	-										
4. Grammar	.58**	.37**	.49**	-									
5. Theory of mind	.38**	.32**	.35**	.25**	-								
6. Metamemory	.34**	.50**	.46**	.31**	.21*	-							
7. Metalinguistic	.21*	.38**	.21*	.24*	.14	.37**	-						
8. Source monitoring	.36**	.35**	.25**	.34**	.13	.59**	.31**	-					
9. Inference making	.45**	.51**	.52**	.49**	.29**	.20*	.19*	.21*	-				
10. Comp monitoring	.21*	.23*	.19*	.21*	.16	.28**	.16	.15	.10	-			
11. Working memory	.36**	.33**	.39*	.32**	.34**	.44**	.16	.32**	.31**	.10	-		
12. Non-verbal ability	.46**	.32**	.45**	.47**	.28**	.43**	.14	.42**	.33**	.18	.28**	-	
13. Age	.22*	.13	.13	.14	.09	.19	.15	.14	.11	.18	-.01	.41**	-

Note. ** = p < .001, * = p < .05

4.3.3.2 SEM analysis

As with the previous time points, the fit of two models for LC was compared; one model included ToM (Model 3a see Figure 4.9[7]) and the other included the other (broad) metacognitive skills (Model 3b see Figure 4.11). First, missing data was imputed using EM. This method of data imputation was used because missing data was minimal (6 cases) and Little's MCAR test reported the data to be missing at random (p =.25).

All Time 3 measures met univariate normality assumption for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (Skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were not violated as confirmed by a Mardia's coefficient of multivariate critical ratio of -.57 for the variables in Model 3a and -.31 for variables in Model 3b (a value < 1.96 demonstrates normality; Gao et al., 2008).

Model 3a was shown to have a good fit; χ^2 (10) = 16.50, p = .09; CFI = .98, TLI = .89, and RMSEA = .08. Model 3b was not shown to have a good fit; χ^2 (28) = 57.64, p = .001; CFI = .92, TLI = .82, and RMSEA = .10. For Model 3a, seven multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05). When these seven participants were removed and the model re-run, the model fit indices

[7] It should be noted that this model is the same as Model 1 reported in Chapter 3, but in the context of this chapter is being compared to a model including broad metacognition.

changed only slightly (to χ^2 (10) = 16.72, p = .10; CFI = .98, TLI = .89, and RMSEA = .08). For Model 3b, five multivariate outliers were identified using Mahalanobis d-squared, when these five participants were removed and the model re-run, the model fit indices changed only slightly (to χ^2 (28) = 65.91, p < .001; CFI = .90, TLI = .75, and RMSEA = .12). Due to these only slight changes the full models with all 107 participants are presented here.

Standardised path coefficients of the models are shown in Figure 4.10 (Model 3a) and Figure 4.12 (Model 3b). For Model 3a, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .28, p =.001), grammatical knowledge (β = .20, p = .03) but not to LC (β = .14, p = .21). Grammatical knowledge was not significantly related to ToM (β = .05, p = .66), or comprehension monitoring (β = .13, p = .23), but was to inference making (β = .30, p = .002). Vocabulary was significantly related to ToM (β = .27, p = .01) and inference (β = .36, p <.001), but not to comprehension monitoring (β = .11, p = .35). Both vocabulary (β = .33, p = .01) and grammatical knowledge (β = .35, p =. 01) were directly related to LC. ToM was significantly independently related to LC (β = .19, p = .04) but inference (β = .23, p = .09) and comprehension monitoring were not (β = .11, p = .31). A total of 40% of variance in Time 3 LC was explained by the concurrent skills in the model.

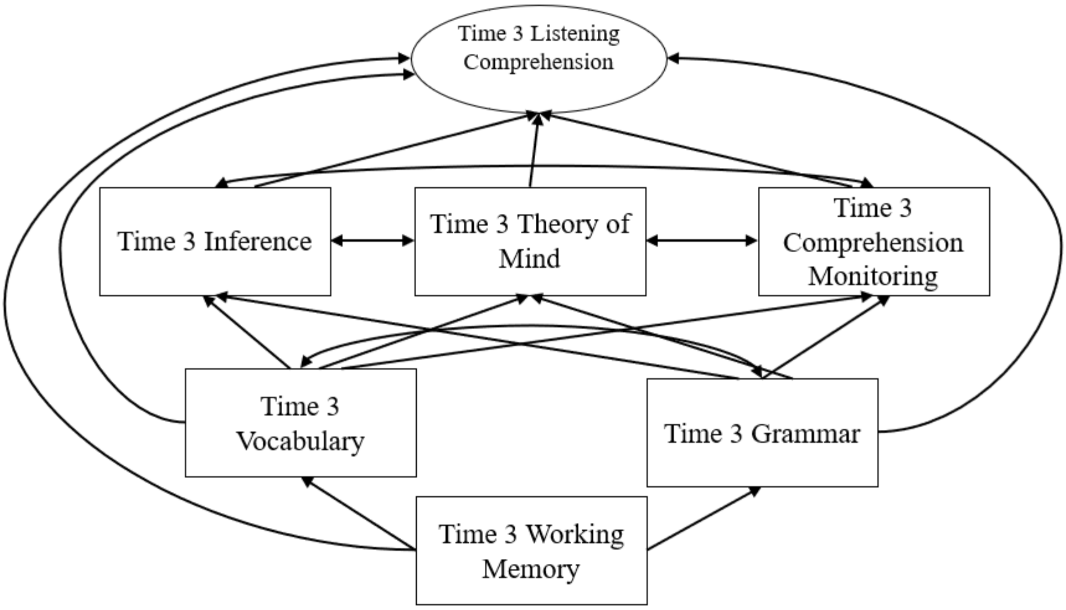


Figure 4.9: Hypothesised Model 3a after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

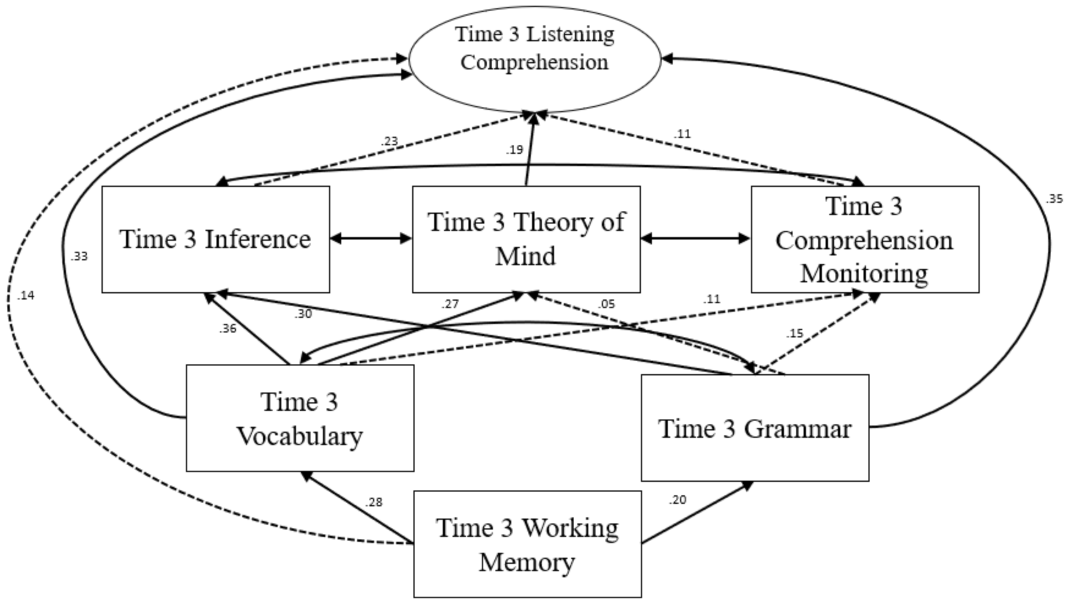


Figure 4.10: Model 3a with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

For Model 3b, after accounting for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .28, p= .001) and grammatical knowledge (β = .20, p = .03) but not directly to LC (β = .09, p = .43). Grammatical knowledge was significantly related to inference making (β = .30, p < .01) but not to metacognition (β = .11, p = .39), or comprehension monitoring (β = .13, p = .26). Vocabulary was significantly related to metacognition (β = .30, p = .02) and inference making (β = .36, p < .001) but not to comprehension monitoring (β = .11, p = .35). Vocabulary was directly related to LC (β = .26, p < .05) but grammatical knowledge (β = .24, p = .07) was not. Metacognition was not independently related to LC (β = .39, p = .06), neither was comprehension monitoring (β = .07, p = .50), but inference making was (β = .33, p < .01). A total of 36.5% of variance in Time 3 LC was explained by the concurrent skills in the model.

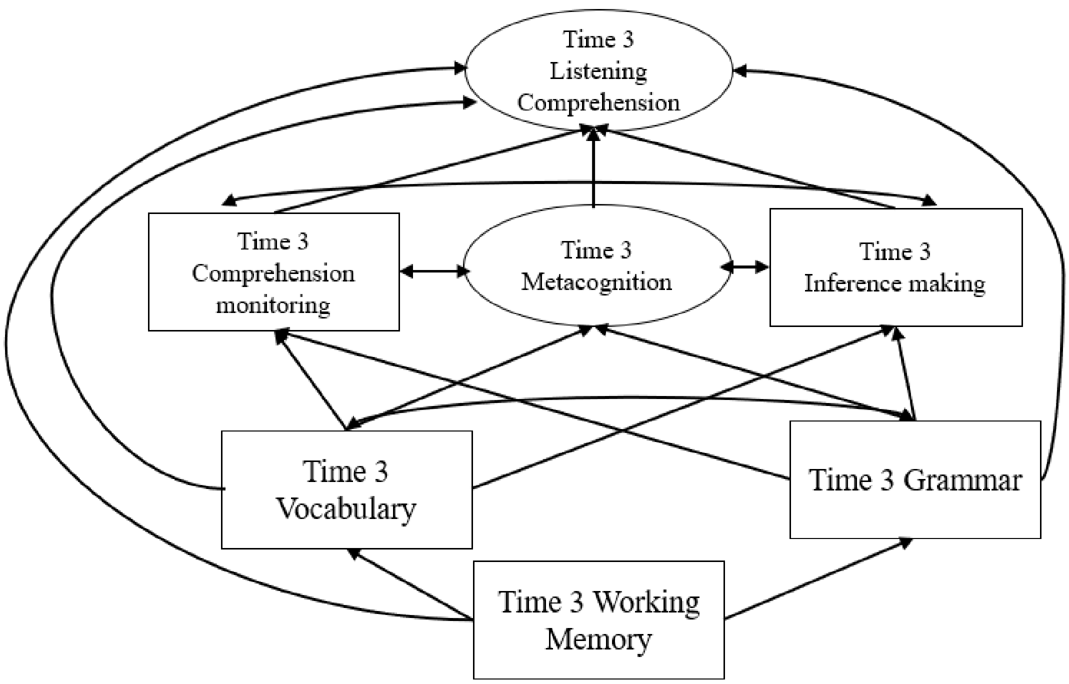


Figure 4.11: Hypothesised Model 3b after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

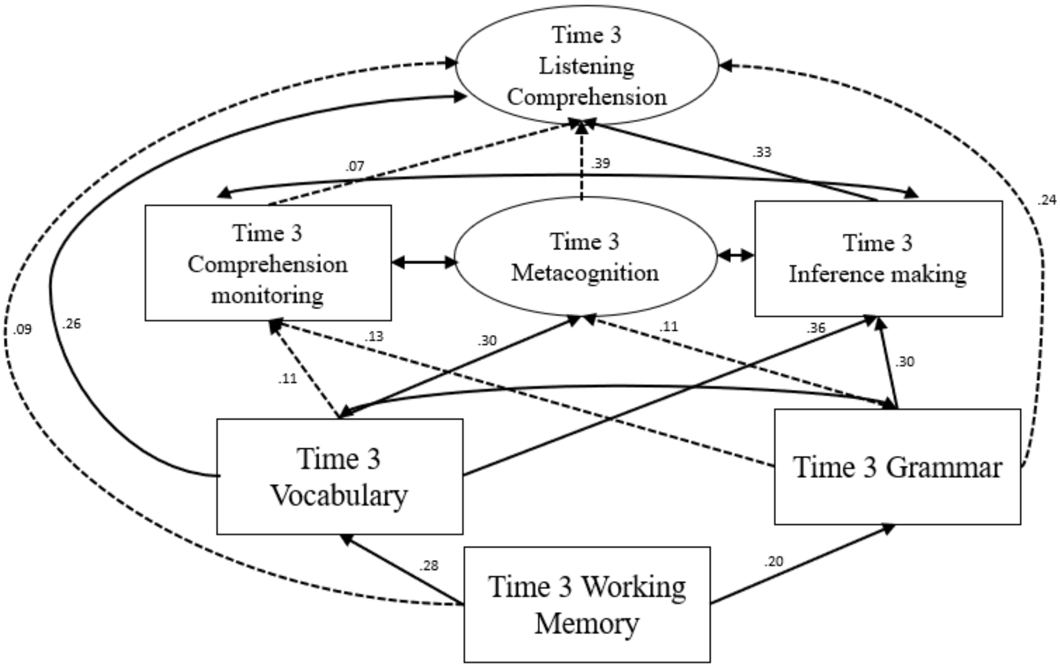


Figure 4.12: Model 3b with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

4.3.4 Further analysis

4.3.4.1 Comprehension monitoring and inference making

Both Time 3 models were run again with the exclusion of comprehension monitoring and inference making because these skills are arguably metacognitive (see Section 4.4.2 for a full discussion on this). When this was done Model 3a (including ToM see Figure 4.9) remained a good fit (χ^2 (6) = 8.07, p = .23; CFI = .99, TLI = .96, and RMSEA = .06) and ToM still significantly predicted LC after accounting for all other skills in the model (β = .23, p = .04). For Model 3b (including broad metacognition see Figure 4.11) the model remained a poor fit (χ^2 (28) = 84.32, p < .001; CFI = .82, TLI = .72, and RMSEA = .14) and still broad metacognition did not significantly predict LC after accounting for all other skills in the model (β = .37, p = .06).

4.3.4.2 Children with English as an additional language

A proportion of the sample had English as an additional; language (EAL). This was the case for 33% of the sample at Time 1, 23% of Time 2 and 30% of Time 3. This could have affected results as EAL children are shown to have weaker oral language skills than their monolingual peers (Burgoyne et al., 2009). Indeed, just at Time 1 the monolingual participants had significantly higher scores on vocabulary (t(202) = 3.30, p =.001) syntax (t(202) = 3.25, p =.001) and LC (t(202) = 2.72, p =.006). Therefore, multigroup analysis was used in AMOS to compare the fit of all models for EAL participants to English only speaking participants. This was done using a chi-square difference test whereby a non-significant chi-square shows that the model fit is the same for both groups (Dimitrov, 2006). For each of the six models non-significant chi-square demonstrated that the fit was no different for EAL and English only speaking participants: Model 1a: χ^2 (16) = 12.57, p = .70, Model 1b: χ^2 (27) = 28.28, p = .40, Model 2a: χ^2 (19) = 29.10, p = .06, Model 2b: χ^2 (29) = 31.97, p = .32, Model 3a: χ^2 (29) = 40.95, p = .07, Model 3b: χ^2 (17) = 23.34, p = .07. This said, some of these non-significant chi-squares were marginal.

4.4 Discussion

This chapter followed from the previous to give a further in-depth understanding of the role played by ToM in LC. The primary aim of the chapter was to understand what it is about the nature of ToM which is useful for early LC. To do this, the concurrent role of ToM in LC was compared to the role of other metacognitive skills not social in nature (source monitoring, metamemory and metalinguistic awareness). This allowed for a direct examination into the specific component of ToM which is important for assisting LC i.e. whether it is the social specificity or the broad metacognitive nature that plays a significant role. SEM analysis was used at each individual time point (aged four, aged five and aged six), comparing the fit of models of LC which included ToM, and models which included a latent variable of these non-social types of metacognition.

4.4.1 Main findings – social specificity or general metacognitive nature?

It was hypothesized that for all time points, models which included ToM would have a better fit than those which included non-social types of metacognition, and that ToM would directly predict LC. This is because it is argued that the social element of ToM which facilitates LC (Dore et al., 2018), in that ToM provides assistance with the social details of a story and in building a mental model of the text. This hypothesis was very partially supported with only limited evidence for a better fit and direct effect of ToM.

SEM analysis showed that only at Time 3, when children were in Year 1 and aged six, did ToM directly predict LC after controlling for other skills in the model (working memory, vocabulary, grammatical knowledge, inference making and comprehension monitoring), age and non-verbal ability. At Time 1 and 2, models which included ToM (Models 1a and 2a) were not good fits and ToM did not directly significantly predict LC. On the other hand, none of the broad metacognitive models (Models 1b, 2b and 3b) had a good fit at any time point. This said, at Time 2 (Model 2b) the latent variable of broad metacognition made a significant direct contribution to LC after accounting for all other variables in the model, but this was not the case at Time 1 (Model 1b) or Time 3 (Model 3b).

These findings are therefore not straightforward. On one hand they suggest that it is the social element of ToM which is important for facilitating LC rather than its broad metacognitive nature because Model 3a, which included ToM, was the only well-fitting model. Yet if this is true, it is also the case that ToM is not important for LC until children are older given that models at the first two time points were not well fitting and there was not a direct significant path from ToM to LC. This is consistent with past work, because although Kim (2015; 2016; 2017) found that ToM made a direct concurrent contribution to LC for children aged six and seven years, Strassser and Rio (2014) found that for preschoolers (aged from four to six years) ToM did not make a significant contribution to LC. This finding also supports the research which shows that early-on (when children are three to five years old) low-level language skills such as vocabulary are most important for comprehension rather than high-order skills (Kendeou et al., 2005; Lynch et al., 2008). It may be that until the age of six children are not encountering oral passages of such complexity that they require high level social understanding in order to comprehend them, therefore younger children are not practiced at using these skills during comprehension.

There is an alternative explanation for the findings and instead, results may be influenced by the sensitivity of LC measures used with younger children. It could be that LC passages administered to preschool children lack complexity as they do not contain advanced social information that requires a ToM understanding, or any other high-order skill, to aid with understanding. For example, the first passage of the NARA administered in this thesis tells the story of a little girl looking after a bird and her babies. Although there is some social information within this passage as the girl helps the bird, this social information is very limited especially in comparison to later passages (administered at later time points) which include emotions, false beliefs, complex and inferred intentions and cognitions. The lack of social information captured by the early LC measures, by both this thesis and past research (e.g. Strasser & Río, 2014), may explain why ToM does not contribute towards these LC measures earlier on. Of course, it could also be said that these simple LC measures used in research with preschool children reflect the type of simple stories that preschool children are exposed to at home and in their classrooms.

This chapter is unable to unequivocally answer the question of whether it is the social element of ToM which is important for LC because broad metacognition made a significant contribution to LC at Time 2 and this relationship was approaching significance at Time 3, so there is some suggestion that broad metacognition is important for LC. This evidence is, however, very limited given that these models were not well fitting and also that metacognition did not make a direct contribution to LC in the Time 1 model (Model 1b). Future research may confirm these findings.

4.4.2 Comprehension monitoring and inference making as metacognitive skills

At Time 3 comprehension monitoring and inference making were measured and therefore were included in both the Time 3 models (Models 3a and 3b). These skills were measured and included in line with the DIET models of Kim (2015; 2016; 2017), and because past research shows them to be important for LC (Cain et al., 2001; Kim & Phillips, 2014; Strasser & Río, 2014; Tompkins et al., 2013). They were not included in the earlier models as it is very difficult to measure these skills before the age of six years old.

Although not explicitly stated by the reading and listening literature, comprehension monitoring and inference making are metacognitive skills because they require awareness of thought processes (L. Baker & Brown, 1984; Flavell, 1979; Halpern, 1998; Kinnunen et al., 1998; Pitts, 1983). Comprehension monitoring calls upon the listener to think about their own understanding of the passage (Dabarera et al., 2014; Kinnunen et al., 1998; Paris & Myers, 1981; Pitts, 1983) to monitor this understanding. Likewise, during inference making the listener must think about their own knowledge to both link together different parts of the passage (local inferences) and use their own existing knowledge to fill in gaps (global inferences) in order to gain a deeper understanding about the passage (Cain & Oakhill, 1999; Cain et al., 2001; Cain et al., 2004). Given that comprehension monitoring and inference making can be seen as non-social metacognitive skills, their inclusion within the models at Time 3 could have muddied the waters when it came to comparing the contribution of ToM to the other non-social metacognitive skills (metamemory, source monitoring and metalinguistic awareness). Therefore, the two Time 3 models were re-run excluding comprehension monitoring and inference making (See Section 4.3.4.1 in further analysis). When this was done model fits remained the same, and ToM still significantly predicted LC after accounting for all other skills in the model, whereas broad metacognition still did not. Therefore, the inclusion of inference making and comprehension monitoring does not seem to affect assessment of the contribution of ToM to other non-social metacognitive skills for LC.

In the ToM model (Model 3a) neither comprehension monitoring nor inference making directly predicted LC, while in the non-social metacognitive model (Model 3b) although comprehension monitoring did not significantly predict LC, inference making did. The reason for these contrasting findings between models suggests that ToM and inference making are very closely linked and in the ToM model (Model 3a) ToM is accounting for the contribution of inference making, something which is not accounted for in Model 3b by any of the non-social metacognitive skills. This is logical because to score well on the strange stories task (the Time 3 measure of ToM) children must make inferences about characters within the stories. Indeed, research shows that the two are related with findings for four to seven year olds showing that inference making predicts concurrent emotion understanding (Farina, Albanese, & Pons, 2007). This finding in this chapter provides some further evidence that ToM is more important than general metacognition, as in the metacognition models inference making is making up for the absence of ToM.

4.4.3 Limitations and further direction

An argument made within this chapter is that ToM is important for LC, but not until the age of five and six (Time 3). Yet this finding could be instead due to issues with reliability of the measures at the first two time points when children were especially young i.e. the skills were not captured reliably preventing the models from fitting well. Research with very young children has inherent problems regarding the reliability of measures (Einarsdóttir, 2007; Fargas-Malet, McSherry, Larkin, & Robinson, 2010). Selecting age appropriate measures is always a challenge with this type of research. Moreover, within this study a large battery of measures was administered to these young children within one sitting, which could have affected reliability due to fatigue or disengagement. This said, as these age findings are consistent with past work (e.g. Strasser & Rio, 2014 also did not find ToM to predict LC in the younger years) and that here steps were taken to choose reliable and age appropriate measures at all time points, there is reason to be confident in these results. In order to confirm this, future research should measure these skills in older children (e.g. in Year 2 aged seven and Year 3 aged eight) to assess if even further on, ToM and broad metacognition directly predict LC. If this is the case this would support the idea that it is only when children begin to experience more complex oral passages that metacognition becomes important for their comprehension, and that before this age comprehension is driven by language skills and working memory only.

In this chapter ToM was found only to predict LC aged six, and not in younger children. However, the previous chapter demonstrated that in these same children earlier ToM (one year previously) could be used in a longitudinal model to directly predict LC aged six. This is still consistent with the idea that it is not until the age of six that children require ToM to facilitate their comprehension with more complex passages, but it is unknown what the longitudinal influence of non-social metacognitive skills might be on LC. Future research should focus on longitudinal models comparing the contribution of ToM to non-social metacognition. This will give a clearer answer to the question of whether it is the socially-specific element of ToM or the broad metacognitive nature which is important for assisting LC.

4.4.4 Conclusions

This chapter aimed to determine the specific nature of ToM which is important for assisting LC at three different ages (four years old; Time 1, five years old; Time 2 and six years old; Time 3). The focus was on whether it is the social specificity of ToM, or its general metacognitive nature which helps LC. Either could be important; the social nature because of its ability to assist with the understanding of the social information within a passage such as character's intentions, and the broad metacognitive nature because of its capacity to support the understanding of non-social information such as space, time and objects. Both could underpin the building of a mental model of text (Graesser et al., 1994; Kim & Pilcher, 2016; Kintsch, 1988).To test this, at each of the three time points the fit of SEM models which included either ToM, or a latent variable of three broad metacognitive skills (source monitoring, metalinguistic awareness and metamemory) were compared.

The findings were not straightforward and were not consistent across time points. Although the only model to fit well was a ToM model at Time 3 (when children were aged six), with ToM making a direct contribution to LC after controlling for all other variables in the model, metacognition also made a direct contribution to LC at Time 2 (Model 2b) and was approaching significance at Time 3 (Model 3b). The provides only very limited evidence for the importance of the social element of ToM in LC. Future research should look at these relationships longitudinally, with the hope that this will give a clearer answer to the question of whether it is the socially specific element of ToM or the broad metacognitive nature which is important for assisting LC, or indeed if both are vital. This was the goal of the next chapter (Chapter 5).

5 Theory of mind versus broad metacognition in longitudinal listening comprehension

5.1 Introduction

Findings from the previous two chapters show that there is a relationship between theory of mind (ToM) and listening comprehension (LC), but that this is not a simple or straightforward relationship. Specifically, the previous chapter (Chapter 4) compared the contribution of other types of metacognition to ToM and their ability to predict LC in hierarchical SEM models based on the DIET model. The types of metacognition examined were non-social broad metacognitive skills and included metamemory, source monitoring and metalinguistic awareness. The rationale for this comparison was to determine whether it is the social element of ToM which is important for LC, or instead if it is the broad metacognitive nature which ToM taps. Findings from this chapter showed that models which included a latent variable of these broad non-social types of metacognition were not good fits when children were aged four, five or six. Whereas a model which included ToM was a good fit when children were six years old and ToM directly predicted LC after controlling for all other skills in the model. However, before this at the earlier two time points (when children were four and five years old) ToM did not make a direct contribution to LC. Although these findings do not clearly answer the question of whether it is the socially-specific element of ToM or the broad metacognitive nature which is important for LC in the early years, they do provide more evidence in favour of the social specificity rather than the broad metacognitive nature. It was concluded that the social specificity of ToM is important for LC, but only when children are older and encountering more complex passages and stories which require an understanding of social details.

The social specificity of ToM may be important for LC because gives it a child a better understanding of the desires, intentions and perspectives of a speaker, or the mental states of a protagonist in an oral passage, so that they can form a more detailed mental representation of the passage to comprehend it to a higher level (Graesser et al., 1994; Kim & Pilcher, 2016). Constructing a mental model of the passage is crucial for proficient comprehension (Graesser et al., 1994; Kim & Pilcher, 2016; Kintsch, 1988) and ToM understanding may directly underpin building mental models of the text and passages during LC. On the other hand, broad metacognitive skills could be important for LC because they inform the listener about non-social details in the passage and aid the formation of a mental representation of the passage which includes details about space, objects and time (Graesser et al., 1994). For a further explanation on the role which ToM and metacognition may play in LC see Section 4.1.1 in Chapter 4.

5.1.1 The present study

Given that findings from the previous chapter are not consistent across age, further work is needed to explore the question of what it is about the nature of ToM which is important for assisting LC. The concurrent models in Chapter 4 seem to suggest that the social specificity of ToM is of importance for LC, but longitudinal models may help to confirm this. Chapter 3 demonstrated that concurrent and longitudinal relationships can differ. Longitudinal studies have advantages over cross-sectional studies such as the ability to document the developmental trajectory of a specific skill, identify precursors of an ability, and examine how the relationship between two (or more) related skills interacts and progresses over time (Grammer et al., 2013; Kendeou et al., 2008). Therefore, longitudinal studies can establish the direction and magnitude of causal relationships (Ebert, 2015). In this case, it was hoped that longitudinal models would be able to compare the magnitude of the causal relationships which social (ToM) versus non-social metacognition may have to LC.

Therefore, this chapter aimed to compare the contribution of ToM to broad metacognition in DIET longitudinal models of LC. Firstly, the ability of ToM versus broad metacognition to predict LC across one year, from ToM and metacognition aged four (Time 1) to LC aged five (Time 2), and secondly their ability to predict LC even further across time (22 months) from ToM and metacognition aged four (Time 1) to LC aged six (Time 3). For each of these longitudinal timeframes the fit of two models were compared; (a) models which just included ToM, (b) models which included a latent variable of metacognition (comprising metamemory, metalinguistic awareness and source monitoring). In addition to the model fits, whether ToM or broad metacognition made a direct significant contribution to LC (after controlling for all other skills in the models) was assessed. As with previous chapters age and non-verbal ability were also controlled for.

It was hypothesized that:

1) Across 12 months, when children were four (Time 1) to when they were five (Time 2), a model which included just ToM (Model 1-2a) would have a better fit than a model which included a latent variable of broad metacognition (consisting of metamemory, metalinguistic awareness and source monitoring; Model 1-2b). Moreover, it was hypothesized that in Model 1-2a ToM would directly predict LC after controlling for all other skills in the model.
2) Across 22 months, when children were four (Time 1) to when they were six (Time 3) a model which included just ToM (Model 1-3a) would have a better fit than a model which included a latent variable of broad metacognition (consisting of metamemory, metalinguistic awareness and source monitoring; Model 1-3b). Moreover, it was hypothesized that in Model 1-3a ToM would directly predict LC after controlling for all other skills in the model.

These hypotheses suggest that it is the social nature of ToM which is important for LC rather than the broad metacognitive nature which it also taps.

5.2 Method

5.2.1 Participants

As with previous chapters the models presented here included different numbers of participants because Cohort 2 was not followed to Time 3 and could therefore not appear in analysis for models predicting Time 3 LC. For more information on the participants including recruitment see Section 2.3 in Chapter 2. For the first set of models (Model 1-2a and Model 1-2b) which used Time 1 skills to predict Time 2 LC there were 162 participants from both cohorts. For the second set of models (Model 1-3a and Model 1-3b) which used Time 1 skills to predict Time 3 LC there were 107 participants from just Cohort 1. Descriptive statistics of the participants included in each of the models is shown in Table 5.1.

Table 5.1

Descriptive statistics of participants in each of the model

Models	**N**	**N male**	**N female**	**Mean age(SD)[a]**
All 1-2 models[b]	162	82	80	Time 1: 4;1 (4.55) Time 2: 5;1 (4.43)
All 1-3 models[c]	107	54	53	Time 1: 4;3 (3.64) Time 3: 6;1 (3.65)

Note. [a] Mean age in Years; Months [b] This includes Model 1-2a and Model 1-2b
[c] This includes Model 1-3a and Model 1-3b.

5.2.2 Materials and measures

Table 5.2 shows the measures administered at each time point. For comprehensive details on each of these measures refer to Section 2.5 in Chapter 2.

Table 5.2

Measures administered to children at each of the three time points

Skill	Measure	Time 1	Time 2	Time 3
Listening comprehension	NARA	-	✓	✓
	OWLS	-	✓	✓
Vocabulary	BPVS-III	✓	-	-
Syntactic Knowledge	Sentence Structure (CELF-Preschool 2uk)	✓	-	-
Theory of mind	Unexpected contents task	✓	-	-
	Unexpected locations task	✓	-	-
Metamemory	Metamemory task	✓	-	-
Source monitoring	Tunnel task	✓	-	-
Metalinguistic awareness	Synonym judgement task	✓	-	-
Working memory	Reverse word span task	✓	-	-
Non-verbal ability	Block design subset of WPPSI-II	✓	-	-

Note: Many of these measures were administered at Time 2 or Time 3 but only the specific time point measures used in the analysis for this chapter are listed here. For further details see Chapter 2

5.2.3 Procedure

Participants in Cohort 1 were administered the measures in Table 5.2 across the three time points within two 20 minute sessions in a quiet place in or just outside their school classroom. Participants in Cohort 2 were administered the measures in Table 5.2 at the first two time points in a single session in a quiet place in their own home. For complete procedural details refer to Chapter 2 Sections 2.7 and 2.8.

5.2.4 Analysis

As with the two previous chapters, the primary data analytical strategy was structural equation modelling (SEM) using AMOS Version 25 (Arbuckle, 2016). For each of the models tested, first descriptive statistics were computed and then initial correlational and regression analysis were carried out. During SEM analysis, a latent variable was created for LC using two measures (NARA and OWLS) at Time 2 and Time 3. For the broad metacognitive models, a latent variable for broad metacognition was also created which included the three types of non-social metacognition (metamemory, source monitoring and metalinguistic awareness). All other language and cognitive skills were assessed by a single measure for each construct, and so therefore observed variables were used. For justification of this analysis and further details of its use see Section 3.2.4 in Chapter 3.

5.3 Results

Results from the two longitudinal time frames are outlined separately. The first two models assessed the ability of Time 1 skills to predict Time 2 LC, comparing the ability of a model which included just ToM (Model 1-2a), to a model which included a latent variable of non-social metacognitive measures (Model 1-2b). The second set of models assessed the ability of Time 1 skills to predict Time 3 LC, comparing the ability of a model which included just ToM (Model 1-3a), to one including a latent variable of non-social metacognitive measures (Model 1-3b).

5.3.1 Time 1 skills predicting Time 2 listening comprehension (1-2 models)

5.3.1.1 Descriptive statistics and preliminary analysis

Table 5.3 shows descriptive statistics for all measures used in the 1-2 models. It should be noted that a ToM composite consisting of the two ToM measures was computed by summing scores from the two ToM measures from Time 1 (unexpected contents and unexpected locations). This composite was used to give a richer measure of ToM and was justified as the individual measures correlated significantly ($r = .39$, $p < .001$) and because this type of composite is often used (Atkinson et al., 2017; Guajardo & Cartwright, 2016; Ruffman et al., 2002). This composite was used in all further analysis.

Table 5.3

Descriptive statistics for all measures in the 1-2 models [a]

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 1								
Vocabulary	BPVS-II	162	168	10-102	54.61	17.15	-.14	-.49
Syntactic knowledge	CELF-Preschool	161	22	0-21	12.14	4.33	-.68	.03
Theory of mind	ToM composite	158	5	0-5	2.03	1.72	.32	-1.10
Metamemory	Metamemory task	161	4	0-3	.84	.97	.80	-.56
Source monitoring	Tunnel task	162	6	0-6	3.65	1.75	-.42	-.66
Metalinguistic awareness	Synonym judgement	161	4	1-4	3.32	.96	-1.02	-.40
Working memory	Reverse word span	160	9	0-9	1.07	2.21	1.94	2.52
Non-verbal ability	Block design	162	40	0-32	16.56	5.86	-.47	.37
Time 2								
Listening comprehension	NARA	161	44	0-13	3.86	3.62	.57	-.79
	OWLS	162	130	24-98	59.98	15.06	.49	.26
Age[b]	-	162	-	50-69	61.18	4.43	-.17	-.69

[a] This includes Model 1-2a and Model 1-2b [b] Age in months

Correlations between all these measures are displayed in Table 5.4. All metacognitive measures at Time 1 (ToM, metamemory, metalinguistic awareness and source monitoring) were weakly to moderately related to both the LC measures (NARA and OWLS) at Time 2. Preliminary regression analyses were then performed to assess whether the Time 1 metacognitive measures had predictive power in calculating LC at Time 2, and therefore whether it was logical to proceed with the models. Separate linear regressions were carried out for each individual metacognitive measure (Time 1) for its ability to predict NARA and OWLS scores (the Time 2 LC outcome measures). All variables significantly predicted Time 2 NARA scores, with ToM shown to predict 7.5% ($p = .001$) of unique variance, metamemory 8.9% ($p < .001$), metalinguistic awareness 13.2% ($p < .001$), and source monitoring 9% ($p < .001$). All variables also significantly predicted Time 2 OWLS scores, with ToM shown to predict 8.9% ($p < .001$) of unique variance, metamemory 9% ($p < .001$), metalinguistic awareness 3.4% ($p = .02$), and source monitoring 11.5% ($p < .001$). Given this, full modelling using these variables and LC as an outcome measure (latent variable of NARA and OWLS) was carried out.

Table 5.4

Correlation matrix of all measures in the 1-2 models[a]

	1	2	3	4	5	6	7	8	9	10	11
1.Time 2 NARA	-										
2. Time 2 OWLS	.56**	-									
3. Time 1 Vocabulary	.35**	.38**	-								
4. Time 1 Syntactic knowledge	.38**	.37**	.64**	-							
5. Time 1 Theory of mind	.27**	.30**	.39**	.48**	-						
6. Time 1 Metamemory	.30**	.30**	.42**	.47**	.32**	-					
7. Time 1 Metalinguistic awareness	.36**	.19*	.31**	.43**	.26**	.30**	-				
8. Time 1 Source monitoring	.30**	.34**	.47**	.62**	.44**	.48**	.43**	-			
9. Time 1 Working memory	.20*	.21**	.42**	.40**	.36**	.42**	.21**	.33**	-		
10. Time 1 Non-verbal ability	.17*	.29**	.61**	.48**	.30**	.36**	.44**	.43**	.37**	-	
11. Age	.01	.06	.09	.15	.06	.21**	.06	.21**	.19*	.10	-

[a] This includes Model 1-2a and Model 1-2b.

5.3.1.2 SEM analysis

The longitudinal models were then fitted to the data using SEM with Time 2 LC as an outcome measure. One model which included just Time 1 ToM (Model 1-2a as shown in Figure 5.1), and one that included a Time 1 latent variable of broad metacognition (Model 1-2b as seen in Figure 5.3; comprising source monitoring, metamemory and metalinguistic awareness). Prior to SEM analysis, missing data (see Table 5.4 for details of which measures) was imputed using expectation maximization (EM) in SPSS. This method of data imputation was used because missing data was minimal (three cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random (p =.09).

5.3.1.2.1 Model fits

All variables in Model 1-2a met the univariate normality assumptions for SEM (as stated by Chou & Bentler, 1995; Kline, 2005). Multivariate normality assumptions were also checked and were not violated as demonstrated by a Mardia's coefficient of multivariate critical ratio of .01. Model 1-2a[8] had a good to excellent fit, $\chi^2(6)$ = 10.42, p = .11; CFI = .99, TLI = .94, and RMSEA = .07. Seven multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05), when these seven participants were removed and the model re-run, the model fit indices changed

[8] Note that Model 1-2a is identical to Model 2 in Chapter 3. Here the model is being compared against models which include other types of metacognition, whereas in Chapter 3 it was used to validate an existing model.

only slightly to $\chi^2(6) = 8.38$, $p = .21$; CFI = .99, TLI = .96, and RMSEA = .05. Therefore, the model with all 162 participants included is presented as the final model.

All variables in Model 1-2b met the univariate normality assumptions for SEM (as stated by Chou & Bentler, 1995; Kline, 2005). Multivariate normality assumptions were also checked and were not violated as demonstrated by a Mardia's coefficient of multivariate critical ratio of .79. Model 1-2b had a poor fit, $\chi^2(16) = 33.01$, $p = .01$; CFI = .97, TLI = .91, and RMSEA = .08. Eight multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05), when these seven participants were removed and the model re-run, the model fit indices changed only slightly to $\chi^2(15) = 30.20$, $p = .01$; CFI = .97, TLI = .90, and RMSEA = .08. Therefore, the model with all 162 participants included is presented as the final model.

5.3.1.2.2 Standardised paths

For Model 1-2a standardised path coefficients of the model are shown in Figure 5.2. As presented in Figures 5.2, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary ($\beta = .22$, $p < .001$) and grammatical knowledge ($\beta = .25$, $p < .001$), but not directly to later LC ($\beta = .02$, $p = .90$). Vocabulary was not significantly related to ToM ($\beta = .12$, $p = .24$). Grammatical knowledge was related to ToM ($\beta = .40$, $p < .001$). Vocabulary ($\beta = .47$, $p = .03$) and grammatical knowledge ($\beta = .48$, $p = .02$) were also directly related to later LC. ToM was significantly independently related to LC (β = .30, p = .04). A total of 47% of variance in Time 2 LC was explained by the Time 1 skills in the model.

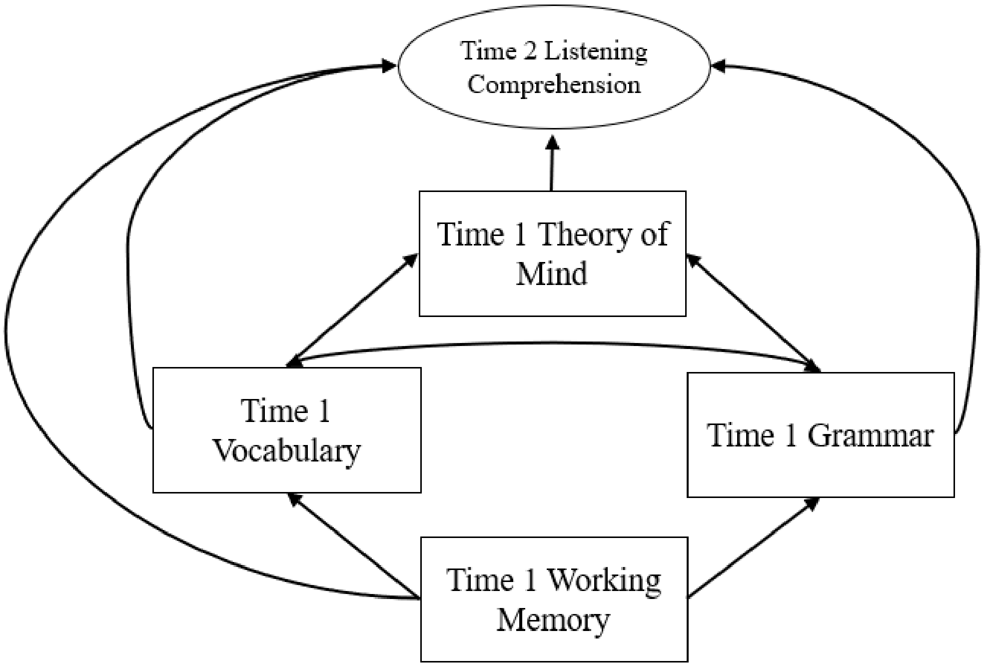


Figure 5.1: Hypothesized Model 1-2a after controlling for age and earlier non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

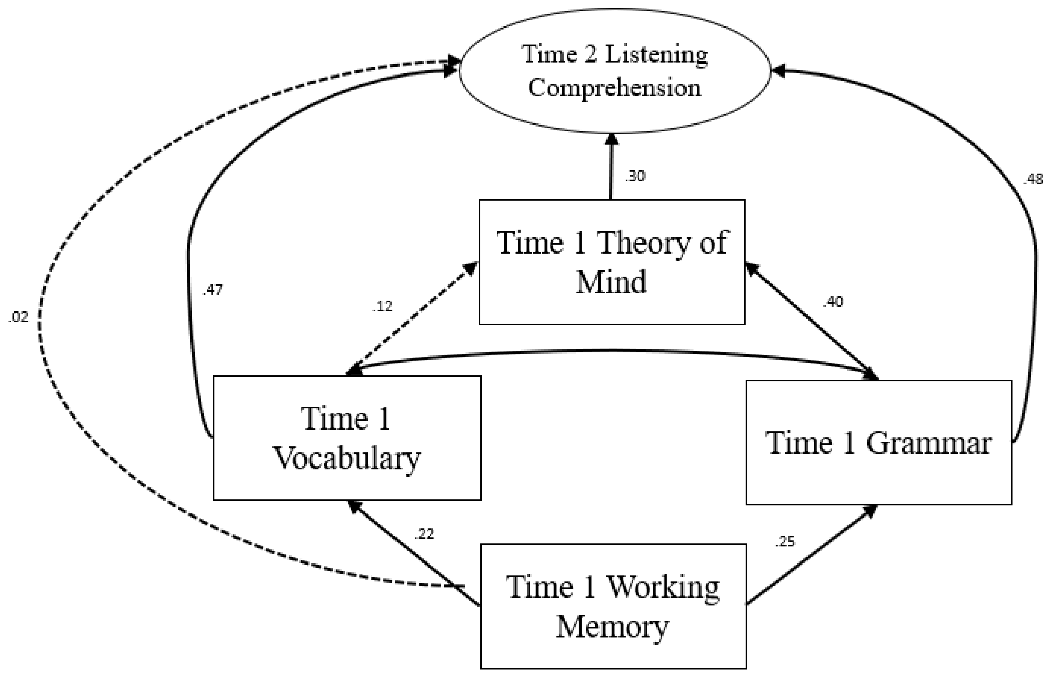


Figure 5.2: Model 1-2a with standardised path coefficient weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

For Model 1-2b standardised path coefficients of the model are shown in Figure 5.4. As presented in Figures 5.4, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary ($\beta = .22$, $p < .001$) and grammatical knowledge ($\beta = .26$, $p < .001$), but not to later LC ($\beta = .02$, $p = .88$). Vocabulary was not significantly related to metacognition ($\beta = .09$, $p = .46$) but grammatical knowledge was related to metacognition ($\beta = .71$, $p < .001$). Vocabulary ($\beta = .38$, $p = .04$) was directly related to later LC, but and grammatical knowledge ($\beta = -.05$, $p = .87$) was not. Metacognition was not significantly independently related to LC ($\beta = .82$, $p = .06$). A total of 24% of variance in Time 2 LC was explained by the Time 1 skills in the model.

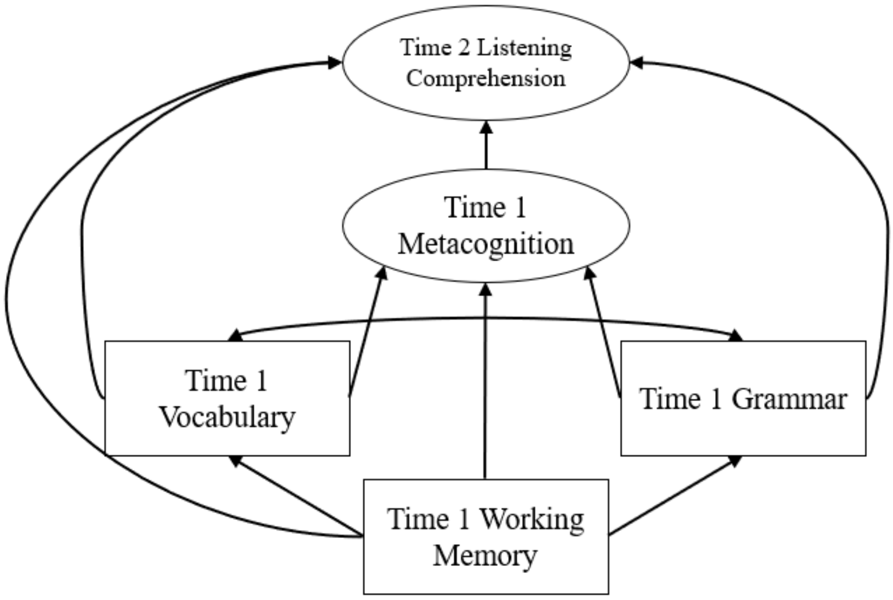


Figure 5.3: Hypothesized Model 1-2b after controlling for age and earlier non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

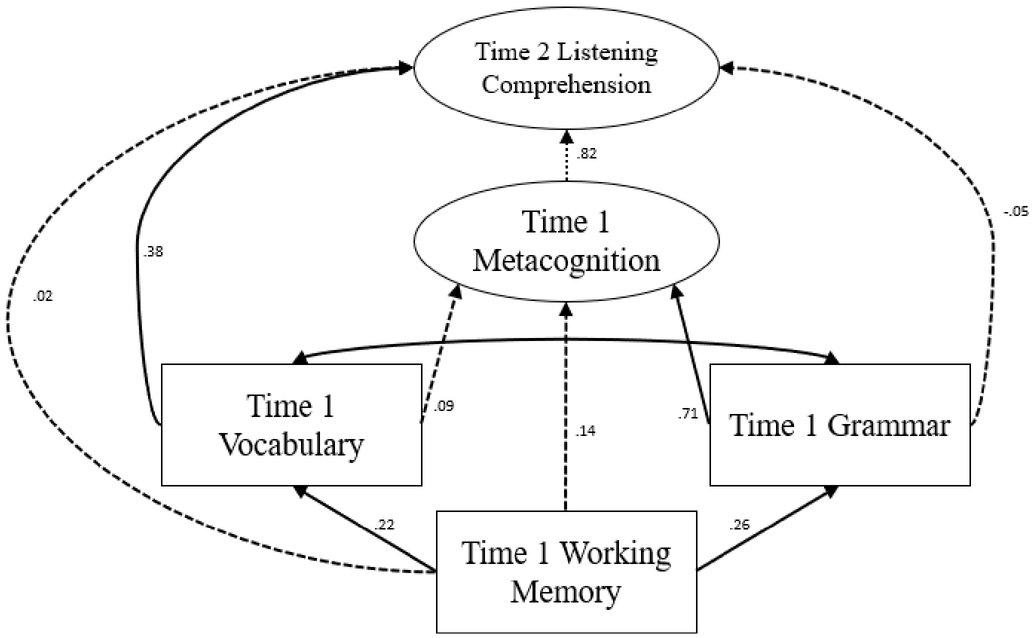


Figure 5.4: Model 1-2b with standardised path coefficient weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

5.3.2 Time 1 skills predicting Time 3 listening comprehension (1-3 models)

5.3.2.1 Descriptive statistics and preliminary analysis

Table 5.5 shows descriptive statistics for all measures used in the 1-3 models. As with the previous models, it should be noted that a ToM composite consisting of the two ToM measures was computed by summing scores from the two ToM measures from Time 1 (unexpected contents and unexpected locations). This composite was used to give a richer measure of ToM.

Correlations between all these measures are displayed in Table 5.6. All metacognitive measures at Time 1 (ToM, metamemory, metalinguistic awareness and source monitoring) were weakly to moderately related to both LC measures (NARA and OWLS) at Time 3. Preliminary regression analyses were then performed to assess whether the Time 1 metacognitive measures had predictive power in calculating LC at Time 3 and therefore whether it was logical to precede with the models. Separate linear regressions were carried out for each individual metacognitive measure (Time 1) for its ability to predict NARA and OWLS scores (the Time 3 LC outcome measures). All variables significantly predicted Time 3 NARA scores, with ToM shown to predict 9.8% ($p = .001$) of unique variance, metamemory 11.6% ($p < .001$), metalinguistic awareness 7.2% ($p = .003$), and source monitoring 13.8% ($p < .001$). All variables also significantly predicted Time 3 OWLS scores, with ToM shown to predict 11.7% ($p < .001$) of unique variance, metamemory 5% ($p = .01$), metalinguistic awareness 8.2% ($p = .002$), and source monitoring 21.9% ($p < .001$). Given this, full modelling using these variables and LC as an outcome measure was carried out.

Table 5.5

Descriptive statistics for all measures in the 1-3 models[a]

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 1								
Vocabulary	BPVS-II	107	168	10-102	51.59	18.12	.13	-.43
Syntactic knowledge	CELF-Preschool	107	22	0-21	11.64	4.61	-.55	-.29
Theory of mind	ToM composite	104	5	0-5	1.94	1.70	.25	-.96
Metamemory	Metamemory task	106	4	0-3	.87	1.00	.76	-.70
Source monitoring	Tunnel task	107	6	0-6	3.64	1.78	-.47	-.60
Metalinguistic awareness	Synonym judgement	107	4	1-4	3.21	1.00	-.76	-.86
Working memory	Reverse word span	106	9	0-8	1.00	2.12	2.02	2.75
Non-verbal ability	Block design	107	40	0-32	15.34	6.50	-.12	-.13
Time 3								
Listening comprehension	NARA	106	44	1-16	6.58	3.38	.69	-.30
	OWLS	107	130	36-94	63.66	12.32	.38	-.55
Age [b]	-	107	-	66-80	72.95	3.68	-.11	-1.10

[a] This includes Model 1-3a and Model 1-2b [b] Age in months

Table 5.6

Correlation matrix of all measures in the 1-3 models[a]

	1	2	3	4	5	6	7	8	9	10	11
1.Time 3 NARA	-										
2. Time 3 OWLS	.62**	-									
3. Time 1 Vocabulary	.51**	.40**	-								
4. Time 1 Syntactic knowledge	.41**	.47**	.67**	-							
5. Time 1 Theory of mind	.33**	.35**	.39**	.46**	-						
6. Time 1 Metamemory	.35**	.24*	.44**	.45**	.30**	-					
7. Time 1 Metalinguistic awareness	.28**	.30**	.26**	.42**	.21*	.27**	-				
8. Time 1 Source monitoring	.38**	.48**	.54**	.69**	.49**	.48**	.39**	-			
9. Time 1 Working memory	.49**	.36**	.51**	.41**	.24*	.31**	.16	.37**	-		
10. Time 1 Non-verbal ability	.46**	.32**	.62**	.50**	.28**	.50**	.44**	.50**	.40**	-	
11. Age	.22*	.13	.36**	.29**	.04	.19*	.20*	.19*	.29**	.41**	-

[a] This includes Model 1-2a and Model 1-2c.

5.3.2.2 SEM analysis

The longitudinal models were then fitted to the data using SEM with Time 3 LC as an outcome measure. One model which included just Time 1 ToM (Model 1-3a as shown in Figure 5.5; also see Model 3 in Chapter 3) and one that included a Time 1 latent variable of broad metacognition (Model 1-3b as seen in Figure 5.7; comprising source monitoring, metamemory and metalinguistic awareness). Prior to SEM analysis, missing data (see Table 5.6 for details of which measures) was imputed using expectation maximization (EM) in SPSS. This method of data imputation was used because missing data was minimal (three cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random (p =.22).

5.3.2.2.1 Model fits

All variables in Model 1-3a met the univariate normality assumptions for SEM (as stated by Chou & Bentler, 1995; Kline, 2005). Multivariate normality assumptions were also checked and were not violated as demonstrated by a Mardia's coefficient of multivariate critical ratio of -.43. Model 1-3a[9] had an excellent fit, χ^2 (6) = 8.51, p = .20; CFI = .99, TLI = .96, and RMSEA = .06. Six multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05), when these six

---

[9] Note that Model 1-3a is identical to Model 3 in Chapter 3. Here the model is being compared against models which include other types of metacognition, whereas in Chapter 3 it was used to validate existing models.

participants were removed and the model re-run, the model fit indices changed only slightly to χ^2 (6) = 6.95, *p* = .33; CFI = 1.00, TLI = .98, and RMSEA = .04. Therefore, the model with all 107 participants included is presented as the final model.

For Model 1-3b all variables met the univariate normality assumptions for SEM (As stated by Chou & Bentler, 1995; Kline, 2005). Multivariate normality assumptions were also checked and were not violated as demonstrated by a Mardia's coefficient of multivariate critical ratio of .45. Model 1-3b had a good fit, χ^2 (16) = 23.01, *p* = .11, CFI = .98, TLI = .95, and RMSEA = .06. Five multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05), when these five participants were removed and the model re-run, the model fit indices changed only slightly to χ^2 (15) = 23.25, *p* = .08; CFI = .98, TLI = .94, and RMSEA = .07. Therefore, the model with all 107 participants included is presented as the final model.

5.3.2.2.2 Standardised paths

Standardised path coefficients for Model 1-3a are shown in Figure 5.6. As presented in Figure 5.6 after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .30, p < .001) and grammatical knowledge (β = .24, p < .01) as well as directly to later LC (β = .45, p = .003). Vocabulary was not significantly related to ToM (β = .17, p = .16). Grammatical knowledge was related to ToM (β = .43, p < .001). Neither vocabulary (β = .25, p =

.20) or grammatical knowledge (β = .14, p = .48) were directly related to LC. Finally, ToM was not significantly independently related to LC (β = .25, p = .11) after accounting for everything else. A total of 30% of variance in Time 3 LC was explained by the Time 1 skills in the model.

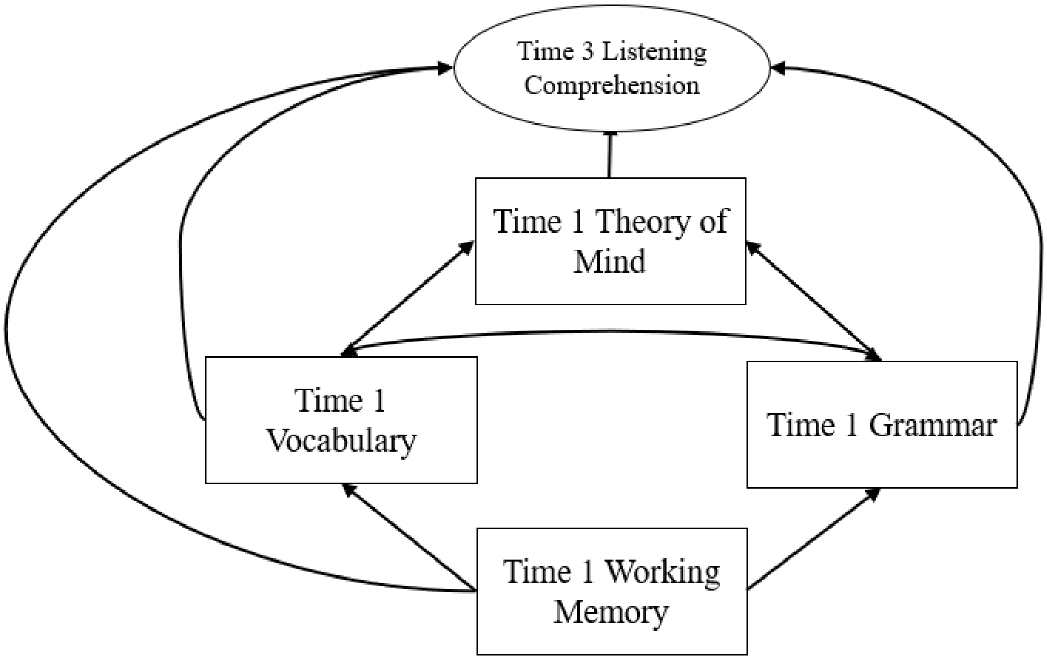


Figure 5.5: Hypothesized Model 1-3a after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

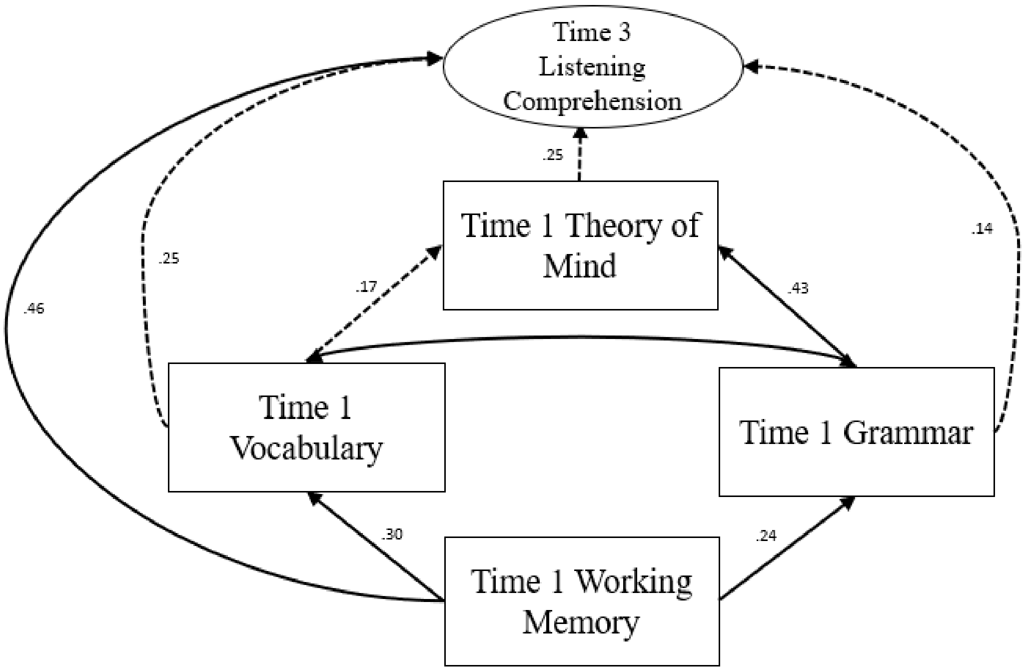


Figure 5.6: Model 1-3a with standardised path coefficient weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

For Model 1-3b standardised path coefficients of the model are shown in Figure 5.8. As presented in Figures 5.8, after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary ($\beta = .30$, $p < .001$) and grammatical knowledge ($\beta = .24$, $p = .01$) and directly to later LC ($\beta = .42$, $p = .01$). Vocabulary was not significantly related to metacognition ($\beta = .06$, $p = .67$) but grammatical knowledge was related to metacognition ($\beta = .76$, $p < .001$). Vocabulary ($\beta = .26$, $p = .20$) was not directly related to later LC and neither was grammatical knowledge ($\beta = -.07$, $p = .82$). Metacognition was not significantly independently related to LC ($\beta = .46$, $p = .21$). A total of 31.5% of variance in Time 3 LC was explained by the Time 1 skills in the model.

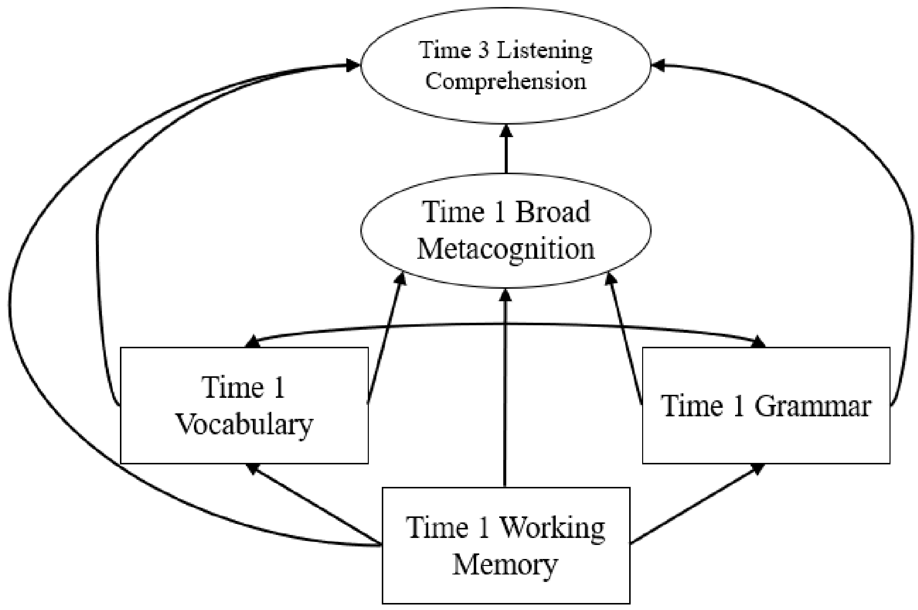


Figure 5.7: Hypothesized Model 1-3b after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

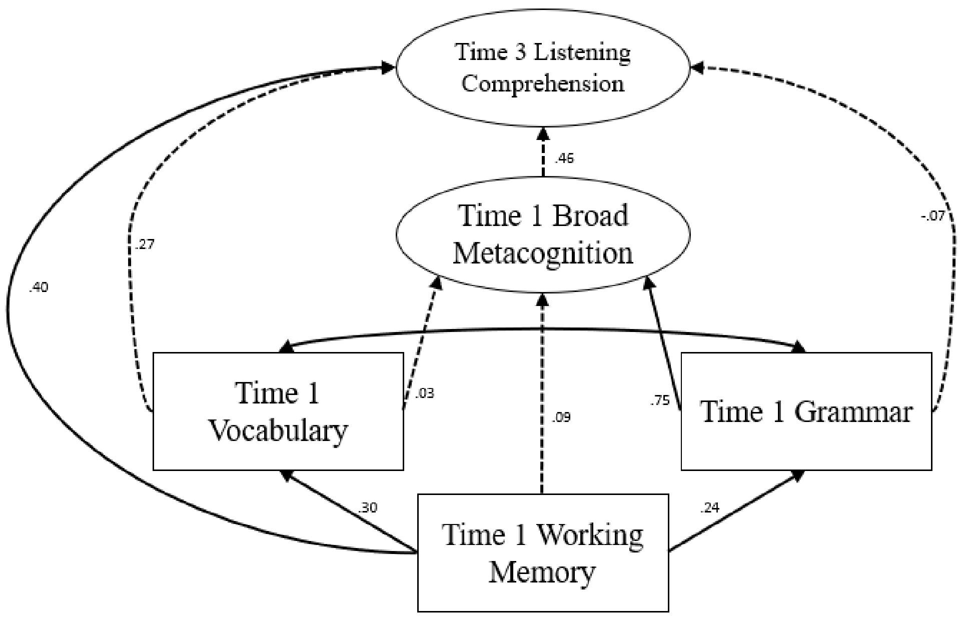


Figure 5.8: Model 1-3b with standardised path coefficient weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

5.3.3 Further analysis

Further analysis was carried out in order to produce further information not gained through the main analysis.

5.3.3.1 Children with English as an additional language

A proportion of the sample had English as an additional language (EAL). This could have affected results as EAL children are shown to have weaker oral language skills than their monolingual peers (Burgoyne et al., 2009). Therefore, multigroup analysis was used in AMOS to compare the fit of all models for EAL participants to English only speaking participants. This was done using a chi-square difference test whereby a non-significant chi-square shows that the model fit is the same for both groups (Dimitrov, 2006). For each of the models a non-significant chi-square demonstrated that the fit was no different for EAL and English only speaking participants. Model 1-2a: $\chi^2(9) = 9.13$, $p = .43$, Model 1-2b: $\chi^2(29) = 32.65$, $p = .29$, Model 1-3a: $\chi^2(9) = 1.59$, $p = .99$, Model 1-3b: $\chi^2(29) = 26.87$, $p = .58$.

5.3.3.2 Power analysis

The models run at different longitudinal time frames had different sample sizes and so power analyses were run to check each had enough power. The first set of models which used Time 1 skills to predict Time 2 LC had 162 participants, whereas the second set of models (Time 1 skills predicting Time 3 LC) only had 107 participants (see Section 5.2.1). As noted in Chapter 3, SEM analysis is sensitive to sample size (E. J. Wolf et al., 2013) therefore a 34% reduction in participant numbers could have influenced model outcome. Power analysis of each model was calculated using an online power calculator (Soper, 2017). This analysis showed each of the four models to have sufficient power as the observed statistical power reached at least .80 for each (the cut off for minimum power as advocated by P. Cohen et al., 2014). However, a calculation of sample size required (Soper, 2015; Westland, 2010) suggested that for these 1-3 models the required minimum sample size was either not quite or was just met, as 97 were required for Model 1-3a, 108 for Model 1-3b (107 were used in actual model).

5.4 Discussion

The aim of this chapter was to use longitudinal direct and indirect models of LC to determine if it is the social nature of ToM which is important for LC, or instead if it is the general metacognitive nature which is of greatest importance. To do this the ability of ToM versus other non-social metacognitive skills (metamemory, source monitoring, metalinguistic awareness) at predicting LC a year later, and a further 10 months after that (22 months later in total) were compared. Longitudinal SEM models, which included either (a) just ToM or (b) a latent variable consisting of non-social metacognitive skills, were fitted to the data. Models were firstly fitted for Time 1 (aged four) skills predicting Time 2 (aged five), and then for Time 1 skills predicting Time 3 LC (when children were six years old).

This longitudinal extension of the previous chapter (Chapter 4) was important because the previous chapter was unable to equivocally answer whether the social specificity of ToM is what is of importance for LC. It was hoped that the longitudinal models would give a clearer answer by comparing the magnitude of the causal relationships which social versus non-social metacognition may have to LC. Longitudinal studies have advantages over cross-sectional studies such as the ability to document the developmental trajectory of a specific skill, identify precursors of an ability, and examine how the relationship between two (or more) related skills interacts and progresses over time (Grammer et al., 2013; Kendeou et al., 2008). Therefore, longitudinal studies can establish the direction and magnitude of causal relationships (Ebert, 2015).

It was hypothesized that it is the social nature of ToM which assists LC and therefore for each longitudinal timeframe (e.g. Time 1 skills predicting Time 2 LC, and Time 1 skills predicting Time 3 LC), models which included just ToM would have better fit than models which included broad metacognition. Furthermore, that earlier ToM would directly predict later LC after accounting for all other skills in the model. This was hypothesised because the understanding of the social details may aid the listener in constructing a mental model of the passage which is crucial for proficient comprehension (Graesser et al., 1994; Kim & Pilcher, 2016; Kintsch, 1988).

5.4.1 Main findings

The hypotheses were only partially supported because as with the cross-sectional analysis in the previous chapter, findings were not clear-cut, and it was not fully possible to declare the social nature of ToM nor the general metacognitive nature as more important than the other. The hypothesis was supported for Time 1 skills predicting Time 2 LC, as here the ToM model (Model 1-2a) was good fitting, and Time 1 ToM directly predicted Time 2 LC after controlling for all other skills whereas the model which included non-social metacognition (Model 1-2b) was not well fitting and broad metacognition did not make a significant direct contribution to LC a year later.

These results were not consistent for predicting LC a further 10 months later (skills aged four predicting LC aged six) as here, despite the ToM model (Model 1-3a) having a good fit, Time 1 ToM did not directly predict Time 3 LC after controlling for all other skills. This same pattern was seen for the metacognitive model too (Model 1-3b) as it too was well fitting but broad metacognition was unable to directly significantly predict LC 22 months later. Therefore, neither models which included ToM nor broad metacognition can be said to be superior to the other. Sample size may explain the contrasting findings across longitudinal time frames as there was a 34% reduction in participant numbers from the models which included Time 1 skills predicting Time 2 LC to the models including Time 1 skills predicting Time 3 skills. Indeed, calculations of sample sizes required (Soper, 2015; Westland, 2010) preformed in Section 5.3.3.2 showed that for the Time 1-Time 3 models which included only 107 participants the required number of participants was either only just met (in the case of Model 1-3a) or were not quite met (in the case of Model 1-3b). Required number of participants is calculated based on the number of variables and number of paths in a model (Soper, 2015; Westland, 2010). Given this, future research should replicate the current study but with an increased number of participants for these longer longitudinal models.

Despite these contrasting findings across longitudinal timeframes what can be said, is that there is slightly more evidence for the social nature of ToM being important for LC than the broad metacognitive nature. The social nature of ToM could be important for LC because it could give a child a better understanding of the desires, intentions and perspectives of a speaker, or the mental states of a protagonist in an oral passage so that the child can form a more detailed mental representation of the passage to comprehend the passage to a higher level (Graesser et al., 1994; Kim & Pilcher, 2016). Further to this, constructing a mental model of the passage is crucial for proficient comprehension and ToM ability may directly underpin building mental models of the text and passages. Yet given that findings were not consistent across longitudinal timeframes, further work is needed to fully determine if it is the social nature which is most important for LC.

5.4.2 Limitations and future direction

As already mentioned, sample size may have been a limitation at the last time point (Time 3 when children were six years old) and so future work should replicate the current study but with an increase of participants. In addition to this, given the discrepancies in findings between longitudinal time frames it would be useful to assess children at different ages, for example aged eight or nine to explore if earlier ToM and metacognition can predict LC at this age. This would give further understanding to how this relationship works longitudinally.

5.4.3 Chapter conclusions

This chapter showed that longitudinally from four years to five years ToM can better predict LC (when controlling for other skills known to be important for LC) than metacognition not social in nature. This provides some evidence that it is the social element of ToM which is important for assisting LC rather than the broad metacognitive nature. This could be because ToM could give children a better understanding of the desires, intentions and perspectives of a speaker, or the mental states of a protagonist in an oral passage and can help a child to build mental models of the text. However, this finding was not replicated further across time as ToM was no better than broad metacognition at predicting LC from aged four to aged six. This could be because of the lower number of participants in these models (162 reduced to 107) therefore more research with a larger sample is needed before further conclusions can be drawn.

5.5 Review of the role of theory of mind in listening comprehension

This chapter and the preceding two sought to gain a deeper understanding of the role played by ToM in LC. Chapter 3 successfully replicated the DIET model of Kim

(2015; 2016; 2017) showing that a concurrent model which included ToM can predict LC at the age of six years old, and importantly ToM can directly predict LC after controlling for all other skills in the model. Chapter 3 was also able to extend the DIET model longitudinally showing that this same model had a good fit for skills aged four predicting LC aged five, with ToM aged four also shown to make a direct contribution to LC aged five. Yet this model was not replicated for predicting LC a further 10 months later (aged six), nor (as Chapter 4 showed) concurrently aged four or five.

Chapter 4 also showed that non-social metacognition (a latent variable consisting of source monitoring, metalinguistic awareness and metamemory) was not as good at predicting LC as ToM at four, five or six years old, as models which included broad metacognition did not have a good fit. This provides evidence that it is the social nature of ToM which is important for facilitating LC rather than the broad metacognitive nature of ToM. Adding to this, the current chapter found that longitudinally from four years to five years ToM can better predict LC (when controlling for other skills in the model) than other types of metacognition not social in nature. Again however, this was not replicated a further 10 months across time. For clarity Table 5.7 summarises the fits of the models in Chapters 3-5.

Table 5.7

A summary of the fit of models in Chapters 3-5

Model	Fit	Ability to predict LC
Cross-sectional models		
T1 ToM predicting T1 LC	Not a good fit	ToM not significantly related to LC
T2 ToM predicting T2 LC	Not a good fit	ToM not significantly related to LC
T3 ToM predicting T3 LC	Good fit	ToM significantly related to LC
T1 metacognition predicting T1 LC	Not a good fit	Metacognition not significantly related to LC
T2 metacognition predicting T2 LC	Not a good fit	Metacognition significantly related to LC
T3 metacognition predicting T3 LC	Not a good fit	Metacognition significantly related to LC
Longitudinal models		
T1 ToM predicting T2 LC	Good fit	ToM significantly related to LC
T1 metacognition predicting T2 LC	Not a good fit	Metacognition not significantly related to LC
T1 ToM predicting T3 LC	Good fit	ToM not significantly related to LC
T1 metacognition predicting T3 LC	Good fit	Metacognition not significantly related to LC

Note. T1 = Time 1, T2 = Time 2, T3 = Time 3. Metacognition = a latent variable of source monitoring, metamemory and metalinguistic awareness

Taking the results from the past three chapters together it can be concluded that there is partial support for the role of the social understanding element of ToM in LC both concurrently and longitudinally. Concurrently ToM seems to only be important later on when children are six years or older. This could be because until this age children are not encountering complex oral language and so they do not require help with social details to comprehend oral passages and only language skills are required. Longitudinally, ToM was only found to predict LC a year later and not 22 months later, but this may be due to a smaller sample size at this later time point. When comparing ToM to non-social metacognition (a latent variable consisting of source monitoring, metalinguistic awareness and metamemory) for its ability to directly predict LC in a model, ToM was found to be a better predictor concurrently aged six and longitudinally from aged four to aged five. This provides some evidence that it is the social element of ToM which is important for assisting LC rather than broad metacognitive skills. A likely explanation for this is that social understanding will help a child with social details in a passage and aid them to construct an advanced mental picture of a passage which includes social details such as character intentions.

The current research is the first to both extend the DIET model longitudinally and to assess the specific component of ToM which is helpful for LC. However, further research is needed to cement these conclusions. Future research should include larger sample sizes to meet with the demands of SEM analysis, and it should follow children even further across time when they are experiencing more complex oral language (e.g. when they are aged eight or nine). Further work should also explore whether these findings can be transferred to reading comprehension. This was the aim of the next chapter.

6 Theory of mind and broad metacognition in models of reading comprehension

6.1 Introduction

The previous three chapters of this thesis have focused on listening comprehension (LC) and the role which theory of mind (ToM) plays in its development, but there is also evidence to suggest that ToM may be important for the development of reading comprehension. Reading comprehension (RC), the ability to read text, process it, and take meaning from it (Snowling & Hulme, 2008), is one of the fundamental aims of primary school education but RC is a complex skill consisting of many components (Tobia & Bonifacci, 2015). Due to the importance of RC for education, employment, social, and cultural purposes (Florit & Cain, 2011) much research has been conducted to determine its component skills.

6.1.1 Component skills of reading comprehension

The most influential model of reading, The Simple View of Reading (SVR), posits that RC is a product of two components: linguistic (or language) comprehension and decoding (Gough & Tunmer, 1986). Linguistic comprehension is defined by Gough and Tunmer as the ability to interpret sentences and discourse presented orally (although see Section 1.2.1 in Chapter 1 for a discussion of this definition and how other phrases are used synonymously and interchangeably within the literature). Decoding is the ability to read isolated words quickly accurately and silently (Gough & Tunmer, 1986). The model states that an individual must be competent in both these skills to be a proficient reader, and the absence of one can result in reading difficulties.

Research into the SVR suggests that linguistic comprehension itself also consists of sub-components, and these may include the oral language skills of listening comprehension, vocabulary, and syntactic knowledge (Atkinson et al., 2017; Cain, 2017; Foorman et al., 2015). Regarding the sub-skill of LC research shows that although LC is an important competency in its own right (Hogan et al., 2014) it also makes a key contribution to RC as a component skill of linguistic comprehension. Indeed, it is well documented that RC is dependent on LC (e.g. Cutting, Materek, Cole, Levine, & Mahone, 2009; Gough & Tunmer, 1986; Hoover & Gough, 1990; Kendeou et al., 2009; Ouellette & Beers, 2010; Roth et al., 2002; Storch & Whitehurst, 2002). Longitudinal studies show that LC contributes to later RC and has been shown to make a stronger contribution to RC than predictors of decoding skills, such as phonological awareness (Bianco, 2012) at the age of four predicting RC two years later. Moreover, the NICHD study which tracked 1,137 typically developing children from three until seven years old found that LC at four years predicted RC aged six (NICHD Early Child Care Research Network, 2005). Vocabulary and syntactic knowledge are other reported sub-skills both LC and linguistic comprehension.

Vocabulary is shown to be a component skill of linguistic comprehension and makes a key contribution to RC. Findings from longitudinal studies show that vocabulary at the age of four predicts RC at six years old (Silva & Cain, 2015). This predictive relationship can be seen even further across time with vocabulary at the age of two shown to predict RC up to five years later (Duff et al., 2015). Syntactic knowledge is also suggested as a component skill of linguistic comprehension and as a result is shown to contribute to RC. For example, syntactic knowledge at the age of four is found to predict RC two years later (M. Bianco et al., 2012; Muter et al., 2004; Silva & Cain, 2015), and even further across time from syntax at the age of five to RC aged eight (Foorman et al., 2015).

Despite much evidence for the SVR, some researchers have suggested that the model is too simple and that an additional skill should be added to the model in order to account for unexplained variance (e.g. Adlof et al., 2006; Cain, 2015; Conners, 2009; Hoien-tengesdal, 2010; Joshi & Aaron, 2000; Kirby & Savage, 2008; Ouellette & Beers, 2010). Additional skills considered include: reading fluency, naming speed, processing speed, working memory, performance IQ and attentional control (for further discussion on this see Section 1.3.1 of Chapter 1). Yet to date no skill has been shown to reliably predict unique variance in RC after accounting for the skills in the SVR model (linguistic comprehension and decoding) therefore there is no strong evidence that an additional skill should be added as a direct predictor of RC.

Given this, others take an alternative stance instead proposing that more skills (in addition to listening comprehension, vocabulary, and syntactic knowledge) are sub-components of linguistic comprehension (Kim, 2015; Kim, 2016; Kim, 2017). Hoover and Gough (1990) write: “The simple view does not deny the complexity of reading, but asserts that such complexities are restricted to either of the two components” (p.150), thus proposing that the model suggests that other skills could be sub-components of linguistic comprehension (or even decoding).

One such sub-skill of linguistic comprehension could be working memory as research finds that it relates to RC in the early years of primary education (Van Den Broek, Kendeou, Lousberg, & Visser, 2011). Longitudinally speaking, working memory at the age of eight has been found to predict RC aged 11 years (Cain et al., 2004) and working memory aged six predicts RC aged nine years (Seigneuric & Ehrlich, 2005). Another suggested sub-skill is comprehension monitoring with findings showing that comprehension monitoring in 8-11 year olds explains variance in RC (Cain et al., 2004). Additionally, longitudinally comprehension monitoring at the age of seven predicts RC at 11 years old (Oakhill et al., 2003; Oakhill & Cain, 2012). This is supported by more recent longitudinal studies finding that comprehension monitoring aged seven makes a significant contribution to RC aged nine (Cain & Yeomans-Maldonado, 2017), and that comprehension monitoring at the age of ten predicts RC a year later (Muijselaar et al., 2017).

Lastly, inference making has been suggested to be a sub-skill of linguistic comprehension as research shows it predicts RC. For example, Oakhill and Cain (2012) found that for typically developing children aged 10-11 years, both concurrent and earlier inference making skills predicted RC. Training studies show that participation in inference interventions improves RC for both typically developing readers (Bos et al., 2016; E. M. Carr et al., 1983) and poor comprehenders (E. M. Carr et al., 1983; McGee & Johnson, 2003; Yuill & Oakhill, 1988). Inference making and RC are related in younger readers too with children aged seven (Casteel & Simpson, 1991). Another skill proposed recently (e.g. Kim, 2015; Kim, 2016; Kim, 2017) as a sub-skill of linguistic comprehension is theory of mind (ToM). However, research investigating its role has been very limited, especially studies examining longitudinal relations.

6.1.2 Theory of mind and reading comprehension

Research already shows a predictive relationship between ToM and other types of language development (Hughes, Ensor, & Marks, 2011; Milligan et al., 2007; Slade & Ruffman, 2005) and despite a recent review (Dore et al., 2018) suggesting that ToM may be the hidden factor in RC, only a small number of research studies have explored the role of ToM in RC. Details of these studies are shown in Table 6.1.

Table 6.1

A summary of past research into the role of theory of mind in reading comprehension listed in chronological order

Author	**N**	**Population**	**Mean Age**[a]	**Design**	**ToM measure**	**Reading comprehension measure**	**Analysis**	**Overview of results**
Ricketts, Jones, Happé, & Charman (2013)	100	ASD English speakers	15;7	Cross-sectional	Strange Stories (Happé, 1994) Frith-Happe Animations (Abell, Happe, & Frith, 2000)	Reading Comprehension subset WORD[b] (Wechsler, 1993)	Regression	ToM predicted RC after controlling for the variance explained by decoding and oral language
Kim (2015)	148	TD Korean speakers	6;1	Cross-sectional	First-order false belief task (Gwon & Lee, 2012) Second-order false belief task (Caillies et al., 2012)	Passage reading tasks (Kim, 2011; Kim, Park, & Wagner, 2014) Passage Comprehension subtest of Woodcock Johnson (Woodcock, McGrew, & Mather, 2001)	Structural equation modelling	In a model of RC ToM in-directly via LC predicted RC (See Figure 6.1)

Guajardo & Cartwright (2016)	31	TD English speakers	T1 = 4;4 T2 = 8;1	Longitudinal	Unexpected change task (Wimmer & Perner, 1983) Deception tasks (Lalonde & Chandler, 1995; Wimmer & Perner, 1983)	Passage Comprehension subtest of Woodcock Johnson-III (Woodcock et al., 2001)	Regression	ToM at T1 predicted phrase and sentence comprehension and reading awareness at T2 but did not account for unique variance in RC.
Kim (2017)	350	TD English speakers	7;6	Cross-sectional	Unexpected contents task (Bartsch & Wellman, 1989; Lewis & Osborne, 1990) Second-order false belief tasks (Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994) Second-order false belief tasks (Kim & Phillips, 2014)	Passage Comprehension subtest Woodcock Johnson (Woodcock et al., 2001) Reading comprehension subset WIAT[c] (Wechsler, 2007)	Structural equation modelling	In the DIER model ToM in-directly via LC predicted RC (See Figure 6.3). In the DIET model ToM directly predicted RC (See Figure 6.2).
Atkinson, Slade, Powell & Levy (2017)	80	TD English speakers	T1 = 3;10 T2 = 6;3	Longitudinal	Unexpected contents task (Hogrefe et al., 1986) The unexpected location task (Wimmer & Perner, 1983)	YARC[d] (Hulme et al., 2009)	Regression and mediation	ToM at T1 indirectly predicted T2 RC via language. Importantly, ToM at Time 1 also directly predicted Time 2 RC.

[a]Age in years;month [b]WORD = Wechsler Objective Reading Dimensions [c]WIAT = Wechsler Individual Achievement Test [d]YARC = The York Assessment of Reading for Comprehension: Passage Reading

As outlined in Table 6.1 the first of the studies into ToM and RC was Ricketts, Jones, Happé, and Charman (2013) who investigated the relationship in adolescence with Autism Spectrum Disorder. Findings showed that ToM directly predicted RC after controlling for word recognition and oral language skills. This led to further work with typically developing younger children, including the work of Kim (2015; 2017). Kim (2015) proposed a SEM cross-sectional hierarchical model of RC in which low-level skills predict high-level skills that in turn predict RC. Importantly, this model includes an indirect role for ToM (See Figure 6.1).

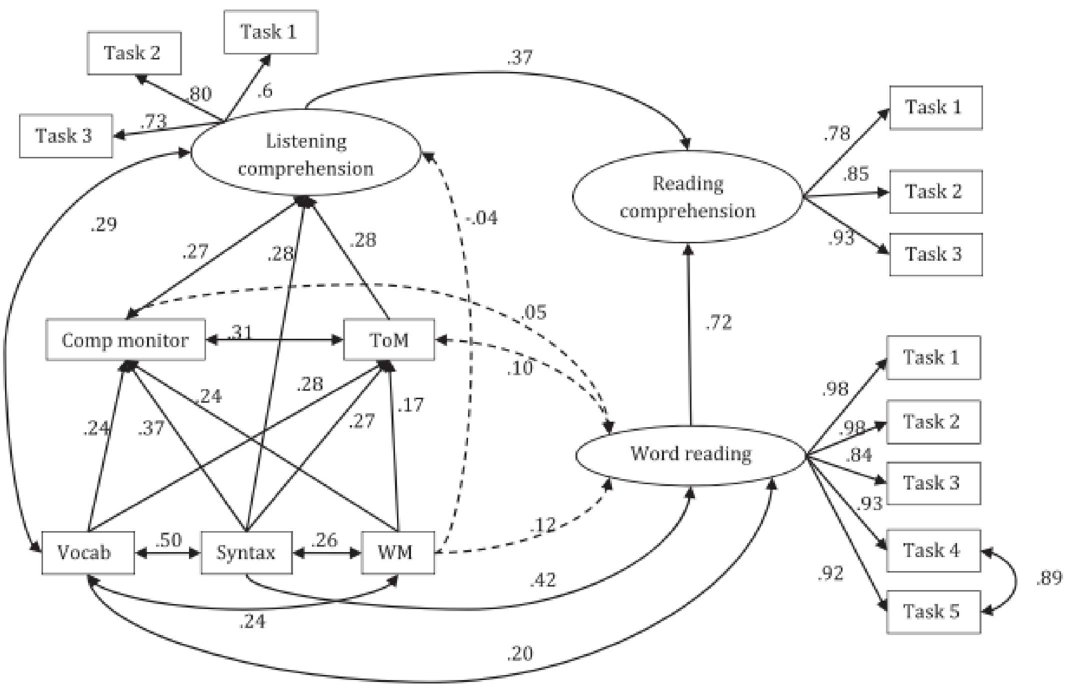


Figure 6.1: Best fitting model of reading comprehension as proposed by Kim (2015) including standardised regression weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

Kim (2017) developed this work further by fitting a model of RC with a direct path from ToM to RC (see Figure 6.2). Results showed that ToM directly predicts RC. This model was based on the DIET model (see Chapter 3, 4 and 5) of LC. The rationale for fitting the DIET model to RC (in the same way it had been fitted to LC) was that both LC and RC have the same theoretical framework (see Section 1.5 in Chapter 1 for further discussion on this), and so it is important to examine whether the structural relations of language and cognitive skills fit well for both RC and LC (Kim, 2017).

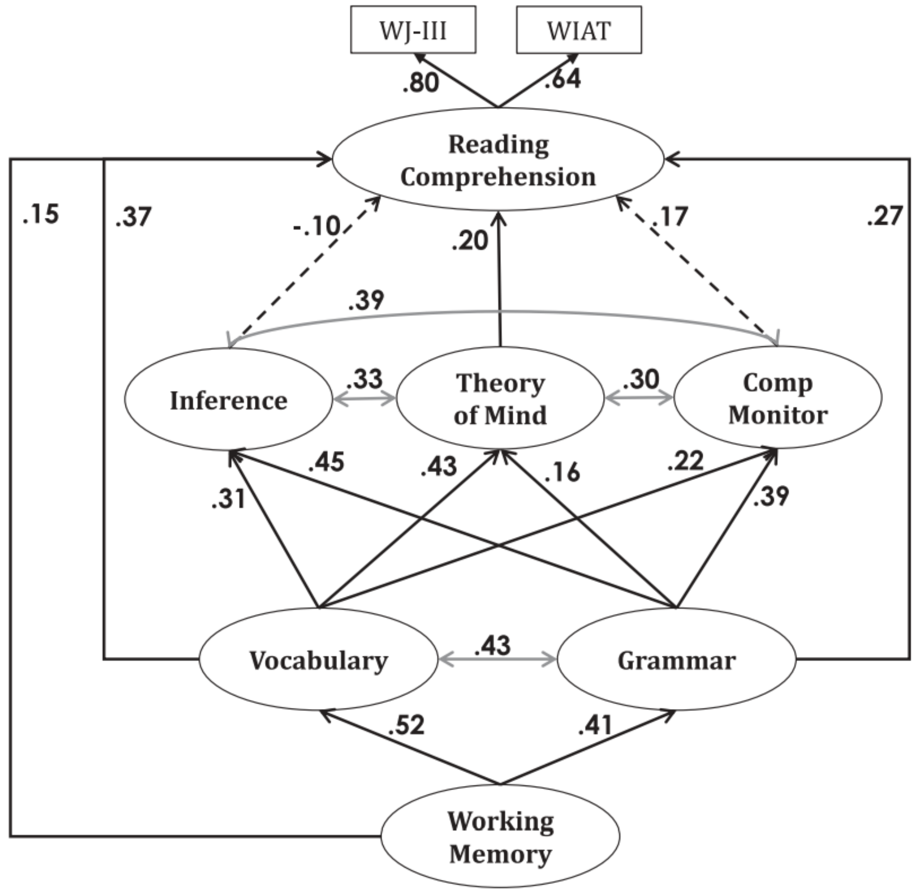


Figure 6.2: A DIET model as proposed by Kim (2017) in which ToM makes a direct contribution to RC. Including standardised regression weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

Within the 2017 paper, Kim also replicates the findings of Kim (2015) with a model that adds word reading and does not show a direct path from ToM to RC. This model is named the Direct and Indirect Effect Model of Reading (DIER; see Figure 6.3). As Figures 6.1 and 6.3 show this DIER model is similar to the earlier 2015 model but the latter includes the additional skill of inference making and expands from a three-tiered model to a four-tiered one (See Section 1.5 in Chapter 1 for further discussion). These models are important because they support the SVR suggesting that RC is a product of only linguistic comprehension and decoding (or word reading), but that linguistic comprehension comprises many sub-skills with LC at the top of the hierarchy of sub-skills.

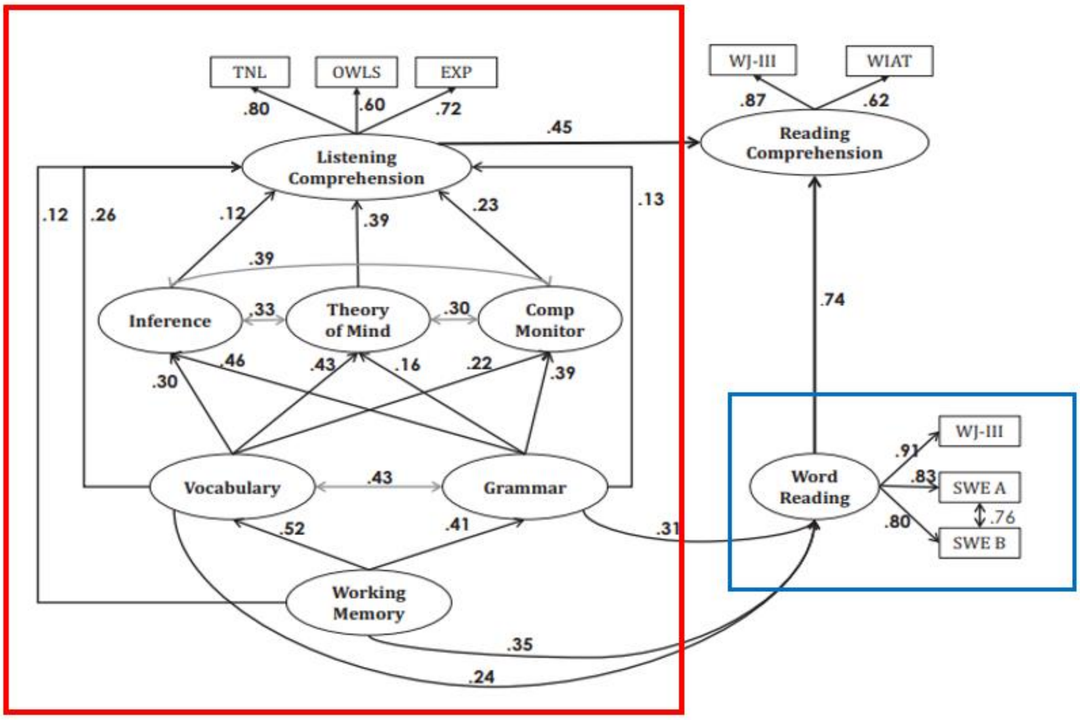


Figure 6.3: Best fitting DIER model of reading comprehension as proposed by Kim (2017) including standardised regression weights. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations. The coloured boxes represent the two components of reading comprehension as suggested by the simple view of reading. Within the blue box is the decoding (or word reading) component. Within the red box is the linguistic comprehension component containing all the sub-components of linguistic comprehension including listening comprehension at the top.

6.1.2.1 Limitations of research into theory of mind and reading comprehension

6.1.2.1.1 Lack of early longitudinal work

A key limitation of the work examining the relationship between ToM and RC is the age range of the children studies. Kim studies these relationships in children in mid-primary school (aged around seven or eight years) meaning that the research does not give an insight into factors that precede RC. This is important because knowing the skills that precede RC has important applied implications for the early years classroom for reading instruction. In the UK classroom, direct reading instruction starts in the Reception year when children are four or five years old and so skills measured before this age (i.e. in the preschool years) can be seen as precursors of RC.

Kim (2015; 2017) are both cross-sectional. As such, the concurrent and correlational nature of these studies means that the direction of relations in the models is based on theory, and for this reason the author highlights the need for longitudinal work (Kim, 2015, p.30). To address this Guajardo and Cartwright (2016) and Atkinson et al. (2017) (see Table 6.1) both conducted longitudinal work. Regression analysis of Guajardo and Cartwright (2016) found that ToM aged four did not predict unique variance in RC aged eight. However, the mediation analysis of Atkinson et al. (2017) found that ToM aged four both indirectly (via language) and directly predicted RC aged six.

As well as their contradictory results, these longitudinal studies had small or modest samples. Guajardo and Cartwright (2016) had a sample of just 31 children, and they conclude that the result that ToM understanding at the age of four years did not contribute uniquely to RC aged eight was derived from the small sample. Although Atkinson et al. (2017) used a much larger sample of 80, this sample still did not allow for structural equation modelling (SEM) analysis. The use of latent variable modelling such as SEM for analysis has grown dramatically over the last three decades (Morrison et al., 2017), and is the increasing choice in developmental psychology (e.g. Devine & Hughes, 2017; Kim, 2015; Kim, 2016; Kim, 2017; Puglisi et al., 2017) as it allows for the interrelationships amongst many variables to be explored (Morrison et al., 2017). The advantage of SEM over regression is that it permits many interrelationships to be tested simultaneously (Von der Embse, 2016). As RC is a complex skill with many components, the use of SEM to test models of RC is of great value. Although the sample size needed for SEM remains a point of contention (Barrett, 2007), a sample of 80, as used by Atkinson et al., (2017), is generally not accepted as enough. For further details on the use of SEM see Section 3.2.4 in Chapter 3. Additionally, thus far studies into the role of ToM in RC have been unable to determine exactly why ToM assists RC.

6.1.2.1.2 Why does theory of mind assist reading comprehension?

One explanation for why ToM facilitates RC is that it is the social element of ToM that is important. If a child has a better awareness and understanding of mental states, then they may be able to use this whilst reading to aid their understanding. This is the argument used by Ricketts et al. (2013) who suggest that for those with autism, deficits in social understanding may affect the ability to make inferences regarding the intentions and desires of characters in a story or the writer's communicative intentions. Guajardo and Cartwright (2016) and Atkinson et al. (2017) when discussing ToM for RC in typically developing children also consider this explanation. Social understanding could be important particularly for young readers as plots of age appropriate story books often revolve around the mental states of the characters (Strasser & Río, 2014; Zunshine, 2019), such as their thoughts, intentions, and feelings, and more complex social situations such as deceptions or misunderstandings. Indeed, Lynch and van den Broek (2007) suggest that characters' goals and mental states are what holds together a coherent story, and Zunshine (2019) describes how the plots of many much-loved children's books (e.g. *Gruffalo*, *Rosie's Walk* and *Winnie the Pooh*) all revolve around characters' mental states.

Given that mental states are central to many storybooks, a child with a better understanding of ToM might also have a better understanding of these stories. This explanation is consistent with the situation model (Kintsch, 1994; Van Dijk et al., 1983; Zwaan, 2016) which states that proficient text comprehension requires the construction of a mental representation of the story (Kintsch, 1988). The understanding of social details may assist the reader in constructing a mental model of the story (Graesser et al., 1994; Kim & Pilcher, 2016; Kintsch, 1988). ToM may directly contribute to model building itself as Perner and colleagues describe the possession of a ToM as mental model building (Johnson-Laird, 1983; Perner, 1991), in that a belief is a mental model of the world. It has been suggested that young children fail false belief tests because they do not understand that mental models or representations can differ from true reality (Lillard & Flavell, 1992; Perner, 1991). As such, developments in ToM ability may directly underpin building mental models of story plots during RC.

The view that ToM is important for RC because of its social specificity and the assistance it gives with the social details of a story, is that taken in the recent review of ToM in RC by Dore, et al. (2018). More specifically Dore, et al. (2018) argue that understanding characters' mental states is what leads to better RC. Yet, as posited by

Atkinson et al., (2017) there is an alternative explanation. Instead, it could be the general metacognitive nature of ToM that is important rather than the socially specific element. As explained in previous chapters, metacognition is defined as knowing about knowledge or thinking about thinking (Flavell, 1976; Flavell, 1979) and ToM is a metacognitive skill. ToM may facilitate RC because of its broad metacognitive nature which informs knowledge, actions, and understandings not necessarily related with story characters or oneself as a reader (Atkinson et al., 2017). As described above the situation model suggests that creating a mental representation of what the text is about is crucial for proficient RC. Constructing this accurate mental model of the passage may also require representation of non-social aspects, such as space, time and objects (Graesser et al., 1994). The broad metacognitive nature that ToM taps could aid with this. Indeed, research emphasises that successful RC requires non-social metacognitive processes such as the ability to monitor one's own knowledge whilst reading e.g. comprehension monitoring (Oakhill & Cain, 2012). Moreover, ToM has been linked to the development of other non-social metacognitive skill.

ToM has been linked to the development of other metacognitive processes, such as metamemory (Lockl & Schneider, 2007), metalinguistic awareness (Doherty, 2000), and source monitoring (Bright-Paul et al., 2008). See Section 1.7.1 in Chapter 1 for further details. Research that shows that these non-social metacognitive skills are related to ToM in the early years suggests that these skills share the same underlying metacognitive nature as ToM (Atkinson et al., 2017). The findings imply that the understanding of memory, language and the source of one's knowledge depend on the same representational insights as ToM does (Perner, 1991). Given this, in order to assess if it is the social nature of ToM that is important for facilitating RC, or instead if it is the general metacognitive nature, research should compare the contribution of ToM and these other types of non-social metacognition for RC. If these non-social metacognitive skills predict RC in the same way that ToM does then this suggests it is because of the broad-metacognitive nature that these skills share. On the other hand, if these other types of metacognition cannot predict RC, but ToM can, then this advocates that it is the social component of ToM that is important for RC. Determining this will give a clearer understanding of the role that ToM plays in RC.

6.1.3 The present study

Recent research has begun to explore the role of ToM in the development of RC (see Table 6.1). However, no study has been longitudinal with a large enough sample size for the use of SEM analysis to allow for a model of RC with multiple latent variables to be tested. Furthermore, past studies have not examined whether it is the socially specific element of ToM, or its general-metacognitive nature that assists RC. This chapter aimed to address these issues by carrying out a larger sample longitudinal study to test models of RC which included the contribution of both ToM and non-social broad metacognition. This chapter sought to assess both DIET and DIER models of RC to evaluate both direct and indirect effects of ToM on RC. The overall goal was to give a clearer understanding of the role that ToM plays in RC, and to assess if the same pattern shown by the three previous chapters with LC could be extended to RC.

Firstly, a concurrent model of RC (at age six) including a direct path from ToM to RC based on Kim's (2017) DIET model (Figure 6.2) was tested (Model 1). Then, this model was compared to a model that included a direct path from broad metacognition (instead of ToM) to RC. This allowed for an assessment of what aspect of ToM is directly important for RC. Next, these models were tested longitudinally (skills aged four predicting skills aged six) to determine if the relationships would hold across 22 months (Model 3 and Model 4). These longitudinal models would also determine whether ToM makes a direct longitudinal contribution to RC to support the work of Atkinson et al. (2017).

Models were then tested (Models 5, 6, 7 and 8) based on Kim's DIER model (see Figure 6.3) in which word reading was added to the model of RC and ToM did not have a direct path to RC. This DIER model supports the SVR because whilst other language and cognitive skills are shown to contribute to LC (are its component skills) only linguistic comprehension and word reading directly predict RC. Model 5 and Model 6 were concurrent (at the age of six) and compared the fit of including ToM (Model 5) to including broad metacognition (Model 6). Model 7 and Model 8 were longitudinal (aged four skills predicting aged six RC) and again compared the fit of including ToM (Model 7) to including broad metacognition (Model 8). As with previous chapters in all eight models, age and non-verbal ability were controlled for.

It was hypothesized that after controlling for age and non-verbal ability:

1) Concurrently a model of RC (at the age of six; Model 1) which included roles for working memory, vocabulary, syntactic knowledge, comprehension

monitoring, inference making and ToM would be a good fit. Moreover, ToM would make a direct significant contribution to RC after controlling for other skills in the model.

2) This Model 1 (including ToM) would be a better fit than a model which instead included a latent variable of broad metacognition (Model 2; consisting of source monitoring, metamemory and metalinguistic awareness) and this is because the social nature of ToM is important for RC.
3) Hypotheses 1 and 2 would be extended longitudinally across 22 months (from aged four to aged six)[10], in that a longitudinal model of RC including ToM (Model 3) would have a better fit than a model which included broad metacognition (Model 4), and ToM aged four would make a direct contribution to RC aged six, whereas broad metacognition would not.
4) These hypotheses would be extended to a model of RC that included word reading (based on the Kim, 2017; DIER model). Both concurrently and longitudinally[11] DIER models that included ToM (Model 5 and Model 7) would have better fits than models which included broad metacognition.

For clarity the models tested in this chapter and their hypotheses are presented in Table 6.2.

---

[10] In contrast to the concurrent models these models did not include comprehension monitoring and inference making as they were not measured at Time 1 due to lack of available UK measure for such young children.

[11] As with the DIET models longitudinal models did not include comprehension monitoring or inference making

Table 6.2

Summary of the models tested in this chapter and their corresponding hypotheses

Model	Design	Model details	Hypothesis
	DIET models[a]		
Model 1	Concurrent	A concurrent model of RC including a direct role for ToM.	Good fit. ToM will predict concurrent RC.
Model 2	Concurrent	A concurrent model of RC including a direct role for broad metacognition.	Not a good fit. Broad metacognition will not predict RC.
Model 3	Longitudinal	A longitudinal model of RC including a direct role for ToM	Good fit. Earlier ToM will directly predict later RC.
Model 4	Longitudinal	A longitudinal model of RC including a direct role for broad metacognition.	Not a good fit. Earlier broad metacognition will not predict RC.
	DIER models[b]		
Model 5	Concurrent	A concurrent model of RC including an indirect role for ToM	Good fit.
Model 6	Concurrent	A concurrent model of RC including an indirect role for metacognition	Poor fit
Model 7	Longitudinal	A longitudinal model of RC including an indirect role for ToM	Good fit.
Model 8	Longitudinal	A longitudinal model of RC including an indirect role for metacognition	Poor fit

Notes. [a] DIET models are based on Kim (2017) in which ToM makes a direct contribution to RC as shown in Figure 6.2 [b] DIER are based on Kim (2017) in which RC is a product of only linguistic comprehension and word reading but linguistic comprehension comprises many sub-skills including ToM (see Figure 6.3).

6.2 Method

6.2.1 Participants

Participants consisted of those from Cohort 1 only. They were tested at Time 1 and again 22 months later at Time 3. At Time 1 there were 151 children in the sample but this dropped to 107 at Time 3. Only participants with a full dataset at both time points (the 107) were used in the analysis. For more information on the participants including recruitment see Section 2.3 in Chapter 2. Table 6.3 shows demographic information for the 107 participants used in the analysis.

Table 6.3

Demographic information at each time point

	N	**N males**	**N females**	**Mean Age**[a]
Time 1	107	54	53	4;3 (SD = 3.59)
Time 3	107	54	53	6;1 (SD = 3.68)

Note. [a] Mean age in Years;Months

6.2.2 Materials and measures

Table 6.4 shows the measures administered at each time point. For comprehensive details on each of these measures, refer to Section 2.5 in Chapter 2.

Table 6.4

Measures administered to children at each of the time points

Skill	Measure	Time 1	Time 3
Vocabulary	BPVS-III	✓	✓
Syntax	Sentence Structure (CELF-Preschool 2^{uk})	✓	-
	Sentence Structure (CELF-4^{uk})	-	✓
Precursors of decoding	Preschool Repetition Task	✓	-
	Letter Sound Knowledge (YARC)	✓	-
Decoding (single word reading)	DTWRP	-	✓
Listening comprehension	NARA-II	✓	✓
	OWLS	-	✓
Theory of mind	Unexpected contents task	✓	-
	Unexpected locations task	✓	-
	Strange Stories	-	✓
Metacognition	Source monitoring tunnel task	✓	-
	Source monitoring events task	-	✓
	Synonym judgment task	✓	-
	Homonym judgment task	-	✓
	Metamemory task	✓	✓
Working memory	Reverse word span task	✓	✓
Comprehension monitoring	Comprehension monitoring stories	-	✓
Inference making	Inference stories	-	✓
Non-verbal ability	Block design (WPPSI – III)	✓	-
Reading comprehension	YARC	-	✓

6.2.3 Procedure

Informed (opt-in) consent was obtained from headteachers and from children's parents (opt-out). For further details refer to Section 2.4 in Chapter 2. The children were initially tested in term three of their preschool year. The testing sessions took place in a quiet area within their classroom and each child took part in two 20-minute sessions. Children were reassessed 22 months later when they were in Year 1. Once again, testing sessions took place in a quiet area within (or just outside) their classroom and each child took part in two 20-minute sessions. For complete procedural details refer to Chapter 2 Section 2.7.

6.2.4 Analysis

The primary data analytical strategy was structural equation modelling (SEM) using AMOS Version 25 (Arbuckle, 2016). For each of the models tested, first descriptive statistics were computed and then initial correlational analysis was carried out. During SEM analysis, for the broad metacognitive models (Models 2, 4, 6 and 8) a latent variable for broad metacognition was created which included the three types of non-social metacognition (metamemory, source monitoring and metalinguistic awareness). All other language and cognitive skills including RC were assessed by a single measure for each construct. For justification of this analysis and further details of its use see Section 3.2.4 in Chapter 3.

6.3 Results

First models which included a direct path from ToM or metacognition to RC were tested (DIET models). Concurrent models of RC were tested first; one that included ToM (Model 1) and one that included broad metacognition (Model 2). Next, these models were extended longitudinally with Time 1 (aged four) skills predicting Time 3 RC (aged six; Model 3 and Model 4).

Then DIER models were tested which included word reading and supported the SVR. As with the previous models, first concurrent models were tested, one that included ToM (Model 5) and one that included broad metacognition (Model 5). Then these models were extended longitudinally with Time 1 (aged four) skills predicting Time 3 RC (aged six), one which included ToM (Model 7) and one which included broad metacognition (Model 8). For further details on each of these models see Table 6.2. In all analysis the outcome measure of RC was the YARC reading comprehension ability score. This was used because its use is in line with recent research (e.g. Dockrell, Connelly, & Arfè, 2019; Hulme, Nash, Gooch, Lervåg, & Snowling, 2015; Joseph & Nation, 2018; Valentini, Ricketts, Pye, & Houston-Price, 2018). As different children read different level passages the ability score reflects both the raw score and the difficulty of the passages they have read. As such, raw scores attained on low passages yield a lower ability score than the same raw scores on more difficult passages. Ability scores are obtained using the Rasch model (Snowling et al., 2011) and do not consider age.

6.3.1 DIET models of reading comprehension

6.3.1.1 Concurrent DIET models

Concurrent DIET models of RC (see Figure 6.2) were tested at Time 3 (when children were aged six). One model included ToM (Model 1) and the other included a latent variable of broad metacognition (Model 2).

6.3.1.1.1 Descriptive statistics and preliminary analysis

Table 6.5 shows descriptive statistics for all measures at Time 3. Table 6.6 shows correlations between these measures.

Table 6.5

Descriptive statistics for all measures at Time 3 and non-verbal ability at Time 1 (all measures used in the concurrent models)

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Vocabulary	BPVS	107	168	48-130	87.03	13.28	.14	1.29
Syntax	CELF-4	107	26	7-26	20.24	3.74	-.99	1.07
Listening comprehension	NARA	107	44	1-16	6.58	3.38	.69	-.31
	OWLS	106	130	36-94	63.66	12.32	.38	-.55
Decoding (word reading)	DTWRP	107	90	3-75	35.87	16.27	.09	-.66
Theory of mind	Strange stories	106	10	0-8	2.33	1.57	.92	1.43
Source monitoring	Source monitoring events	105	6	2-6	4.46	1.23	-.20	-.66
Metalinguistic awareness	Homonym judgment task	107	8	0-8	5.28	2.17	-.34	-.69
Metamemory	Metamemory task	106	4	0-4	2.70	1.22	-.65	-.50
Working memory	Reverse word span task	107	11	0-11	6.69	3.78	-.25	-1.42
Comprehension monitoring	Monitoring stories	107	5	0-5	2.67	.98	.09	-.37
Inference making	Inference stories	106	8	0-8	4.21	1.89	-.20	-.53
Time 1 non-verbal ability	Block design	107	40	0-32	15.34	6.50	-.12	-.13
Reading comprehension	YARC ability score	107	91	19-55	39.78	7.74	-.46	-.36

Table 6.6

Correlation matrix for all measures used in the concurrent models

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.BPVS	-														
2.CELF-4	.49**	-													
3.NARA	.54**	.58**	-												
4.OWLS	.50**	.37**	.62*	-											
5.DTWRP	.24*	.32**	.20*	.17	-										
6.Strange stories	.35**	.25**	.38**	.32**	.11	-									
7.Source monitoring events	.25**	.34**	.36**	.35**	.29**	.13	-								
8.Homonym judgment task	.21*	.24*	.21*	.38**	.24*	.14	.32**	-							
9.Metamemory task	.47**	.31**	.34**	.50**	.47**	.21*	.56**	.37**	-						
10.Reverse word span task	.39**	.32**	.37**	.33**	.45**	.30**	.28**	.56**	.44**	-					
11.Comprehension monitoring	.19*	.21*	.21*	.23*	.23*	.16	.15	.28**	.28**	.10	-				
12.Inference stories	.52**	.49**	.45**	.51**	.04	.29**	.21*	.15	.20*	.32**	.10	-			
13.Block design	.45**	.46**	.46**	.32**	.43**	.29**	.41**	.21*	.44**	.28**	.18	.33**	-		
14.Age	.13	.22*	.22*	.13	.20*	.10	.16	.41**	.19	-.01	.18	.11	.41**	-	
15.YARC	.51**	.41**	.43**	.38**	.45**	.14	.39**	.22*	.53**	.35**	.20*	.29**	.34**	.23*	-

Note. * $p < .05$ ** $p < .001$

6.3.1.1.2 SEM analysis

First a concurrent DIET model which included a direct path from ToM to RC was fitted (Model 1; see Figure 6.4). Prior to SEM analysis missing data (see Table 6.5 for details of which measures) was imputed using expectation maximization (EM) in SPSS. This method of data imputation was used because missing data was minimal (two cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random (p =.08).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were confirmed by a Mardia's coefficient of multivariate critical ratio of -.40 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). This model had a good fit: χ^2 (3) = 3.82, p = .28; CFI = 1.00, TLI = .95, and RMSEA =.05. Four multivariate outliers were identified using Mahalanobis d-squared (those with a p < .05). When these four participants were removed and the model re-run the model fit became excellent: χ^2 (3) = 1.57, p = .67; CFI = 1.00, TLI = 1.01, and RMSEA < .001, therefore this model with 103 participants is reported.

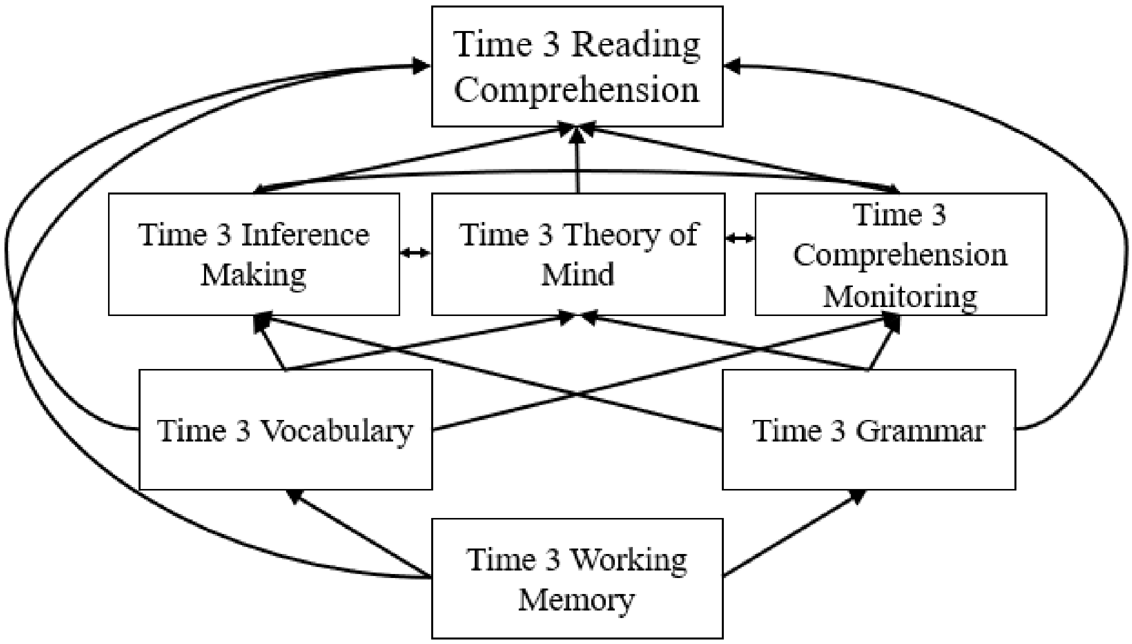


Figure 6.4: Hypothesised Model 1. A concurrent DIET model including ToM. Controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.5. As presented in Figures 6.5 after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .30, p <.001) and to RC (β = .19, p = .04), but not to grammatical knowledge (β = .16, p = .09). Grammatical knowledge was not significantly related to ToM (β = -.04, p = .701), or comprehension monitoring (β = .15, p = .16), but was to inference making (β = .26, p = .01). Vocabulary was significantly related to ToM (β = .40, p <.001) and inference (β = .35, p < .001), but not to comprehension monitoring (β = .12, p = .30). Vocabulary (β = .35, p =.002) was directly related to RC but grammatical knowledge (β = .14, p =.18) was not. Neither ToM (β = .13, p =.17), inference (β = .04, p =. 70) or comprehension monitoring (β = .05 p =. 59) were significantly related to RC. A total of 44% of variance in Time 3 RC was explained by the concurrent skills in the model.

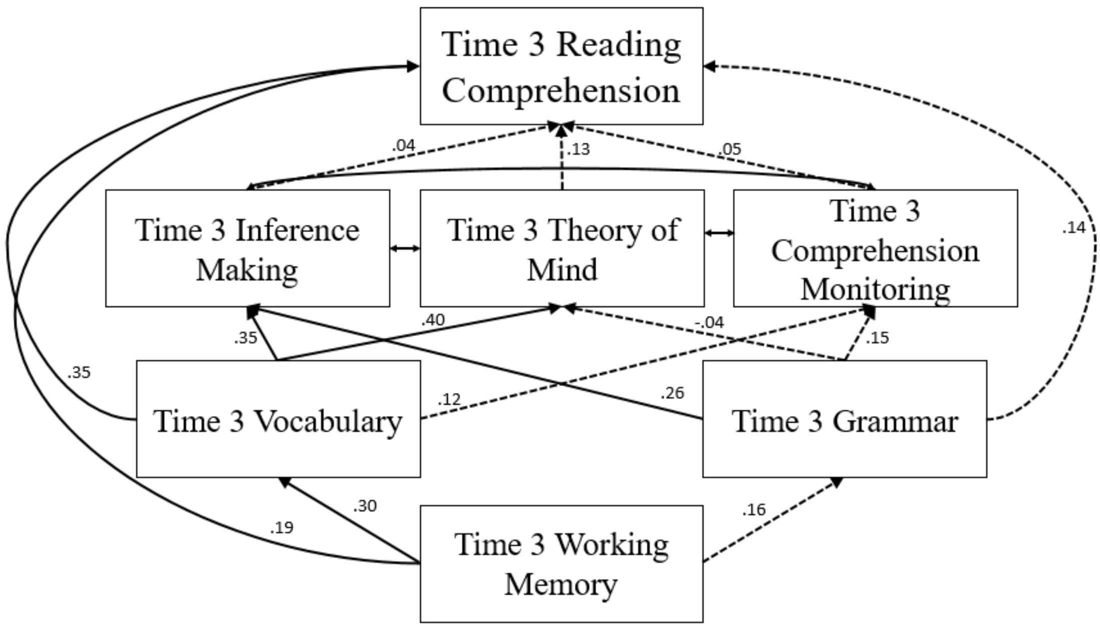


Figure 6.5: Model 1 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

Next a concurrent DIET model which included a direct path from broad metacognition to RC was fitted (Model 2; see Figure 6.6). Broad metacognition was entered into the model in the form of a latent variable consisting of source monitoring, metamemory and metalinguistic awareness. Prior to SEM analysis, missing data (see Table 6.5 for details of which measures) was imputed using EM in SPSS. This method of data imputation was used because missing data was minimal (four cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random (p =.06).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were appropriate as confirmed by a Mardia's coefficient of multivariate critical ratio of -.07 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). This model had a good fit: χ^2 (19) = 28.49, p =.07; CFI = .97, TLI = .91, and RMSEA = .07. Nine multivariate outliers were identified using Mahalanobis d-squared (those with a p < .05). When these nine participants were removed and the model re-run the model fit changed dramatically becoming poor: χ^2 (19) = 42.48, p = .002; CFI = .93, TLI = .79, and RMSEA = .11, therefore this model with 98 participants is reported.

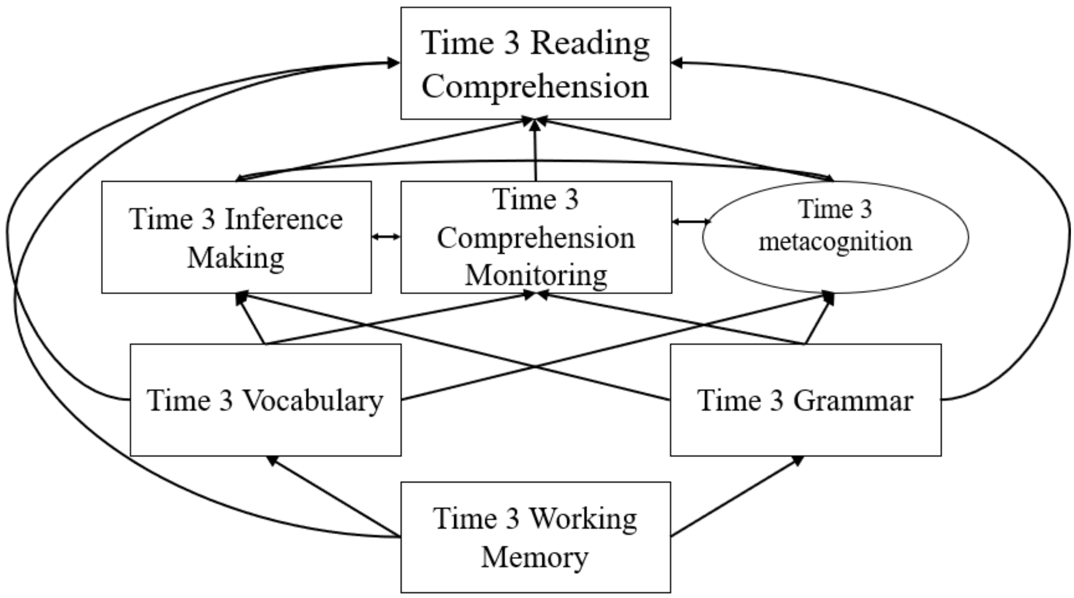


Figure 6.6: Hypothesised Model 2. A concurrent DIET model including broad metacognition. After controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.7. As presented in Figure 6.7 after controlling for age and non-verbal ability for all paths, working memory was significantly related to vocabulary (β = .25, p = .006), but not to grammatical knowledge (β = .16, p = .07) or to RC (β = .11, p = .19). Grammatical knowledge was not significantly related to metacognition (β = .13, p = .28), or comprehension monitoring (β = .12, p = .33), but was to inference making (β = .31, p = .003). Vocabulary was significantly related to metacognition (β = .33, p = .003) and inference (β = .36, p < .001), but not to comprehension monitoring (β = .09, p = .43). Vocabulary (β = .27, p =.004) was directly related to RC but grammatical knowledge (β = .14, p =. 12) was not. Neither inference (β = -.01, p =. 93) or comprehension monitoring (β = .03, p =.68) were significantly related to RC but metacognition was (β = .43 p <.001). A total of 47% of variance in Time 3 RC was explained by the concurrent skills in the model.

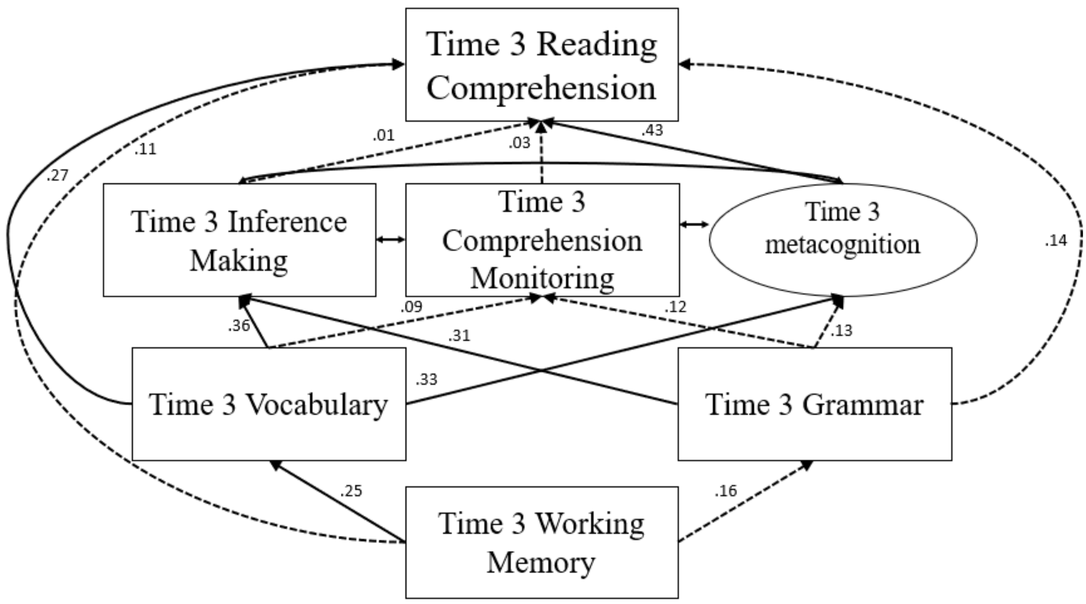


Figure 6.7: Model 2 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

6.3.1.2 Longitudinal DIET models

Next longitudinal DIET models of RC were tested for Time 1 skills (when children were aged four) predicting Time 3 RC. One model included ToM (Model 3) and the other included a latent variable of broad metacognition (Model 4).

6.3.1.2.1 Descriptive statistics and preliminary analysis

Table 6.7 shows descriptive statistics for all measures used in the longitudinal models. It should be noted that a Time 1 ToM composite consisting of the two ToM measures was computed by summing scores from the two ToM measures from Time 1 (unexpected contents and unexpected locations). This composite was used to give a richer measure of ToM and was justified as the individual measures correlated significantly ($r = .39$, $p < .001$). Table 6.8 shows correlations between the measures used in the longitudinal models.

Table 6.7

Descriptive statistics for all measures at Time 1 and 3 used in the longitudinal models

Skill	**Measure**	**N**	**Max**	**Range**	**Mean**	**SD**	**Skewness**	**Kurtosis**
Time 1 vocabulary	BPVS	107	168	10-102	51.59	18.12	.13	-.43
Time 1 syntax	CELF-preschool	107	22	0-21	11.64	4.61	-.55	-.29
Time 1 listening comprehension	NARA	107	44	0-9	1.49	2.34	1.79	1.87
Time 1 precursors to decoding	Repetition task	104	36	4-36	26.74	6.53	-.1.12	.89
	Letter sound knowledge	98	26	0-26	7.19	7.53	.94	-.28
Time 1 theory of mind	ToM composite	104	5	0-5	2.16	1.63	.21	-.97
Time 1 source monitoring	Source monitoring tunnel	107	6	0-6	3.64	1.78	-.47	.60
Time 1 metalinguistic awareness	Synonym judgment task	107	4	1-4	3.21	1.00	-.76	-.86
Time 1 metamemory	Metamemory task	106	4	0-3	.87	1.00	.74	-.72
Time 1 working memory	Reverse word span task	106	9	0-8	1.00	2.16	2.02	2.72
Time 1 non-verbal ability	Block design	107	40	0-32	15.34	6.50	-.12	-.13
Time 3 reading comprehension	YARC ability score	107	91	19-55	39.78	7.74	-.46	-.36

Table 6.8

Correlation matrix for all measures at Time 1 and 3 used in the longitudinal models

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.BPVS	-												
2.CELF	.67**	-											
3.NARA	.42**	.32**	-										
4.Word repetition	.36**	.33**	.28**	-									
5.Letter sound	.25*	.35**	.21*	.27**	-								
6.ToM composite	.37**	.44**	.25**	.22*	.19	-							
7.Source monitoring	.54**	.69**	.28**	.33**	.19	.47**	-						
8.Synonym judgment task	.26**	.42**	.15	.25*	.28**	.19	.39**	-					
9.Metamemory task	.44**	.45**	.23*	.10	.10	.29**	.48**	.27**	-				
10.Reverse word span task	.51**	.41**	.30**	.30**	.19	.24*	.37**	.16	.31**	-			
11.Block design	.62**	.50**	.34**	.44**	.21*	.26**	.50**	.44**	.38**	.40**	-		
12.Age	.36**	.29**	.20*	.07	.02	.04	.19*	.20*	.21*	.29**	.41**	-	
13.YARC	.38**	.31**	.26**	.13	.25*	.21*	.19	.30**	.32**	.32**	.34**	.24*	-

Note. * p < .05 ** p < .001

6.3.1.2.2 SEM analysis

A longitudinal DIET model[12] which included a direct path from ToM aged four to RC aged six was fitted (Model 3; see Figure 6.8). Prior to SEM analysis missing data (see Table 6.7 for details of which measures) was imputed using EM in SPSS. This method of data imputation was used because missing data was minimal (four cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random (p =.06).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and was fine as confirmed by a Mardia's coefficient of multivariate critical ratio of -.832 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). This model had an excellent fit: χ^2 (1) = .18 p = .67; CFI = 1.00, TLI = 1.08, and RMSEA < .001. Two multivariate outliers were identified using Mahalanobis d-squared (those with a p < .05). When these two participants were removed and the model re-run the model only changed slightly fit changed: χ^2 (1) = .02, p = .883; CFI = 1.00, TLI = 1.10, and RMSEA = < .001, therefore the original model with all 107 participants is reported here.

---

[12] As noted previously the longitudinal models did not include comprehension monitoring and inference making as these were not measured at Time 1

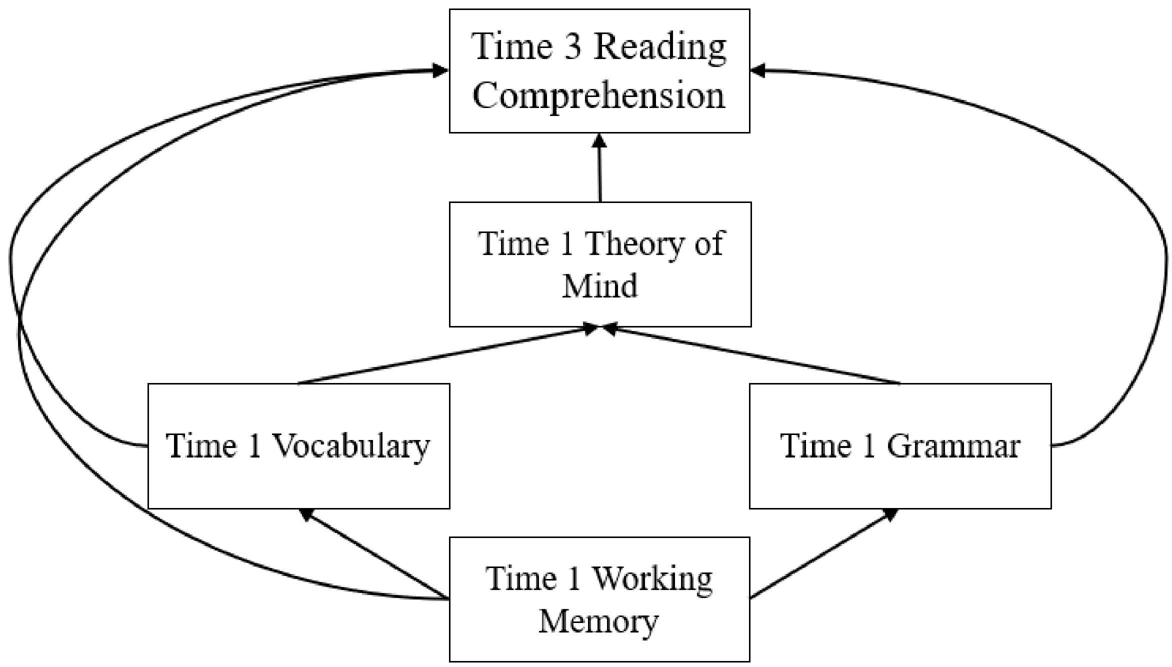


Figure 6.8: Hypothesised Model 3. A longitudinal DIET model including ToM,. Controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.8. As presented in Figures 6.9, after controlling for age and non-verbal ability for all paths, Time 1 working memory was significantly related to concurrent vocabulary (β = .21, p = .01) and grammatical knowledge (β = .18, p = .007) but not to Time 3 RC (β = .21, p = .16). Grammatical knowledge was significantly related to ToM (β = .38, p = .001). Vocabulary was not significantly related to ToM (β = .07, p = .23). Neither Time 1 vocabulary (β = .22, p =.22) nor grammatical knowledge (β = -.05, p =.68) were directly related to Time 3 RC. Time 1 ToM was not significantly related to Time 3 RC (β = .19, p = .07). A total of 28% of variance in Time 3 RC was explained by the Time 1 skills in the model.

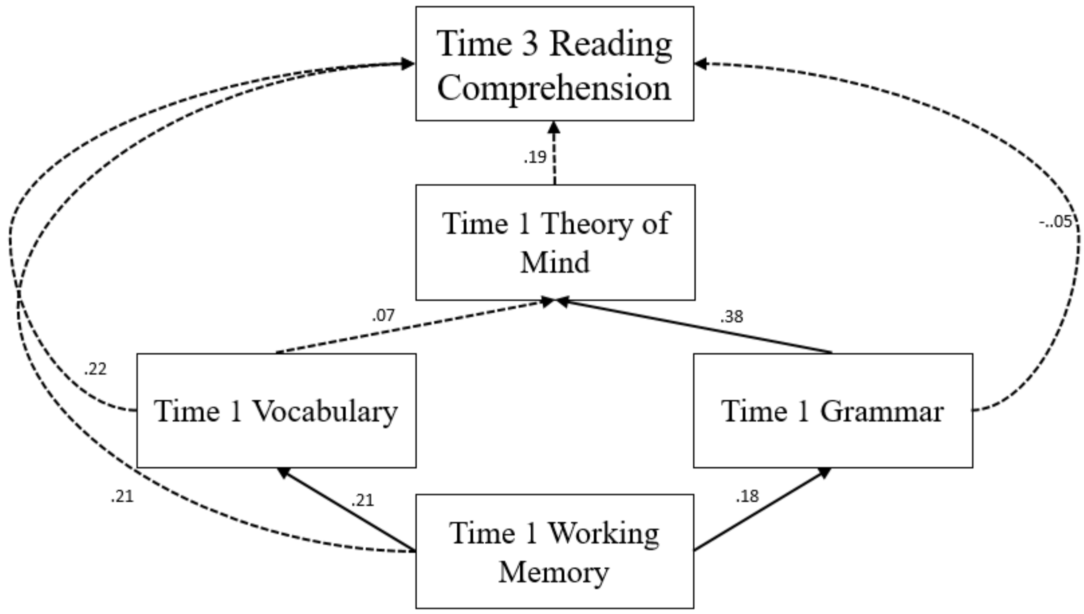


Figure 6.9: Model 3 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

Next a longitudinal DIET model which included a direct path from Time 1 broad metacognition to Time 3 RC was fitted (Model 4; see Figure 6.14). Broad metacognition was entered into the model in the form of a latent variable consisting of source monitoring, metamemory and metalinguistic awareness. Prior to SEM analysis, missing data (see Table 6.7 for details of which measures) was imputed using EM in SPSS. This method of data imputation was used because missing data was minimal (two cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little’s MCAR test reported the data to be missing at random ($p = .06$).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were within range as confirmed by a Mardia's coefficient of multivariate critical ratio of -.80 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). This model had a good fit: χ^2 (13) = 18.17, *p* = .15; CFI = .98, TLI = .96, and RMSEA = .06. No multivariate outliers were identified using Mahalanobis d-squared (those with a p1< .05) therefore the model with 107 participants is presented here.

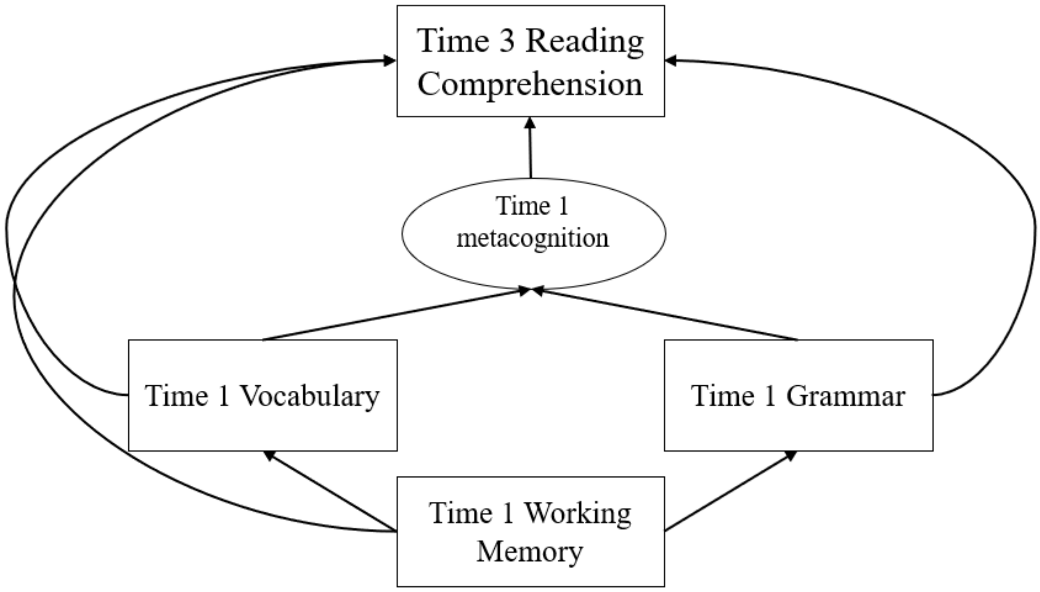


Figure 6.10: Hypothesised Model 4. A longitudinal DIET model including broad metacognition, after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.11. As presented in Figures 6.11 , after controlling for age and non-verbal ability for all paths, Time 1 working memory was significantly related to concurrent vocabulary (β = .23, p = .004) but not to grammatical knowledge (β = .14, p = .11) nor Time 3 RC (β = .17, p

= .09). Grammatical knowledge was significantly related to metacognition (β = .65, p < .001). Vocabulary was not significantly related to metacognition (β = -.004, p = .98). Neither Time 1 vocabulary (β = .16, p =.24) nor grammatical knowledge (β = -.19, p =. 43) were directly related to Time 3 RC. Time 1 metacognition was not significantly related to Time 3 RC (β = .37, p = .27). A total of 29% of variance in Time 3 RC was explained by the Time 1 skills in the model.

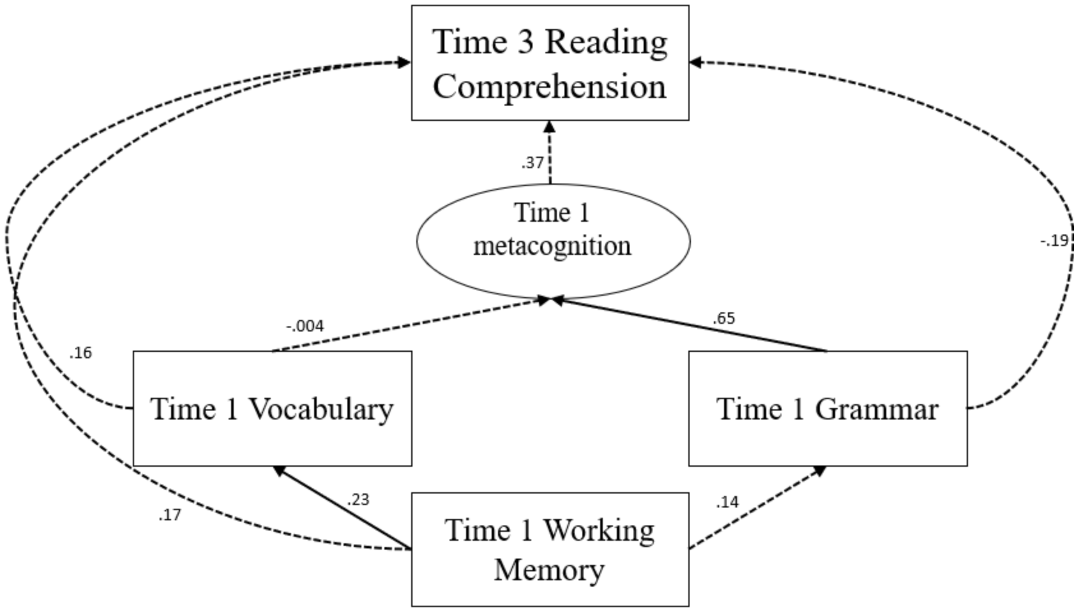


Figure 6.11: Model 4 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

6.3.2 DIER models of reading comprehension

Next DIER models of RC based on the model of Kim (2017) were tested (see Figure 6.2). These models included word reading and supported the SVR.

6.3.2.1 Concurrent DIER models

First concurrent models were tested for Time 3 (aged six skills) predicting Time 3 RC. One model included ToM (Model 5) and the other included a latent variable of broad metacognition (Model 6).

6.3.2.1.1 Descriptive statistics and preliminary analysis

Table 6.5 (above) shows descriptive statistics for all measures used in the concurrent DIER models. Table 6.6 (also above) shows correlation between these measures.

6.3.2.1.2 SEM analysis

A model based on the DIER model (Kim, 2017) was fitted to the data (Model 5; see Figure 6.12). Prior to SEM analysis missing data (see Table 6.5 for details of which measures) was imputed using EM in SPSS. This method of data imputation was used because missing data was minimal (three cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little’s MCAR test reported the data to be missing at random ($p = .06$).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were within range as confirmed by a Mardia's coefficient of multivariate critical ratio of -.24 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). The model had a moderately good fit: χ^2 (21) = 36.27, p = .05; CFI = .96, TLI = .89 and RMSEA = .08. Seven multivariate outliers were identified using Mahalanobis d-squared (those with a p < .05). When these seven participants were removed and the model re-run the model only changed slightly fit changed: χ^2 (21) = 35.87, p = .05; CFI = .96, TLI = .89, and RMSEA = .08, therefore the original model with all 107 participants is reported here.

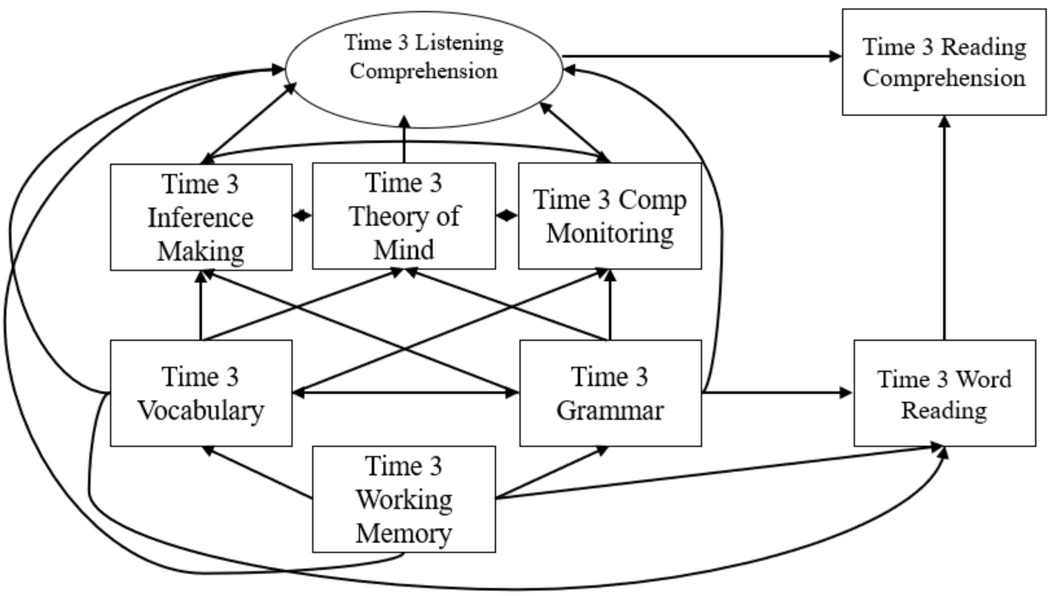


Figure 6.12: Hypothesised Model 5. A concurrent DIER model including ToM. Controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.13. As presented in Figure 6.13, after controlling for age and non-verbal ability the SVR was supported as both linguistic comprehension ($\beta = .52$, $p < .001$) and word reading ($\beta = .31$, $p < .001$) significantly related to RC. Working memory ($\beta = .38$, $p < .001$) but not grammatical knowledge ($\beta = .10$, $p = .30$) nor vocabulary ($\beta = -.09$, $p = .35$) was significantly related to word reading. Working memory was significantly related to vocabulary ($\beta = .28$, $p = .001$) and grammatical knowledge ($\beta = .20$, $p = .03$) but not to LC ($\beta = .12$, $p = .19$). Grammatical knowledge was not significantly related to ToM ($\beta = .05$, $p = .65$), or comprehension monitoring ($\beta = .13$, $p = .26$), but was to inference making ($\beta = .28$, $p = .002$). Vocabulary was significantly related to ToM ($\beta = .26$, $p =.02$) and inference ($\beta = .36$, $p < .001$), but not to comprehension monitoring ($\beta = .11$, $p = .35$). Both vocabulary ($\beta = .38$, $p = .001$) and grammatical knowledge ($\beta = .27$, $p =. 01$) were directly related to LC. ToM was significantly independently related to LC ($\beta = .23$, $p = .04$), but inference ($\beta = .19$, $p = .09$) nor comprehension monitoring were ($\beta = .09$, $p = .31$). A total of 64% of variance in Time 3 RC was explained by LC and word reading.

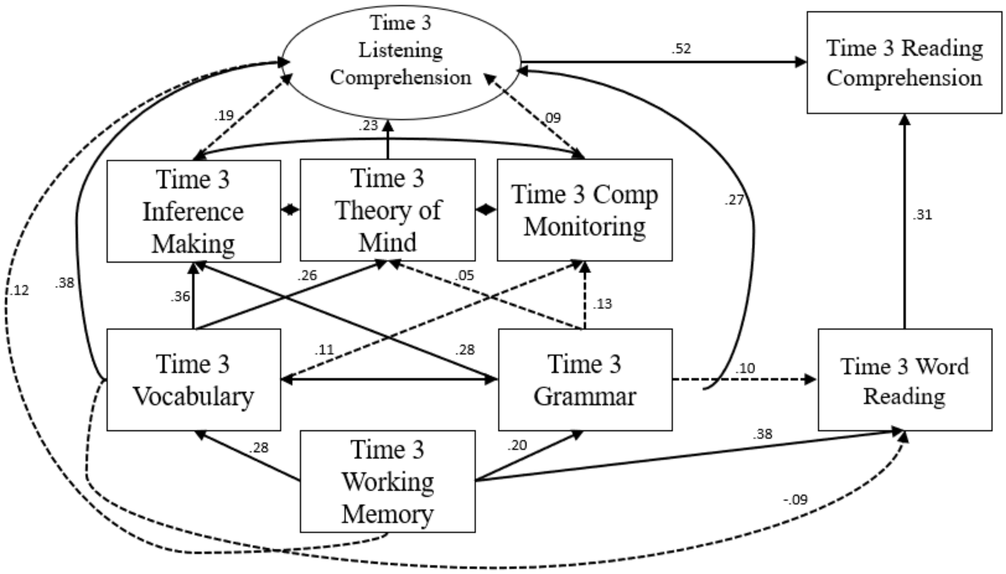


Figure 6.13: Model 5 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

Next, a model based on the DIER model (Kim, 2017) and the above but with a latent variable of broad metacognition instead of ToM was fitted to the data (Model 6; see Figure 6.14). Prior to SEM analysis, missing data (see Table 6.5 for details of which measures) was imputed using EM in SPSS. This method of data imputation was used because missing data was minimal (five cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random ($p = .06$).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-

10). Multivariate normality assumptions were also checked and were met as confirmed by a Mardia's coefficient of multivariate critical ratio of .256 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008. This model had a poor fit: χ^2 (44) = 83.03, *p* < .001; CFI = .92, TLI = .84, and RMSEA = .09. Six multivariate outliers were identified using Mahalanobis d-squared (those with a p < .05). When these six participants were removed and the model re-run the model only changed slightly fit changed: χ^2 (44) = 103.45, *p* < .001 ; CFI = .88, TLI = .75, and RMSEA = .12, therefore the original model with all 107 participants is reported here.

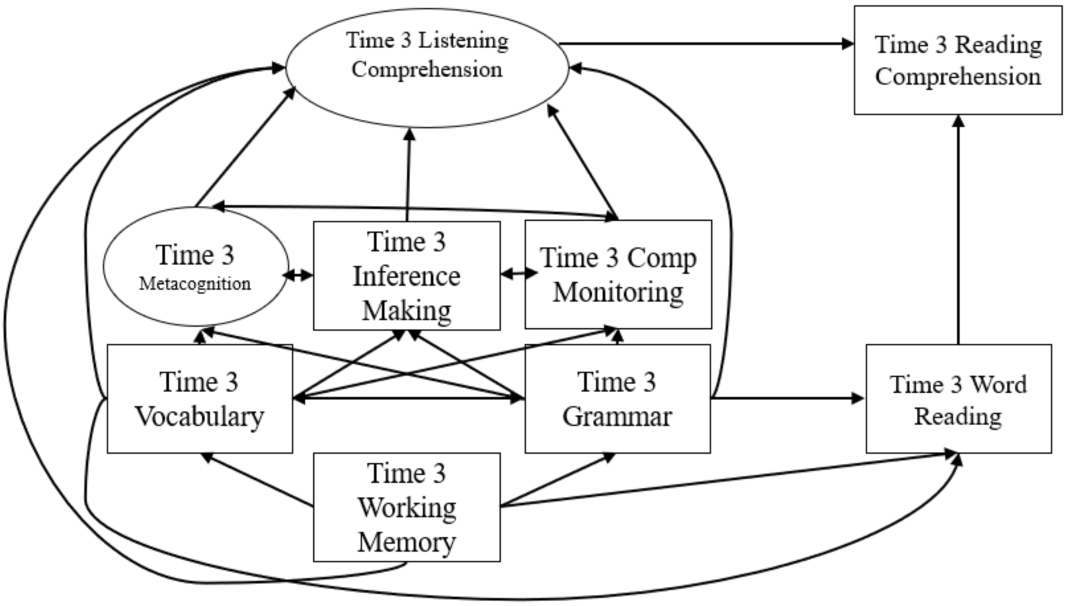


Figure 6.14: Hypothesised Model 6. A concurrent model including broad metacognition. Controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.15. As presented in Figure 6.15, after controlling for age and non-verbal ability the SVR was supported as both LC (or linguistic comprehension) (β = .63, p < .001) and word reading (β = .30, p < .001) significantly related to RC. Working memory (β = .38, p

< .001) but not grammatical knowledge (β = .10, p = .30) nor vocabulary (β = -.09, p = .35) was significantly related to word reading. Working memory was significantly related to vocabulary (β = .28, p = .001) and to grammatical knowledge (β = .20, p = .03) but not to LC (β = .05, p = .54). Grammatical knowledge was not significantly related to metacognition (β = .10, p = .38), or comprehension monitoring (β = .13, p = .26), but was to inference making (β = .30, p = .002). Vocabulary was significantly related to metacognition (β = .30, p =.008) and inference (β = .36, p <.001), but not to comprehension monitoring (β = .11, p = .35). Both vocabulary (β = .32, p =.003) and grammatical knowledge (β = .23, p =. 03) were directly related to LC. Metacognition was significantly independently related to LC (β = .50, p = .002), as was inference (β = .24, p = .03), but comprehension monitoring was not (β = .02, p = .82). A total of 47% of variance in Time 3 RC was explained by LC and word reading.

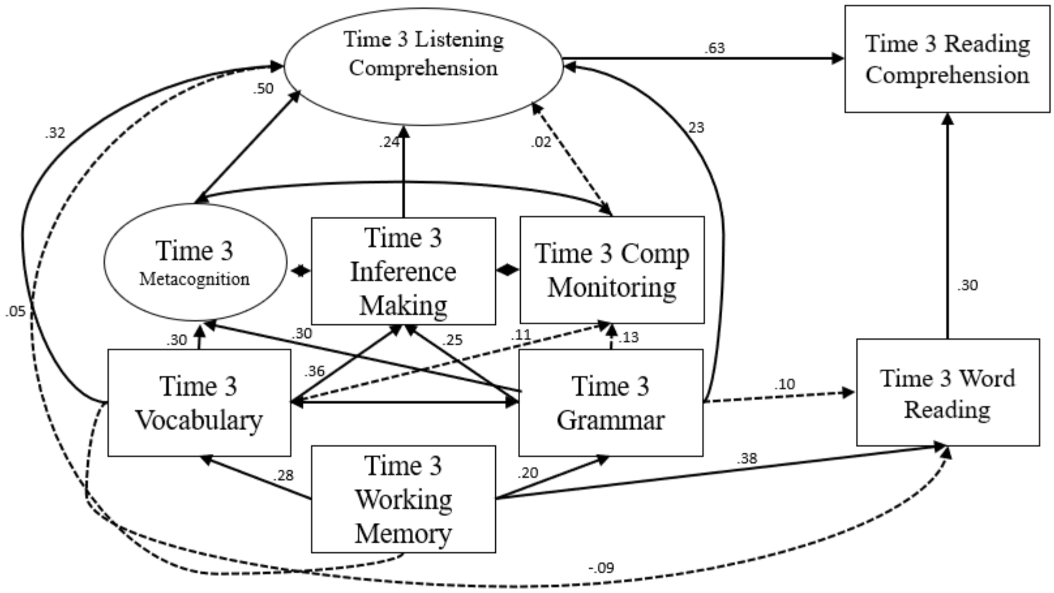


Figure 6.15: Model 6 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

6.3.2.2 Longitudinal DIER models of reading comprehension

Next longitudinal DIER models were tested for Time 1 (skills at the age of four) predicting Time 3 RC (at the age of six). One model included ToM (Model 7) and the other included a latent variable of broad metacognition (Model 8).

6.3.2.2.1 Descriptive statistics and preliminary analysis

Table 6.7 (see above) shows descriptive statistics for all measures used in the longitudinal DIER models. As with the DIET longitudinal models, a ToM composite consisting of the two ToM measures was computed by summing scores from the two ToM measures from Time 1 (unexpected contents and unexpected locations). This composite was used to give a richer measure of ToM and was justified as the individual measures correlated significantly (r = .39, p <.001). Table 6.8 (also above) shows correlations between the measures used in the DIER longitudinal models.

6.3.2.2.2 SEM analysis

A longitudinal model based on the DIER model[13] (Kim, 2017) was fitted to the data (Model 7; see Figure 6.16) for Time 1 skills predicting Time 3 RC. Prior to SEM analysis, missing data (see Table 6.7 for details of which measures) was imputed using EM in SPSS. This method of data imputation was used because missing data was minimal (seventeen cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random (p =.32).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were met as confirmed by a Mardia's coefficient of multivariate critical ratio of .869 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). This model had an excellent: χ^2 (15) = 13.85, p = .54; CFI = 1.00, TLI = 1.01, and RMSEA <

[13] As noted before the longitudinal models did not include comprehension monitoring and inference making as these were not measured at Time 1

.001. Five multivariate outliers were identified using Mahalanobis d-squared (those with a p < .05). When these five participants were removed and the model re-run the model fit only changed slightly: χ^2 (14) = 12.68, *p* = .55; CFI = 1.00, TLI = 1.02, and RMSEA < .001, therefore the original model with all 107 participants is reported here.

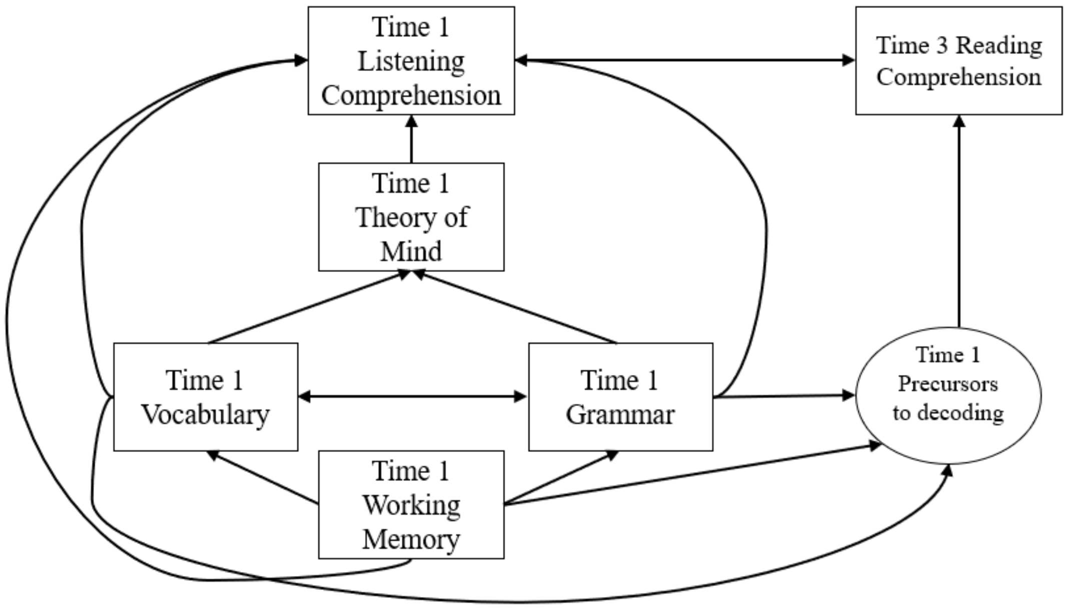


Figure 6.16: Hypothesised Model 7. A longitudinal model including ToM, after controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.17. As presented in Figures 6.17 after controlling for age and non-verbal ability the SVR was supported as both LC (or linguistic comprehension) (β = .19, p =.04) and decoding precursors (β = .38, p =.03) significantly related to RC. Neither working memory (β = .16, p =.10) nor grammatical knowledge (β = .32, p = .07), or vocabulary (β = .06, p = .75) was significantly related to precursors to decoding. Working memory was significantly related to vocabulary (β = .23, p =.005) and grammatical knowledge (β = .22, p = .02) but not to LC (β = .16, p = .10). Grammatical knowledge was significantly related to ToM (β = .36, p = .002). Vocabulary was not significantly related to ToM (β = .06, p = .66). Vocabulary (β = .29, p = .03) was directly related to LC, but grammatical knowledge (β = -.003, p = .98) was not. ToM was not significantly independently related to LC (β = .18, p = .06). A total of 44% of variance in Time 3 RC was explained by Time 1 LC and word reading.

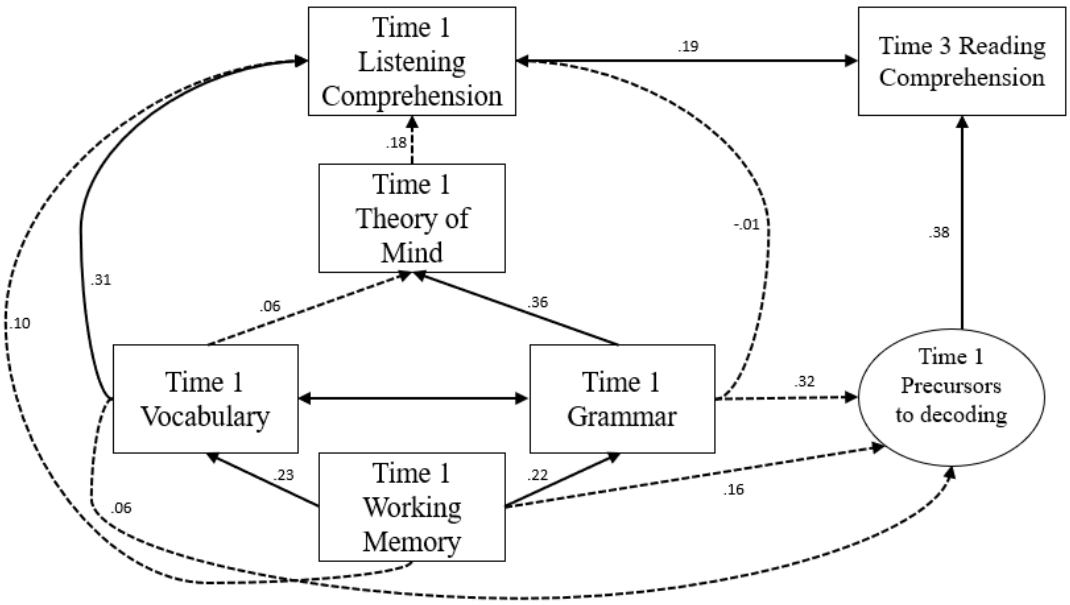


Figure 6.17: Model 7 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

Next, a longitudinal model based on the DIER model (Kim, 2017) and the above but with a latent variable of broad metacognition instead of ToM, was fitted to the data for Time 1 skills predicting Time 3 RC (Model 8; see Figure 6.18). Prior to SEM analysis, missing data (see Table 6.7 for details of which measures) was imputed using EM in SPSS. This method of data imputation was used because missing data was minimal (fourteen cases), and Hair, Black, Babin, Anderson, and Tatham (2006) recommend its usage with data sets with less than 250 participants. Furthermore, Little's MCAR test reported the data to be missing at random ($p = .17$).

For each of the variables in the model univariate normality assumptions were appropriate for SEM analysis as indicated by skewness and kurtosis values proposed by Chou and Bentler (1995) and Kline (2005) (skewness < +3/-3 and kurtosis < 10/-10). Multivariate normality assumptions were also checked and were met as confirmed by a Mardia's coefficient of multivariate critical ratio of .242 (a value < -/+ 1.96 demonstrates normality; Gao, Mokhtarian, & Johnston, 2008). This model had a good fit: χ^2 (32) = 40.97, p = .13; CFI = .98, TLI = .95, and RMSEA = .05. Nine multivariate outliers were identified using Mahalanobis d-squared (those with a $p < .05$). When these nine participants were removed and the model re-run, the model fit changed to poor: χ^2 (32) = 50.12, p = .02; CFI = .95, TLI = .90, and RMSEA = .08, therefore this model without the outliers removed is presented here.

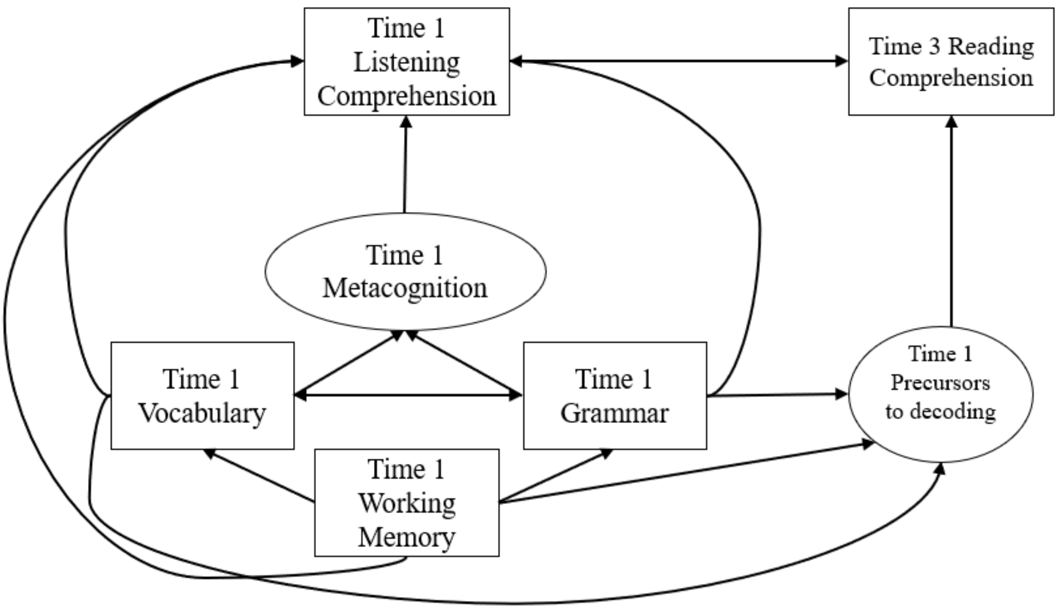


Figure 6.18: Hypothesised Model 8. A longitudinal model including broad metacognition. Controlling for age and non-verbal ability. *Note.* Two-sided arrows represent covariances. Grammar assessed by the knowledge of syntax.

Standardised path coefficients of the model are shown in Figure 6.19. As presented in Figures 6.19, after controlling for age and non-verbal ability the SVR was not supported as neither decoding precursors ($\beta = .38, p = .32$) nor LC (or linguistic comprehension) ($\beta = .02, p = .88$) were significantly related to RC. Grammatical knowledge was related to decoding precursors ($\beta = .38, p = .02$) but working memory ($\beta = .21, p = .11$) and vocabulary ($\beta = .09, p = .61$) were not. Working memory was significantly related to vocabulary ($\beta = .22, p = .008$) and to LC ($\beta = .25, p = .02$) but not grammatical knowledge ($\beta = .19, p = .06$). Grammatical knowledge was significantly related to metacognition ($\beta = .72, p < .001$). Vocabulary was not significantly related to metacognition ($\beta = -.02, p = .90$). Vocabulary ($\beta = .29, p = .04$) was directly related to LC, but grammatical knowledge ($\beta = -.12, p = .75$) was not. Metacognition was not significantly independently related to LC ($\beta = .28$, p =

.60). A total of 27% of variance in Time 3 RC was explained by Time 1 LC and word reading.

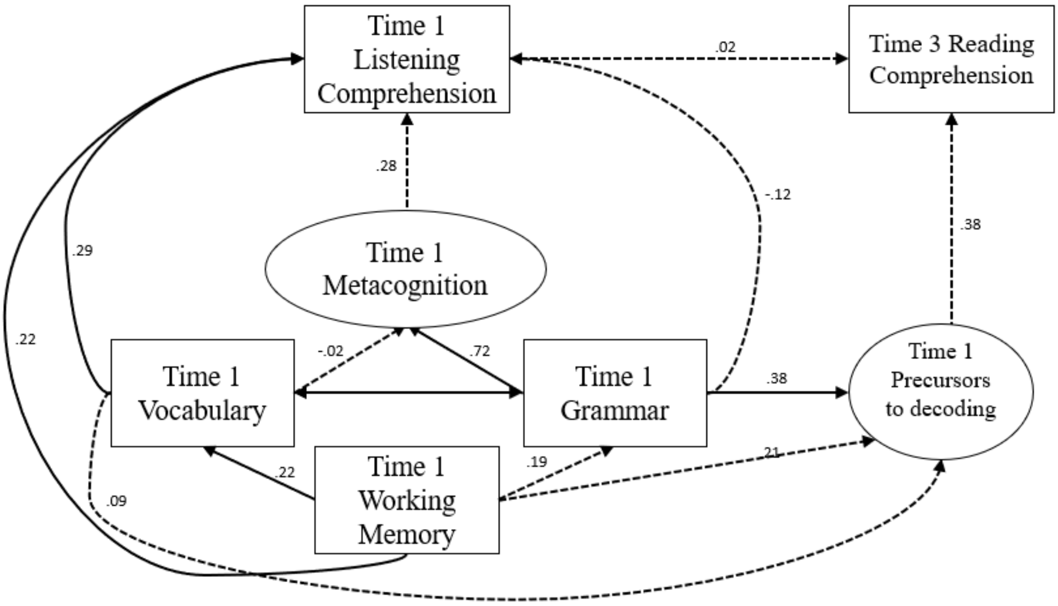


Figure 6.19: Model 8 with standardised path coefficient weight. *Note.* Two-sided arrows represent covariances. Complete lines represent significant relations and dashed lines represent non-significant relations.

6.3.3 Further analysis

Further analysis was carried out in order to produce further information not gained through the main analysis.

6.3.3.1 Children with English as an additional language

A proportion of the sample (31%) had English as an additional language (EAL). This could have affected results as EAL children show RC difficulties (Bowyer-Crane et al., 2017). Therefore, multigroup analysis was used in AMOS to compare the fit of all models for EAL participants to English only speaking participants. This was preformed using a chi-square difference test whereby a non-significant chi-square shows that the model fit is the same for both groups (Dimitrov, 2006). For each of the eight models a non-significant chi-square demonstrated that the fit was no different for EAL and English only speaking participants, Model 1: $\chi^2(32) = 23.64$, $p = .60$, Model 2: $\chi^2(32) = 36.36$, $p = .27$, Model 3: $\chi^2(16) = 15.67$, $p = .48$, Model 4: $\chi^2(3) = 3.39$, $p = .35$, Model 5: $\chi^2(36) = 42.80$, $p = .20$, Model 6: $\chi^2(44) = 52.95$ $p = .17$, Model 7: $\chi^2(26) = 28.10$, $p = .35$, Model 8: $\chi^2(28) = 23.77$, $p = .59$.

6.3.3.2 Power analysis

Power analysis of each model was calculated using an online power calculator (Soper, 2017). This analysis showed each of the eight models to have sufficient power as the observed statistical power reached at least .80 for each (.80 is the cut off for minimum power as advocated by Cohen, West, & Aiken, 2014). The observed statistical powers were: Model 1 = .99, Model 2 = 1.00, Model 3 = .98, Model 4 = .98, Model 5 = 1.00, Model 6 = 1.00, Model 7 = .99 and Model 8 = .98.

However, a calculator of sample size required (Soper, 2015; Westland, 2010) suggested that for all of the models tested within this chapter the required minimum sample size was either not quite or was just met. Model 1 had 103 participants and the minimum required was 97. Model 2 had 98 and the minimum required was 108. Model 3 had 107 and the minimum required was 84. Model 4 had 107 and the minimum required was 97. Model 5 had 107 participants and the minimum required was 108. Model 6 had 107 participants and the minimum required was 118. Model 7 had 107 participants and the minimum required was 97. Model 8 had 98 participants and the minimum required was 108.

6.4 Discussion

The aim of this chapter was to gain a deeper understanding of the role played by ToM in RC by exploring both concurrent and longitudinal models of RC which included roles for ToM or metacognition. This chapter compared the contribution of ToM to other metacognitive skills non-social in nature in models of RC. The rationale for making this comparison was to determine the nature of ToM which is important for RC. This was important as work has not addressed the question over why ToM assists RC. Further, longitudinal research into RC and ToM is, thus far, lacking.

The chapter tested both DIET and DIER models (See Kim, 2017). The DIET models (Direct and Indirect Effects models of Text comprehension) included a path straight from ToM or broad metacognition to RC. Whereas, the DIER models (Direct and Indirect Effect Model of Reading) included the SVR with RC as a product of just two component skills (word reading and linguistic comprehension), but with linguistic comprehension comprising of many sub-components including metacognition. In testing both types of models this chapter sought to assess both the direct and indirect effects of ToM on RC.

It was hypothesized that across both concurrent and longitudinal DIET and DIER models, models which included ToM over models which included broad metacognition would have a better fit. Moreover, it was anticipated that DIET models would show that ToM makes a direct contribution to RC whereas broad metacognition does not. Put simply, it was predicted that ToM assists RC both concurrently and longitudinally, directly and indirectly, but broad metacognition does not because it is the social element of ToM which is important for RC. The social element may be central for RC because a better awareness and understanding of mental states will help with the understanding of thoughts, desires and perspectives of a character in a story, or the intentions of an author of a story.

Following SEM analysis these hypotheses were partially supported. Table 6.9 summarises the findings of the eight models in this chapter, and the nature of this partial support is further detailed in the following sections.

Table 6.9

A summary of the fit of models in Chapter 6

Model	Fit	Prediction of RC
Model 1 – Concurrent DIET at Time 1 (ToM)	Excellent fit	No direct significant path from ToM to RC
Model 2 – Concurrent DIET Time 1 (metacognition)	Poor fit	No direct significant path from metacognition to RC
Model 3 – Longitudinal DIET Time 1 to Time 3 (ToM)	Excellent fit	No direct significant path from ToM to RC
Model 4 – Longitudinal DIET Time 1 to Time 3 (metacognition)	Good fit	No direct significant path from metacognition to RC
Model 5 – Concurrent DIER Time 3 (ToM)	Good fit	Indirect significant path from ToM to RC via LC
Model 6 – Concurrent DIER Time 3 (metacognition)	Poor fit	Indirect significant path from metacognition to RC via LC
Model 7 – Longitudinal DIER Time 1 to Time 3 (ToM)	Excellent fit	No indirect significant path from ToM to RC via LC
Model 8 – Longitudinal DIER Time 1 to Time 3 (metacognition)	Good fit	No indirect significant path from metacognition to RC via LC

6.4.1 DIET models of reading comprehension

First, both concurrent and longitudinal DIET models of RC were tested. These models included a direct path from ToM or broad metacognition to RC. Findings showed that concurrently the model which included ToM (Model 1) had an excellent fit, whereas the model which included broad metacognition (Model 2) was a poor fit. These findings are consistent with Kim (2017) showing that a model which includes working memory, vocabulary, syntactic knowledge, ToM, comprehension monitoring and inference making fits well for concurrent RC.

In contrast to Kim (2017), and against the hypothesis, the ToM model (Model 1) did not show a significant direct path from ToM to RC. Reasons for this contrasting finding could be due to sample size (100 participants compared to 350 in Kim, 2017), younger aged participants (mean age in years and months 6;1 here and 7;3 for Kim, 2017), and that here age and non-verbal memory were controlled for. This said, this path between ToM and RC was approaching significance and perhaps a larger sample size would have led to a significant path. Although power calculations indicate sufficient power overall, the specific sample size calculations indicated that the required sample size was not, or was only just, met. Given also that the sample size here is markedly smaller that Kim (2015; 2016; 2017), then increasing sample size for these specific analyses would strengthen the claims that can be made from these findings. Despite the finding that ToM did not directly relate to RC contrasting with the recent work of Kim (2017) and Atkinson et al., (2017), this finding does support the SVR which states that only linguistic comprehension and word reading directly relate to RC and any other skills which contributes towards RC is a sub-skill of either linguistic comprehension or word reading.

When the DIET models were extended longitudinally for skills aged four (Time 1) predicting RC aged six (Time 3), results were consistent with the concurrent findings. Again, a model which included ToM (Model 3) had a better fit than one which included broad metacognition (Model 4). Also comparable with the concurrent findings, neither the ToM model (Model 3) or the broad metacognition (Model 4) showed a significant direct path from ToM aged four to RC aged six, but yet again for the ToM model (Model 3) this was marginal. This was in stark contrast to the path from broad metacognition aged four to RC aged six in the metacognition model (Model 4), which was considerably non-significant, highlighting again that ToM has a stronger predictive relationship with RC than broad metacognition.

Overall the DIET models provide slightly more evidence for the role of the social element of ToM in RC (both concurrently and longitudinally) than they do for the role of broad metacognition as these models were better fitting and the paths from ToM to RC were marginally not significant next to the paths from metacognition to RC which were considerably non-significant. Further work is required with a larger sample to cement these conclusions.

6.4.2 DIER models of reading comprehension

Secondly, DIER models were tested, both concurrently and longitudinally. These models did not include a direct path from metacognition and ToM to RC, instead these were housed within the sub-component of linguistic comprehension, and so the path from broad metacognition or ToM to RC was indirect. With the exception of Model 8 (the longitudinal model including broad metacognition), all models supported the SVR, in that the skills of linguistic comprehension and word reading significantly predicted RC. This is consistent with the vast amount of empirical evidence which supports the SVR (e.g. Catts et al., 2015; Chen & Vellutino, 1997;

Cutting, & Scarborough, 2006; Gough, Hoover, Peterson, Cornoldi, & Oakhill, 1996b; Johnston & Kirby, 2006; Joshi & Aaron, 2000; Proctor, Carlo, August, & Snow, 2005; Savage, 2001; Savage, 2006; Storch & Whitehurst, 2002; Tiu et al., 2003; Vellutino et al., 2007). In Model 8 as the linguistic comprehension component did not significantly predict RC, this can be seen as evidence that metacognition does not fit well within a DIER model or the SVR.

Concurrent DIER findings showed that the model which included ToM (Model 5) had a good fit, whereas the model which included broad metacognition (Model 6) was a poor fit. This is consistent with Kim (2017) as a model which includes working memory, vocabulary, syntactic knowledge, ToM, comprehension monitoring, inference making and listening comprehension as a linguistic comprehension component, along with word reading predicts concurrent RC. Further, this supports the hypothesis that a model which includes ToM is superior to one which includes broad metacognition. Within Model 5 there was a significant path from ToM to LC, again supporting Kim (2017).

When the DIER models were extended longitudinally for skills aged four (Time 1) predicting RC aged six (Time 3) results were consistent with the concurrent findings; the model which included ToM (Model 7) had a good fit whereas the broad metacognition model (Model 8) was a poor fit. This again supports the hypothesis that the social element of ToM is important for RC rather than the general metacognitive nature. However, Model 7 was unable to show a significant path from ToM to LC, and therefore it cannot be claimed that longitudinally ToM indirectly via

LC predicts RC. Due to this it cannot be said that the findings from this chapter fully support Atkinson et al., (2017) who did find ToM to predict RC two years later indirectly via LC. Yet, as with the DIET models this path (between ToM and LC) was approaching significance, and so if the sample size was larger (closer to that of Kim, 2017) this may have become significant. More research with a larger sample size is needed to explore whether this is the case.

6.4.3 Limitation and further direction

Taken together results from DIET and DIER models tested here provide some evidence that ToM is important for RC both concurrently at the age of six, and longitudinally from four years old across 22 months to 6 years old. The findings also show that ToM is somewhat better (within a SEM model) at predicting RC both concurrently and longitudinally than other non-social metacognitive skills (a latent variable of metamemory, source monitoring and metalinguistic awareness), perhaps suggesting that it is the social nature of ToM which facilitates RC. However, when it comes to direct and indirect paths from ToM to RC for all but Model 5 (which did show a significant indirect path via LC from ToM to concurrent RC) these paths were marginally non-significant. Therefore, as with all previous chapters the main limitation of this work was that the sample size was only just large enough for SEM analysis. SEM analysis is sensitive to sample size (E. J. Wolf et al., 2013) and as demonstrated in Section 6.3.3.2 for most of the models tested within this chapter the required minimum sample size was either not quite or was only just met. Indeed, sample size has already been suggested as an issue by similar work (Guajardo & Cartwright, 2016). Future work needs to address this in order to assess if a larger sample size would ensure these paths became significant. This future research could support the work of Kim (2015; 2017) and Atkinson et al., (2017).

6.4.4 Conclusions

Concurrent and longitudinal DIET and DIER models tested in this chapter provide some evidence that ToM is important for RC because of its social nature and not because of its metacognitive nature. The social element of ToM may assist with the understanding of social information in a story and underpin mental model building of the plot. These findings support very recent work (Atkinson et al., 2017; Kim, 2015; Kim, 2017; Ricketts et al., 2013) and the theory of Dore et al., (2018). The findings from this chapter are novel as they are the first to confirm this is the case longitudinally using SEM analysis, and they were also able to partially determine that it is the social nature of ToM not its general metacognitive nature which facilitates RC. The findings support the SVR because in the DIET models ToM or metacognition were not shown to make a direct contribution to RC suggesting, in line with the SVR, that RC is a product of linguistic comprehension (and its sub-components) and word reading only. Findings also support the situation model as the situation model theoretically underpins the DIET and DIER models.

These novel findings have important applied implications because although in line with the SVR a strong focus is given in the UK early years curriculum to phonics (Rose, 2009), the current findings suggest that social understanding should perhaps also be fostered within the early years classroom as it too may be important for emerging RC. However, further work is needed with larger sample sizes to confirm these conclusions as some direct/indirect paths from ToM to RC were marginally not significant.

7 Mental state talk, theory of mind, and listening comprehension

7.1 Introduction

7.1.1 Theory of mind and mental state talk

The social account of theory of mind (e.g. Carpendale & Lewis, 2006; Heyes & Frith, 2014) focuses on the role of socially mediated processes in underpinning theory of mind (ToM) development (Nelson, 2004). This is the approach also taken by Dunn et al. (1991) who conducted the pioneering study into maternal mental state talk and ToM. Maternal mental state talk is mother's talk with her child involving mental states such as, cognition e.g. "*think*" "*know*", emotions e.g. "*sad*" "*happy*" and desires e.g. "*want*" "*dislike*". The research with fifty mothers and their children, showed that that the frequency of maternal mental state talk at 33 months old predicted individual differences in a child's ability to pass a false belief test aged 40 months over and above other types of talk such as talk about causality (i.e. where the cause of events was discussed). Dunn et al. (1991) concluded that through family talk about mental states, thoughts and memories, children's attention is focused on these mental states and as a result a stronger ToM awareness is developed. This has been said to be consistent with the work Vygotsky's zone of proximal development (Vygotsky, 1978) as it suggests that mothers' talk scaffolds ToM understanding (Taumoepeau & Ruffman, 2008).

There is now a vast body of support for the findings of Dunn et al. (1991). Indeed, a recent meta-analysis found that across 28 studies including 1,914 children aged two to five years old from seven different countries, there was a modest but statistically significant relationship between maternal mental state talk and children's ability to pass a false belief test (r = .21), which held even when child's verbal ability was accounted for (r =.19) (Devine & Hughes, 2016a). Yet it is interesting to note that this meta-analysis found that more recent studies show less of an effect size than older research. Typically, these studies have taken place in the home whereby conversations between mothers and their child engaged in normal activities (often during play, mealtimes or book sharing) are recorded, transcribed and coded for different types of mental state talk.

The relationship between maternal mental state talk and child ToM remains across time, with maternal mental state talk at three to five years predicting ToM a year later (Adrián et al., 2007), maternal mental state talk at two years predicting ToM at six years (Ensor et al., 2014), and maternal mental state talk at six years predicting ToM at 10 years (Ensor et al., 2014). This holds even when controlling for a child's earlier ToM and their general language ability (e.g. Ruffman et al., 2002). The relationship endures across childhood with maternal mental state talk at the age of three and four shown to relate to ToM seven years later when children are ten years old (A. Carr et al., 2018). Importantly, maternal mental state talk is considered causal of ToM, as cross-lagged studies show that earlier maternal mental state talk predicts ToM but not vice versa (Ruffman et al., 2002).

The content of mothers' mental state talk changes over time. Mothers use a higher frequency of desire language than cognitive and emotion terms when children are two years old (Taumoepeau & Ruffman, 2008), but a shift occurs so that by the age of six years mothers are making twice as many cognitive (think/know) references than emotion and desire references (Ensor et al., 2014; Jenkins et al., 2003). Generally, research suggests that total number of all maternal mental state talk utterances relates to ToM (J. Dunn et al., 1991; Meins et al., 2003; Ruffman et al., 2002), but there is some recent research which suggest that that for older children, talk about cognitive terms is most important. For example, at three years old mothers talk about cognition was shown to be the best predictor of ToM at five years old (Adrián et al., 2007).

Research also suggests that children's *own* mental state talk relates to their ToM understanding both concurrently and longitudinally. This is perhaps because children who reference mental states habitually, do so because of a preconscious understanding of the mind (Bartsch & Wellman, 1995). Research supporting this shows that for five to seven year olds, a child's use of mental state terms during a story telling task was shown to strongly relate to their concurrent ability to pass a false belief task (Symons et al., 2005).

An alternative explanation for the relationship between child's own mental state talk and ToM is the inverse of the above account, that instead talking about mental states frequently helps children to form a solid understanding of one's own and others' mental states (Garner, Jones, Gaddy, & Rennie, 1997). This is supported by the original study by Dunn et al. (1991) which found that children's own mental state talk aged 33 months (two years and nine months) was related to their ToM seven months later aged 40 months (three years and four months). This finding is supported by Ruffman et al., (2002) who showed that children's own mental state talk aged three years related to ToM a year later.

While the majority of research has been conducted using recorded conversational methods in order to gain rich data some studies have opted for a self-report method whereby mothers report on their usage and preferences of mental state words and phrases (C. Peterson & Slaughter, 2003). Although self-reports do have their biases, they are important because it is possible that during recorded sessions both children and mothers do not act naturalistically in this unfamiliar situation where they are aware of being recorded. Findings from self-report studies show a relationship between mothers' reports of their own mental state talk usage and their child's ToM development both concurrently and longitudinally (Ebert, Peterson, Slaughter, & Weinert, 2017; Farrant et al., 2012; C. Peterson & Slaughter, 2003). There is however limited research in which mothers report on their child's mental state term understanding and usage especially in conjunction with direct observation of mental state talk.

7.1.2 Theory of mind and general language

The body of work described above demonstrates a relationship between mental state talk (both maternal and child's) and ToM. There is also strong evidence for a link between ToM and a child's own general language ability with numerous studies, including correlational and longitudinal, showing that language ability is related to ToM understanding. For example, Jenkins and Astington (1996) found a correlation between the ability to pass a false belief test and general language ability (as measured by a standardized language test) in typically developing three to five year olds, a finding supported by others (e.g. Cutting, & Dunn, 1999; Hughes & Dunn, 1997). Indeed, a meta-analysis of 104 studies indicated a moderate to large correlation between language and false belief understanding independent of age (Milligan et al., 2007). This correlation is reported across different language skills including vocabulary, semantics, syntactic knowledge, and pragmatics (Milligan et al., 2007).

Longitudinal studies also support a relationship between ToM and general language, showing that language ability aged two predicts false belief understanding at the age of four (Farrar & Maag, 2002; Watson et al., 2001), and that vocabulary at the age of four correlates with later ToM understanding (Hughes, 1998b). Longitudinal studies also show that general language (including semantics and syntax) at the age of three predicts later false belief understanding measured at several subsequent time points (Astington & Jenkins, 1999; Ruffman et al., 2003). Importantly, as noted in previous chapters, ToM is also important for later language development (Milligan et al., 2007; Slade & Ruffman, 2005) with research showing a bi-directional relationship between language and ToM, with early language ability being important for later ToM and also, importantly, early ToM being important for later language, including comprehension of stories (e.g. Strasser & Río, 2014).

7.1.2.1 Theory of mind and listening comprehension

Research indicates that there is a relationship between ToM and listening comprehension (LC) in that concurrent ToM predicts LC from the ages of four to twelve years (Kim, 2015; Kim, 2016; Kim, 2017; Pelletier & Beatty, 2015). Indeed, previous chapters of this thesis have found that concurrent ToM at the age of six years predicts LC, and that longitudinally ToM aged four predicts LC a year later at the age of five within a SEM model of LC (Chapter 3). Furthermore, it has been argued elsewhere in this thesis that ToM is a better predictor of LC than broad metacognition (including metamemory, source monitoring and metalinguistic knowledge) both concurrently (Chapter 4) and longitudinally (Chapter 5). This is maybe because the social nature of ToM aids the listener in understanding the viewpoint, desires and intentions of the speaker and could lead to better awareness of social information and details within the spoken passage (Dore et al., 2018).

7.1.3 Mental state talk, theory of mind, and listening comprehension

Given that there is a wealth of evidence to show that mental state talk (both maternal and child's) predicts both concurrent and future ToM, and that an increasing number of recent studies show that ToM predicts LC both concurrently and longitudinally, a mediational relationship between these three factors is possible, in which mental state talk relates to LC directly and indirectly via ToM.

Research already shows that talk around a book during shared reading has an influence on a child's LC. For example, Zucker, Cabell, Justice, Pentimonti, and

Kaderavek (2013) found that extratextual talk by preschool teachers during shared reading related to children's LC both concurrently and a year later when they were five years old, whereas the frequency of the shared reading did not. In this study extratextual talk was measured by coding literal, inferential and phonological talk during class shared reading. Similar findings are seen during shared book reading at home with mothers with the same age group. Children assigned to an extratextual talk group outperformed children in a control group (in which mothers simply read a story with no talk) on LC measures (Collins, 2016). Likewise, mothers talk (including describing the pictures and extending the topic) during book reading was found to predict both child vocabulary and reading comprehension scores a year later when children were six years old (Demir, Applebaum, Levine, Petty, & Goldin-Meadow, 2011). Until now research has not focused specifically on mental state talk during book sharing in relation to LC.

Book sharing is a great opportunity for talk about mental states (Symons et al., 2005). Very recently, Zunshine (2019) argued that many children's stories, including those aimed at very young children, contain mental states and as children progress onto more advanced books mental states within stories becomes more complex too. For example, Julia Donaldson's popular storybook "Gruffalo" (aimed at three to seven year olds) tells the tale of a big scary monster who believes a mouse's claims that she is the most powerful animal in the forest. Preschoolers are 'let-into' the understanding that the Gruffalo does not realise that when he is walking behind the mouse in the forest, other animals are scattering because they are afraid of him, and not of the tiny mouse (Zunshine, 2019). Such a story plot provides the opportunity for much talk about mental states between mother and child. It is possible that both maternal and child mental state talk during book sharing of these sorts of picture books facilitates ToM understanding, which in turn improves children's comprehension of such stories. Investigation into the mediational relationship between these three factors will give a deeper understanding of the relationship between ToM and LC.

Figure 7.1 and Figure 7.2 illustrate the hypothesised direct and indirect paths between both maternal and child mental state talk, ToM and LC. Here, and consistent with past literature, mental state talk predicts ToM, and ToM predicts LC. In addition, these models also suggest that maternal/child mental state talk directly (and indirectly via ToM) predicts LC. These direct/indirect relationships could be true both concurrently and longitudinally.

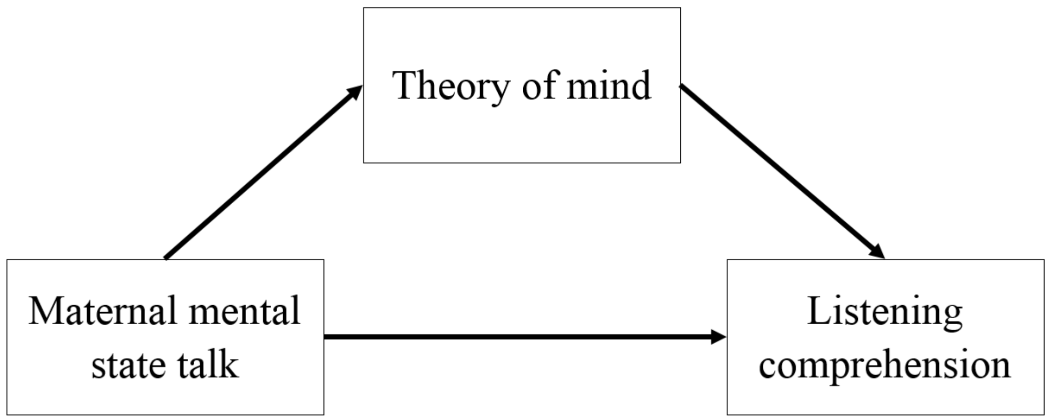


Figure 7.1: The hypothesised relationship between ToM, LC and maternal mental state talk in which ToM mediates the relationship between the two other factors. It is hypothesized that this will be the case both concurrently and longitudinally with earlier maternal mental state talk predicting both later ToM and LC.

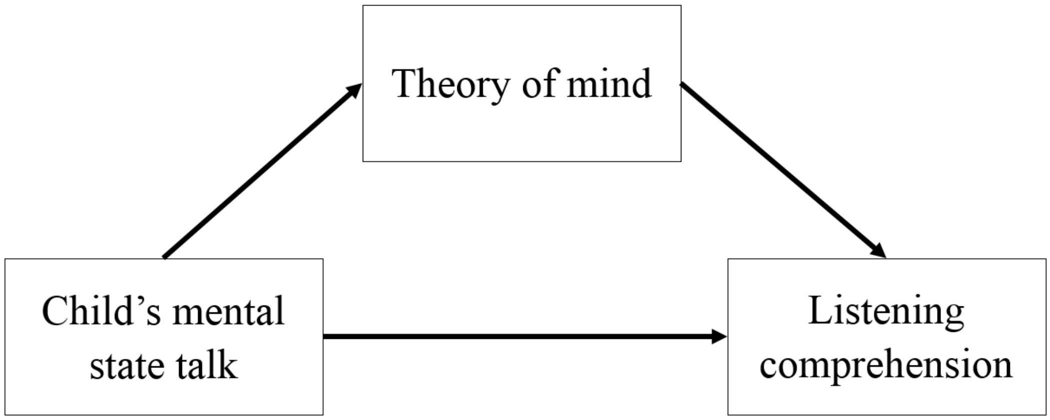


Figure 7.2: The hypothesised relationship between ToM, LC and child mental state talk in which ToM mediates the relationship between the two other factors. It is hypothesized that this will be the case both concurrently and longitudinally with earlier child mental state talk predicting both later ToM and LC.

7.1.4 The present study

Existing concurrent and longitudinal research links both maternal and child mental state talk to ToM, and ToM to LC. However, no study has explored these three factors together to examine whether ToM mediates the relationship between the other two factors, concurrently and longitudinally. Therefore, this was the main aim of this chapter. This exploration gives a deeper understanding of the relationship between ToM and LC investigated in Chapter 3 and Chapter 4. Longitudinal work, in addition to concurrent work, is important because it provides evidence of consistent developmental patterns and causal relationships can be suggested (Caruana et al., 2015).

To address this aim, a longitudinal study with two time points assessing maternal and child mental state talk, ToM, LC and other skills known to be important for LC (vocabulary, syntactic knowledge and working memory) was implemented. Maternal mental state talk was measured by coding conversations between children and their mothers during a picture book sharing activity. Child mental state talk was measured in two ways, through mother self-report of their child's mental state understanding and through coding conversations between children and their mothers during a shared reading activity. Measuring child mental state talk in two ways allowed for the biases in each method to be acknowledged and addressed. It also gave a larger sample size, as more participants took part in the self-report element of the research.

All skills were measured at Time 1 when children were three years old, and again at Time 2 when they were four years old. This age group was important because this is a crucial age for both ToM and LC development (Strasser & Río, 2014; Wellman et al., 2001). ToM and mental state talk were examined for their ability to predict LC after controlling for other language skills known to be important for LC.

It was hypothesized that:

1) Concurrently, at both the ages of three (Time 1) and four (Time 2) children's mental state talk (as measured by parental self-report) would directly predict both ToM and LC. ToM would also mediate the relationship between mental state talk and LC.
2) Longitudinally, child mental state talk (as measured by parental self-report) at the age of three (Time 1) would directly predict both ToM and LC a year later at the age of four (Time 2). Time 2 ToM would also mediate the relationship between Time 1 mental state talk and Time 2 LC.

3) Concurrently, at both the ages of three (Time 1) and four (Time 2) children's mental state talk (as coded from dialogue with their mother) would directly predict both ToM and LC. ToM would also mediate the relationship between mental state talk and LC.
4) Longitudinally, child mental state talk (as coded from dialogue with their mother) at the age of three (Time 1) would directly predict both ToM and LC a year later at the age of four (Time 2). Time 2 ToM would also mediate the relationship between Time 1 mental state talk and Time 2 LC.
5) Concurrently, maternal mental state talk (as coded from dialogue with their child) when children are both three years old (Time 1) and four years old (Time 2) would directly predict both a child's ToM and LC. ToM would also mediate the relationship between maternal mental state talk and child LC.
6) Longitudinally, maternal mental state talk (as coded from dialogue with their child) when children are age of three (Time 1) would directly predict both ToM and LC a year later at the age of four (Time 2). Time 2 ToM would also mediate the relationship between Time 1 maternal mental state talk and Time 2 LC.

7.2 Method

7.2.1 Participants

Participants consisted of all of those from Cohort 2 and those from Cohort 1 whose parents chose to participate in the home measures. Table 7.1 gives demographic information for the two home-based measures. More participants took part in the self-report questionnaire as many mothers from Cohort 1 agreed to complete the questionnaire but not take part in the book sharing activity. For more information on the participants including recruitment see Section 2.3 in Chapter 2. It should be noted that the same participants from Cohort 1 did not always participate at both Time 1 and Time 2. For example, some participants did not complete the questionnaire at Time 1 but did at Time 2.

Table 7.1

Demographic information for the self-report measure and book sharing task

	Measure	N	Males	Females	Mean age	Mothers education[b]
Time 1	Self-report	110	55	55	3;11[a] (SD=4.39)	2.94 (SD = 1.07)
	Book sharing	57	28	29	3;10[a] (SD=3.79)	3.16 (SD = .94)
Time 2	Self-report	91	37	54	4;11[a] (SD=4.30)	2.87 (SD = .90)
	Book sharing	46	22	24	4;9[a] (SD=3.40)	3.08 (SD =.77)

[a] Age in years;months [b] Mother's highest level of education graded as: School leavers certificate = 0, GSCEs = 1, Alevels or GNVQs or BTECs = 2, University degree = 3, Postgraduate degree = 4, Doctorate = 5.

7.2.2 Materials and measures

Materials consisted of maternal measures (self-report and book sharing activity) and language and ToM tasks administered directly to the child by the researcher.

7.2.2.1 Maternal materials

A questionnaire (Appendix 1) collected information about parental perception of child mental state word comprehension and production. This took the form of a vocabulary checklist whereby parents indicated whether their child understood but did not say the word yet, understood and said the word, or did not know the word at all. The checklist included 43 mental state words from a list of mental state words and phrases used to code mental state talk activities by Ruffman and colleagues (Ruffman et al., 2002; Ruffman et al., 2006) which was based on the criteria of Bartsch and Wellman (1995), e.g. "*Hope*", "*Think*" and "*Angry*". To control for demand characteristics so that parents did not guess the aim of the checklist it also included 40 general words typical in children's vocabularies, e.g. "*Big*" "*Empty*" "*Red*". These general words were taken from the descriptive words section of the Oxford Communicative Development Inventory (A. Hamilton et al., 2000), which is a UK adaptation of the MacArthur-Bates Communicative Development Inventory. For further details see Section 2.7.1 in Chapter 2.

Further to this, a book sharing activity was administered in which children and mothers sat together whilst viewing ten photographs of people and families in everyday situations (Appendices 19a and 19b) whilst their conversations were recorded. A different set of ten pictures were used at each of the two time points and all based on those used by Ruffman et al., (2002) which were adapted from the Thorpe Interaction Measure (Thorpe, 1996). Each utterance from the conversations was coded for mental state talk with both mother and child's dialogue coded. An utterance referred to a string of words identified by a grammatical mark of completeness or a pause (Golinkoff & Ames, 1979). The coding system was based on that of Ruffman, Slade and Crowe (2002) and coded utterances into one of five mental states categories (Cognitive terms e.g. "*think"* or *"know*", desire terms e.g. "*hope*" or "*want*", modulation of assertions e.g. "*could be*" or "*maybe*", emotion terms e.g. "*happy*" or "*sad*", and other mental state terms e.g. "*remember*" or "*pretending*") or one of the thirteen non-mental state categories (simple description, physical state terms, causal utterances, elaboration of a theme, links to child's life, factorial utterances, orientation responses, self-repetition, repetition of other, unrelated utterances, prompt question, short responses and nonsensical utterances). Each utterance could be coded more than once, for example if it contained both reference to cognitive terms and desire. In the main analysis, only the total of all types of mental state talk were used as here the interest was in the overall use of mental state talk and the relationship between ToM and LC. However, it was also interesting to look at the frequency of the subtypes of mental state, and the frequency of other types of talk which did not reference mental states. For further details, including the full coding system with examples, refer to Chapter 2 Section 2.7.2.

7.2.2.2 Measures administered to children by researcher

During their one-to-one session with the researcher, children participated in language and ToM measures. Table 7.2 gives the measures used. For comprehensive details on each of these measures refer to Section 2.5 in Chapter 2. In all analyses which included ToM, a composite of the two measures per timepoint was used (unexpected contents and unexpected locations for Time 1 and belief-desire reasoning and unexpected locations second-order for Time 2).

Table 7.2

Measures administered to children at each of the time points

Skill	Measure	Time 1	Time 2
Listening comprehension	NARA	✓	✓
	OWLS	-	✓
Theory of mind	Unexpected contents	✓	-
	Unexpected locations	✓	-
	Belief desire reasoning	-	✓
	Unexpected locations second-order	-	✓
Vocabulary	BPVS	✓	✓
Syntactic Knowledge	CELF-Preschool	✓	✓
Working memory	Reverse word	✓	✓
Non-verbal ability	Block design	✓	-

7.2.3 Procedure

Parents provided opt-in consent for both the questionnaire and book sharing activity. For Cohort 1 the questionnaire was sent home by class teachers and parents returned it to school once completed. For Cohort 2 parents completed the questionnaire in the presence of the researcher whilst the researcher was administering the one-to-one measures to their child.

For the book sharing activity, Cohort 1 parents received activity details after completion of the questionnaire and gave their consent and contact details if they wished to participate. The researcher then emailed the parents the instructions and materials and they completed the activity themselves in their own home. For this reason, not all of those in Cohort 1 who completed the questionnaire participated in the book sharing activity (see Table 7.1). Participation included recording their conversation on a phone or tablet device and sending the recording back to the researcher via email. From Cohort 1 there was only six participants who took part in the book sharing activity at Time 1 and one participant at Time 2. These were used in the cross-sectional analyses but not in the longitudinal analyses as the one participant from Time 2 did not take part at Time 1.

For Cohort 2 parents and children completed the activity in the presence of the researcher whilst the researcher recorded the activity using a portable recording device. For both cohorts the pictures were the same and parents were encouraged to "*look at the pictures and talk about them as they would a storybook at bedtime*". For each cohort the same procedure was used at both time points. For complete procedural details refer to Chapter 2 Section 2.7.

7.2.4 Analysis

First, data from mothers' self-report of their child's mental state use and understanding was analysed; concurrently and then longitudinally. Next, data from the book sharing activities was analysed. The book sharing conversations were transcribed by a transcriber and then coded for mental state talk using the coding system outlined in Tables 2.4 and 2.5 in Chapter 2, Section 2.7.2. First, correlation analysis was conducted then mediation analysis was planned. Mediation analysis was selected rather than SEM (like previous chapters) due to a lower sample size which was not appropriate for SEM analysis. The mediation analysis was performed with PROCESS in SPSS (Hayes, 2012) using 1000 bootstrap samples to compute bias-corrected and accelerated confidence intervals around the indirect effect.

7.3 Results

7.3.1 Self-report results

First the self-report data was analysed. Here the aim was to assess if a child's mental state talk (as reported by their mothers) predicted their LC both directly and indirectly via ToM.

7.3.1.1 Descriptive statistics

Only mother's scores for their child's production of mental state talk (not children's comprehension) were used in the analysis. Scores were computed by summing the mental state words a mother reported her child to know. Descriptive statistics were computed separately for Time 1, Time 2 and longitudinally as the participants differed slightly as these time points (see Section 7.2.1 for more information). These descriptive statistics, including mental state talk scores and ToM and language score, are shown in tables 7.3-7.5. As shown in the tables skewness and kurtosis values showed some measures were not normally distributed so Spearman's correlation was used rather than Pearson's which is more appropriate for non-normal data (Field, 2013). Estimate likelihood was also performed for all missing data; this was justified as Little's MCAR test reported the data to be missing at random ($p = .28$). Correlation analysis was then conducted. Correlations between all variables at each time point are shown in tables 7.6-7.8.

Table 7.3

Descriptive statistics for measures at both Time 1

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 1 child mental state talk	Parental self-report	110	43	7-43	27.84	8.94	-.40	-.42
Time 1 listening comp	NARA	110	44	0-12	1.65	2.44	2.02	3.84
Time 1 theory of mind	ToM composite	101	5	0-5	2.35	1.67	.06	-1.09
Time 1 vocabulary	BPVS	110	168	10-102	56	16.82	-.08	-.24
Time 1 syntactic knowledge	CELF	109	26	0-21	12.26	4.32	-.71	.17
Time 1 working memory	Reverse word	109	9	0-9	1.30	2.40	1.68	1.47
Time 1 non-verbal ability	Block design	110	40	0-30	17.48	2.40	-.91	1.41

Table 7.4

Descriptive statistics for measures at Time 2

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 2 child mental state talk	Parental self-report	91	43	12-40	31.58	6.84	-1.21	1.23
Time 2 listening comp	NARA	90	44	0-13	4.10	3.75	.64	-.62
	OWLS	91	130	24-98	54.07	16.23	.27	-.01
Time 2 theory of mind	ToM composite	91	5	0-5	2.18	1.68	.21	-1.19
Time 2 vocabulary	BPVS	91	168	24-105	79.59	12.96	-.97	1.97
Time 2 syntactic knowledge	CELF	91	26	1-22	16.29	3.82	-1.30	2.71
Time 2 working memory	Reverse word	91	9	0-9	3.91	2.72	.19	-.80
Time 1 non-verbal ability	Block design	91	40	0-30	17.48	2.40	-.91	1.41

Table 7.5

Descriptive statistics for longitudinal analysis

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 1 child mental state talk	Parental self-report	74	43	7-43	28.50	8.82	-.61	-.05
Time 2 listening comp	NARA	73	44	0-13	4.31	3.71	.53	-.67
	OWLS	74	130	24-98	54.76	16.74	.27	.01
Time 2 theory of mind	ToM composite	69	5	0-5	1.97	1.60	.32	-1.08
Time 2 vocabulary	BPVS	74	168	24-105	79.66	13.88	-.99	-.15
Time 2 syntactic knowledge	CELF	73	26	5-22	16.30	3.62	-.92	.98
Time 2 working memory	Reverse word	73	9	0-9	3.86	2.53	.11	-.64
Time 1 non-verbal ability	Block design	74	40	0-30	17.65	5.36	-1.01	2.17

Table 7.6

Correlations for measures at Time 1

	1.	2.	3.	4.	5.	6.	7.	8.
1.Time 1 MST	-							
2.Time 1 NARA	.18	-						
3.Time 1 ToM	.40**	.23*	-					
4.Time 1 BPVS	.36**	.41**	.35**	-				
5. Time 1 CELF	.35**	.30**	.50**	.62**	-			
6.Time 1 Reverse word	.16	.18	.36**	.32**	.46**	-		
7.Time 1 Block design	.07	.24*	.26**	.46**	.36**	31**	-	
8.Time 1 Age	.02	-.08	.26**	.05	.23*	.28*	.12	-

Note. ** = p < .01, * = p < .05. MST = mental state talk

Table 7.7

Correlations for measures at Time 2

	1.	2.	3.	4.	5.	6.	7.	8	9.
1.Time 2 MST	-								
2.Time 2 NARA	.12	-							
3. Time 2 OWLS	.10	.58**	-						
4.Time 2 ToM	.21*	.24*	.33**	-					
5.Time 2 BPVS	.27**	.42**	.45**	.39**	-				
6. Time 2 CELF	.09	.35**	.54**	.33**	.43**	-			
7.Time 2 Reverse word	.26*	.27**	.40**	.34*	.38**	.46**	-		
8.Time 1 Block design	.07	.26*	.21	.26	.36**	.28*	.25*	-	
9.Time 2 Age	-.03	-.07	.06	.12	.18	.17	.18	.14	-

Note. ** = p < .01, * = p < .05. MST = mental state talk

Table 7.8

Longitudinal correlations

	1.	2.	3.	4.	5.	6.	7.	8	9.
1.Time 1 MST	-								
2.Time 2 NARA	.14	-							
3. Time 2 OWLS	.07	.56**	-						
4.Time 2 ToM	.25*	.25*	.31**	-					
5.Time 2 BPVS	.46**	.44*	.43**	.33**	-				
6. Time 2 CELF	.24*	.35**	.57**	.31**	.48**	-			
7.Time 2 Reverse word	.11	.28*	.45**	.27*	.39**	.43**	-		
8.Time 1 Block design	.18	-.04	.26*	.24*	.23*	.33**	.41**	-	
9.Time 2 Age	.07	-.04	.14	-.02	-.02	.14	.16	-.03	-

Note. ** = p < .01, * = p < .05. MST = mental state talk

7.3.1.2 Predictive relationships at Time 1

Although mental state talk did not correlate significantly with LC, mental state talk correlated with ToM and ToM also correlated with LC. Given these relationships a mediation analysis could be carried out using Hayes' PROCESS (Hayes, 2012) as, unlike the traditional causal step Baron and Kenny approach (Baron & Kenny, 1986), this modern approach does not require the direct path to be significant and allows just the indirect effect to be tested. Therefore, mediation analysis was conducted to investigate whether there was an indirect effect of child mental state talk on LC via ToM. PROCESS computes bias-correct and accelerated (BCa) confidence interval (CI) around the indirect effect and a significant indirect effect is indicated if a zero does not fall between the confidence interval (Preacher & Hayes, 2008). The analysis showed neither a significant direct effect of MST on LC (b =.17, p = .74) nor a significant indirect effect via ToM (b =.02, BCa CI (-.003, .06). This is shown in Figure 7.3.

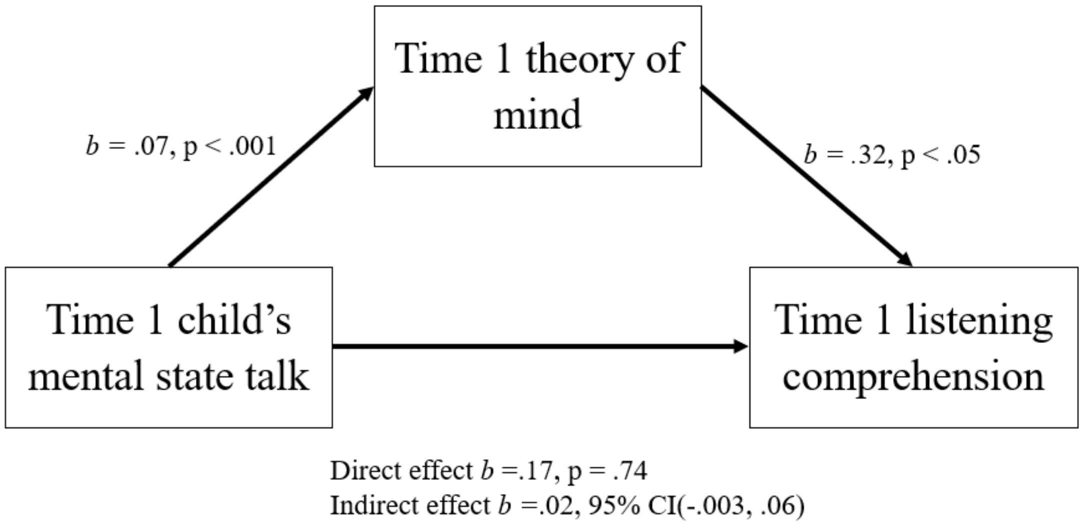


Figure 7.3: Mediation analysis showing neither a direct effect of MST on LC nor an indirect effect via ToM. Unstandardized estimates are presented with significance based on absence of zero in bootstrapped confidence intervals; the confidence interval for each indirect effect is a bias-corrected and accelerated bootstrapped confidence interval based on 1000 samples.

7.3.1.3 Predictive relationships at Time 2

Correlations were the same at Time 2. Mental state talk still did not correlate significantly with LC (neither the OWLS nor NARA measure) and mental state talk again correlated with ToM. At this time point ToM also correlated with LC for both the NARA and the OWLS measure. Given these relationships a mediation analysis could be carried out using Hayes' PROCESS (Hayes, 2012). The LC variable was computed by creating a composite of the two LC measures (NARA and OWLs). Again, the analysis showed neither a significant direct effect of MST on LC (*b* =.19, p = .51) nor a significant indirect effect via ToM (*b* =.17, BCa CI (-.01, .37). This is shown in Figure 7.4.

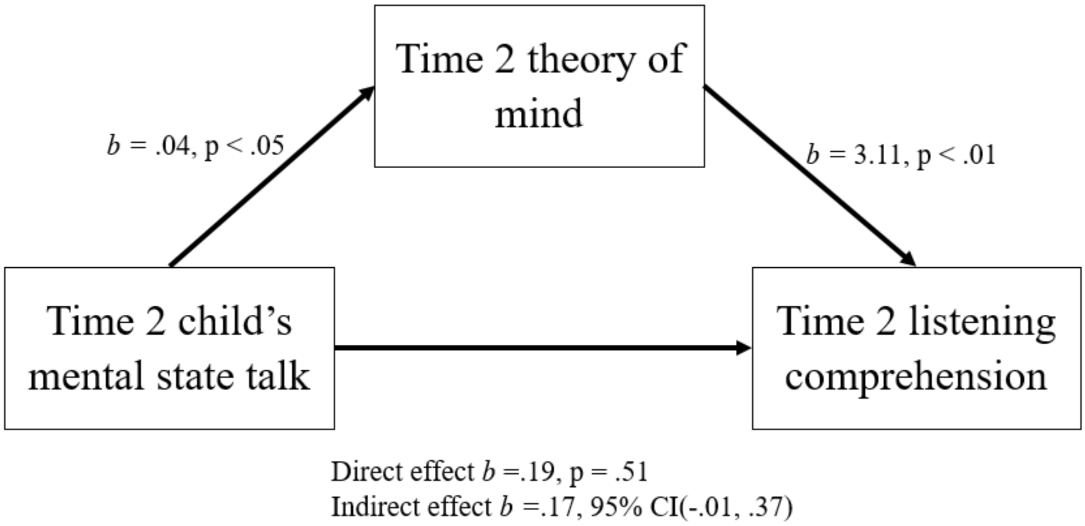


Figure 7.4: Mediation analysis showing neither a direct effect of MST on LC nor an indirect effect via ToM. Unstandardized estimates are presented with significance based on absence of zero in bootstrapped confidence intervals; the confidence interval for each indirect effect is a bias-corrected and accelerated bootstrapped confidence interval based on 1000 samples.

7.3.1.4 Longitudinal predictions

Longitudinally, the relationships were the same as mental state talk at Time 1 correlated with ToM at Time 2, and Time 2 ToM correlated with both measures of concurrent LC but mental state talk at Time 1 did not correlate with LC. Again, mediation analysis could be performed so analysis was conducted to investigate whether there was an indirect effect of Time 1 child mental state talk on Time 2 LC via Time 2 ToM. Again, the LC variable was computed by creating a composite of the two LC measures (NARA and OWLs). The analysis showed a significant indirect effect via ToM (*b* =.17, BCa CI (.01, .42) but no direct effect of MST on LC (*b* =.23, p = .38). This is shown in Figure 7.5.

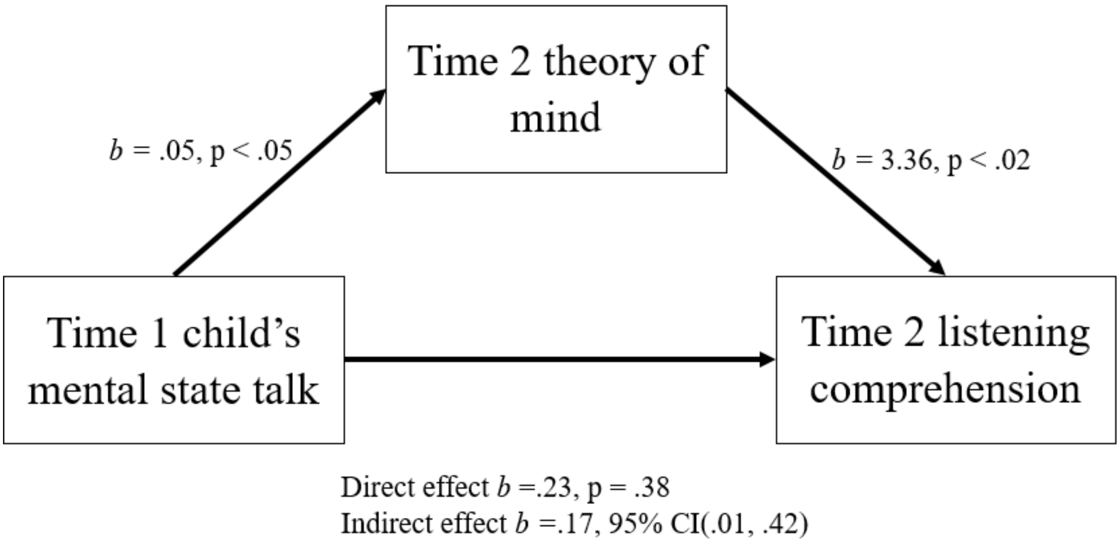


Figure 7.5: Mediation analysis showing an indirect effect of Time 1 MST on LC via ToM but no direct effect of MST on LC. Unstandardized estimates are presented with significance based on absence of zero in bootstrapped confidence intervals; the confidence interval for each indirect effect is a bias-corrected and accelerated bootstrapped confidence interval based on 1000 samples.

This significant indirect effect however did not hold when Time 2 vocabulary, syntactic knowledge, working memory and Time 1 non-verbal ability were controlled for (b = .03, CI (-.002, .009)).

7.3.2 Book sharing activity results

Next mental state talk (both child and maternal) from the book sharing activity was analysed. Recordings were transcribed and then coded.

7.3.2.1 Descriptive statistics

The mean and standard deviation for the length of recording, total number of utterances per recording and mean length (number of words) of utterances (MLU) are shown in Table 7.9.

Table 7.9

Descriptive statistics for the book sharing recordings.

	Time 1	Time 2
Length of recording	467.02 (22.92)	410.58 (198.81)
Total utterances by mothers	136.70 (74.35)	167.54 (84.50)
Total utterances by children	72.82 (38.84)	80.13 (34.25)
Mean length of mothers' utterances	5.44 (1.07)	3.92 (.49)
Mean length of children's utterances	2.90 (.84)	2.77 (.55)

Note. All times in seconds.

Following Ruffman et al., (2002) the mean number of each type of utterance at Time 1 and Time 2 for both children and mothers is shown in Figure 7.6. It is interesting to note that at both time points both mothers' and children's most frequently used mental state talk was cognitive terms. Regarding non-mental state talk at Time 1 for mothers the most frequent category of utterance was prompting question but at Time 2 it was description. For children the most frequent was simple description at both time points.

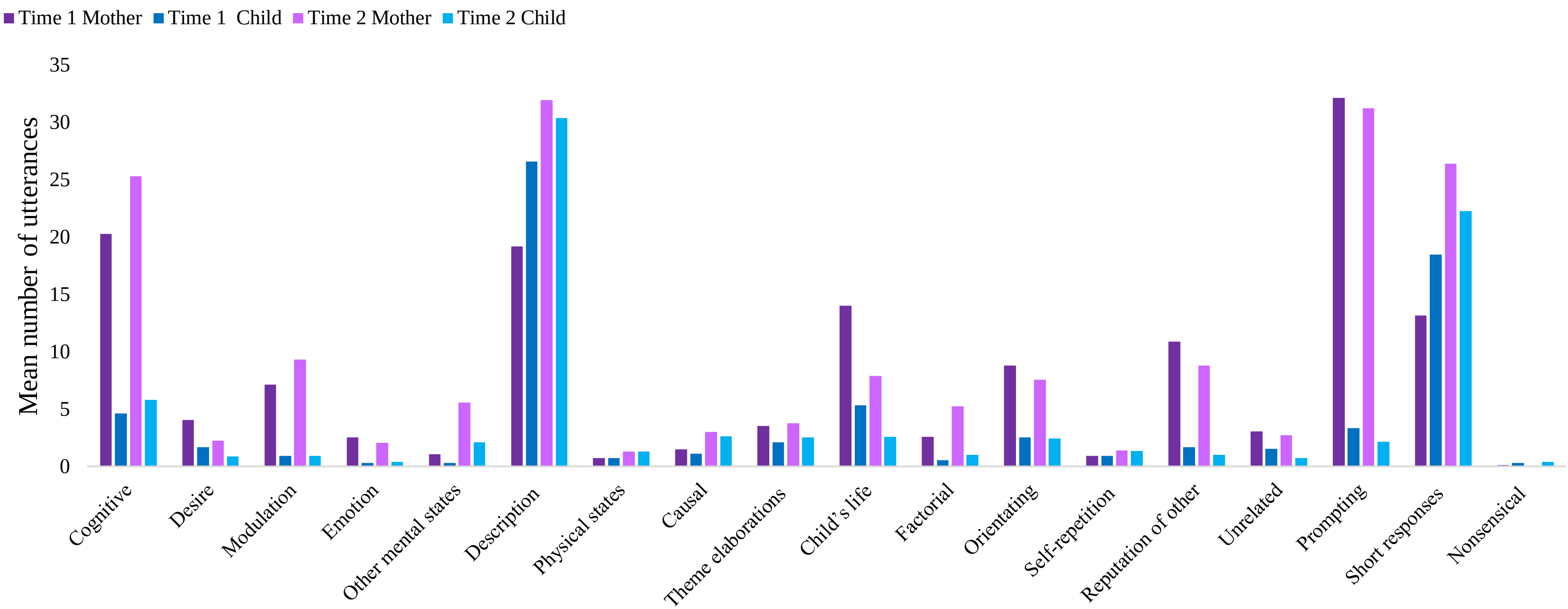


Figure 7.6: Mean number of each type of utterance at Time 1 and Time 2 for both mother and child. Cognitive terms, desire terms, modulations of assertions, emotion terms and other mental state terms were all coded as mental state utterance

In all further analysis a sum of all mental state talk (including cognitive terms, emotion terms, desire terms, modulation of assertion and other mental state terms) was used for both children and mothers, as well as individual frequencies for cognitive, emotion and desire terms for mothers only. This is in line with other similar research (e.g. A. Carr et al., 2018). Descriptive statistics were computed separately for Time 1, Time 2 and longitudinally as the participants differed slightly as these time points (see Section 7.2.1 for more information). These descriptive statistics, including mental state talk scores and ToM and language score, are shown in tables 7.10-7.12. As shown in the tables, there was no missing data, but skewness and kurtosis values show that some measures were not normally distributed (most of the mental state variables and NARA and reverse word at Time 1 were positively skewed). These skewness and kurtosis values as well as the large ranges and standard deviations for the mental state talk variables suggest that there may be some outliers.

Outliers were identified by converting both mother and child total mental state talk score into z-scores and removing data for all participants whose z-score were less than -2.68 or greater than 2.68 as this cut off demonstrates that this data is more than 1.5 times the inter-quartile range (Ghasemi & Zahediasl, 2012; Walfish, 2006). Boxplots were used to confirm these outliers. At Time 1 two participants were removed, at Time 2 three were removed and longitudinally one participant was removed. All these participants were from Cohort 2. When this was done the skewness and kurtosis scores were much improved. However, normality was still not quite met and so Spearman's correlation was used rather than Pearson's which is more appropriate for non-normal data (Field, 2013). Correlations for each separate time point are shown in tables 7.13-7.15.

Table 7.10

Descriptive statistics for measures at Time 1

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 1 mother MST	Coded Bk-Share	57	-	5-130	35.09	23.89	1.61	3.58
Time 1 mother cognitive	Coded Bk-Share	57	-	1-63	20.28	14.83	1.21	1.12
Time 1 mother desire	Coded Bk-Share	57	-	0-23	4.07	3.93	2.45	9.11
Time 1 mother emotion	Coded Bk-Share	57	-	0-7	1.05	1.52	1.85	3.65
Time 1 child MST	Coded Bk-Share	57	-	0-31	7.56	1.55	1.98	2.78
Time 1 listening comp	NARA	57	40	0-12	1.98	2.45	2.00	4.38
Time 1 theory of mind	ToM composite	57	5	0-5	1.76	1.75	.06	-1.25
Time 1 vocabulary	BPVS	57	168	26-84	59.39	15.33	-.53	-.67
Time 1 syntactic knowledge	CELF	57	26	4-19	13.14	3.24	-.57	.35
Time 1 working memory	Reverse word	57	9	0-9	1.36	2.49	1.69	1.61
Time 1 non-verbal ability	Block design	57	40	10-26	18.54	3.69	-.29	-.06

Note. MST = mental state talk total

Table 7.11

Descriptive statistics for measures at Time 2

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 2 mother MST	Coded Bk-Share	46	-	5-162	45.52	29.94	1.94	5.06
Time 2 mother cognitive	Coded Bk-Share	46	-	1-91	25.85	17.78	1.92	5.13
Time 2 mother desire	Coded Bk-Share	46	-	0-12	2.26	2.89	1.63	2.50
Time 2 mother emotion	Coded Bk-Share	46	-	0-21	5.70	4.80	1.43	2.36
Time 2 child MST	Coded Bk-Share	46	-	0-78	25.73	16.13	1.26	1.70
Time 2 listening comp	NARA	46	40	0-13	4.04	3.58	.62	-.35
	OWLS	46	130	26-98	53.91	16.60	.74	.72
Time 2 theory of mind	ToM composite	46	5	0-5	2.07	1.61	.29	-1.02
Time 2 Vocabulary	BPVS	46	168	54-101	79.54	9.89	-.01	.64
Time 2 Syntactic knowledge	CELF	46	26	7-22	16.39	3.14	-.69	1.12
Time 2 Working memory	Reverse word	46	9	0-8	4.24	2.17	-.04	-.36
Time 1 non-verbal ability	Block design	46	40	12-26	19.28	3.29	-.15	.04

Note. MST = mental state talk total

Table 7.12

Descriptive statistics for longitudinal analysis

Skill	Measure	N	Max	Range	Mean	SD	Skewness	Kurtosis
Time 1 mother MST	Coded Bk-Share	43	-	5-130	34.06	22.71	1.95	6.90
Time 1 mother cognitive	Coded Bk-Share	43	-	1-57	19.37	12.50	.97	1.14
Time 1 mother desire	Coded Bk-Share	43	-	0-23	3.98	4.11	2.79	10.64
Time 1 mother emotion	Coded Bk-Share	43	-	0-7	1.09	1.52	2.03	4.87
Time 1 child MST	Coded Bk-Share	43	-	0-31	10.40	9.29	1.98	5.03
Time 2 listening comp	NARA	43	40	0-12	4.30	3.56	.54	-.38
	OWLS	43	130	28-98	55.23	16.19	.80	.81
Time 1 theory of mind	ToM composite	43	5	0-5	2.31	1.75	.07	-1.28
Time 1 vocabulary	BPVS	43	168	28-84	62.56	13.44	-.80	.03
Time 1 syntactic knowledge	CELF	43	26	4-19	13.58	3.19	-.65	.90
Time 1 working memory	Reverse word	43	9	0-9	1.48	2.56	1.60	1.47
Time 1 non-verbal ability	Block design	43	40	12-26	19.28	3.29	-.15	.04

Note. MST = mental state talk total

Table 7.13

Correlations for measures at Time 1 with outliers removed

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**	**9.**	**10.**	**11.**	**12.**
1.T1 Mother MST	-											
2.T1 Mother Cognitive	.96**	-										
3.T1 Mother Emotion	.54**	.56**	-									
4.T1 Mother Desire	.19	.06	-.04	-								
5.T1 Child MST	.50**	.45**	.10	.34*	-							
6.T1 NARA	.20	.18	.18	-.20	.14	-						
7.T1 ToM	.21	.19	.15	.01	.25	.16	-					
8.T1 BPVS	.15	.11	.02	.09	.15	.30*	.19	-				
9.T1 CELF	.14	.15	.09	-.13	.22	.20	.38**	.42**	-			
10.T1 Reverse word	.24	.26	.08	.07	.20	.27*	.49**	.30*	.49**	-		
11.T1 Block design	.09	.12	.01	.07	.25	.08	.28*	.31*	.18	.29*	-	
12.T1 Age	.30*	.30*	.10	-.12	.13	.09	.34*	.11	.34*	.26	.11	-

Note. ** = p < .01, * = p < .05. T1 = Time 1

Table 7.14

Correlations for measures at Time 2 with outliers removed

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1.T2 Mother MST	-												
2.T2 Mother Cognitive	.90**	-											
3.T2 Mother Emotion	.48**	.35*	-										
4.T2 Mother Desire	.34*	.20	.30	-									
5.T2 Child MST	.84**	.62**	.54**	.50**	-								
6.T2 NARA	.12	.14	.14	.13	.09	-							
7. T2 OWLS	.06	.09	.17	.06	.13	.53**	-						
8.T2 ToM	.12	.11	.09	.03	.04	.28	.27	-					
9.T2 BPVS	.30	.28	.15	.08	.28	.24*	.33*	.36*	-				
10.T2 CELF	.07	-.02	.24	.27	.28	.29	.45**	.22	.30	-			
11.T2 Reverse word	.20	.27	.05	-.03	.08	.35*	.38*	.24	.38*	.30	-		
12.T1 Block design	.12	.05	-.10	-.05	.20	-.06	.13	.16	.41**	.23	.22	-	
13.T2 Age	.32*	.31*	.24	-.12	.20	-.13	.13	.02	.17	-.05	.29	.23	-

Note. ** = p < .01, * = p < .05. T1 = Time 1, T2 = Time 2

Table 7.15

Correlations for longitudinal analysis with outliers removed

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1.T1 Mother MST	-												
2.T1 Mother Cognitive	.95**	-											
3.T1 Mother Emotion	.64**	.67**	-										
4.T1 Mother Desire	.21	.09	-.01	-									
5.T1 Child MST	.51**	.39*	.13	.38*	-								
6.T2 NARA	.19	.19	.13	.01	.10	-							
7.T2 OWLS	.16	16	-.02	.01	.32*	.51**	-						
8.T1 ToM	.14	.14	.13	-.09	.25	.17	.12	-					
9.T1 BPVS	.12	.03	.01	.11	.22	.33*	.12	.17	-				
10.T1 CELF	.01	-.01	.03	-.19	.15	.33*	.26	.46**	.39*	-			
11.T1 Reverse word	.08	.11	-.06	-.02	.11	.20	.24	.60**	.18	.49**	-		
12.T1 Block design	.15	.18	.06	.09	.32*	-.07	.12	.32*	.25	.26	.39*	-	
13.T1 Age	-.12	-.06	-.15	-.36*	-.10	-.19	.06	.19	.17	.33	.20	.15	-

Note. ** = p < .01, * = p < .05. T1 = Time 1, T2 = Time 2

7.3.2.2 Predictive analysis

Correlations shown in tables 7.13-7.15 demonstrate that for either time points nor longitudinally, for both child and mother, mental state talk did not correlate with ToM. Therefore, the planned mediation analysis was abandoned because no relationships was shown between ToM and mental state talk. Correlations also showed there to be no relationship between LC and maternal mental state talk. There was however a moderate correlation between child mental state talk and the OWLS LC measure longitudinally, i.e. between Time 1 child mental state talk and Time 2 OWLS score. As this relationship was found with only one of the LC measures no further analyses were carried out.

7.3.3 Further Analysis

The procedure for the book sharing activity for the participants from Cohort 1 (six participants for Time 1, and one participant for Time 2) differed to the procedure for those from Cohort 2, in that these mothers recorded the activity themselves without the researcher present. Although these Cohort 1 data sets were not shown to be outliers (see Section 7.3.2.1) the different procedure used between cohorts may have affected the amount of mental state talk used by the mothers and children and thus affected the pattern of the overall relationship between mental state talk LC and ToM. Given this, the data for these Cohort 1 participants was removed and correlations between mental state talk and ToM and LC were run again to check that the patterns remained the same. As shown in Table 7.16 these new correlations showed a very similar relationship and no changes in significance was found.

Table 7.16

Correlation comparisons between the full data set and with Cohort 1 participants removed

Relationship	Including Cohort 1 participants	Cohort1 participant removed
Time 1 mother total MST and NARA LC	.20	.24
Time 1 mother total MST and ToM	.21	.20
Time 1 mother cognitive terms and NARA LC	.18	.20
Time 1 mother cognitive terms and ToM	.19	.21
Time 1 mother emotion terms and NARA LC	.18	.24
Time 1 mother emotion terms and ToM	.15	.15
Time 1 mother desire terms and NARA LC	-.20	-.21
Time 1 mother desire terms and ToM	.01	-.08
Time 1 child total MST and NARA LC	.14	.16
Time 1 child total MST and ToM	.25	.19
Time 2 mother total MST and NARA LC	.12	.12
Time 2 mother total MST and OWLS LC	.06	.06
Time 2 mother total MST and ToM	.12	.12
Time 2 mother cognitive terms and NARA LC	.14	.15
Time 2 mother cognitive terms and OWLS LC	.09	.09
Time 2 mother cognitive terms and ToM	.11	.11
Time 2 mother emotion terms and NARA LC	.14	.14
Time 2 mother emotion terms and OWLS LC	.17	.17
Time 2 mother emotion terms and ToM	.09	.09
Time 2 mother desire terms and NARA LC	.13	.13
Time 2 mother emotion terms and OWLS LC	.06	.06
Time 2 mother desire terms and ToM	.03	.03
Time 2 child total MST and NARA LC	.09	.09
Time 2 child total MST and OWLS LC	.13	.13
Time 2 child total MST and ToM	.04	.04

7.4 Discussion

The aim of this chapter was to gain a deeper understanding into the relation between LC and ToM by exploring the relationship the two have with mental state talk (both maternal and child's own). It was predicted that ToM would mediate the relationship between LC and mental state talk. It was hypothesized that this would be the case for both child mental state talk and maternal mental state talk when mental state talk was measured both by self-report and in recorded dialogue between mothers and child, and that this would be true both concurrently (when children were three and four years old) and also longitudinally for mental state talk aged three predicting LC a year later aged four. Theoretically, mental state talk may predict LC because talking about mental states frequently could help a child to form a solid understanding of the social information in a story thus improving their comprehension of the story, and ToM could act as a mediator between the two because past research has already shown that ToM facilities LC (e.g. Kim, 2017) and that mental state talk facilities ToM (e.g. A. Carr et al., 2018).

With the exception of the self-report data, findings did not support the hypothesizes as many of the predicted initial correlations were not found. Therefore, the planned mediation analyses were abandoned. However, results still highlight several noteworthy aspects and suggest a direction for future research.

7.4.1 Self –report of child mental state talk

Self-report data showed that for Time 1, Time 2 or longitudinally child mental state talk showed no relation with LC. However, at both time points and longitudinally relationships were found both between mental state talk and ToM, and ToM and LC. Therefore, mediation analysis was carried out to assess an indirect relationship of mental state talk on LC via ToM. This mediation analysis showed that concurrently at Time 1 and Time 2 there was no indirect relation, but longitudinally there was a significant indirect effect of Time 1 mental state talk on Time 2 LC via Time 2 ToM. Reasons for this indirect effect could be that frequent use of mental state talk focuses a child on mental states which improved their ToM which in turn aids children with the understanding of social information in an oral story thus improving their comprehension of the story. It is difficult to explain why there was no effect concurrently but that a longitudinal indirect effect was found, however inspections of the confidence intervals show that for the longitudinal model the indirect effect was only just significant as the confidence intervals only just avoided passing zero, and for the concurrent models they only just passed zero. Moreover, the fact that this indirect longitudinal relationship did not hold when language skills were controlled for, perhaps suggests that in this self-report measure mothers were actually reporting on their child's general language ability rather than their mental state talk. Further research is therefore needed to confirm these findings. Despite these differing findings across timeframes, this is the first research to suggest a preliminary indirect effect of mental state talk on LC via ToM.

These results support past research that child mental state talk at the age of three (Hughes & Dunn, 1998) age of four (Barreto, Osório, Baptista, Fearon, & Martins, 2018) and longitudinally from age three years to four years (Barreto et al., 2018) are related to ToM. This is also the first research to show this relationship using data from mothers self- reporting of their child's mental state talk as past work has only measured maternal mental state talk using self-reports (Ebert et al., 2017; Farrant et al., 2012; C. Peterson & Slaughter, 2003) and has only used coded conversations between mothers and children to measure children's mental state talk (J. Dunn et al., 1991; Symons et al., 2005) . It is encouraging that these self-report findings match those of recorded coded conversations and suggests that future work should use this method.

7.4.2 Book sharing activity

The book sharing activity showed no relationship between frequency of child mental state talk and ToM at any time point or longitudinally. This is in contrast to past work which has shown a relationship between child mental state talk and ToM in this age group (Barreto et al., 2018; A. Carr et al., 2018; Symons et al., 2005). This is also in contrast to the findings from the self-report data where mothers' reports of their child's mental state usage did relate to ToM. Differences may be due to the lower number of participants in the book sharing activity compared with the self-report data. Regarding maternal mental state talk, again findings did not support past work that there is a relationship between maternal mental state talk and ToM (Adrián et al., 2007; Ruffman et al., 2002). However despite not being significant, correlation confidence were approaching significance this therefore may reflect the finding by the recent metanalysis that more recent research into ToM and mental state talk shows a smaller effect size than older research (Devine & Hughes, 2016a).

There was also no relationship found between maternal mental state talk and LC either concurrently at either timepoint, or longitudinally. Although child mental state talk was shown to have a moderate relationship with the OWLS LC measure longitudinally only. Unlike the ToM finding, this is not particularly surprising given there is no suggestion in the literature, or theoretical explanation, that mental state talk has a direct relationship. Moreover, for child mental state talk and LC this is inconsistent with the self-report findings.

Despite these somewhat surprising findings, this is the first research to explore the relationship between the three variables of mental state talk, LC and ToM. The research adds to the existing research into the effects of extratextual talk during shared book reading (Collins, 2016; Demir et al., 2011; Zucker et al., 2013). Here no relationship was found between maternal mental talk to LC, but existing work has found that adult descriptive and inferential talk during book sharing has a relationship to LC (Collins, 2016; Demir et al., 2011; Zucker et al., 2013). Therefore, it could be that other "quality" talk during shared reading is of importance for LC rather than talk which references mental states. It was hypothesized that maternal mental state talk may aid comprehension because it focuses children's attention to the social elements of stories during book sharing. Instead though, it might be suggested that during shared reading it is not necessarily that talk is focused on mental states, rather that the talk is of good quality (i.e. indexing linguistically rich talk and interaction) and that any extensive reflection and chat during shared reading helps children develop as proficient listening comprehenders. Indeed, findings from Collins (2016) showed children assigned to a group in which they engaged in shared reading where extratextual talk was encouraged had significantly higher LC scores than children assigned to a non-talking group, and home literacy research shows that quality home literacy practices lead to improved literacy outcomes (e.g. Bingham, 2007). In order to explore why mental state talk may not assist with LC but other types of talk do, future work should focus on other types of talk during shared reading (for example exploration of themes), and compare their influence on LC to talk with references mental states.

7.4.3 Limitations

The result that maternal mental state talk did not significantly relate to ToM either at the age of three, the age of four or longitudinally was surprising. It is worth noting though that the recent meta-analysis found that effects size for maternal mental state talk and ToM are smaller for more recent studies than for older research (Devine & Hughes, 2016a). Moreover, despite not being significant, effect sizes were encouraging as they ranged from $r = .12$ to $r = .25$ for total maternal mental state talk across time frames compared to $r = .21$ reported by the meta-analysis.

One explanation for the contradictory findings could be the sample size. Here, there were 57 child and mother pairs at Time 1 and 46 at Time 2, which does not compare favourably to other studies. For example, Ensor et al., (2013) had 105, Ruffman et al., (2002) 82 and Carr et al., (2018) had 73 even at their last time point. The much smaller sample size could have led to the non-significant results. Indeed, some correlations were approaching significance for example for Time 2 ToM and concurrent mothers' reference to cognition and as stated above effect sizes were encouraging, so with a larger sample size significant relation may have been found. This said, the original work by Dunn et al., only had 50 participants and Adrián et al., (2007) included 41 dyads at their first time point and 37 at Time 2, and findings still showed that mothers' use of cognitive terms related to child ToM concurrently and a year later. Yet, the self-report data presented here had a sample almost twice as large as the book sharing sample (N = 110 for Time 1 and N = 91) and showed a significant relationship between ToM and child mental state word usages which suggest the book sharing activity sample could have been too small. Future research should explore this by replicating the study with a larger sample.

7.4.4 Future direction and conclusions

Although, in most cases, the planned mediation analysis was abandoned, this chapter gives some useful insights into the relationship between mental state talk and LC. Firstly, both data from both mothers' self-reports of their child's mental state talk and a shared book reading activity suggest that children's mental state usage is not important for LC directly at the ages of three to four years. However, the research provides some evidence for an indirect effect of child mental state talk on LC via ToM. Secondly, data from the book sharing activity suggests that maternal mental state talk does not predict LC (concurrently or longitudinally) directly or via ToM. Further research is needed to corroborate these early exploratory findings.

This future work should replicate these findings using a larger sample. A smaller sample than past research may explain why maternal mental state talk did not correlate with ToM. Additionally, future research should compare mental state talk to other forms of quality child-mother talk, for example exploration of themes and causal discussion during book sharing. This would explore further whether it is vital for children to understand and engage with social elements within story plots to comprehend proficiently. Lastly, this chapter was only concerned with the relationships between mental state, ToM and LC, and it would be of use to now look at mental state talk and ToM in relation to reading comprehension (RC). This research is important as RC is one of the fundamental aims of primary school education (Lervåg et al., 2018) and therefore any research which explores facilitators of RC is valuable.

8 General discussion and conclusion

8.1 Key aims and purpose

The primary aim of this longitudinal thesis was to gain a deeper understanding of the role played by theory of mind (ToM) in both listening comprehension (LC) and reading comprehension (RC) in children aged three to six years from preschool into Year 1. Recent research suggests that ToM facilitates both LC (Kim, 2016; Kim, 2017) and RC (Atkinson et al., 2017; Kim, 2015; Kim, 2017; Ricketts et al., 2013) but, until now, no study has examined this longitudinally using a large sample to test direct and indirect models. Testing whether ToM impacts on LC and RC because of its socially specific ability (Dore et al., 2018) as opposed to its general metacognitive nature was also a key novel aim addressed in this thesis.

The majority of research into ToM, LC and RC has been conducted by Kim and colleagues (e.g. Kim & Phillips, 2014; Kim, 2015; Kim & Pilcher, 2016; Kim, 2016; Kim, 2017) who propose hierarchical models of LC (the DIET model) and RC (DIET and DIER models) with a specific crucial role for ToM. This thesis extended these models longitudinally and explored other types of metacognition within these models to assess if the social contribution of ToM is vital for LC and RC. Lastly, this thesis investigated the role of mental state talk within the home as a facilitator of both ToM and LC to further address the question over why and how ToM helps LC.

Chapter 3 aimed to validate the concurrent DIET model (Kim, 2016; Kim, 2017) for predicting LC longitudinally. Chapters 4 and 5 explored the specific element of ToM which facilities LC. Again, using the DIET model of LC, these chapters examined whether the social specificity of ToM assists LC, or instead whether the broad metacognitive nature of ToM is of more importance. To do this the contribution of ToM to other non-social metacognitive skills was compared, firstly concurrently (Chapter 4) and then longitudinally (Chapter 5). Chapter 6 extended these findings to RC by validating the DIET model (in which cognitive and language skills direct predict RC) and DIER model (in which cognitive and language skills indirectly predict RC via linguistic comprehension) for predicting RC concurrently and longitudinally, and then by comparing the contribution of ToM to other non-social metacognitive skills for predicting RC. Lastly, Chapter 7 aimed to assess the family contribution of ToM in LC by exploring the relationship that mothers' talk about mental states has with ToM and LC. This current chapter summarises the findings of the thesis and considers them in relation to theories and previous findings. It discusses the implications and limitations and suggest directions for future work before concluding.

8.2 Key findings

Findings showed that a concurrent DIET model of LC fitted well for a sample of UK six year olds, with ToM making a direct contribution to LC within this model. This supports work by Kim (2015; 2016; 2017) with Korean and American English speaking children aged seven to nine years. Findings also suggested that this model fitted well longitudinally for skills at the age of four predicting LC at the age of five with ToM again making a direct contribution. However, although the model also fitted well for predicting LC further across time (skills aged four predicting LC age six) here there was no significant evidence that ToM makes a direct contribution to LC. However, it is worth noting that the contribution of ToM was approaching significance in this two year longitudinal model and this marginal in-significant path may reflect a smaller sample size than that used in the studies of Kim (e.g. 107 compared to 350 by Kim, 2017).

When comparing ToM to other forms of metacognition to address why and how ToM helps LC, findings suggested overall ToM was a better predictor of LC than other forms of metacognition not social in nature e.g. source monitoring (Bright-Paul et al., 2008), metamemory (Ebert, 2015) and metalinguistic awareness (Doherty & Perner, 1998). This was because concurrently at the age of six and longitudinally across 12 months for skills at four years predicting LC aged five, the fit of a DIET model including ToM rather than a broad metacognition latent variable, was better. This supports the view taken by the literature that ToM helps a listener to understand the social information within a story plot (Dore et al., 2018) such as filling-in missing information regarding social information such as intentions, thoughts and emotions (Kim & Pilcher, 2016).

When considering the home environment and references to mental states made between mothers and children during book sharing as a predictor of LC, findings from a coded book sharing activity showed that maternal mental state talk did not predict LC directly or indirectly (via ToM) either at four year old, five years old or longitudinally. These results were mirrored for children's own mental state talk when measured through the book sharing activity, however, when measured through mothers' self-report of their child's mental state talk usage, longitudinally only, an indirect effect of child mental state talk on LC via ToM was found. This, however, did not hold when general language was controlled for, perhaps suggesting that in this self-report measure mothers were actually reporting on their child's general language ability rather than their mental state talk.

Regarding RC, both concurrently (aged six) and longitudinally (aged four skills predicting RC aged six) findings did not support the past work that ToM directly predicts RC (Atkinson et al., 2017) as a DIET model of RC did not show ToM to make a significant direct contribution to RC. Instead concurrent findings aged six back-up a DIER model of RC in which ToM make an indirect significant contribution to RC. Theoretically these findings are in line with the SVR and therefore support claims made by the SVR (Gough & Tunmer, 1986; Hoover & Tunmer, 2018) that RC is a product of only decoding and linguistic comprehension and all other skills are subskills of these two components. Longitudinally, the DIER model of RC also had a good fit for skills aged four predicting RC aged six although here ToM did not make an in-direct contribution to RC via LC. Yet, as with the LC models this indirect effect was approaching significance and could perhaps be explained by the small sample size. Further to this, the model including ToM was a better fit when compared to one including broad metacognition which, as with LC, suggests that the social specificity of ToM is what is important for aiding comprehension.

Overall, the findings of this thesis are not clear-cut and there are inconsistencies across time points. This may be due to limitations of the research discussed in Section 8.3 particularly sample size. However, the results do provide some evidence for a role of ToM for LC and RC. They suggest that ToM *directly* predicts LC both longitudinally and concurrently, and *indirectly* (via linguistic comprehension) predicts RC, at least, concurrently. Findings also suggest that this is probably because the social specificity of ToM, rather than its general metacognitive nature, which helps children to understand social information within a story plot. Mothers' self -report data also showed that there may be an indirect relationship between a child's earlier mental state talk and later LC via ToM, but this finding did not hold when language was controlled for and was not supported by data from a live coded book sharing activity. Additionally, the finding was not shadowed for maternal mental state talk.

8.3 Theoretical implications

Theoretically, the findings of this thesis support models of reading including the SVR, the situation model, the DIET and DIER models. One of the main aims of the thesis was to validate the DIET and DIER models (Kim, 2017) in a UK population with a slightly younger aged sample and longitudinally. The DIET model of LC and the DIER model of RC were supported concurrently at the age of six in that these models had a good fit. However, not all paths were replicated in these models. For example, in the DIET model of LC the significant path from syntax to ToM was not replicated. Such differences may reflect the differing measures used here to Kim (2017) and the slight age difference of the children. Longitudinally, the models were also replicated. The DIET model of LC was validated both for skills aged four predicting LC aged five, and for skills aged four predicting LC aged six, and the DIER model of RC was replicated for skills aged four predicting RC aged six, as all models fitted well. As with the concurrent models some paths were not directly replicated, including in some models a significant path from ToM, but this is still the first research conducted outside of Kim's research group and outside the US to confirm both the DIET and the DIER models.

Validation of the DIET and DIER models is important theoretically because these models support both the situation model and the SVR model. Although the SVR is influential and had much evidence to support it, a well cited criticism of the model is that it is too simple (Ouellette & Beers, 2010). There are two ways of seeking to expand the SVR and make it more complex. The first is taken by Kim and others (Massonnié et al., 2018) who attempt to unpack the component skill of linguistic comprehension. The alternative approach is to look for an additional factor to explain variance in RC not covered by the SVR (e.g. Atkinson et al., 2017; Conners, 2009). This thesis gives evidence to support the former approach because the DIER models of RC was able to unpack the component of linguistic comprehension supporting that it has many sub-skills, whereas the DIET model did not show direct paths of the subskills to RC. Therefore, the validation of the DIET and DIER models provided by this thesis, particularly the validation longitudinally, expands understanding into the SVR model.

The findings also support the situation model. The situation model states that successful comprehension is ultimately achieved by the construction of a mental model (Graesser et al., 1994; Zwaan et al., 1995) which is a mental representation of what a passage is about (Kintsch, 1988). The situation model is hierarchical including three levels: the surface code level, the text-base level and the situation level (Kintsch, 1994; Van Dijk et al., 1983). At each level a more advanced portrayal is gained of the passage until an accurate and thorough representation is achieved at the top level (the situation level). The DIET and DIER models are theoretically based on the situation model with the skills required for each level hypothesized e.g. basic language skills and working memory at the surface code level, high-order cognitive skills at the text-base level and LC at the situation model level. In validating the DIET and DIER models longitudinally, this thesis also supports the situation model and gives evidence to confirm that these are the skills needed at each of the levels of the situation model.

Additionally, this thesis verifies the view taken by the literature that ToM is important for LC and RC because of the social element (Dore et al., 2018). In their review Dore et al., (2018) suggest that ToM may be the missing piece in accounts of RC stating that ToM is important for RC because it helps with the understanding of characters' mental states in story books. This theoretical stance can also be extended to LC as and it is the explanation taken by others to explain their findings (e.g. Guajardo & Cartwright, 2016; Ricketts et al., 2013). However, although acknowledging this explanation Atkinson et al., (2017) argued that instead of social specificity of ToM being of importance for RC, it may be ToM's broad metacognitive nature which facilitates comprehension as it could help inform knowledge, actions, and understanding not necessarily related with mental states or social information. By comparing other non-social types of metacognition to ToM in models of LC and RC this thesis was able to test if the social element of ToM is what assist LC and RC rather than the broad metacognitive nature of it. The findings confirm the theory of Dore et al., (2018) providing early evidence that the social element of ToM seems to be important for LC and RC as models including ToM were a better fit for both LC and RC both concurrently and longitudinally. However, within models some of the paths from broad metacognition to LC and RC were approaching significance and so more research is needed to confirm these preliminary findings.

Findings could be taken further to help explain the process in which the social element of ToM assists LC and RC. ToM may assist with the understanding of social information in a story in that it underpins mental model building of the plot. The building of a mental model is required for both proficient LC and RC (Graesser et al., 1994; Zwaan et al., 1995). ToM may directly contribute to model building itself as Perner and colleagues describe the possession of a ToM as mental model building (Johnson-Laird, 1983; Perner, 1991), in that a belief is a mental model of the world. However, the findings cannot verify for certain if the social element assists with mental models building in this way and this may be an avenue for future work.

8.4 Educational implications

The findings in this thesis can inform early years literacy instruction. The current UK literacy curriculum is based on the SVR (Rose, 2005) with a big drive towards phonics in order to foster early decoding skills (Department for Education, 2013). This thesis supports the SVR and therefore also supports the literacy curriculum. However, it also suggests that the linguistic side of reading is important too and this is something that is perhaps not stressed in the current literacy curriculum. The finding that ToM, and particularly the social element of ToM, is important for LC and RC suggests that supporting and encouraging children's metacognition development may also improve their LC and RC. Research shows that false belief training can improve children's ToM (Lecce et al., 2014). Therefore, training such as this used in the preschool setting has the potential to improve concurrent and future LC and RC. Helping children with their social understanding is not a large focus of early years education in the UK (Department for Education, 2017), but given that this thesis provides evidence that a better ToM understanding leads to improved LC and RC perhaps more attention should be given to this.

Further to this, the preliminary finding that childrens' own mental state talk aged four predicts LC aged five (via ToM) suggests that parents should be informed of the importance of child's own talk has for literacy outcomes. There is much research which shows that child mental state talk predicts ToM (e.g. J. Dunn et al., 1991; Ruffman et al., 2002) , but until now mental state talk has not been shown to indirectly predict LC. Recent research shows that preschoolers can be trained to use more mental state talk (Grazzani & Ornaghi, 2011; Ornaghi, Grazzani, Cherubin, Conte, & Piralli, 2015), so training like this could be used in the preschool classroom to improve LC. Regarding maternal mental state talk, there was no evidence that this directly or indirectly relates to LC. Past research shows that extratextual talk by preschool teachers during shared class reading improves LC in children up to a year later (Zucker et al., 2013) and these current findings suggest that this extratextual talk by parents should possibly not give a large focus to discussing and referencing mental states.

8.5 Limitations and considerations

There are a number of noteworthy limitations and considerations to this research. Many of these limitations have already been debated within individual chapters, however here they are discussed in relation to the broad general conclusions of this thesis.

8.5.1 Sample characteristics

This research did not exclude children who did not have English as their first language. In fact, 34% of the Time 1 sample had EAL with 24 different languages represented. Including these EAL children in the sample is an advantage as the Department for Education suggests that 20.1% of the UK classroom has EAL (Department for Education, 2016) and therefore the models tested in this thesis can potentially be generalised to the UK classroom. However, because the majority of literacy research only uses monolingual children (e.g. Atkinson et al., 2017; Cain & Chiu, 2018; Nation et al., 2010; Ricketts et al., 2013) caution may need to be taken when comparing these findings to other research. This said, within each chapter comparisons were made between EAL and English only groups with no differences found in the fit of models. It should be noted that a further possible issue regarding child language is the method used to obtain this information. In part this reflects the opt-in approach taken by for the questionnaire data. Where parents completed the questionnaire, language information was gained this way. However, for those who did not participate in the questionnaire, information on what languages the child spoke at home was gained via the class teacher.

In addition, there may have been some selection bias as parents with an interest in literacy development may have been more likely to agree to participation. For Cohort 1, this was controlled somewhat by the use of opt-out consent in schools. However, for Cohort 2 and for returns of the questionnaire and participation in the book sharing activity in Cohort 1, this may have been an issue. This is reflected by the SES of Cohort 2 who can be best described as a middle class population, with 67% of this cohort reporting their family income to be £70,000 or more, and 83% of mothers reporting to have at least an undergraduate degree. This is however an inherent problem in all developmental psychological research (Braver & Bay, 1992) and is also reflected in other very similar research (e.g. Ruffman et al., 2002). This said, some researchers do endeavour to use enriched and representative samples such as the work by Hughes and colleagues (e.g. Hughes, White, Sharpen, & Dunn, 2000; Hughes, Lecce, & Wilson, 2007) who have conducted developmental research in deprived areas of London. Future work should attempt to follow suit.

Sample size is an issue which has been cited several times within this thesis. This is in relation to the sample size needed for structural equation modelling for Chapters 3-6, the fact that some models included smaller sample sizes than others, and a smaller sample size used in Chapter 7 than similar mental state talk research. For full discussions on these refer to Section 3.4.2 in Chapter 3, Section 5.4.2 in Chapter 5, and Section 6.4.3 in Chapter 6 regarding SEM, and Section 7.4.3 of Chapter 7 regarding mental state talk research. Again, it should be noted that many statistical tests within chapters were approaching statistical significance and this should be considered in the context of the sample size.

8.5.2 The measures

Assessing young children is challenging. Finding age appropriate measures is difficult in developmental research. This is particularly the case for longitudinal work when attempting to track the same skill across time as it is not always appropriate to use the same measure across a span of two or three years due to ceiling and floor effects (Faden et al., 2004). Therefore, here, in some cases new more cognitively advanced measures had to be used at later time points, or existing tasks had to be modified. Careful consideration was always taken when choosing measures and the use of different measures at different time points in longitudinal studies is consistent with similar research (e.g. Atkinson et al., 2017; Caravolas, Hulme, & Snowling, 2001; Hughes, Marks, Ensor, & Lecce, 2010; Lervåg, Bråten, & Hulme, 2009). Yet, when reviewing the longitudinal pattern of results the use of different measures at progressive time points needs to be noted.

Furthermore, children are influenced by the environment in which they are assessed (Shepard, Kagan, & Wurtz, 1998). Here assessment sessions were relatively long, and this also varied between the two cohorts. For Cohort 1, children took part in two 20-25 minute sessions per timepoint, whereas for Cohort 2 all measures were administered in a single session which was up to an hour in duration. Although counterbalancing of task order was used, and tasks were ordered so that content and presentation of consecutive tasks was varied and engaging, it is possible (particularly at the first time point when children were only three years old) that children could have become fatigued and disengaged affecting the reliability of scores. Indeed, research suggests that children have relatively short concentration spans at this age and that there is much individual difference in self-regulation and ability to pay attention (McClelland, Acock, Piccinin, Rhea, & Stallings, 2013).

The environment where the assessment sessions took place may also have influenced the reliability of measures (Faden et al., 2004). For Cohort 1 at Time 1, all children were tested within their preschool classroom and in some cases this was a loud and busy environment that could have affected children's engagement. For future time points the Reception and Year 1 classrooms were much quieter and testing sessions often took place out of the classroom but still within proximity to it. The Cohort 2 sessions were conducted in children's homes which in most cases were quiet and undisturbed environments however in some cases there were instances of disturbances from siblings or other family members.

Chapters 4-6 sought to compare the contribution towards LC and RC of ToM to other forms of metacognition that were not social in nature to determine if the social nature of ToM is of importance to LC and RC. Source monitoring, metamemory and metalinguistic awareness were chosen as broad non-social measures of metacognition because these metacognitive skills have been linked to ToM in the preschool years (Bright-Paul et al., 2008; Doherty, 2000; Lockl & Schneider, 2007) and because measures of these metacognitive skills can be administered in a structurally similar way to false belief tests and have similar language demands. However, it should be noted that these measures are in some respect social. For example, the metamemory measure used at all time points (See Chapter 2, Section 2.5.7.3 for a full description of this task) required children to think about memory and how it works (Ebert, 2015). In the measure children must choose between two dolls, stating which doll would find it easier to remember a group of pictures (e.g. in one trial one doll had six pictures to remember whilst the other doll had only three). The purpose of the task was to assess a child's understanding of metamemory; the understanding of how best human memory works (Ebert, 2015) but as this task involved characters and required the child to think about the dolls and how best their memory works, it could be said that the task had some social aspects. In a similar way some of the metalinguistic measures also involved characters. Therefore, perhaps the claim cannot be fully made that this thesis assessed social versus non-social metacognition and conclusions should be drawn with caution.

Comprehension monitoring and inference making were not measured until Time 3 despite these skills known importance for LC and RC (Cain & Oakhill, 1999; Cain & Yeomans-Maldonado, 2017; Lepola et al., 2012) and their inclusion in the original DIET and DIER models for older children (Kim, 2017). These skills were not measured at the first two time points because they are difficult to measure before the age of six years old and, to the knowledge of the researcher, no UK based measures exist for children this young. As comprehension monitoring and inference making were not included in the longitudinal models in Chapters 3 and Chapters 5-6 these models are not exact replicas of the DIET and DIER models (Kim, 2017) and therefore caution should be taken when making direct comparisons.

A more specific methodological issue which may have important implications relates to the measures of LC used in this thesis (the NARA and the OWLS). These measures were chosen to reflect the work of Kim (2015; 2016; 2017) so that DIET and DIER models were comparable. The measures were also chosen as they were age appropriate to use at each of the three time points as the OWLS is a standardised measure normed at three to 21 years (Carrow-Woolfolk, 2011), and the NARA has been used by other research as a measure of LC in similar aged children (e.g. Bowyer-Crane et al., 2008; Nation et al., 2010). However, these measures administered to such young children, as in the case of this thesis, may not have tapped LC in the same way that the Kim (2015; 2016; 2017) does when administering similar measures to older children. For example, the OWLS measure, which Kim also uses, begins with items which have lower LC demands but later items are more advanced as they measure comprehension of inferences, metaphors and complex embedded sentences. Due to the discontinuation rule, the younger children in the current sample did not reach the higher more complex items as the older children in Kim (2015; 2016; 2017) did, and therefore their LC was perhaps not being measured in the same way. Similarly, in the NARA where stories were read aloud to children and questions asked about their content, again due to the discontinuation rule children in the current research were only administered very simple and short stories compared to those administered to the older children in Kim (2015; 2016; 2017). Therefore, it is possible that the NARA and OWLS measures of

LC here were measuring nothing more than general language ability in these younger children and did not tap the same construct of LC as Kim (2015; 2016; 2017).

8.6 Avenues for future research

There are several important areas to explore in the future. Many of these have already been discussed in individual chapters; here they are synthesised. Firstly, exploring these research questions in older children would be useful. The current research investigated RC and LC development beginning when children were three years old and in preschool, across a span of nearly two years when they were six years old in Year 1. The rationale for the focus on this age group was that at the first time point children had not yet had any formal training in reading and so skills that precede reading could be measured. Secondly, the age of 3-4 years is an important time for ToM development as here children begin to be able to pass a false belief test (Wellman et al., 2001; Wimmer & Perner, 1983). It would now be useful to explore these relationships and particularly the fit the DIET and DIER models in older children and further across time. This is important because the findings of this thesis have been, in some cases, inconsistent across time and therefore exploring further age groups will help to give a bigger picture. Moreover, it would also be interesting to explore whether the contribution of cognitive skills such as ToM, comprehension monitoring and inference making become more important for comprehension as children advance on to more difficult stories which include more complex mental states and inferential points. In the original work on the DIER and DIET Kim also recognises that future research should assess if the model can be generalised across different developmental phases (Kim, 2015, p. 30; Kim, 2017). It would be helpful to follow the current sample further across time and then preschool skills for predicting RC and LC even further across time (e.g. five years) could be assessed.

Further research into the DIET and DIER models could also concentrate on the structure of the models. As explained in Section 3.1.2.1 in Chapter 3 across publications Kim's models have developed slightly (Kim, 2015; Kim, 2016; Kim, 2017) For example, the addition of inference making and moving from a three-tiered model to a four-tiered one. This thesis focused on validating the most current model (Kim, 2017) as this reflects most current thinking. However, Kim does not explain the reason for the structural changes of the models across publication. Therefore, future research should explore whether this most recent model is the best fitting. Given that across chapters many of the statistical paths were marginal, it could be that the structure and order of the model needs development.

To explore further the question over why ToM facilitates LC and RC and if it is because of its assistance with the understanding of social information in a story, future work should investigate the effect of ToM on the comprehension of fiction stories compared with non-fiction stories. If ToM solely assists comprehension with the understanding of mental states in a story plot, then either a weaker or no relationship should be seen between ToM and non-fiction comprehension. Research shows that children's storybook plots revolve around mental states (Zunshine, 2019) but there should be less or no mental state referencing in non-fiction books or passages. In this thesis measures of LC only assessed the comprehension of fiction passages as all the NARA stories were fiction based and the OWLs items were all fiction related too. Although the YARC, which measured RC, did include both types of passages with children taking part in one non-fiction passage and one fiction passage a separate score for each story type was impossible to compute. This is because children take part in different level stories based on their SWRT score (single word reading) and a standardised score based on their age and the level of story is computed, so just taking scores from an individual story would not be a direct comparison. Future research should use measures of LC and RC which can be split into non-fiction and fiction sub-scores so that comparisons can be made. Understanding that different skills required for different types of comprehension has important educational implication for the early years and primary classroom and this research would also give a better understanding into why ToM assists LC and RC. In general reading research has concentrated on fiction comprehension and in comparison only a small number of studies have focused on non-fiction comprehension (L. Baker et al., 2017).

There is also scope for future research into the relationship between mental state talk and LC. The finding that maternal mental state talk did not relate to ToM was surprising, yet this relationship was approaching significance so replication with a larger sample size would be useful. Moreover, regarding children's own mental state talk, the self-report data and the book sharing activity data gave contradictory findings with the self-report data showing that earlier mental state talk predicted LC via ToM but book sharing activity data not supporting this. This inconsistent finding is further evidence for the need for a replication with a larger sample. It would also be useful to look at the effects that these types of talk have on RC. This future work would give a deeper understanding into the relationship between ToM and RC and

LC, and would also extend past home literacy environment research into the contribution of extra-textual talk during shared reading for LC and RC (e.g. Collins, 2016; Demir et al., 2011).

8.7 Conclusions

The findings of this thesis were not always straightforward or consistent across time points, however they do add to the growing body of research that demonstrates that ToM is important for LC and RC in the early years. Further to this, they provide preliminary evidence that ToM facilitates LC and RC due to its social specificity in that ToM assists with the understanding of social information in a story. They also suggest children’s own mental state talk may indirectly predict LC via ToM. This thesis also supports a DIET model of LC and a DIER model of RC and extends these models to a younger population and longitudinally. Although further work is needed to cement findings, this research has important implications not only for updating understanding of models of reading but also for informing early years’ instruction and future intervention.

Appendices

Appendix 1: Parental questionnaire

University of Roehampton London

Reading and Family Questionnaire

This questionnaire will ask you questions about you and your child's reading habits at home, about other activities that you take part in together, and a few basic details about you and your family. Please fill it in as honestly as you can. Try to answer all the questions, but remember that if there are any questions that you feel uncomfortable in answering (or you do not know the answer to) you can leave these blank and go onto the next one.

Thank you for your help!

Questionnaire Consent Form

As you are aware your child is already taking part in our research at school. We thank you for your support of the study so far. We are also interested in finding out about you and your child's shared reading habits. Therefore, we would like to ask you to participate in an additional part of the research by completing the attached questionnaire.

Below is some information about this additional part of the research. If you are interested in taking part you can do so by filling in the attached questionnaire or by going to the website below.

<u>What will happen?</u>
Participation will include filling out a questionnaire. The questionnaire will tell us a little bit more about your family and about the kinds of activities you and your family enjoy. This should take you no longer than 20 minutes to complete. You can do this on paper and return to school, or online at the below website, whichever is easiest for you!

<u>What about confidentiality?</u>
Remember you are not obliged to take part so if you do not wish to that is fine. If you do take part all data will be treated entirely confidentially. The collected data will be securely stored in confidential computer files and in locked filing cabinets at Roehampton University and will be accessible only to the study investigators. We are not interested in any individual's data rather we are interested in general trends, as such this will not include any identifying details of individual children. This is a long-term project, and if you or your child should wish to withdraw from the study at any later date, please contact us and we will remove all your child's scores from our dataset. Please note however that despite withdrawing from the study, data may already have been used in publications relating to this research, though only in aggregate form as part of larger datasets used for statistical analysis.

There will, of course, be no adverse consequences if you choose not to take part or to withdraw your data.

If you wish to take part and if you would like to do the questionnaires on paper please fill in the questionnaire attached to this letter and return to your child's classroom teacher once completed.

Alternatively, if you would like to complete the questionnaire online instead please go to www.XXXX.com. This online version of the questionnaire can be completed on any internet linked device e.g. your phone or computer.

Consent Statement:

I confirm that:
I agree to take part in this research, and am aware that we are free to withdraw at any point without giving reason, although if we do so I understand that the data might still be used in a collated form. I understand that the information provided will be treated in confidence by the investigator and that mine or my child's identity will be protected in the publication of any findings, and that data will be collected and processed in accordance with the Data Protection Act 1999, and with the University's Data Protection Policy.

Name of child: …………………………………..................................

Name of parent/guardian: ....

Signature: …………………………..……

Date: ……………………………………

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator or the Director of Studies. However if you would like to contact an independent party please contact the head of department.

Director of Studies contact details:

Dr Lance Slade
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
L.Slade@roehampton.ac.uk
Tel: 020 8392 3576

The Head of Department's contact details:

Dr Diane Bray
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
d.bray@roehampton.ac.uk
Tel: 020 8392 3617

For words that your child **understands but does not yet say,** place a mark in the yellow column, labelled **"U"**

	X	U	U/S
Upset	○	●	○

For the words that your child **understands and also says**, place a mark in the green column, labelled **"U/S"**

	X	U	U/S
Upset	○	○	●

For the words that your child **neither understands or says**, place a mark in the red column, labelled **"X"**

	X	U	U/S
Upset	●	○	○

	X	U	U/S
All gone	○	○	○
Want	○	○	○
Asleep	○	○	○
Hope	○	○	○
Wish	○	○	○
Bad	○	○	○
Dark	○	○	○
Afraid	○	○	○
Dirty	○	○	○
Dream	○	○	○
Like	○	○	○
Prefer	○	○	○
Dry	○	○	○
Love	○	○	○
Keen on	○	○	○
Think	○	○	○
Believe	○	○	○
Empty	○	○	○
Wonder	○	○	○
Fast	○	○	○
Fine	○	○	○
Gentle	○	○	○
Know	○	○	○
Expect	○	○	○
Annoyed	○	○	○
Good	○	○	○
Mad	○	○	○
Fun	○	○	○
Hurtful	○	○	○
Interested	○	○	○
Green	○	○	○
Happy	○	○	○
Hard	○	○	○

	X	U	U/S
Big	○	○	○
Broken	○	○	○
Bored	○	○	○
Frightened	○	○	○
Careful	○	○	○
Frustrated	○	○	○
Hot	○	○	○
Afraid	○	○	○
Soft	○	○	○
Missed	○	○	○
Feel	○	○	○
Hungry	○	○	○
Hurt	○	○	○
Little	○	○	○
Feel bad	○	○	○
Nasty	○	○	○
Worried	○	○	○
Distinguished	○	○	○
Naughty	○	○	○
Better	○	○	○
Shocked	○	○	○
Tired	○	○	○
Okay	○	○	○
Nice	○	○	○
Upset	○	○	○
Shy	○	○	○
Old	○	○	○
Good	○	○	○
Fed up	○	○	○
Yellow	○	○	○
Surprised	○	○	○
Sad	○	○	○
Scared	○	○	○

	X	U	U/S
Miserable	○	○	○
Pleased	○	○	○
Clean	○	○	○
Cross	○	○	○
Blue	○	○	○
Cold	○	○	○
Grumpy	○	○	○
Sick	○	○	○
Enjoy	○	○	○
Angry	○	○	○
Sleepy	○	○	○
Excited	○	○	○
Pretty	○	○	○
Red	○	○	○
Wet	○	○	○
Thirsty	○	○	○

About you and your family

What is your marital status? ______________________________

What is your occupation? ______________________________

How many hours a week do you work? ______________________________

What is your highest educational achievement? (please tick one)

School leavers certificate	GCSEs	A levels	GNVQs or BTECs	University degree	Postgraduate degree	Doctorate

What is your partner's occupation? ______________________________

How many hours a week do they work? ______________________________

What is your partner's highest educational achievement? (please tick one)

School leavers certificate	GCSEs	A levels	GNVQs or BTECs	University degree	Postgrad degree	Doctorate	Not applicable or unsure

What is your household annual income? (please tick one)

Less than £15,000	£15,000 - £24,999	£25,000- £34,999	£35,000- £49,999	£50,000- £69,999	£70,000 or more	Prefer not to say

Last year (before they started school) did your child attend nursery or pre-school? (please tick one)

Yes, nursery	Yes, nursery school (attached to a Primary school)	Yes, playgroup	No, they were at home all the time	Other

What is your child's main language? (which the main caregiver has spoken to your child from birth until now) ____

Does your child speak or read in any other language? ______________________________

If yes, what languages? ____

What languages is mainly spoken at home? ____________________________________

Appendix 2: Cohort 2 recruitment poster

Exploring early book sharing and reading

University of Roehampton
London

Is your child aged 3-4 years? Or do you know a child of this age?

Take part in a 1 hour study on reading with them to earn a £10 Amazon voucher

Reading is a really essential skill so it is important we understand all the factors which affect it. Because young children can't "tell" us what they know, we have designed simple games and puzzles that allow them to "show" us what they know. You and your child will come into the university and take part in a number of fun games and activities together.

For more information contact Sophie at jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Reading Study
Contact:
Sophie at
jacksons1@roehampton.ac.uk

Appendix 3: Cohort 2 recruitment flyer

Is your child aged 3-4 years? Will they start school in September 2017?

Take part in a 1 hour study on reading with them and receive a £10 Amazon voucher

Aim of the research: Reading is a complex task and so it is important that we understand all factors which affect reading development. At the University of Roehampton we are addressing this by working with children the year before they start school and tracking their performance in reading and reading-related skills as they progress through the first two years of primary school. It is hoped that approximately 200-250 children across the UK will be taking part in this project.

Taking part: If you choose to take part the researcher will visit you in your home for a one hour session. During this session the researcher will sit with your child and together they will work through a number of different games and puzzles. Because young children can't "tell" us what they know, we have designed simple games and puzzles that allow them to "show" us what they know. These games are fun and children really enjoy taking part. We will also ask you to fill in a quick questionnaire which will give us some information about the reading habits of you and your family. Lastly, we ask you to take part in a book sharing activity with your child, in which you will look at some pictures together and talk about them as you would a book at bedtime.

What next? We would ideally like to look at these abilities at three time points; preschool, Reception and Year 1, and so we would like to carry out similar sessions with your child again when they are in Reception and Year 1. However, you do not need to commit to these future sessions now. Now you are only committing to this one preschool session. We will give you a £10 Amazon voucher for each session as a "thank you" for your time.

If you are interested and would like more info please email Sophie Jackson at jacksons1@roehampton.ac.uk

Appendix 4: Headteacher consent form

<u>Title of Research Project:</u> Exploring book sharing and reading over early years

<u>Aim of the research:</u>
This longitudinal study will specifically address factors which promote later reading comprehension, by assessing children during their nursery year, before they experience any formal reading instruction, and then track their performance in reading, language skills, social and non-social reasoning as they progress through the first two years of primary school. Recent evidence suggests that as well as children's decoding and language skills, children's social reasoning (their ability to reason about the thoughts and feelings of others) may also contribute to later reading comprehension. This may be because this social understanding draws on broad reasoning abilities or because it is more socially specialised to work out the intentions of others, including story characters. It is also known that the home literacy environment is a powerful influence on children's early reading ability. Therefore, in addition to these school-based tasks we will invite childrens' parents to answer some questions about their child's reading, and to take part in a short book sharing activity at home.

<u>How will it be carried out?</u>
We will be carrying out a wide range of tasks to index the range of skills thought to impact on literacy. These include measures of language ability (e.g. understanding vocabulary, understanding of complex sentences, early letter sound knowledge and later decoding abilities). We will also be looking at children's social (theory of mind ability) and non-social understanding (understanding of memory and reasoning tasks). We aim to look at these abilities at three time points; Nursery, Reception and Year 1. During Year 1 reading comprehension will also be assessed. The researcher will be very happy to talk you and the class teachers through each of the tasks.

The first tasks will be carried out with the nursery children in the summer term. It is anticipated that this will involve two individual test sessions for each child lasting between 20-25 minutes. These sessions will take place in a quiet area of the school under the supervisor of the teacher. The children will then be re-assessed during Reception and Year 1 and will take part in similar tasks. Tasks will involve a combination of standardised pen-and-paper task, customised computer or hands on measures, which will be presented in the form of enjoyable games and puzzles. In general, the tasks will be largely typical of children's normal classroom activities.

We are also interested in the way in which parents and guardians read and interact with their child. Therefore we will also ask for letters to be sent home each year in which we will invite the childrens' parents to fill out a reading habits questionnaire. Those parents who return the questionnaire will be invited to take part in a further book sharing activity at home which will be audio recorded by the parent and returned to the researcher.

<u>What about parental consent?</u>
For activities that take place on the school premises we will be using opt out consent. This means that before the beginning of the research sessions parents' of the children in the nursery class will be sent letters informing them of the study. They will only need to reply to this letter if they <u>do not wish</u> their child to take part in the research thus opting out. After a child's first testing session an additional letter will be sent to parents thanking them for their support and inviting them to complete the attached reading habits questionnaire. This will involve parents opting in to give this consent. Parents who return the questionnaire will be further invited to take part in the home based book sharing activity. Again, this will require parents to opt in to this part of the study.

<u>What about confidentiality?</u>
The collected data will be treated entirely confidentially. It will be securely stored in confidential computer files and in locked filing cabinets at Roehampton University and will be accessible only to the study investigators. We aim to use the aggregated data in future academic publications; however this will not include any identifying details of individual children. This is a long-term project, and should you wish to withdraw your school from the study at any later date, please contact us at the address given below and we will remove the childrens' data from the study. Please note however that despite withdrawing from the study, data may already have been used in publications relating to this research, though only in aggregate form as part of larger datasets used for statistical analysis. All researchers working on the project have full Disclosure and Barring Service clearance.

Investigator Contact Details:
Miss Sophie Jackson
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
jacksons1@roehampton.ac.uk

Consent Statement:
I confirm that:

- I have read and understand the brief description of the research project.
- The above study has been fully explained to me and I have had the opportunity to ask questions.
- Parents/guardians of each child participating in the study will be fully informed about the nature if the research by letter sent home to them.
- I understand that the parents (or guardians) of the children who are identified as potential participants, and, who are interested in taking part in the study, will be sent an opt-out consent form.
- Parents/guardians will be given a reasonable period of time (at least one week) to withdraw their child from participating in the study.
- I understand that my school's participation is voluntary and that we are free to withdraw ant time without penalty.

I agree for XXXX School to take part in this research, and am aware that we are free to withdraw at any point without giving reason, although if we do so I understand that the data might still be used in a collated form. We understand that the information provided will be treated in confidence by the investigator and that all identities will be protected in the publication of any findings, and that data will be collected and processed in accordance with the Data Protection Act 1999, and with the University's Data Protection Policy.

Name ……………………………………………….....................

Signature ………………………………………....................

Date …………………………………………….....................

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator or the Director of Studies. However if you would like to contact an independent party please contact the head of department.

Director of Studies:
Dr Lance Slade
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
L.Slade@roehampton.ac.uk
Tel: 020 8392 3576

Head of Department:
Dr Diane Bray
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
d.bray@roehampton.ac.uk
Tel: 020 8392 3617

Appendix 5: Cohort 1 Time 1 parental information letters and opt-out consent

Reading Research Parent Information Sheet (Opt out consent form)

We are currently running a research study looking at early book sharing. This project has been discussed and agreed with [insert headteacher name]. However, it is also important that parents/guardians understand why the study is being conducted and what it will involve.

Title of Research Project : Exploring book sharing and reading over early years

Why is the research important?
Reading is a highly complex task. It is essential that we fully understand all factors which affect reading development. We are working with children to specifically address this by assessing children during their nursery year, before they experience any formal reading instruction, and tracking their performance in reading and reading-related skills as they progress through the first two years of primary school. It is hoped that approximately 100-150 children across the UK will be taking part.

What will happen?
We will be carrying out a wide range of tasks to gain an index of the range of skills thought to impact on literacy. These include measures of language ability (e.g. understanding vocabulary, complex sentences, early letter sound knowledge and later decoding abilities). We will also be looking at children's social (theory of mind ability) and non-social understanding (understanding of memory and reasoning tasks).The first sessions will be carried out with your nursery child in the summer term. It is anticipated that this will involve two individual sessions lasting no longer than 25 minutes. Your child will complete similar tasks during Reception and Year 1. Tasks will involve a combination of standardised pen-and-paper assessments and customised computer or hands on measures, which will be presented in the form of enjoyable games or puzzles. In general, the tasks will be largely typical of children's normal classroom activities. We may also be inviting you to take part in a follow up activity. Please note however, that we will seek your consent for this in a separate letter. This is entirely voluntary.

What about confidentiality?
All data will be treated entirely confidentially and your child's name will never be directly linked with his or her scores on any of the tasks that they complete. The collected data will be securely stored in confidential computer files and in locked filing cabinets at Roehampton University and will be accessible only to the study investigators. We are not interested in any particular individual's data rather we are interested in general trends, as such this will not include any identifying details of individual children. This is a long-term project, and if you or your child should wish to withdraw from the study at any later date, please contact us at the address given below and we will remove all your child's scores from our dataset. Please note however that despite withdrawing from the study, data may already have been used in publications relating to this research, though only in aggregate form as part of larger datasets used for statistical analysis. There will, of course, be no adverse consequences if you choose not to take part or to withdraw your data. All researchers working on the project have full Disclosure and Barring Service clearance.

If you would like any further information in the meantime please contact us at the addresses below.

Investigator Contact Details:

Miss Sophie Jackson
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
jacksons1@roehampton.ac.uk

Dr Lance Slade
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
l.slade@roehampton.ac.uk

Please note: if you have a concern about any aspect of your child's participation or any other queries please raise this with the investigator. However if you would like to contact an independent party please contact the head of department.

The Head of Department's contact details:
Dr Diane Bray
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
d.bray@roehampton.ac.uk
Tel: 020-8392 3617

If you are willing for your child to take part in the study you do not need to contact us; however if you <u>do not wish your child to take part</u> please complete the attached form and return it to your child's class teacher.

- -

I would prefer my child **<u>not</u>** to take part in the reading research project.

(Please <u>only</u> complete this form if you <u>object</u> to your child taking part in the research study; you do not need to respond if you are happy for your child to participate.)

Name of child: …………………………………………………………………………

Name of parent/guardian: ……………………………………………………………

Signature: ……………………………………………………………………………

Appendix 6: Cohort 1 Time 3 and 3 parental courtesy letters

Reading Research
Exploring book sharing and reading over early years

Dear parent,

This is a reminder that your child will soon be taking part in the next session of the above project at school. Last year your child participated in two individual sessions for this project during their nursery year. These next sessions will be similar to last year, and will involve language activities and reasoning tasks, which will be presented in the form of enjoyable games and puzzles. Your child will take part in another similar session this time next year. Like last year we may also invite you to take part in a follow up activity. Please note however, that we will seek your consent for this in a separate letter and participation is entirely voluntary.

We would also like to remind you that data will be treated entirely confidentially and your child's name will never be directly linked with his or her scores on any tasks they complete. Please also note we are not interested in any particular individual's data, rather we are interested in general trends. All researchers working in the school on the project have full Disclosure and Barring Service clearance.

If you would like any further information, or wish to withdraw your child from the project please contact the researchers at the addresses below. Alternatively, the researchers will be at school from Tuesday 2nd May until Friday 5th May if you wish to speak to them in person. There will, of course, be no adverse consequences if you choose to withdraw your child's data.

Investigator contact details:

Miss Sophie Jackson
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
jacksons1@roehampton.ac.uk

Dr Samantha McCormick
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
Samantha.McCormick@roehampton.ac.uk

Please note: if you have a concern about any aspect of your child's participation or any other queries please raise this with the investigator. However, if you would like to contact an independent party please contact the head of department.

The head of department's contact details:
Dr Diane Bray
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
d.bray@roehampton.ac.uk
Tel: 020-8392 3617

Appendix 7: Cohort 1 book sharing activity consent form

Exploring book sharing and reading over early years

Thank you for your support with the project so far and thank you for your interest in helping us further. We are also want to find out about the types of conversations you and your child have during reading and book sharing. If you choose to participate we will ask you to make an audio recording of you and your child during a picture book activity. Below is some information about this activity. If you are interested in taking part you can do so by following the attached instructions.

<u>What will happen?</u>
We have emailed you a presentation which includes 10 different pictures. These pictures depict children and their families' taking part in everyday activities. We will ask you to make an audio recording on your phone (or similar device such as a tablet) as you sit with your child and talk together about these pictures. You should talk about the pictures as if you would if you were reading a picture book at bedtime. You can take as long as you like on each picture, there is no maximum or minimum time. Once you have completed the recording you will send it back to the researchers by uploading it on to a secure site within the university website. Detailed step-by-step instructions have been given to you so that you can easily make the recording and upload it to the website.

<u>What about confidentiality?</u>
Remember you are not obliged to take part so if you do not wish to that is fine. If you do take part all data will be treated entirely confidentially. The website in which the recordings are uploaded is secure and recordings will only be accessible to the study investigators. The recordings will only be listened to by those working on the project. We are not interested in any individual's data rather we are interested in general trends, as such this will not include any identifying details of individual children. This is a long-term project, and if you or your child should wish to withdraw from the study at any later date, please contact us and we will remove all your child's scores from our dataset. Please note however that despite withdrawing from the study, data may already have been used in publications relating to this research, though only in aggregate form as part of larger datasets used for statistical analysis.

Consent Statement:
I confirm that:
I agree to take part in this research, and am aware that we are free to withdraw at any point without giving reason, although if we do so I understand that the data might still be used in a collated form. I understand that the information provided will be treated in confidence by the investigator and that mine or my child's identity will be protected in the publication of any findings, and that data will be collected and processed in accordance with the Data Protection Act 1999, and with the University's Data Protection Policy.

Name of child: …………………………………...............................

Name of parent/guardian: ....

Signature: ………………………...……

Date: ……………………………………

Please note: if you have a concern about any aspect of your participation or any other queries please raise this with the investigator or the director of study. However if you would like to contact an independent party please contact the head of department.

Director of study:
Dr Lance Slade
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
L.Slade@roehampton.ac.uk
Tel: 020 8392 3576

The Head of Department:
Dr Diane Bray
Department of Psychology
University of Roehampton
Whitelands College
Holybourne Avenue
London
SW15 4JD
d.bray@roehampton.ac.uk
Tel: 020 8392 3617

Appendix 8: Cohort 2 parental consent form

Reading Research Parent Information Sheet and Consent Form

Title of Research Project : Exploring book sharing and reading over early years

Why is the research important?
Reading is a highly complex task and so it is important that we fully understand all factors which affect reading development. We are working with children to specifically address this by assessing children the year before they start school when they have not yet experienced any formal reading instruction, and tracking their performance in reading and reading-related skills as they progress through the first two years of primary school. It is hoped that approximately 200-250 children across the UK will be taking part in this project.

What will happen?
We will be carrying out a wide range of tasks with your child to gain an index of the range of skills thought to impact on literacy. These include measures of language ability (e.g. understanding vocabulary, complex sentences, letter sound knowledge and decoding ability). We will also be looking at their social (theory of mind ability) and non-social understanding (understanding of memory and reasoning tasks). Additionally we are interested in the activities that you and your child take part in at home, including the types of conversations you and your child have during reading and book sharing.

During this first session the researcher will sit with your child and together they will work through a number of different games and puzzles. These tasks will involve a combination of standardised pen-and-paper assessments and hands on measures presented in the form of enjoyable games or puzzles. In general, the tasks will be typical of children's normal activities (that they may experience in the home with you or in a nursery or playgroup setting). During the session we will also ask you to fill in a questionnaire. This questionnaire will tell us a little bit more about your family and about the kinds of activities you and your family enjoy. Lastly, we will ask you and your child to take part in an activity together; you will sit with your child and talk together about ten pictures. You should talk about the pictures as if you would if you were reading a picture book at bedtime. Your conversation and some of the other tasks will be audio recorded for the researcher's reference.

We estimate that all these tasks together should take about an hour and will take part either in your home or in a quiet room within the university (depending on your preference).

These first sessions will be carried out with your pre-school child in late 2016 or early 2017 when your child is 3 or 4 years old. This is a long term study and so we will be in touch with you again a year later so that you and your child can take part in similar sessions during their Reception year, and again when they are in Year 1.

What about confidentiality?
All data will be treated entirely confidentially and yours or your child's name will never be directly linked with his or her scores on any of the tasks that they complete. We are not interested in any particular individual's data rather we are interested in general trends. The collected data will be securely stored in confidential computer

files and in locked filing cabinets at Roehampton University and will be accessible only to the study investigators. Additionally, the audio recordings made during the sessions will also be stored securely and only listened to by those working on the project. This is a long-term project, and if you or your child should wish to withdraw from the study at any later date please contact us at the addresses given below and we will remove all your child's scores from our dataset. Please note however that despite withdrawing from the study, data may already have been used in publications relating to this research, though only in aggregate form as part of larger datasets used for statistical analysis. There will, of course, be no adverse consequences if you choose not to take part or to withdraw your data. All researchers working on the project have full Disclosure and Barring Service clearance.

Consent Statement:

I confirm that:
I agree to take part in this research, and am aware that we are free to withdraw at any point without giving reason, although if we do so I understand that the data might still be used in a collated form. I understand that the information provided will be treated in confidence by the investigator and that mine or my child's identity will be protected in the publication of any findings, and that data will be collected and processed in accordance with the Data Protection Act 1999, and with the University's Data Protection Policy.

Name of child: …………………………………..........................

Name of parent/guardian: ....

Signature: …………………………...…

Date: ………………………………………………………………….…

Appendix 9: NARA stories and questions

Bird (Level 1)

A bird hoped up to my window. I gave her some bread. She made a nest in my garden. Now I look after her little ones.

Questions:
1) Where did the bird hop to? (Answer = to my/the window).
2) What did the little girl give the bird? (Answer = bread).
3) What did the bird do in the garden? (Answer = build a nest).
4) What does the little girl do now for the bird (Answer = look after the little ones).

Road Safety (Level 2)

Kim stopped on her way to school. In the middle of the traffic lay two children. Their bicycles had crashed into each other. Kim ran quickly to help. She saw that no-one was hurt. The children pointed to a television camera "We are taking part in a road safety lesion" they said.

Questions:
1) Where was Kim going? (Answer = to school).
2) Why did Kim stop? (Answer = she saw two children lying on the road/ she saw an accident.)
3) What had happened to the bikes? (Answer = they had crashed).
4) How do you think Kim felt? (Answer = frightened / curious / anxious / scared / worried / upset).
5) What did Kim do? (Answer = she ran to help them).
6) Were the children hurt? (Answer = no).
7) What were the children really doing? (Answer = taking part in a road safety lesson).
8) How did Kim find out what was happening? (Answer = the children told her and pointed to the television camera).

Ali (Level 3)

As Ali sheltered in an old temple, his shoulder knocked a secret spring. Instantly, he was thrown into an underground room. In the darkness the walls seemed to be covered with jewels. Ali rested awhile. He knew that dessert travellers often imagined strange things. Later, he explored the place for a way to escape. To his amazement, the jewels were still there. He had found a palace that had been buried long ago.

Questions:
1) Why did Ali go into the temple? (Answer = to shelter).
2) How did he find the secret spring? (Answer = knocked it with his shoulder).
3) What happened when he touched the spring? (Answer = he was thrown into an underground room).
4) What did he see there? (Answer = Jewels)
5) Why did Ali not rush to look at the jewels? (Answer = he did not think they were real / he thought he was imagining them).

6) After he had rested, what did Ali try to find? (Answer = a way out / a way to escape).
7) Why was he so surprised? (Answer = the jewels were still there / the jewels were real).
8) How did the jewels come to be there? (Answer = they belonged to a buried palace of long ago).

Jan (Level 4)

Jan buckled in her diving belt of mental weights and dropped from the launch. Skipper Kells supervised her air-hose to present tangling. Leo, following the bubbles, guided the dinghy above the diver, as she searched the mysterious underwater world. Jan surfaced frequently clutching crayfish. The required number of specimens was almost obtained when the grey nurse shark advanced directly towards her. Jan retreated cautiously without signalling for assistance. The creature brushed by, ignoring her, as baby sharks emerged from some rocky grooves. Their welfare was more important to the shark than the diver's now motionless figure.

Questions:
1) What equipment assisted Jan in her exploration under water? (Answer = *two of:* Diving belt/ weights/ air-hose).
2) What did Skipper Kells do to help Jan? (Answer = supervised her air-hose to prevent tangling).
3) How did Leo know where the diver was? (Answer = following her bubbles)
4) What do you think Jan was diving for? (Answer = specimens and/or crayfish).
5) Why did it seem that the shark might attack her? (Answer = it advanced directly towards her).
6) How did Jan avoid trouble with the shark? (Answer = retreated cautiously and remained motionless / she kept still).
7) What kind of home protected the baby sharks from enemies? (Answer = rocky grooves).
8) Why was the shark not interested in Jan? (Answer = it was more concerned with the welfare of its babies / wanted to protect its babies).

Appendix 10: Preschool Repetition words

Regular words:

Ladder
Egg
Magic
Arm
Computer
Machine
Mouse
Cigarette
Banana
Police
Magazine
Jar
Holiday
Lamb
Toe
Dinosaur
Balloon
Person

Non-words:

Jai
Tur (silent ‘r’)
Oog (short ‘oo’ put)
Aum
Lom
Mees
Dalla
Serpon
Jamick
Lopice
Shameen
Laboon
Sinodaw (like dinosaur)
Lodihay
Nanaba
Tongkupa
Gazameen
Rigaset

Appendix 11: DTWRP stimuli

11a: Card 1 (alien words)

11b: Card 2 (non- words)

sus	gouse	netrich	piclin	gobner
cortue	turmness	chimpister	stroise	marzentrate
statnic	banifice	sacranzee	anecoil	audimental
concipan	wilderdote	ostant	elephaps	experorium

11c: Card 4 (regular words)

up	sun	them	went	us
made	dragon	well	mouse	gave
elephant	street	corner	kettle	noise
ostrich	chimpanzee	picnic	perhaps	goblin
banister	statue	marzipan	experimental	turmoil
concentrate	sacrifice	wilderness	auditorium	anecdote

11d: Card 3 (exception words)

his	come	ball	some	who
there	monkey	half	ghost	know
many	sugar	want	giant	island
station	soup	cousin	machine	stomach
vehicle	restaurant	parachute	reservoir	mosquito
sovereign	treacherous	horizon	speciality	miscellaneous

Appendix 12: Time 2 Strange Stories Stimuli

12a: Script

"Now I'm going to play you some stories. I want you to listen carefully because afterwards I am going to ask you some questions to see what you think of them. Are you ready?"

Give positive comments throughout testing, but ***do not*** *provide feedback about correctness of answers. Do not provide prompts. Complete all the stories and questions. Each story may be repeated if requested by the child or if they answer 'I don't know' to question 1.*

Simon and Gemma

Q: Why will Gemma look in the cupboard for the paddle?

2 points: reference to Gemma knowing Simon lies
1 point: reference to facts (that's where it really is, Simon's a big liar) or Simon hiding it without reference to implications of lying
0 points: reference to general nonspecific info

Army

Q: Why did the prisoner say that?

2 points: reference to fact that other army will not believe and hence look in other place, reference to prisoner's realization that that's what they'll do, or reference to double bluff
1 point: reference to outcome (to save his army's tanks) or to mislead them
0 points: reference to motivation that misses the point of double bluff (he was scared)

Hungry Brian

Q: Why does Brian say this?

2 points: reference to fact that he's trying to elicit sympathy, being deceptive
1 point: reference to his state (greedy), outcome (to get more sausages) or factual
0 points: reference to a motivation that misses the point of sympathy elicitation/deception, or factually incorrect

Mrs Peabody

Q: Why did she say that?

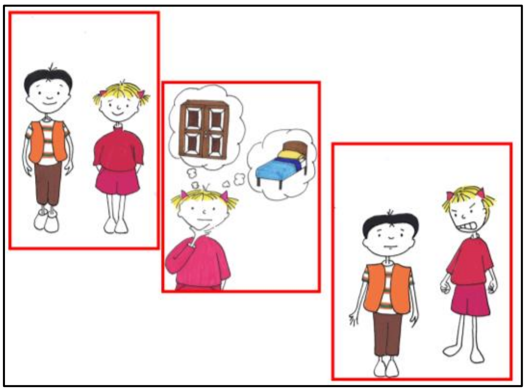
2 points: reference to her belief that he was going to mug her or her ignorance of his real intention
1 point: reference to her trait (she's nervous) or state (she's scared) or intention (so he wouldn't hurt her) without suggestion that fear was unnecessary
0 points: factually incorrect/irrelevant answers; reference to the man actually intending to attack her

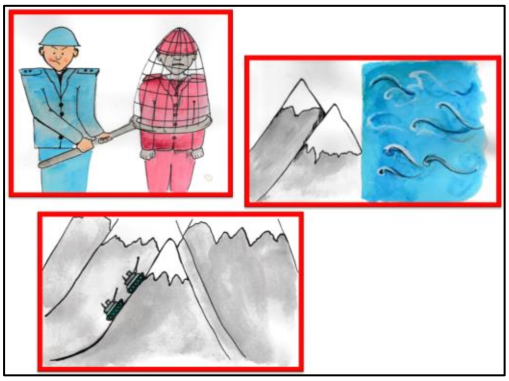
Burglar and Policeman

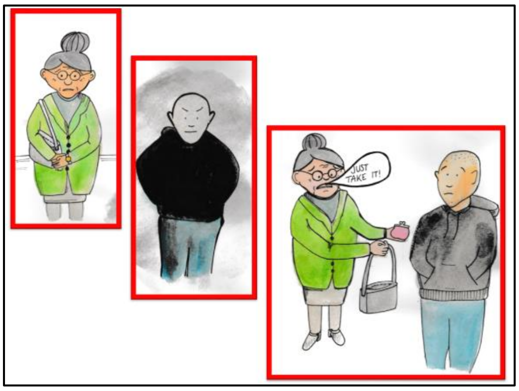
Q: Why did the burglar do that?

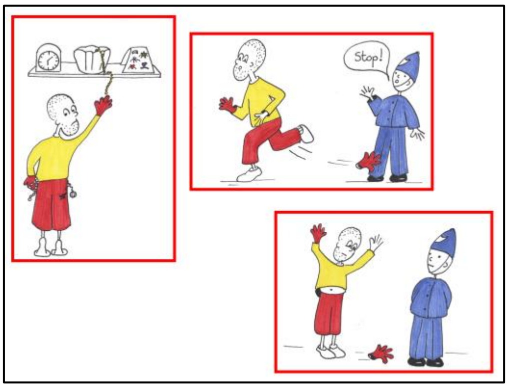
2 points: reference to belief that policeman knew that he'd burgled the shop
1 point: reference to something factually correct in story
0 points: actually incorrect/irrelevant answers

12b: pictures



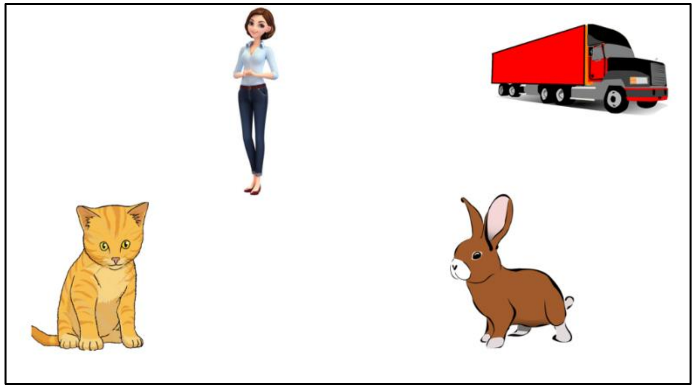


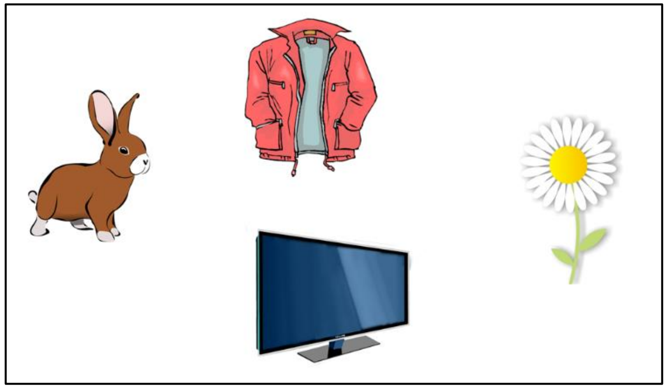


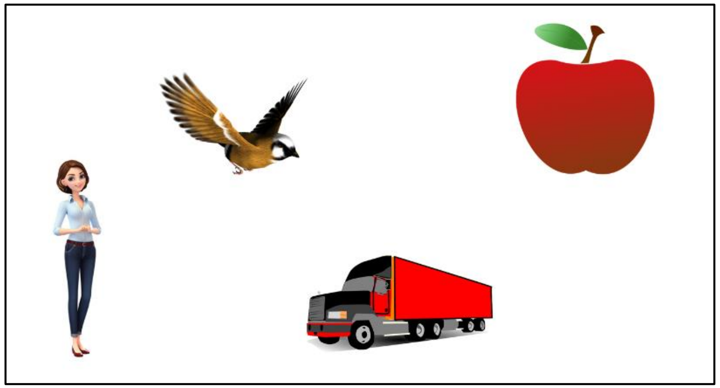


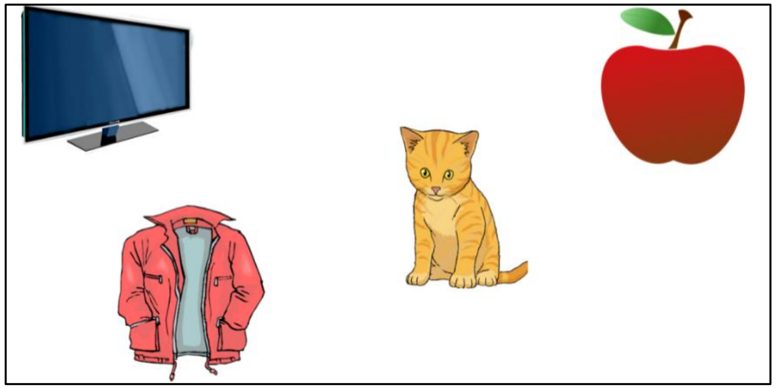
Appendix 13: Time 1 metalinguistic awareness stimuli

13a: Vocabulary picture cards

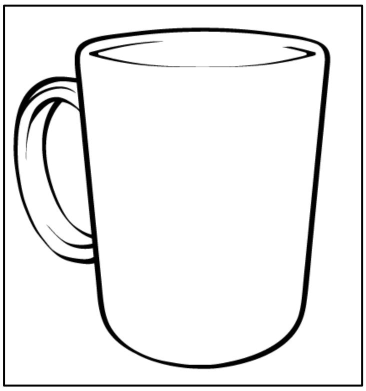


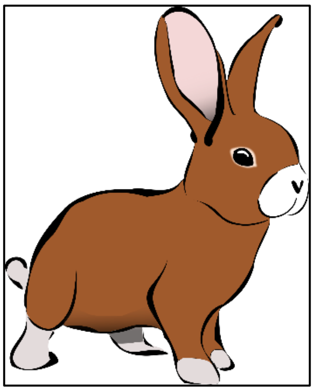




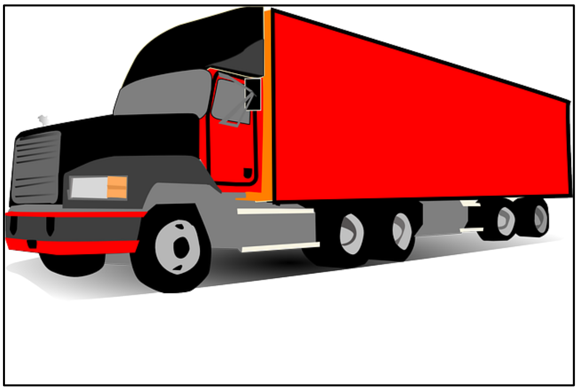


13b: Modelling picture cards



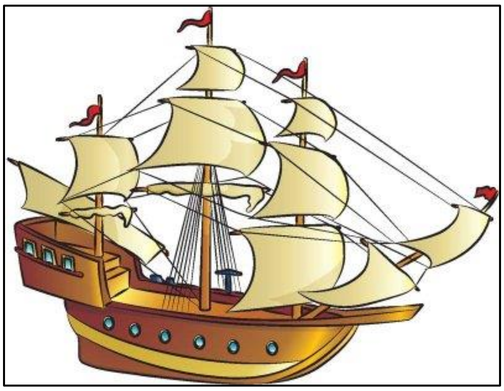


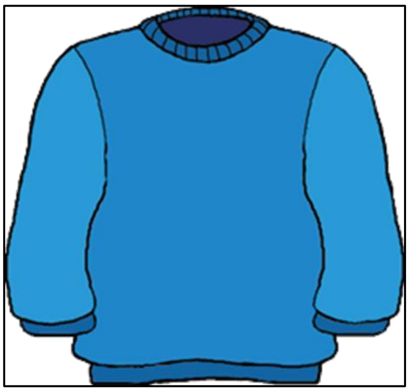
13c: Test picture cards

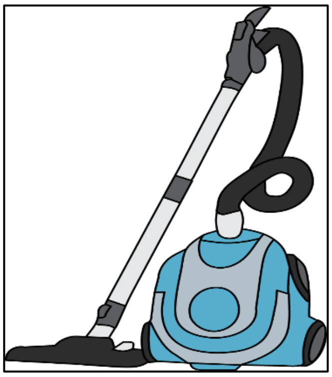


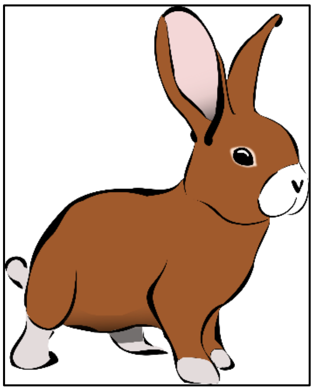
Appendix 14: Time 2 metalinguistic awareness stimuli

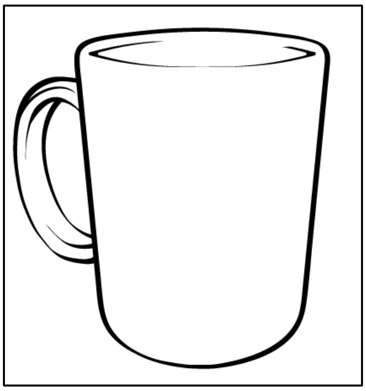
14a: Picture test cards

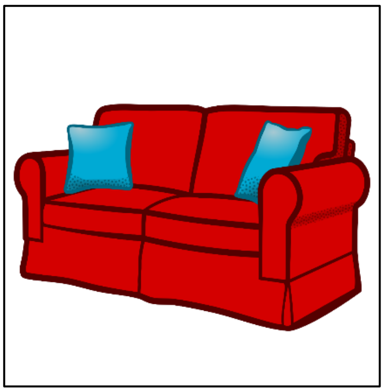


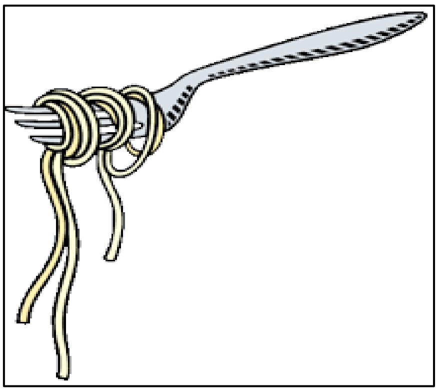


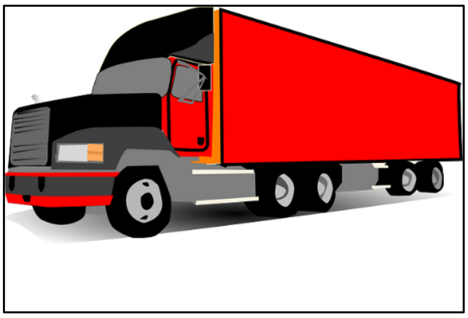


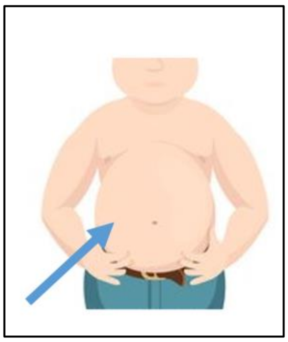












14b: Script

Training: "Now let's play a picture game" *Present the rabbit/bunny picture* "Some objects have two names, see this can be called a rabbit, but it can also be called a bunny" *Now present the Cup/Mug* "This is a mug. Can I also call this a bike?" *wait a few seconds for response* "NO! That is not the same, is it? A mug and a bike aren't the same thing; I can't call this a bike"

Testing "Right let's play the game. In the game, I am going to show you some picture and tell you what they are. Then I want you to tell me if I can also call it another name. Ok? Let's have a go"

1. *Present picture of Jumper/Sweater* "This is a **jumper**; can I also call this a **sweater**?"

Response: ________________________________ (correct answer: yes)

2. *Present picture of Lady/Woman* "This is a **lady**; can I also call this a **jacket**?"

Response: ________________________________ (correct answer: no)

3. *Present picture of Boat/Ship* "This is a **boat**; can I also call this a **ship**?"

Response: ________________________________ (correct answer: yes)

4. *Present picture of Ocean/Sea* "This is the **ocean**; can I also call this the **sea?**"

Response: ________________________________ (correct answer: yes)

5. *Present picture of Television/TV* "This is a **television**; can I also call this **pasta?**"

Response: ________________________________ (correct answer: no)

6. *Present picture of Lorry/Truck* "This is a **lorry**; can I also call this **truck?**"

Response: ________________________________ (correct answer: yes)

7. *Present picture of Spaghetti/Pasta* "This is **spaghetti;** can I also call this a **pushchair?**"

Response: ________________________________ (correct answer: no)

8. *Present picture of Coat/Jacket* "This is a **coat;** can I also call this a **TV?**"

Response: ________________________________ (correct answer: no)

9. *Present picture of Belly/Tummy* "This is a **belly;** can I also call this a **tummy?**"

Response: ________________________________ (correct answer: yes)

10. *Present picture of Buggy/Pushchair* "This is **buggy;** can I also call this a **vacuum cleaner?**"

Response: ________________________________ (correct answer: no)

11. *Present picture of Settee/Sofa* "This is **settee;** can I also call this a **sofa?**"

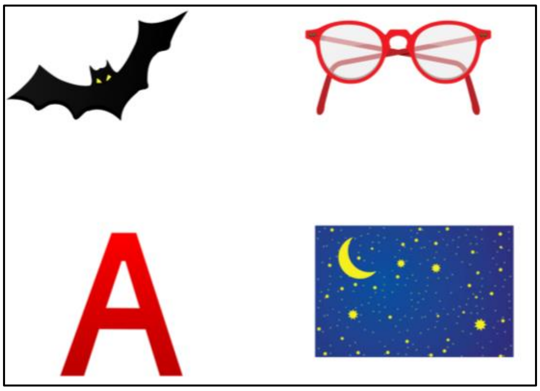
Response: ________________________________ (correct answer: yes)

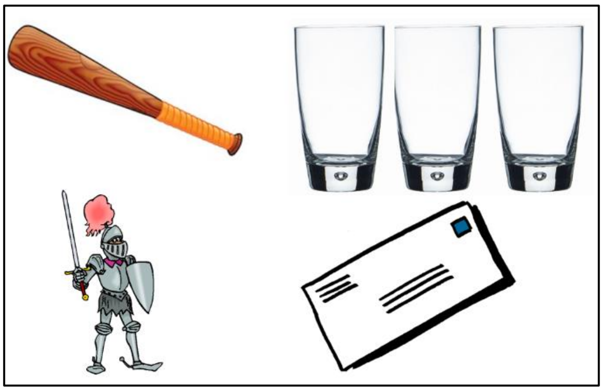
12. *Present picture of Hoover/Vacuum cleaner* "This is **hoover;** can I also call this a **lady?**"

Response: ________________________________ (correct answer: no)

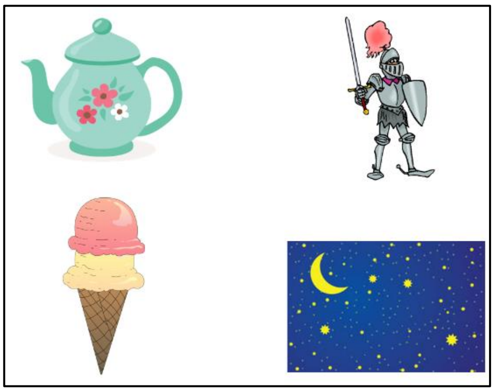
Appendix 15: Time 3 metalinguistic awareness stimuli

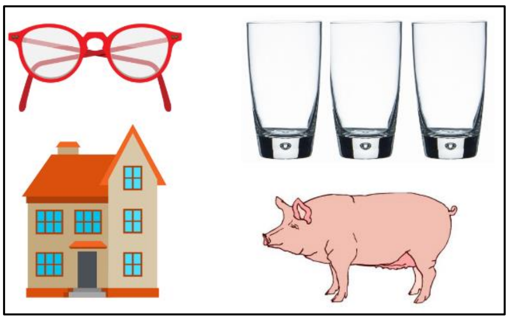
15a: Vocabulary picture cards





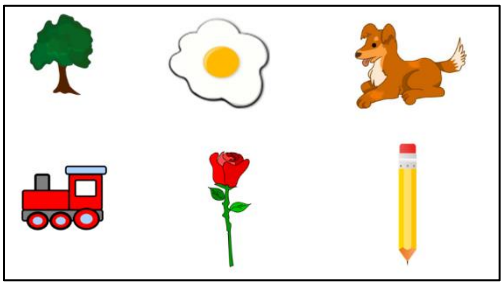
15b: Test picture cards



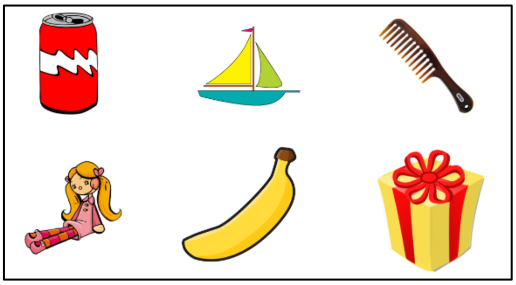


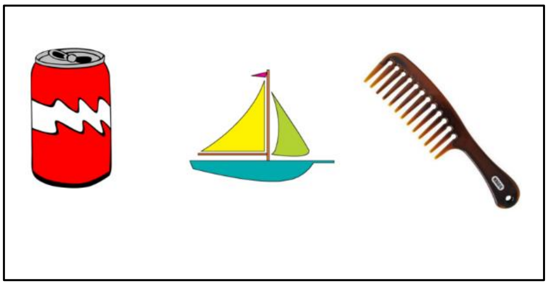
Appendix 16: Metamemory stimuli

16a: Trials 1 and 4 picture card

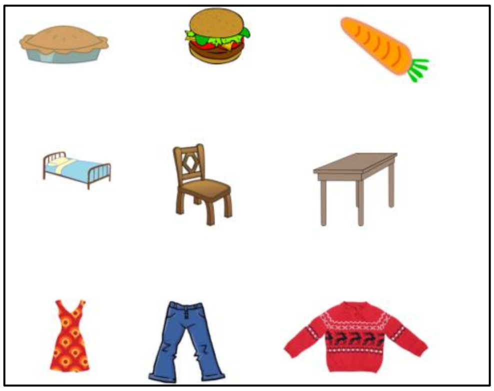


16b: Trial 2 picture cards

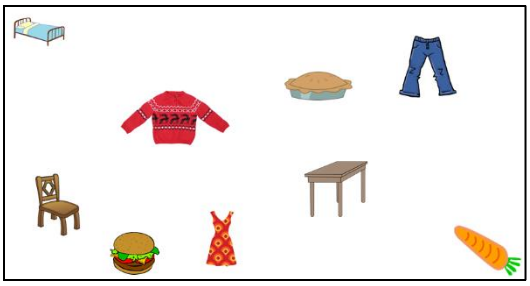




16c: Trial 3 Tom's picture card



16d: Polly's picture card



Appendix 17: Inference making story and answers

Tim had a new pet called Sparky. Sparky was soft, furry, and very playful. At first, Sparky slept indoors in a cardboard box with a nice soft blanket. Sparky soon grew very big. Tim decided to build a kennel and a tall wooden fence around the garden. Tim went to the shop. He already had a hammer and a saw, but he needed some wood and some nails. Tim built the kennel first. His friend Jack helped him to build the fence. Jack held the wood and Tim banged in the nails. The fence was soon finished. Even though Tim's thumb was bruised and sore, he was smiling. He put the hammer that had caused the pain away in his toolbox. He was very pleased with his hard work. That evening, Tim moved Sparky into his new home. But, Sparky did not like his new home. His old cardboard box was still indoors and Sparky missed his nice soft blanket.

Questions

1. What sort of animal was Sparky?
   Answer: Dog
2. What did Tim buy at the shop?
   Answer: wood and nails
3. Who put up the fence?
   Answer: Tim & Jack, Tim & his friend, the man & his friend
4. Why did Tim need a tall fence?
   Answer: because Sparky could jump/so Sparky didn't run away
5. Why did Tim have a sore thumb?
   Answer: banged/hit his thumb with hammer etc.
6. Where was Sparky's kennel?
   Answer: in yard, outside in garden
7. Why did Sparky no longer sleep in the cardboard box?
   Answer: he was too big, he had grown too big, outgrown it
8. Where was Sparky's blanket?
   Answer: (still) in his box, in the house

Appendix 18: Comprehension monitoring stories

Practice story

Today, it is Katie's birthday.
She is six years old.
She is having a party.
Katie is ten today so there are ten candles on her cake.
All of Katie's friends come to her party.

Test stories

Once there was a cat named Bob.
His fur was all brown and as soft as could be. He was very fluffy and had a beautiful tail. All the other cats wished they had his snow white fur. Bob liked to play in Farmer Smith's garden.

Jack has a rabbit called Floppy.
He has had Floppy for 3 years now.
Floppy never goes outside.
Jack feeds Floppy carrots.
Every day Floppy plays in the garden.
Jack really likes Floppy.

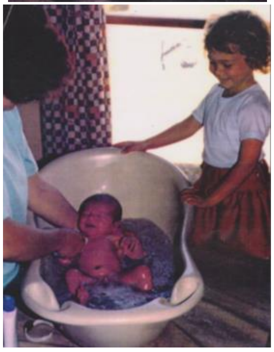
George has planted some sunflower seeds. He got the seeds for his birthday. George waters the garden every day for many months.
In summer he has a row of pretty sunflowers. George wants to plant some poppies next year.

Julie wants to go out to play.
Her dad says that she must tidy her room first.
She folds up her clothes and puts all her toys away. Then she asks dad if she can go out to play. Her dad is very pleased because her room looks so clean. Julie likes to play with her friend Mary.

Tina has just started a new school.
Every day, her dad drives her home in the car. The journey takes about 15 minutes. Tina is always tired after her long walk home.
Tina likes her new school very much.

Appendix 19: Book sharing activity pictures

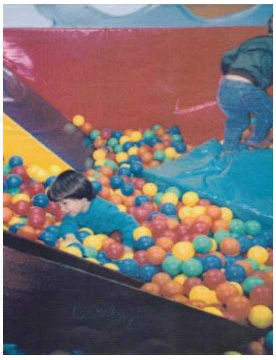
19a: Time 1 book sharing pictures

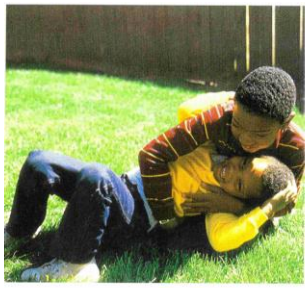


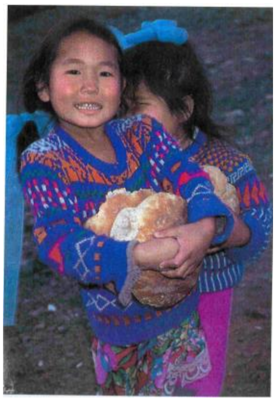


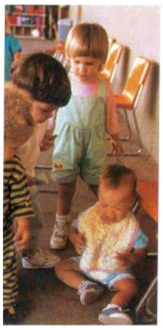




19b: Time 2 book sharing pictures







Appendix 20: Cohort 1 parental instructions for book sharing activity

20a: Apple device instruction

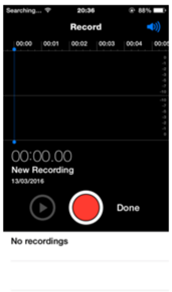
Instructions for Apple phones

The book sharing activity

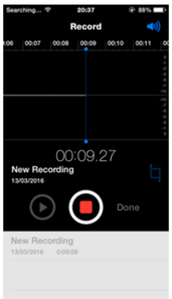
This activity requires you and your child to sit together in a quiet place as you record yourself having a chat about some pictures on your phone or tablet screen. It is important that only you and your child can be heard on the recording so please try and ensure that you are the only two people in the room. Make sure that you are sat closely together so that both of you can be heard clearly. Try and talk about the pictures as you would a picture book at bedtime. You can take as long as you need- there is no minimum or maximum time limit.

Below are some instructions that should help you to download the pictures and make your recording. If you are having problems please either email Sophie at roehamptonbooksharing@hotmail.com or ring or text her.

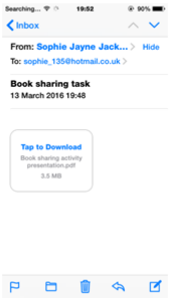
1. Open the voice memo on your phone. You may have other voice recorders on your phone but this one works the best for this activity.



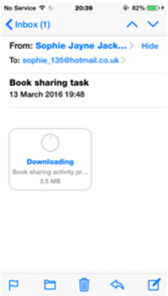
2. Begin recording.



3. Now go to the email that you have been sent by the research team and tap to download the attachment.



4. Wait while the attachment downloads.

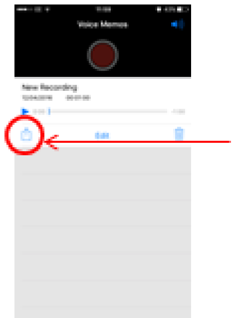


5. Once the attachment has downloaded open it. It is a presentation which includes 10 different pictures.

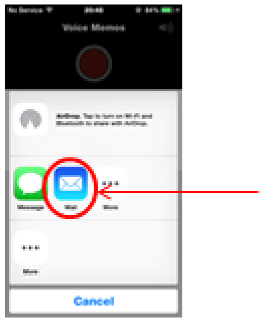
6. You will be able to zoom in on each picture so that only that picture covers your screen. Now you can begin talking about each picture with your child. Remember to talk about them as you would a picture book at bedtime. Move onto the next picture whenever you are ready until you have talked about all ten of the pictures. When moving onto the next picture please say **"next picture"** loudly.

7. When you have finished return to voice memo click done and save as "Book Sharing Activity Recording".

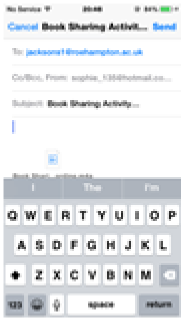
8. To return the recording back to the research team you will need to email them to us. To do so click the share icon.



9. Choose to share via mail.



10. You should then email the recording to Sophie at roehamptonbooksharing@hotmail.com



Once received by Sophie your recording will be saved under an anonymous number (not by your or your child's name) and will only be accessible to members of the research team.

20b: Android device instructions

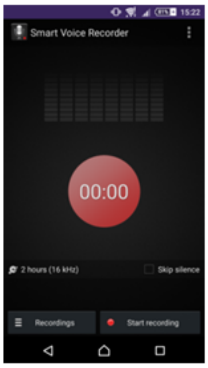
Instructions for Android phones

The book sharing activity

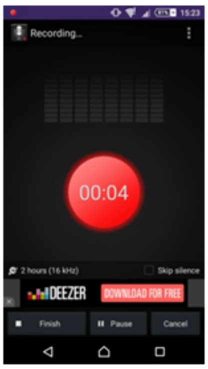
This activity requires you and your child to sit together in a quiet place as you record yourself having a chat about some pictures on your phone or tablet screen. It is important that only you and your child can be heard on the recording so please try and ensure that you are the only two people in the room. Make sure that you are sat closely together so that both of you can be heard clearly. Try and talk about the pictures as you would a picture book at bedtime. You can take as long as you need- there is no minimum or maximum time limit.

Below are some instructions that should help you to download the pictures and make your recording. These screenshots have been taken from a Sony phone and so may look slightly different to your phone, but the general steps will still be the same for all Android phones. If you are having problems please either email Sophie at roehamptonbooksharing@hotmail.com or ring or text her on 07791370506.

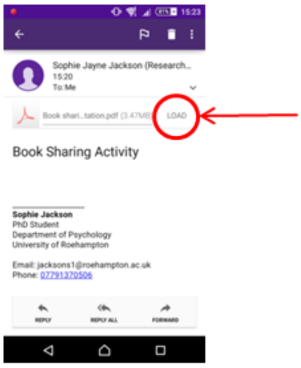
1. Open the voice recorder on your phone. If you do not have a voice recorder you can download an app for free from the Play Store.



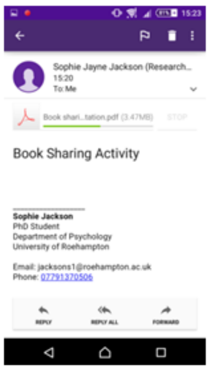
2. Begin recording.



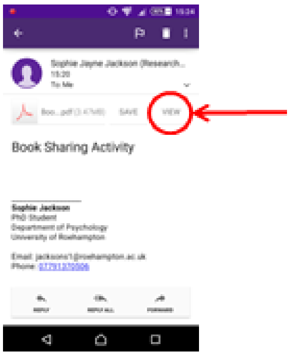
3. Now go to the email that you have been sent by the research team. Press load to download the attachment.



4. Wait while the attachment loads.



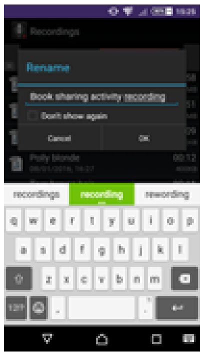
5. Once the attachment has loaded press view to begin viewing it.



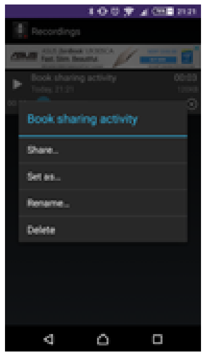
6. This will open up the presentation which includes 10 different pictures.

7. You will be able to zoom in on each picture so that only that picture covers your screen. Now you can begin talking about each picture with your child. Remember to talk about them as you would a picture book at bedtime. Move onto the next picture whenever you are ready until you have talked about all ten of the pictures. When moving onto the next picture please say **"next picture"** loudly.

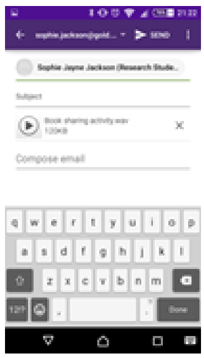
8. When you have finished return to the voice recorder, stop the recording and rename it as "Book Sharing Activity Recording".



9. To return the recording back to the research team you will need to email them to us. To do this click on share and choose to share via email.

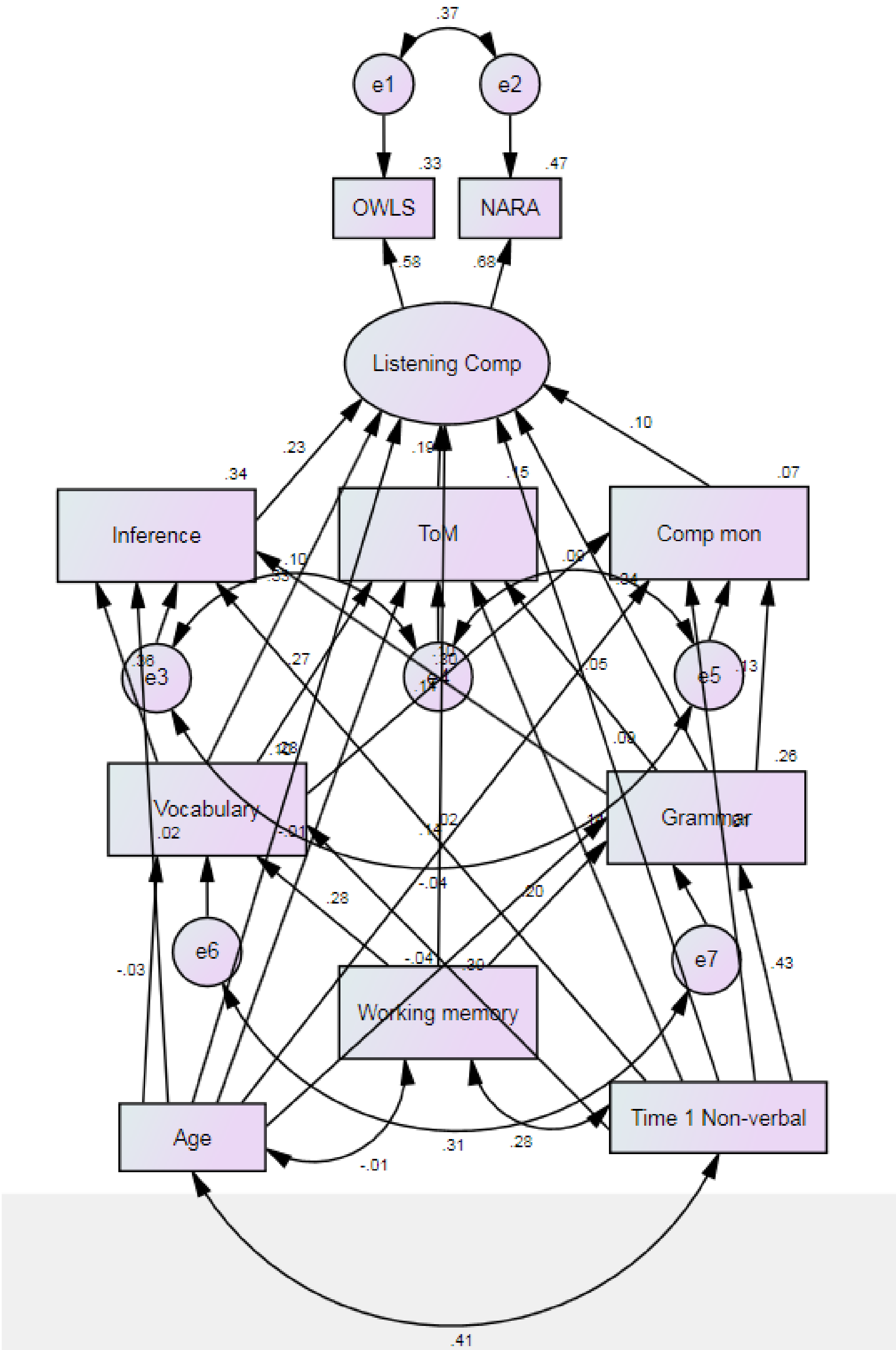


10. You should then email the recording to Sophie at roehamptonbooksharing@hotmail.com



Once received by Sophie your recording will be saved under an anonymous number (not by your or your child's name) and will only be accessible to members of the research team.

Appendix 21: Measurement model for Model 1 from Chapter 3



*All raw models available on request

References

Aaron, P., Joshi, R. M., Gooden, R., & Bentum, K. E. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities, 41*(1), 67-84.

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*(1), 1-16.

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*(9), 933-958.

Adrian, J. E., Clemente, R. A., Villanueva, L., & Rieffe, C. (2005). Parent–child picture-book reading, mothers' mental state language and children's theory of mind. *Journal of Child Language, 32*(03), 673-686.

Adrián, J. E., Clemente, R. A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development, 78*(4), 1052-1067.

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., & Bevens, B. (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Language Disorders, 36*(4), 312-333.

Altman, D. G. (1990). *Practical statistics for medical research*. London: Chapman & Hall.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., . . . Wittrock, M. C. (2001). A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives, abridged edition. *White Plains, NY: Longman,*

Annevirta, T., Laakkonen, E., Kinnunen, R., & Vauras, M. (2007). Developmental dynamics of metacognitive knowledge and text comprehension skill in the first primary school years. *Metacognition and Learning, 2*(1), 21-39.

Arbuckle, J. L. (2016). *Amos 24 user's guide*. USA: Amos Development Corporation.

Arslan, B., Verbrugge, R., Taatgen, N., & Hollebrandse, B. (2015). Teaching children to attribute second-order false beliefs: A training study with feedback. Paper presented at the *Cogsci*, Pasadena, California.

Asadi, I. A. (2018). Reading comprehension subgroups in arabic: A simple but not a multiplicative model. *Reading & Writing Quarterly, 34*(4), 281-290.

Astington, J. W. (2001). The future of Theory-of-Mind research: Understanding motivational states, the role of language, and Real-World consequences. *Child Development, 72*(3), 685-687.

Astington, J. W. (2004). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi, & V. Slaughter (Eds.), *Individual differences in theory of mind implication for typical and atyipical development* (pp. 24-49). New York: Psychology Press.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*(5), 1311.

Atkinson, L. (2014). *A longitudinal investigation of the social, cognitive and social cognitive predictors of reading comprehension* (PhD). Available from EThOS.

Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology, 16,* 225-238.

Avis, J., & Harris, P. L. (1991). Belief-Desire reasoning among baka children: Evidence for a universal conception of mind. *Child Development, 62*(3), 460-467.

Babayiğit, S. (2014). The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of english language. *Journal of Research in Reading, 37*(S1), S22-S47.

Babayiğit, S. (2015). The relations between word reading, oral language, and reading comprehension in children who speak english as a first (L1) and second language (L2): A multigroup structural analysis. *Reading and Writing, 28*(4), 527-544.

Baker, L. (1984a). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology, 76*(4), 588.

Baker, L. (1984b). Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology, 38*(2), 289-311.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 353-394). New Jersey: Routledge.

Baker, L., Dreher, M. J., Shiplet, A. K., Beall, L. C., Voelker, A. N., Garrett, A. J., . . . Finger-Elam, M. (2017). Children's comprehension of informational text: Reading, engaging, and learning. *International Electronic Journal of Elementary Education, 4*(1), 197-227.

Baker, S. T., Leslie, A. M., Gallistel, C., & Hood, B. M. (2016). Bayesian change-point analysis reveals developmental change in a classic theory of mind task. *Cognitive Psychology, 91*, 124-149.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, 30*(2), 285-297.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*(2), 113-125.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46.

Barreto, A. L., Osório, A., Baptista, J., Fearon, P., & Martins, C. (2018). Association between theory of mind and mental state talk in preschoolers and later social competence and behaviour. *Infant and Child Development, 27*(2), e2060.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815-824.

Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*(4), 946-964.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.

Beattie, R. L., & Manis, F. R. (2014). The relationship between prosodic perception, phonological awareness and vocabulary in emergent literacy. *Journal of Research in Reading, 37*(2), 119-137.

Bianco, F., & Lecce, S. (2016). Translating child development research into practice: Can teachers foster children's theory of mind in primary school? *British Journal of Educational Psychology, 86*(4), 592-605.

Bianco, M., Pellenq, C., Lambert, E., Bressoux, P., Lima, L., & Doyen, A. (2012). Impact of early code-skill and oral-comprehension training on reading achievement in first grade. *Journal of Research in Reading, 35*(4), 427-455.

Bingham, G. E. (2007). Maternal literacy beliefs and the quality of mother-child book-reading interactions: Assoications with children's early literacy development. *Early Education and Development, 18*(1), 23-49.

Bishop, D. V. (1989). *Test for the reception of grammar:(TROG)*. Manchester: Medical Research Council.

Bishop, D. V., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain and Behavior, 5*(2), 158-169.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. *Systems Under Indirect Observation: Causality, Structure, Prediction, 2*(3), 149-173.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*(2), 229-242.

Borella, E., & De Ribaupierre, A. (2014). The role of working memory, inhibition, and processing speed in text comprehension in children. *Learning and Individual Differences, 34*, 86-92.

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.

Bosco, F. M., & Gabbatore, I. (2017). Sincere, deceitful, and ironic communicative acts and the role of the theory of mind in childhood. *Frontiers in Psychology, 8*

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., . . . Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry, 49*(4), 422-432.

Bowyer-Crane, C., Fricke, S., Schaefer, B., Lervåg, A., & Hulme, C. (2017). Early literacy and comprehension skills in children learning english as an additional language and monolingual children with language weaknesses. *Reading and Writing, 30*(4), 771-790.

Brand-Gruwel, S., Aarnoutse, C., & Van Den Bos, K. (1998). Improving text comprehension strategies in reading and listening settings. *Learning and Instruction, 8*(1), 63-81.

Braver, S. L., & Bay, R. C. (1992). Assessing and compensating for self-selection bias (non-representativeness) of the family research sample. *Journal of Marriage and the Family, 54*(4), 925-939.

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities, 40*(3), 226-243.

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2008). Theory-of-mind development influences suggestibility and source monitoring. *Developmental Psychology, 44*(4), 1055.

Burgoyne, K., Whiteley, H., & Spooner, A. (2009). The comprehension skills of children learning english as an additional language. *British Journal of Educational Psychology, 79*(4), 735-747.

Burgoyne, K., Duff, F. J., Clarke, P. J., Buckley, S., Snowling, M. J., & Hulme, C. (2012). Efficacy of a reading and language intervention for children with down syndrome: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 53*(10), 1044-1053.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.

Cabell, S. Q., Justice, L. M., Konold, T. R., & McGinty, A. S. (2011). Profiles of emergent literacy skills among preschool children who are at risk for academic difficulties. *Early Childhood Research Quarterly, 26*(1), 1-14.

Cabell, S. Q., Justice, L. M., Logan, J. A., & Konold, T. R. (2013). Emergent literacy profiles among prekindergarten children from low-SES backgrounds: Longitudinal considerations. *Early Childhood Research Quarterly, 28*(3), 608-620.

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., do Céu Cosme, M., & Ribeiro, I. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Reading and Writing, 30*(3), 591-611.

Caillies, S., Hody, A., & Calmus, A. (2012). Theory of mind and irony comprehension in children with cerebral palsy. *Research in Developmental Disabilities, 33*(5), 1380-1388.

Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory, 14*(5), 553-569.

Cain, K. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly, 50*(2), 151-169.

Cain, K. (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research, 60*(5), 1273-1284.

Cain, K., & Bignell, S. (2014). Reading and listening comprehension and their relation to inattention and hyperactivity. *British Journal of Educational Psychology, 84*(1), 108-124.

Cain, K., & Chiu, Y. (2018). The simple view of reading across development: The prediction of grade 3 reading comprehension by prekindergarten skills. *Remedial and Special Education, 39*(5), 289-303.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*(5-6), 489-503.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76*(4), 683-696.

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique, 114*(4), 647-662.

Cain, K., Oakhill, J., Barnes, M. A., & Bryant, P. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*(6), 850-859.

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

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671.

Cain, K., & Yeomans-Maldonado, G. (2017). Development of comprehension monitoring in beginner readers. *Reading and Writing, 30*(9), 2039-2067.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*(5), 431-443.

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., . . . Singh, S. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science, 16*(5), 378-384.

Campbell, R., & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology, 13*(1), 61-68.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language*, *45*(4), 751-774.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*(1), 186-197.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? contributions of inhibitory control and working memory. *Infant and Child Development, 11*(2), 73-92.

Carpendale, J., & Lewis, C. (2006). *How children develop social understanding.* Padstow, Cornwall: Blackwell Publishing.

Carr, A., Slade, L., Yuill, N., Sullivan, S., & Ruffman, T. (2018). Minding the children: A longitudinal study of mental state talk, theory of mind and behavioural adjustment from age 3 to age 10. *Social Development*, *27*(4), 826–840.

Carr, E. M., Dewitz, P., & Patberg, J. P. (1983). The effect of inference training on children's comprehension of expository text. *Journal of Reading Behavior, 15*(3), 1-18.

Carretti, B., Caldarola, N., Tencati, C., & Cornoldi, C. (2014). Improving reading comprehension in reading and listening settings: The effect of two training programmes focusing on metacognition and working memory. *British Journal of Educational Psychology, 84*(2), 194-210.

Carrow-Woolfolk, E. (2011). *Oral and written language scales*. Bloomington: Western Psychological Services.

Cartwright, K. B., Marshall, T. R., Dandy, K. L., & Isaac, M. C. (2010). The development of graphophonological-semantic cognitive flexibility and its contribution to reading comprehension in beginning readers. *Journal of Cognition and Development*, *11*(1), 61-85.

Caruana, E. J., Roman, M., Hernandez-Sanchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease*, *7*(11), 537-540.

Casteel, M. A., & Simpson, G. B. (1991). Textual coherence and the development of inferential generation skills. *Journal of Research in Reading, 14*(2), 116-129.

Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, *28*(9), 1407-1425.

Catts, H. W., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*(2), 278-293.

Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities*, *35*(6), 509-524.

Chard, D., Pikulski, J., & Templeton, S. (2000). From phonemic awareness to fluency: Effective decoding instruction in a research-based reading program. *Current Research in Reading*, *20*, 1-12.

Chen, R., & Vellutino, F. R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research, 29*(1), 1-24.

Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research, 50*(2), 429-443.

Chou, C., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37-55). Thousand Oaks, USA: Sage Publications, Inc.

Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., . . . Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: A latent variable analysis. *Journal of Experimental Psychology: General*, *141*(3), 470.

Clark, N. B., McRoberts, G. W., Van Dyke, J. A., Shankweiler, D. P., & Braze, D. (2012). Immediate memory for pseudowords and phonological awareness are associated in adults and pre-reading children. *Clinical Linguistics & Phonetics, 26*(7), 577-596.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*(8), 1106-1116.

Cohen, J. (Ed.). (1988). *Statistical power analysis for the behavioral sciences. 2nd*. New York: Lawrence Erlbaum Associates.

Cohen, P., West, S. G., & Aiken, L. S. (2014). *Applied multiple regression/correlation analysis for the behavioral sciences*. New Jersey: Psychology Press.

Collins, M. F. (2016). Supporting inferential thinking in preschoolers: Effects of discussion on children's story comprehension. *Early Education and Development, 27*(7), 932-956.

Conners, F. A. (2009). Attentional control and the simple view of reading. *Reading and Writing, 22*(5), 591-613.

Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory-based view of text processing. *Discourse Processes, 26*(2-3), 109-129.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cunningham, A. J., & Carroll, J. M. (2015). Early predictors of phonological and morphological awareness and the link with reading: Evidence from children with different patterns of early deficit. *Applied Psycholinguistics, 36*(3), 509-531.

Cunningham, A. J., Witton, C., Talcott, J. B., Burgess, A. P., & Shapiro, L. R. (2015). Deconstructing phonological tasks: The contribution of stimulus and response type to the prediction of early decoding skills. *Cognition, 143*, 178-186.

Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57-75.

Cutting, & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development, 70*(4), 853-865.

Cutting, Materek, A., Cole, C. A., 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.

Cutting, & 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.

Dabarera, C., Renandya, W. A., & Zhang, L. J. (2014). The impact of metacognitive scaffolding and monitoring on reading comprehension. *System, 42*, 462-473.

Davies, C., Andrés-Roqueta, C., & Norbury, C. F. (2016). Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology, 144*, 98-113.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*(1), 25-31.

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading, 6*(1), 51-77.

Demir, O. E., Applebaum, L., Levine, S. C., Petty, K., & Goldin-Meadow, S. (2011). The story behind parent-child book-reading interactions: Specific relations to later language and reading outcomes. *Proc Annu Boston Univ Conf Lang Dev, 1*, 157-169.

Department for Education (Ed.). (2013). *English programmes of study: Key stages 1 and 2. national curriculum in england*. London: Department for Education.

Department for Education. (2016). *Schools, pupils and their characteristics: January 2016*. London: Department for Education.

Department for Education (Ed.). (2017). *Statutory framework for the early years foundation stage. setting the standards for learning, development and care for children from birth to five*. London: Department for Education.

Devine, R. T., & Hughes, C. (2016a). Family correlates of false belief understanding in early childhood: A Meta-Analysis. *Child Development, 89*(3), 971-987.

Devine, R. T., & Hughes, C. (2016b). Measuring theory of mind across middle childhood: Reliability and validity of the silent films and strange stories tasks. *Journal of Experimental Child Psychology, 149*, 23-40.

Devine, R. T., & Hughes, C. (2017). Let's talk: Parents' mental talk (not Mind-Mindedness or mindreading capacity) predicts children's false belief understanding. *Child Development, 90*(4), 1236-1253.

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology, 52*(5), 758.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.

Diaz, V., & Farrar, M. J. (2017). The missing explanation of the false-belief advantage in bilingual children: A longitudinal study. *Developmental Science, 21*(4), 1-14.

Diaz, V., & Farrar, M. J. (2018). Do bilingual and monolingual preschoolers acquire false belief understanding similarly? the role of executive functioning and language. *First Language, 38*(4), 382-398.

Dilnot, J., Hamilton, L., Maughan, B., & Snowling, M. J. (2016). Child and environmental risk factors predicting readiness for learning in children at high risk of dyslexia. *Development and Psychopathology, 29*(1), 235-244.

Dimitrov, D. M. (2006). Comparing groups on latent variables: A structural equation modeling approach. *Work, 26*(4), 429-436.

Dispaldro, M., Deevy, P., Altoé, G., Benelli, B., & Leonard, L. B. (2011). A cross-linguistic study of real-word and non-word repetition as predictors of grammatical competence in children with typical language development. *International Journal of Language & Communication Disorders, 46*(5), 564-578.

Dockrell, J. E., Connelly, V., & Arfè, B. (2019). Struggling writers in elementary school: Capturing drivers of performance. *Learning and Instruction, 60,* 75-84.

Doherty, M. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language, 27*(2), 367-392.

Doherty, M. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Hove, East Sussex: Psychology Press.

Doherty, M., & Itakura, S. (1995). Metalinguistic awareness and theory of mind: A study from japan. *Bulletin of Oita Prefectural College of Arts and Culture, 33*, 13-24.

Doherty, M., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development, 13*(3), 279-305.

Dore, R. A., Amendum, S. J., Golinkoff, R. M., & Hirsh-Pasek, K. (2018). Theory of mind: A hidden factor in reading comprehension? *Educational Psychology Review, 30*(3), 1067–1089.

Dreyer, L., & Katz, L. (1992). An examination of “the simple view of reading.”. *Haskins Laboratories Status Report on Speech Research, 112*, 161-166.

Duff, F. J., Hulme, C., Grainger, K., Hardwick, S. J., Miles, J. N., & Snowling, M. J. (2014). Reading and language intervention for children at risk of dyslexia: A randomised controlled trial. *Journal of Child Psychology and Psychiatry, 55*(11), 1234-1243.

Duff, F. J., Mengoni, S. E., Bailey, A. M., & Snowling, M. J. (2015). Validity and sensitivity of the phonics screening check: Implications for practice. *Journal of Research in Reading, 38*(2), 109-123.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848-856.

Dunn, J., Brown, J., & Beardsall, L. (1991). Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology, 27*(3), 448.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62*(6), 1352-1366.

Dunn, J., Cutting, A. L., & Fisher, N. (2002). Old friends, new friends: Predictors of children’s perspective on their friends at school. *Child Development, 73*(2), 621-635.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British picture vocabulary scale* (2nd ed.). London: NFER-Nelson.

Dunn, L., Dunn, D., & National Foundation for Educational Research (Eds.). (2009). *The british picture vocabulary scale (3rd ed)*. London: GL Assessment Limited.

Ebert, S. (2015). Longitudinal relations between theory of mind and metacognition and the impact of language. *Journal of Cognition and Development, 16*(4), 559-586.

Ebert, S., Peterson, C., Slaughter, V., & Weinert, S. (2017). Links among parents' mental state language, family socioeconomic status, and preschoolers' theory of mind development. *Cognitive Development, 44*, 32-48.

Einarsdóttir, J. (2007). Research with children: Methodological and ethical challenges. *European Early Childhood Education Research Journal, 15*(2), 197-211.

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.

Emery, D. W. (1996). Helping readers comprehend stories from the characters' perspectives. *The Reading Teacher, 49*(7), 534-541.

Ensor, R., Devine, R. T., Marks, A., & Hughes, C. (2014). Mothers' cognitive references to 2-Year-Olds predict theory of mind at ages 6 and 10. *Child Development, 85*(3), 1222-1235.

Ereky-Stevens, K. (2008). Associations between mothers' sensitivity to their infants' internal states and children's later understanding of mind and emotion. *Infant and Child Development, 17*(5), 527-543.

Evans, A. D. (2005). *Why do young children forget where they learned information? the relation between source monitoring, theory-of-mind understanding and suggestibility* (Masters).

Faden, V. B., Day, N. L., Windle, M., Windle, R., Grube, J. W., Molina, B. S., . . . Jackson, K. M. (2004). Collecting longitudinal data through childhood, adolescence, and young adulthood: Methodological challenges. *Alcoholism: Clinical and Experimental Research, 28*(2), 330-340.

Fargas-Malet, M., McSherry, D., Larkin, E., & Robinson, C. (2010). Research with children: Methodological issues and innovative techniques. *Journal of Early Childhood Research, 8*(2), 175-192.

Farina, E., Albanese, O., & Pons, F. (2007). Making inferences and individual differences in emotion understanding. *Psychology of Language and Communication, 11*(2), 3-19.

Farrant, B. M., Maybery, M. T., & Fletcher, J. (2012). Language, cognitive flexibility, and explicit false belief understanding: Longitudinal analysis in typical development and specific language impairment. *Child Development, 83*(1), 223-235.

Farrar, M. J., & Maag, L. (2002). Early language development and the emergence of a theory of mind. *First Language, 22*(2), 197-213.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Lawrence Erlbaum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906-911.

Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris & D. R. Olson (Eds.), *Developing theories of mind* (pp. 244-267). Cambridge: Cambridge University Press.

Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development, 24*(1), 15-23.

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology, 1*(4), 324-340.

Florit, E., Levorato, M. C., & Roch, M. (2008). Verba volant, scripta manent. developmental changes in reading and listening comprehension. *Giornale Italiano Di Psicologia, 35*(3), 641-664.

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review, 23*(4), 553-576.

Florit, E., & Levorato, M. C. (2012). A longitudinal study on listening text comprehension and receptive vocabulary in preschoolers. *Rivista Di Psicolinguistica Applicata, 12*(1/2), 65-80.

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.

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.

Florit, E., Roch, M., & Levorato, M. C. (2013). The relationship between listening comprehension of text and sentences in preschoolers: Specific or mediated by lower and higher level components? *Applied Psycholinguistics, 34*(2), 395-415.

Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in kindergarten through grade 2. *Reading and Writing, 28*(5), 655-681.

Forum for Research into Language and Literacy. (2012). *Diagnostic test of word reading processes (DTWRP)*. London: GL Assessment.

Freed, J., & Cain, K. (2017). Assessing school-aged children's inference-making: The effect of story test format in listening comprehension. *International Journal of Language & Communication Disorders, 52*(1), 95-105.

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280-290.

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.

Gao, S., Mokhtarian, P., & Johnston, R. (2008). Nonnormality of data in structural equation models. *Transportation Research Record: Journal of the Transportation Research Board, 2082*(1), 116-124.

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in english. *Review of Educational Research, 84*(1), 74-111.

Garner, P. W., Jones, D. C., Gaddy, G., & Rennie, K. M. (1997). Low-Income mothers' conversations about emotions and their children's emotional competence. *Social Development, 6*(1), 37-52.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200-213.

George, D., & Mallery, P. (Eds.). (2003). *SPSS for windows step by step: A simple guide and reference*. Boston: Pearson Education.

Georgiou, G. K., Das, J. P., & Hayward, D. (2009). Revisiting the "simple view of reading" in a group of children with poor reading comprehension. *Journal of Learning Disabilities, 42*(1), 76-84.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486-489.

Gillam, R. B., & Pearson, N. A. (2004). *TNL: Test of narrative language*. Austin: Pro-ed.

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.

Goh, C., & Taib, Y. (2006). Metacognitive instruction in listening for young learners. *ELT Journal*, *60*(3), 222-232.

Golinkoff, R. M., & Ames, G. J. (1979). A comparison of fathers' and mothers' speech with their young children. *Child Development, 50*(1), 28-32.

Gooch, D., Hulme, C., Nash, H. M., & Snowling, M. J. (2014). Comorbidities in preschool children at family risk of dyslexia. *Journal of Child Psychology and Psychiatry, 55*(3), 237-246.

Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., & Hulme, C. (2016). The development of executive function and language skills in the early school years. *Journal of Child Psychology and Psychiatry, 57*(2), 180-187.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*(1), 26-37.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development, 59*(5), 1366-1371.

Gottardo, A., Mirza, A., Koh, P. W., Ferreira, A., & Javier, C. (2018). Unpacking listening comprehension: The role of vocabulary, morphological awareness, and syntactic knowledge in reading comprehension. *Reading and Writing, 31*(8), 1741-1764.

Gough, P., Hoover, W., Peterson, C., Cornoldi, C., & Oakhill, J. (1996a). Some observations on a simple view of reading. In C. Cornoldi, & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, N.J: Lawrence Erlbaum Associates.

Gough, P., Hoover, W. A., Peterson, C. L., Cornoldi, C., & Oakhill, J. (1996b). Some observations on a simple view of reading. In C. Cornoldi, & J. Oakhill (Ed.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah: Lawrence Erlbaum Associates.

Gough, P., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371.

Grammer, J. K., Coffman, J. L., Ornstein, P. A., & Morrison, F. J. (2013). Change over time: Conducting longitudinal studies of children's cognitive development. *Journal of Cognition and Development, 14*(4), 515-528.

Gravetter, F. J., & Wallnau, L. B. (2010). *Essentials of statistics for the behavioral sciences*. Belmont: Cengage Learning.

Grazzani, I., & Ornaghi, V. (2011). Emotional state talk and emotion understanding: A training study with preschool children. *Journal of Child Language, 38*(5), 1124-1139.

Guajardo, N. R., & Cartwright, K. B. (2016). The contribution of theory of mind, counterfactual reasoning, and executive function to pre-readers’ language comprehension and later reading awareness and comprehension in elementary school. *Journal of Experimental Child Psychology, 144*, 27-45.

Guo, Y., Piasta, S. B., & Bowles, R. P. (2015). Exploring preschool children's science content knowledge. *Early Education and Development, 26*(1), 125-146.

Gustafson, S., Samuelsson, C., Johansson, E., & Wallmann, J. (2013). How simple is the simple view of reading? *Scandinavian Journal of Educational Research, 57*(3), 292-308.

Gwon, E., & Lee, H. (2012). Children’s development of lying, false belief and executive function. *The Korean Journal of Developmental Psychology, 25*, 165-184.

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing, 16*(6), 505-539.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (Eds.). (2006). *Multivariate data analysis*. Harlow: Cengage Learning.

Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61*(3), 653-663.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449.

Hamilton, L., Hayiou-Thomas, M. E., Hulme, C., & Snowling, M. J. (2016). The home literacy environment as a predictor of the early literacy development of children at family-risk of dyslexia. *Scientific Studies of Reading, 20*(5), 401-419.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a british communicative development inventory. *Journal of Child Language, 27*(03), 689-705.

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*(2), 129-154.

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843-855.

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition & Emotion, 3*(4), 379-400.

Hayashi, H. (2007). Young children’s understanding of second-order mental states. *Psychologia, 50*(1), 15-25.

Hayes, A. F. (2009). Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408-420.

Hayes, A. F. (2012). My macros and code for SPSS and SAS. Retrieved from http://afhayes.com/spss-sas-and-mplus-macros-and-code.html?m

He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: Evidence from violation-of-expectation change-of-location and unexpected-contents tasks. *Developmental Science, 14*(2), 292-305.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1357-1364.

Highman, C., Leitão, S., Hennessey, N., & Piek, J. (2012). Prelinguistic communication development in children with childhood apraxia of speech: A retrospective analysis. *International Journal of Speech-Language Pathology, 14*(1), 35-47.

Ho, C. S., Chow, B. W., Wong, S. W., Waye, M. M., & Bishop, D. V. (2012). The genetic and environmental foundation of the simple view of reading in chinese. *PloS One, 7*(10), e47872.

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology, 16*(3), 199-207.

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing Services in Schools, 36*(4), 285-293.

Hogrefe, G., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*(3), 567-582.

Hoien-tengesdal, I. (2010). Is the simple view of reading too simple? *Scandinavian Journal of Educational Research, 54*(5), 451-469.

Holliman, A. J., Wood, C., & Sheehy, K. (2010). Does speech rhythm sensitivity predict children's reading ability 1 year later? *Journal of Educational Psychology, 102*(2), 356.

Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology, 100*(2), 252.

Hoover, W. A., & Gough, P. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160.

Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education, 39*(5), 304-312.

Howard, A. A., Mayeux, L., & Naigles, L. R. (2008). Conversational correlates of children's acquisition of mental verbs and a theory of mind. *First Language, 28*(4), 375-402.

Hu, L., & Bentler, P. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Hughes, C. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*(2), 233-253.

Hughes, C. (1998b). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34*(6), 1326-1339.

Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good Test—Retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry, 41*(04), 483-490.

Hughes, C., & Devine, R. T. (2015). A social perspective on theory of mind. In R. Lerner (Ed.), *Handbook of child psychology and developmental science* (pp. 1-46). New Jersey: Wiley.

Hughes, C., & Dunn, J. (1997). "Pretend you didn't know": Preschoolers' talk about mental states in pretend play. *Cognitive Development, 12*(4), 477-497.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology, 34*(5), 1026.

Hughes, C., Ensor, R., & Marks, A. (2011). Individual differences in false belief understanding are stable from 3 to 6 years of age and predict children's mental state talk with school friends. *Journal of Experimental Child Psychology, 108*(1), 96-112.

Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture? *Child Development, 76*(2), 356-370.

Hughes, C., Lecce, S., & Wilson, C. (2007). "Do you know what I want?" preschoolers' talk about desires, thoughts and feelings in their conversations with sibs and friends. *Cognition and Emotion, 21*(2), 330-350.

Hughes, C., Marks, A., Ensor, R., & Lecce, S. (2010). A longitudinal study of conflict and inner state talk in children's conversations with mothers and younger siblings. *Social Development, 19*(4), 822-837.

Hughes, C., White, A., Sharpen, J., & Dunn, J. (2000). Antisocial, angry, and unsympathetic:“Hard-to-manage” preschoolers' peer problems and possible cognitive influences. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(2), 169-179.

Hulme, C., Stothard, S., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Snowling, M. J. (2009). *YARC: York assessment of reading for comprehension early reading*. York: GL Publishers.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science, 26*(12), 1877-1886.

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science, 23*(6), 572-577.

Hutchins, T. L., Prelock, P. A., & Chace, W. (2008). Test-retest reliability of a theory of mind task battery for children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 23*(4), 195-206.

Iao, L., Leekam, S., Perner, J., & McConachie, H. (2011). Further evidence for nonspecificity of theory of mind in preschoolers: Training and transferability in the understanding of false beliefs and false signs. *Journal of Cognition and Development, 12*(1), 56-79.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology, 32*(1), 70.

Jenkins, J. M., Turrell, S. L., Kogushi, Y., Lollis, S., & Ross, H. S. (2003). A longitudinal investigation of the dynamics of mental state talk in families. *Child Development, 74*(3), 905-920.

Jiang, H., & Farquharson, K. (2018). Are working memory and behavioral attention equally important for both reading and listening comprehension? A developmental comparison. *Reading and Writing, 31*(7), 1449-1477.

Johnson, G. (2015). The invention of reading and the evolution of text. *Journal of Literacy and Technology, 16*(1), 107-128.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing, 19*(4), 339-361.

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology, 166*, 190-211.

Joshi, R. M., & Aaron, P. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21*(2), 85-97.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437.

Juel, C., Griffith, P. L., & Gough, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243.

Justice, L. M., Kaderavek, J. N., Fan, X., Sofka, A., & Hunt, A. (2009). Accelerating preschoolers' early literacy development through classroom-based teacher–child storybook reading and explicit print referencing. *Language, Speech, and Hearing Services in Schools, 40*(1), 67-85.

Kang, J. Y. (2012). Do bilingual children possess better phonological awareness? investigation of korean monolingual and korean-english bilingual children. *Reading and Writing, 25*(2), 411-431.

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading, 31*(3), 259-272.

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.

Kendeou, P., Papadopoulos, T. C., & Kotzapoulou, M. (2013). Evidence for the early emergence of the simple view of reading in a transparent orthography. *Reading and Writing, 26*(2), 189-204.

Kendeou, P., Van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765.

Kieffer, M. J., Vukovic, R. K., & Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Reading Research Quarterly, 48*(4), 333-348.

Kim, Y. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development, 86*(1), 128-144.

Kim, Y., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly, 49*(3), 269-281.

Kim, Y. (2011). Proximal and distal predictors of reading comprehension: Evidence from young korean readers. *Scientific Studies of Reading, 15*(2), 167-190.

Kim, Y. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology, 141*, 101-120.

Kim, Y. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading, 21*(7), 310-333.

Kim, Y., Park, C. H., & Wagner, R. K. (2014). Is oral/text reading fluency a “bridge” to reading comprehension? *Reading and Writing, 27*(1), 79-99.

Kim, Y., & Petscher, Y. (2016). Prosodic sensitivity and reading: An investigation of pathways of relations using a latent variable approach. *Journal of Educational Psychology, 108*(5), 630.

Kim, Y., & Pilcher, H. (2016). What is listening comprehension and what does it take to improve listening comprehension? In R. Schiff, & R. M. Joshi (Eds.), *Interventions in learning disabilities: A handbook on systematic training programs for individuals with learning disabilities* (pp. 159-173). Berlin: Springer.

Kim, Y., & Wagner, R. K. (2015). Text (oral) reading fluency as a construct in reading development: An investigation of its mediating role for children from grades 1 to 4. *Scientific Studies of Reading, 19*(3), 224-242.

Kinnunen, R., Vauras, M., & Niemi, P. (1998). Comprehension monitoring in beginning readers. *Scientific Studies of Reading, 2*(4), 353-375.

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

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294-303.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy, 42*(2), 75-82.

Kirk, E., Pine, K., Wheatley, L., Howlett, N., Schulz, J., & Fletcher, B. C. (2015). A longitudinal investigation of the relationship between maternal mind-mindedness and theory of mind. *British Journal of Developmental Psychology, 33*(4), 434-445.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*(6012), 1830-1834.

LaBounty, J., Wellman, H. M., Olson, S., Lagattuta, K., & Liu, D. (2008). Mothers' and fathers' use of internal state talk with their young children. *Social Development, 17*(4), 757-775.

Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. *Cognition & Emotion, 9*(2-3), 167-185.

Language and Reading Research Consortium. (2015). The dimensionality of language ability in young children. *Child Development, 86*(6), 1948-1965.

Language and Reading Research Consortium, & Muijselaar, M. M. (2018). The dimensionality of inference making: Are local and global inferences distinguishable? *Scientific Studies of Reading, 22*(2), 117-136.

Laranjo, J., Bernier, A., Meins, E., & Carlson, S. M. (2014). The roles of maternal mind-mindedness and infant security of attachment in predicting preschoolers' understanding of visual perspective taking and false belief. *Journal of Experimental Child Psychology, 125*, 48-62.

Lavoie, J., Leduc, K., Arruda, C., Crossman, A. M., & Talwar, V. (2017). Developmental profiles of children's spontaneous lie-telling behavior. *Cognitive Development, 41*, 33-45.

Lecce, S., & Bianco, F. (2018). The role of false-belief understanding in preschoolers' development of metamemory: A training study. *European Journal of Developmental Psychology, 16*(6), 697-711.

Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014). Training preschoolers on first-order false belief understanding: Transfer on advanced ToM skills and metamemory. *Child Development, 85*(6), 2404-2418.

Lecce, S., Bianco, F., Devine, R. T., Hughes, C., & Banerjee, R. (2014). Promoting theory of mind during middle childhood: A training program. *Journal of Experimental Child Psychology, 126*, 52-67.

Lecce, S., Caputi, M., & Pagnin, A. (2014). Long-term effect of theory of mind on school achievement: The role of sensitivity to criticism. *European Journal of Developmental Psychology, 11*(3), 305-318.

Lecce, S., Zocchi, S., Pagnin, A., Palladino, P., & Taumoepeau, M. (2010). Reading minds: The relation between children's mental state knowledge and their metaknowledge about reading. *Child Development*, *81*(6), 1876-1893.

Lei, P., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice, 26*(3), 33-43.

Lepola, J., Lynch, J., 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), 259-282.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A norwegian latent variable longitudinal study. *Developmental Psychology, 45*(3), 764.

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development, 89*(5), 1821-1838.

Lewis, C., & Osborne, A. (1990). Three-Year-Olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development, 61*(5), 1514-1519.

Lillard, A. S., & Flavell, J. H. (1992). Young children's understanding of different mental states. *Developmental Psychology, 28*(4), 626.

Lloyd, P., Boada, H., & Forns, M. (1992). New directions in referential communication research. *British Journal of Developmental Psychology, 10*(4), 385-403.

Lockl, K., & Schneider, W. (2007). Knowledge about the mind: Links between theory of mind and later metamemory. *Child Development, 78*(1), 148-167.

Logan, J. (2017). Pressure points in reading comprehension: A quantile multiple regression analysis. *Journal of Educational Psychology, 109*(4), 451.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596.

Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the simple view of reading with elementary school children: Still simple after all these years. *Remedial and Special Education, 39*(5), 260-273.

Lonigan, C. J., Purpura, D. J., Wilson, S. B., Walker, P. M., & Clancy-Menchetti, J. (2013). Evaluating the components of an emergent literacy intervention for preschool children at risk for reading difficulties. *Journal of Experimental Child Psychology, 114*(1), 111-130.

Lundy, B. L. (2013). Paternal and maternal mind-mindedness and preschoolers' theory of mind: The mediating role of interactional attunement. *Social Development, 22*(1), 58-74.

Luoni, C., Balottin, U., Rosana, L., Savelli, E., Salini, S., & Termine, C. (2015). Confrontation naming and reading abilities at primary school: A longitudinal study. *Behavioural Neurology, 10*

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.

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.

MacLeod, C. M. (2007). The concept of inhibition in cognition. In D. S. Gorfein, & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 3-23). Washington, DC: American Psychological Association.

Mahon, M., & Crutchley, A. (2006). Performance of typically-developing school-age children with english as an additional language on the british picture vocabulary scales II. *Child Language Teaching and Therapy, 22*(3), 333-351.

Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin–English bilingual children. *Clinical Linguistics & Phonetics, 24*(4-5), 387-400.

Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. *Journal of Speech, Language, and Hearing Research, 44*(4), 925-940.

Massonnié, J., Bianco, M., Lima, L., & Bressoux, P. (2018). Longitudinal predictors of reading comprehension in french at first grade: Unpacking the oral comprehension component of the simple view. *Learning and Instruction, 60*, 166-179.

McBride–Chang, C., & Kail, R. V. (2002). Cross–cultural similarities in the predictors of reading acquisition. *Child Development, 73*(5), 1392-1407.

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly, 28*(2), 314-324.

McGee, A., & Johnson, H. (2003). The effect of inference training on skilled and less skilled comprehenders. *Educational Psychology, 23*(1), 49-59.

Meins, E., Fernyhough, C., Arnott, B., Leekam, S. R., & Rosnay, M. (2013). Mind-Mindedness and theory of mind: Mediating roles of language and perspectival symbolic play. *Child Development, 84*(5), 1777-1790.

Meins, E., Fernyhough, C., Wainwright, R., Clark-Carter, D., Das Gupta, M., Fradley, E., & Tuckey, M. (2003). Pathways to understanding mind: Construct validity and predictive validity of maternal mind-mindedness. *Child Development, 74*(4), 1194-1211.

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind–mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*(6), 1715-1726.

Metsala, J. L., & Ehri, L. C. (2013). *Word recognition in beginning literacy*. New York: Routledge.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622-646.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Colorado: Cambridge University Press.

Moeller, M. P., & Schick, B. (2006). Relations between maternal input and theory of mind understanding in deaf children. *Child Development, 77*(3), 751-766.

Mokhtari, K., & Niederhauser, D. S. (2012). Vocabulary and syntactic knowledge factors in 5th grade students' reading comprehension. *Journal of Elementary Education, 5*(2), 157-170.

Morrison, T. G., Morrison, M. A., & McCutcheon, J. M. (2017). Best practice recommendations for using structural equation modelling in psychological research. *Psychology, 8*(9), 1326.

Muijselaar, M. M., Swart, N. M., Steenbeek-Planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). Developmental relations between reading comprehension and reading strategies. *Scientific Studies of Reading, 21*(3), 194-209.

Müller, K., & Brady, S. (2001). Correlates of early reading performance in a transparent orthography. *Reading and Writing, 14*(7-8), 757-799.

Muter, V., Hulme, C., & Snowling, M. J. (1997). *The phonological abilities test*. London: The Psychological Corporation.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology, 72*(4), 549-560.

Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 36*(7), 911-919.

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*(9), 1031-1039.

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.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development, 69*(4), 996-1011.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342-356.

Neale, M. D. (1999). *Neale analysis of reading ability: Manual*. London: GL Assessments.

Nelson, K. (2004). Toward a collaborative community of minds. *Behavioral and Brain Sciences, 27*(1), 119-120.

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41*(2), 428-442.

Nilsen, E. S., & Fecica, A. M. (2011). A model of communicative perspective-taking for typical and atypical populations of children. *Developmental Review, 31*(1), 55-78.

Nouwens, S., Groen, M. A., & Verhoeven, L. (2017). How working memory relates to children’s reading comprehension: The importance of domain-specificity in storage and processing. *Reading and Writing, 30*(1), 105-120.

Nunnally, J. C., & Bernstein, I. H. (1967). *Psychometric theory*. New York: McGraw-Hill.

O’Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of Autism and Developmental Disorders, 39*(6), 916-928.

Oakhill, J., & Cain, K. (2007). Introduction to comprehension development. In K. Cain, & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perscepctive* (pp. 3-40). New York: Guilford Press.

Oakhill, J., & 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.

Oakhill, J., Cain, K., & Bryant, P. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18*(4), 443-468.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing, 18*(7), 657-686.

Oh, E. (2016). Comparative studies on the roles of linguistic knowledge and sentence processing speed in L2 listening and reading comprehension in an EFL tertiary setting. *Reading Psychology, 37*(2), 257-285.

O'Neill, D. K., & Gopnik, A. (1991). Young children's ability to identify the sources of their beliefs. *Developmental Psychology, 27*(3), 390-397.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-258.

Ontai, L. L., & Thompson, R. A. (2008). Attachment, parent–child discourse and theory-of-mind development. *Social Development, 17*(1), 47-60.

Ornaghi, V., Grazzani, I., Cherubin, E., Conte, E., & Piralli, F. (2015). 'Let's talk about emotions!'. the effect of conversational training on preschoolers' emotion comprehension and prosocial orientation. *Social Development, 24*(1), 166-183.

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, *23*(2), 189-208.

Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development, 12*(2), 199-230.

Paris, S. G., & Myers, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. *Journal of Reading Behavior, 13*(1), 5-22.

Paris, S. G., & Stahl, S. A. (2005). *Children's reading comprehension and assessment*. Mahwah, New Jersey: Routledge.

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8*(1), 3-26.

Paslawski, T. (2005). The clinical evaluation of language fundamentals (CELF-4): A review. *Canadian Journal of School Psychology, 20*(1/2), 129-134.

Pears, K. C., & Moses, L. J. (2003). Demographics, parenting, and theory of mind in preschool children. *Social Development, 12*(1), 1-20.

Pelletier, J., & Astington, J. (2004). Action, consciousness and theory of mind: Children's ability to coordinate story characters' actions and thoughts. *Early Education and Development, 15*(1), 5-22.

Pelletier, J., & Beatty, R. (2015). Children's understanding of aesop's fables: Relations to reading comprehension and theory of mind. *Frontiers in Psychology, 6*, 1448.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling, C. Hulme (Ed.), *The science of reading: A handbook* (Eds ed., pp. 227). Malden: Blackwell Publishing.

Perner, J. (1991). *Understanding the representational mind.* Cambridge: The MIT Press.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*(4), 1228-1238.

Perner, J., & Wimmer, H. (1985). "John thinks that mary thinks that…" attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology, 39*(3), 437-471.

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or asperger syndrome. *Child Development, 83*(2), 469-485.

Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development, 18*(3), 399-429.

Piasta, S. B., Groom, L. J., Khan, K. S., Skibbe, L. E., & Bowles, R. P. (2018). Young children's narrative skill: Concurrent and predictive associations with emergent literacy and early word reading skills. *Reading and Writing, 31*(7), 1479-1498.

Pitts, M. M. (1983). Comprehension monitoring: Definition and practice. *Journal of Reading*, *26*(6), 516-523.

Poe, M. D., Burchinal, M. R., & Roberts, J. E. (2004). Early language and the development of children's reading skills. *Journal of School Psychology, 42*(4), 315-332.

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.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native spanish-speaking children reading in english: Toward a model of comprehension. *Journal of Educational Psychology, 97*(2), 246.

Puglisi, M. L., Hulme, C., Hamilton, L. G., & Snowling, M. J. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children’s language and literacy development. *Scientific Studies of Reading, 21*(6), 498-514.

Resches, M., & Pereira, M. P. (2007). Referential communication abilities and theory of mind development in preschool children. *Journal of Child Language, 34*(1), 21-52.

Reynolds, R. E. (2000). Attentional resource emancipation: Toward understanding the interaction of word identification and comprehension processes in reading. *Scientific Studies of Reading, 4*(3), 169-195.

Ricketts, J. (2011). Research review: Reading comprehension in developmental disorders of language and communication. *Journal of Child Psychology and Psychiatry, 52*(11), 1111-1123.

Ricketts, J., Davies, R., Masterson, J., Stuart, M., & Duff, F. J. (2016). Evidence for semantic involvement in regular and exception word reading in emergent readers of english. *Journal of Experimental Child Psychology, 150*, 330-345.

Ricketts, J., Jones, C. R., Happé, F., & Charman, T. (2013). Reading comprehension in autism spectrum disorders: The role of oral language and social functioning. *Journal of Autism and Developmental Disorders, 43*(4), 807-816.

Riggio, M. M., & Cassidy, K. W. (2009). Preschoolers' processing of false beliefs within the context of picture book reading. *Early Education and Development, 20*(6), 992-1015.

Rose, J. (2005). *Independent review of the teaching of early reading: Interim report* Department for education and skills.

Rose, J. (2009). Identifying and teaching children and young people with dyslexia and literacy difficulties: An independent report.

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research, 95*(5), 259-272.

Ruffman, T. (2014). To belief or not belief: Children’s theory of mind. *Developmental Review, 34*(3), 265-293.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers’ mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734-751.

Ruffman, T., Slade, L., Devitt, K., & Crowe, E. (2006). What mothers say and what they do: The relation between parenting, theory of mind, language and conflict/cooperation. *British Journal of Developmental Psychology, 24*(1), 105-124.

Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development, 18*(2), 139-158.

Savage, R. (2001). The'simple view'of reading: Some evidence and possible implications. *Educational Psychology in Practice, 17*(1), 17-33.

Savage, R. (2006). Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading, 10*(2), 143-164.

Savage, R., Cornish, K., Manly, T., & Hollis, C. (2006). Cognitive processes in children's reading and attention: The role of working memory, divided attention, and response inhibition. *British Journal of Psychology, 97*(3), 365-385.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*(1), 3-15.

Schmitt, N. (2000). *Vocabulary in language teaching*. New York: Cambridge University Press.

Scrimin, S., Patron, E., Florit, E., Palomba, D., & Mason, L. (2017). The role of cardiac vagal tone and inhibitory control in pre-schoolers' listening comprehension. *Developmental Psychobiology, 59*(8), 970-975.

Seigneuric, A., & Ehrlich, M. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing, 18*(7-9), 617-656.

Semel, E., Wiig, E., & Secord, W. (2004). *CELF 4: Clinical evaluation of language fundamentals 4 screening test*. San Antonio: Pearson: Psychological Corporation.

Semel, E., Wiig, E., & Secord, W. (2006). *CELF 4: Clinical evaluation of language fundamentals*. San Antonio: Pearson: Psychological Corporation.

Sénéchal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445-460.

Sénéchal, M., Pagan, S., Lever, R., & Ouellette, G. (2008). Relations among the frequency of shared reading and 4-year-old children's vocabulary, morphological and syntax comprehension, and narrative skills. *Early Education and Development*, *19*(1), 27-44.

Shahaeian, A. (2015). Sibling, family, and social influences on children's theory of mind understanding: New evidence from diverse intracultural samples. *Journal of Cross-Cultural Psychology, 46*(6), 805-820.

Shanahan, T., & Shanahan, S. (1997). Character perspective charting: Helping children to develop a more complete conception of story. *The Reading Teacher, 50*(8), 668-677.

Shepard, L., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments.* Washington, DC: National Education Goals Panel.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition, 38*(1), 1-12.

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.

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*(1), 117-141.

Slaughter, V., Peterson, C. C., & Mackintosh, E. (2007). Mind what mother says: Narrative input and theory of mind in typical children and those on the autism spectrum. *Child Development, 78*(3), 839-858.

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2016). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry, 57*(12), 1360-1369.

Snowling, M. J., Duff, F., Petrou, A., Schiffeldrin, J., & Bailey, A. M. (2011). Identification of children at risk of dyslexia: The validity of teacher judgements using 'Phonic phases'. *Journal of Research in Reading, 34*(2), 157-170.

Snowling, M. J., & Hulme, C. (2008). *The science of reading: A handbook*. Padstow, Cornwall: Blackwell.

Snowling, M. J., Stothard, S., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2011). *York assessment of reading for comprehension: Early reading and passage reading primary (YARC primary)*. London: GL Assessment.

Sontag, L. W. (1971). The history of longitudinal research: Implications for the future. *Child Development, 42*(4), 987-1002.

Soper, D. (2015). A-priori sample size calculator for structural equation models. Retrieved from https://www.danielsoper.com/statcalc/calculator.aspx?id=89

Soper, D. (2017). Post-hoc statistical power calculator for multiple regression . Retrieved from https://www.danielsoper.com/statcalc/calculator.aspx?id=9

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*(5), 640-656.

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.

Strasser, K., & Río, F. d. (2014). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Reading Research Quarterly, 49*(2), 169-187.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*(3), 395.

Swank, L. K., & Catts, H. W. (1994). Phonological awareness and written word decoding. *Language, Speech, and Hearing Services in Schools, 25*(1), 9-14.

Symons, D. K., Peterson, C. C., Slaughter, V., Roche, J., & Doyle, E. (2005). Theory of mind and mental state discourse during book reading and story-telling tasks. *British Journal of Developmental Psychology, 23*(1), 81-102.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*(4), 381-398.

Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development, 77*(2), 465-481.

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months. *Child Development, 79*(2), 284-302.

Thorpe, K. (1996). *Thorpe interaction measure (TIM).* Unpublished manuscript.

Tiu, R. D.,Jr, Thompson, L. A., & Lewis, B. A. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities, 36*(5), 424-436.

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing, 28*(7), 939-957.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209-253.

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.

Valentini, A., Ricketts, J., Pye, R. E., & Houston-Price, C. (2018). Listening while reading promotes word learning from stories. *Journal of Experimental Child Psychology, 167*, 10-31.

Van Den Broek, P., Kendeou, P., Lousberg, S., & Visser, G. (2011). Preparing for reading comprehension: Fostering text comprehension skills in preschool and early elementary school children. *International Electronic Journal of Elementary Education, 4*(1), 259.

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.

Van Dijk, T. A., Kintsch, W., & Van Dijk, T. A. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first-and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology, 96*(1), 19-30.

van Wingerden, E., Segers, E., van Balkom, H., & Verhoeven, L. (2018). Cognitive constraints on the simple view of reading: A longitudinal study in children with intellectual disabilities. *Scientific Studies of Reading, 22*(4), 321-334.

Vandergrift, L. (2002). 'It was nice to see that our predictions were right': Developing metacognition in L2 listening comprehension. *Canadian Modern Language Review, 58*(4), 555-575.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*(1), 3-32.

Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 22*(3), 407-423.

Vinden, P. G. (1999). Children's understanding of mind and emotion: A multi-culture study. *Cognition & Emotion, 13*(1), 19-48.

Von der Embse, N. P. (2016). What school psychologists need to know about structural equation modelling. *School Psychologists as Consumers of Research, 44*, 10-12.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Massachusetts: Harvard University Press.

Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly, 35*(4), 554-566.

Walfish, S. (2006). A review of statistical outlier methods. *Pharmaceutical Technology, 30*(11), 82.

Wang, Z., Devine, R. T., Wong, K. K., & Hughes, C. (2016). Theory of mind and executive function during middle childhood across cultures. *Journal of Experimental Child Psychology, 149*, 6-22.

Watson, A. C., Painter, K. M., & Bornstein, M. H. (2001). Longitudinal relations between 2-year-olds' language and 4-year-olds' theory of mind. *Journal of Cognition and Development, 2*(4), 449-457.

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence the psychological corporation*. San Antonio: TX: The Psychological Corporation.

Wechsler, D. (2007). *Wechsler individual achievement test—second edition, australian standardised edition (WIAT–II australian)*. San Antonio: TX: The Psychological Corporation.

Wechsler, D. (1993). *Wechsler objective reading dimensions*. London: TX: The Psychological Corporation.

Wechsler, D. (2014). *WISC-V: Technical and interpretive manual*. Bloomington: Pearson.

Wellman, H. M. (1977). Preschoolers' understanding of memory-relevant variables. *Child Development, 48*(4), 1720-1723.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523-541.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*(6), 476-487.

Weston, R., & Gore Jr, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34*(5), 719-751.

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development, 80*(4), 1097-1117.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*(3), 848-872.

Wiig, E., Secord, W., & Semel, E. (2006). *Clinical evaluation of language fundamentals preschool-australian and new zealand standardised edition (CELF P-2 australian and new zealand)*. Sydney: Pearson.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913-934.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5*(3), 211-239.

Wolff, U. (2014). RAN as a predictor of reading skills, and vice versa: Results from a randomised reading intervention. *Annals of Dyslexia, 64*(2), 151.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-johnson tests of achievement* Itasca, IL: Riverside Publishing.

Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: Do infants have a true understanding of false belief? *British Journal of Developmental Psychology, 30*(1), 156-171.

Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology, 2*(1), 33-45.

Zucker, T. A., Cabell, S. Q., Justice, L. M., Pentimonti, J. M., & Kaderavek, J. N. (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental Psychology*, *49*(8), 1425.

Zunshine, L. (2019). What mary poppins knew: Theory of mind, children's literature, history. *Narrative*, *27*(1), 1-29.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review, 23*(4), 1028-1034.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292-297.