Investigating reading comprehension in English as additional language learners:

children's processing of metaphors and internal inconsistencies when reading text



Annina K. Hessel

Brasenose College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity Term, 2018

Table of Contents

Acknowledgements i

Abstract iii

List of Figures v

List of Tables vii

Chapter 1: General Introduction 1

Chapter 2: The Processing of Verbal and Nominal Metaphors 23

Chapter 3: Individual Differences in the Processing of Nominal Metaphors 90

Chapter 4: Individual Differences in Comprehension Monitoring 144

Chapter 5: General Discussion 178

References 201

Appendices 216

Acknowledgements

First, I would like to thank my supervisors, Vicki Murphy and Kate Nation, for their unwavering support throughout my time in Oxford. I can't thank you both enough for your patience and guidance. I would like to also thank the wonderful REAL group for all the feedback and input these last years, in particular Adam Unthiah for his never-ending jokes and Faidra Faitaki for her enthusiasm. Special thanks also to the ReadOxford crew, particularly Niina Tamura and Helen Norris for continuing support, and Yaling Hsiao and Megan Bird for the many sunny lunches. In addition, I would like to thank Sam Forbes and particularly Nick Cooper for their input on all things statistics.

I also wish to thank Brasenose college for their wonderful support, particularly my college advisor Simon Smith and Elspeth Garment for all their advice along the way. For financial support, I would like to thank the Economic and Social Research Council and the Scatcherd European Scholarship. I also wish to thank the children, parents, teachers, and staff of St. Ebbe's, Old Marston, St. Nicholas', and Our Lady's, who made this research possible.

Outside of work, I want to thank the YKBG family for hours of dancing together, for being the most patient participants in all my norming studies, and for making my years in Oxford really special. In particular, I want to thank Becka and Mairi for many shared jokes, Han for all her support with nasty software and everything else, and Yuka for making YKBG what it is. I am extremely happy to have shared my time in Brasenose with Fleur and Victoria who have been and remain to be stellar friends and the best company I could wish for. I am also grateful for those friends who have supported me from a distance, in particular Christina, Katy, Anna, and Gemma for many wonderful times. My thanks also go out to my office family and to my brilliant housemates, in particular Ross, Ping, Ella, and Ilze.

Finally, I would like to thank my family for their endless support and, of course, Saaru, for always being there with a song, an open ear, and a happy thought.
Let's get out of the library and into the sunshine now!

Abstract

Investigating reading comprehension in English as additional language learners: children's processing of metaphors and internal inconsistencies when reading text

Annina K. Hessel, **Brasenose College, University of Oxford**
Thesis submitted for the degree of Doctor of Philosophy in Trinity Term, 2018

English as an additional language (EAL) learners form a large group of students in UK schools and are known to differ from monolinguals, particularly in vocabulary and reading comprehension, but little is known about the way they process texts when reading. The three experiments presented in this thesis had three overarching objectives: first, to establish how children process different metaphors; second, to investigate how children monitor their comprehension; and third, to compare these reading processes between EAL learners and monolinguals. In Experiments 1 and 2, metaphor processing was compared across metaphor types (nominal and verbal) and between children with varying word knowledge. In Experiment 3, comprehension monitoring was associated with word knowledge and strategic awareness. Across all three experiments, language group differences were identified. Nominal metaphors were more challenging to understand for children than literal controls, a difficulty that stemmed from increased demands on post-lexical processing, as manifest in increased rereading. Children were able to monitor their comprehension on internal inconsistencies, which was demonstrated through selective rereading. Language group differences were evident relative to word knowledge. EAL learners were behind monolinguals in their metaphor processing relative to performance on standardised tests. Given sufficient word knowledge, EAL learners and monolinguals were, however, on a par in comprehension monitoring. For future research, the findings indicate that online tasks of reading processing hold valuable information about EAL learners' reading comprehension. For teaching, results indicate the importance of supporting word knowledge to close language group differences and to support skilful reading processing.

List of Figures

Figure 1: Language group differences on assessments of reading and language. 53
Figure 2: Children's processing and comprehension of metaphors and literal controls. 57
Figure 3: Significant interactions between metaphorness and metaphor type on reading metaphors and answering comprehension questions. 63
Figure 4: Vocabulary, language group and metaphorness interactions on reading of metaphors and post-metaphor region. 73
Figure 5: Children's processing and comprehension of metaphors and literal controls. 115
Figure 6: Monolingual and EAL children with low and high reading fluency reading the second word of metaphors or literal controls. 121
Figure 7: Children's actual and self-perceived comprehension by language groups. .. 126
Figure 8: Schema of increasingly skilful metaphor processing in children's reading.. 136
Figure 9: Children's reading processing of plausible and implausible words. 157
Figure 10: Children with small and large vocabulary reading plausible and implausible words, plotted for predicted means at +/- 1 SD of vocabulary size 162
Figure 11: EAL and monolingual children with small and large vocabulary reading plausible and implausible words, plotted for predicted means at +/- 1 SD of vocabulary size 163
Figure 12: EAL and monolingual children's actual and reported selective rereading. 165

List of Tables

Table 1: Language groups on age and non-verbal reasoning. 33
Table 2: Example stories and questions. 36
Table 3: Example of interest areas. 43
Table 4: Language groups compared on assessments of reading and language. 52
Table 5: Means for children reading the first word of metaphors and literal controls. .. 54
Table 6: Means for children reading the second word of metaphors and literal controls, and for answering comprehension questions. 54
Table 7: Means for children reading the post-metaphor region. 55
Table 8: Models using language group, metaphorness and metaphor type to predict reading the first word of the metaphor. 58
Table 9: Models using language group, metaphorness and metaphor type to predict reading the second word of metaphors and answering questions. 59
Table 10: Models using language group, metaphorness and metaphor type to predict reading the post-metaphor region. 60
Table 11: Model means for children reading the first word of the metaphor. 66
Table 12: Model means for children reading the second word of the metaphor and answering comprehension questions. 67
Table 13: Model means for children reading the post-metaphor region. 68
Table 14: Models using individual differences, language group and metaphorness to predict reading the first word of the metaphor. 74
Table 15: Models using individual differences, language group and metaphorness to predict reading the second word of metaphors and answering comprehension questions. 75
Table 16: Models using individual differences, language group and metaphorness to predict reading the post-metaphor region. 76
Table 17: Language groups on age and non-verbal reasoning. 102
Table 18: Example story and question. 104
Table 19: Matching of replaced word in metaphors and literal controls. 106
Table 20: Example of interest areas. 107
Table 21: Language groups compared on reading fluency and vocabulary size. 110

Table 22: Means for children reading the second word of metaphors and answering comprehension questions. 111
Table 23: Means for children reading the first word of metaphors and the post-metaphor region. 111
Table 24: Results of models using language group and metaphorness to predict reading the second word in the metaphor and answering comprehension questions. 113
Table 25: Models using language group and metaphorness to predict reading the first word of the metaphor and the post-metaphor region. 114
Table 26: Predicted means for children reading the second word of the metaphor and answering comprehension questions. 117
Table 27: Predicted means for children reading the first word of the metaphor and the post-metaphor region. 118
Table 28: Models using individual differences, language group and metaphorness to predict reading the second word of metaphors and answering comprehension questions. 123
Table 29: Models using individual differences, language group and metaphorness to predict reading the first word of metaphors and the post-metaphor region. 124
Table 30: Language groups compared on their actual and self-perceived comprehension. 125
Table 31: Reading fluency of children in Experiments 1 and 2. 135
Table 32: Example story and question. 152
Table 33: Matching of implausible words and plausible controls. 153
Table 34: Means for children reading the target word. All reading times in milliseconds. 155
Table 35: Models using language group and implausibility to predict reading the target word. 156
Table 36: Predicted means for children reading the target word by language group and consistency condition. 159
Table 37: Results of models using individual differences, language group and consistency to predict reading of the target word. 160
Table 38: Language groups compared on their self-perceived comprehension and rereading. 164

Chapter 1: **General Introduction**

Imagine that we were to sneak into an early literacy classroom in a British primary school. We would likely see a group of children gathered around a teaching assistant reading with them, or engaged in some other literacy-based activity that will support their development to become proficient readers. Children are surprisingly good at mastering the difficult task of learning to read – as they need to be, since reading will form the basis of many future activities, including most learning at school. Nonetheless, almost one in five students in the UK and other OECD countries may not reach the basic level of reading comprehension that they need in daily life (OECD, 2016). To support these students effectively, it is vital to understand how children's reading develops, and reading research has made great progress in this endeavour (Castles, Rastle, & Nation, 2018).

Most research, however, has focused on monolingual children, whose learning contexts and outcomes often differ from those of bilingual language learners (Murphy, 2014). Understanding bilingual learners' development is particularly relevant in the UK, where one in five primary school students is a minority language learner learning English as an additional language (EAL) (NALDIC, 2016). Growing up bilingual is associated with economic and social advantages (Callahan & Gandara, 2014; García & Baetens Beardsmore, 2009; Heller, 2010), not to mention the importance of dual-language proficiency for identity, self-expression and maintaining family relationships (Gao, 2012; Pearson, 2007; Poplack, 1980). And while it is very possible to reap these benefits of bilingualism without negative impacts on educational attainment (Murphy,

2018), school systems in many countries still struggle to fully integrate children from ethnic minorities (OECD, 2006).

Multiple factors determine the academic outcomes of EAL children, including ethnicity, age of arrival in the UK, and socioeconomic status (SES) (Strand, Malmberg, & Hall, 2015). However, a number of studies have suggested that the key determinant of EAL pupils' academic success is their fluency in English (Demie & Strand, 2006; Strand & Demie, 2005; Strand & Hessel, 2018; Whiteside, Gooch, & Norbury, 2017). Indeed, Whiteside, Gooch and Norbury (2016) demonstrated that when EAL children are matched on English language fluency to non-EAL children, the EAL children either matched or outperformed the non-EAL children on a range of behavioural and cognitive measures. While the directionality of these relationships is hard to establish based on correlations alone, it is plausible that difficulties with English would impact EAL learners' personal and academic development (Whiteside et al., 2017). This is also supported by a recent analysis of local authority data that showed that teacher-rated English proficiency of EAL learners predicted up to 20% more variance in EAL children's attainment, beyond gender, SES, and ethnicity (Strand & Hessel, 2018). This thesis thus focuses on English proficiency as an important area of EAL learners' development.

Achieving this proficiency can be problematic since within each of their languages, bilingual learners may not always have the same language proficiency as a monolingual learner (Bialystok, Luk, Peets, & Yang, 2009; Melby-Lervåg & Lervåg, 2014; Murphy, 2014; Pearson, Fernandez, & Oller, 1993; Spencer & Wagner, 2016). This is likely due to differences in English language exposure between EAL learners and monolinguals: their exposure may differ in amount, for example when an EAL

learner first encounters English at an older age, or when they only hear and use English outside the home (Murphy, 2018). The exposure can also differ qualitatively, for example when EAL learners use English for only certain parts of life but not others (Oller, Pearson, & Cobo-Lewis, 2007), or when their parents speak English at a lower level of proficiency (Paradis, 2016). Since exposure forms the basis of language learning (Tomasello, 2008), such differences in the amount and quality of English language exposure can make it harder for EAL learners to develop the same English proficiency as their monolingual peers, if they are not otherwise supported.

Second language acquisition research has for a long time recognised that expecting bilinguals to have equal language skills to monolinguals within each language is not only unrealistic (as discussed in light of the empirical evidence on input and acquisition above), but that between-group comparisons can also be misleading (Grosjean, 1989). They can be misleading since a bilingual's abilities in their two or more languages are not independent from each other; bilinguals possess skills and knowledge that feed into all their languages alike, building a common knowledge base (Cummins, 2000). Examples are phonological awareness or meta-linguistic knowledge about text representation, in both of which bilinguals tend to be on par with or outperform monolingual speakers (Bialystok, Majumder, & Martin, 2003; Bialystok, Shenfield, & Codd, 2000). Furthermore, comparing monolinguals and bilinguals in a single language may give the false impression that bilinguals know less when in fact, their combined knowledge across languages may be superior to that of monolinguals – as is the case for the conceptual vocabulary knowledge of bilingual compared to monolingual toddlers (Pearson et al., 1993). Finally, group comparisons between bilinguals and monolinguals seem to suggest that differences between groups are clear-

cut when in fact, the language skills of both monolingual and bilingual learners exist on continua which partially overlap (Marinova-Todd, Marshall, & Snow, 2000). Thus, the purpose of the current thesis cannot be to simply compare the English language skills of EAL learners and monolinguals without considering individual differences or for the simple sake of finding differences. However, the research clearly shows that depending on their language learning opportunities, EAL learners' language skills may differ greatly from monolingual peers, and that such a lag in language learning is linked to lower likelihood to do well in school, both academically and personally (Strand & Hessel, 2018; Whiteside et al., 2017). In this context, it is of practical relevance to understand which EAL learners differ from monolinguals and in which aspects of their language skills. Only then, targeted support can be provided to those children who need it in order to assure that all children have the same opportunity to succeed, no matter their background. It is for this reason that the current thesis adopted a design that compares EAL learners to monolinguals, while recognising the theoretical debates surrounding this approach. Individual differences are considered throughout the thesis to acknowledge the continuity within each language group.

Importantly, if EAL and monolinguals differ in their language skills, these differences are likely to show particularly in reading comprehension (see overleaf for a discussion of the evidence). A better understanding of how EAL children learn to read can help support those who may struggle and contributes to our general understanding of reading learning. The current thesis thus investigated EAL children's reading comprehension in English, focusing in particular on under-researched aspects of reading. This chapter will provide a broad overview of the literature to introduce the

topics investigated in this thesis. More detailed discussions of specific areas of research are provided in subsequent chapters.

1.1 EAL learners' reading comprehension and vocabulary knowledge

There is ample evidence to show that if an EAL learner differs from monolinguals in their oral language skills, they are also likely to do so in reading comprehension, as supported by the large discrepancy measured between EAL learners and monolinguals in their reading comprehension (see Melby-Lervåg & Lervåg, 2014; Spencer & Wagner, 2016 for recent meta-analyses, as well as Burgoyne, Whiteley, & Hutchinson, 2013; Droop & Verhoeven, 2003; Hutchinson et al., 2003 for longitudinal evidence). In a recent meta-analysis of 16 studies, Spencer and Wagner (2016) showed that bilingual children compared to monolinguals had weaker comprehension skills both in listening and reading, but that the gap was particularly large in reading comprehension, as manifested in an effect size of d = 2.5. This indicates that language group differences in reading comprehension are not only frequent, but also substantial. However, there is also some variation in the size of the effect across studies, with some reporting no or only negligible differences in the reading comprehension of monolingual and bilingual learners (Bowyer-Crane, Fricke, Schaefer, Lervåg, & Hulme, 2017; D'Angiulli, Siegel, & Serra, 2001; Lesaux, Rupp, & Siegel, 2007; Verhoeven, 2000). Melby-Lervåg and Lervåg (2014) suggested that such variation can be explained by methodological factors, such as measuring reading comprehension through multiple choice (Lesaux et al., 2007) and cloze tests (D'Angiulli et al., 2001) rather than more sensitive open-ended questions, or comparing language groups on specific reading-related abilities where bilinguals tend to excel, such as metalinguistic awareness of script (Bialystok et al., 2000) and phonological awareness (Westman, Korkman, Mickos, & Byring, 2008).

Another explanation for the variation between studies relates to differences between samples: the extent of differences in language skills between monolinguals and EAL children depends on a combination of child-internal factors, such as their language aptitude, or environmental factors, such as richness of language input (Paradis, 2011), and different language learning contexts (Murphy, 2014). These influences may lead to language group differences appearing either smaller or larger, depending on the background of the sample. In the Dutch context, for example, some studies found significant differences in reading comprehension between Dutch monolinguals and minority language learners (Droop & Verhoeven, 2003), while these differences remained numerical in others (Spätgens & Schoonen, 2018). On the other hand, Spencer and Wagner (2016) also point out that many studies have primarily sampled EAL learners with poor comprehension skills, which would make language group differences appear larger. However, despite some unresolved questions on the extent and pervasiveness of group differences, there is an overarching trend that at least a proportion of EAL learners have weaker reading comprehension skills than their monolingual peers.

A framework for analysing the origins of these difficulties is provided by the *Simple View of Reading*, a theory of reading that is well supported by evidence across languages (Babayigit & Stainthorp, 2014; Joshi, Ji, Breznitz, Amiel, & Yulia, 2016; Tilstra, McMaster, Van Den Broek, Kendeou, & Rapp, 2009). The Simple View of Reading states that skilful reading comprehension requires children to develop two sets of skills: the ability to recognise printed words, and the ability to construct meaning (Gough & Tunmer, 1986), often referred to as linguistic comprehension. If either of the two abilities is compromised, so too is reading comprehension. Which of these two

areas, then, is problematic for those EAL learners who struggle with reading comprehension? In general, it is not decoding (Babayiğit, 2015; Burgoyne, Whiteley, & Hutchinson, 2011; Hutchinson et al., 2003; Melby-Lervåg & Lervåg, 2014; Oller et al., 2007; Spencer & Wagner, 2017). For example, Babayiğit (2015) found that 9-year-old EAL learners differed from their monolingual peers in their reading comprehension, but not in word reading. In a longitudinal study, Hutchinson et al. (2011) found that UK EAL learners in Year 3 and Year 4 numerically outperformed monolinguals on tests of word and non-word reading. In the first years of reading, when reading comprehension is most strongly limited by word reading skills, EAL learners can thus even perform equal to monolinguals on tests of reading comprehension (Burgoyne et al., 2011). Why then, would EAL learners be on par with monolinguals in their word reading, if they tend to lag behind in other areas of language learning? One reason may be that to be a good decoder in English, children need to acquire only a finite sound and letter system, which is usually mastered during the early primary school years (Castles et al., 2018). Furthermore, decoding relies on phonological awareness, in which bilinguals have been found to outperform their monolingual peers (Bialystok et al., 2003). For these reasons, decoding may be less susceptible to differences in rich exposure to English.

However, EAL learners tend to have smaller English vocabularies than monolinguals, and these differences in vocabulary seem to underpin differences in reading comprehension (Bialystok, Luk, Peets, & Yang, 2009; Melby-Lervåg & Lervåg, 2014; Murphy, 2014; Pearson, Fernandez, & Oller, 1993; Spencer & Wagner, 2016). For example, EAL learners have repeatedly been found to perform lower on tests of vocabulary size such as the Peabody Picture Vocabulary Test (Bialystok et al., 2009; Pearson et al., 1993), a difference that has also been replicated in UK samples (Hessel

& Murphy, 2018; Smith & Murphy, 2015). Bilinguals' often smaller vocabularies can then become the "bottleneck of comprehension" which limits the information accessed in the text, thus limiting reading comprehension (Lervåg & Aukrust, 2010). As a result, EAL learners tend to fall behind their monolingual peers in their reading comprehension in the later years of reading instruction, when comprehension skills become more important for children's reading comprehension (Babayigit & Stainthorp, 2014; Lervåg & Aukrust, 2010; Nakamoto, Lindsey, & Manis, 2007). The reason may be that unlike decoding, vocabulary may be particularly vulnerable to the influences of reduced exposure. Much of word learning happens implicitly during everyday reading and language use and continues throughout people's lifetime (Brysbaert, Stevens, Mandera, & Keuleers, 2016). For EAL children, less exposure to language might result in limited opportunities for learning new words. Once a difference in vocabulary manifests itself, it can be hard to catch up since vocabulary supports reading (Verhoeven & van Leeuwe, 2012). Better readers are more likely to create further learning opportunities for word learning (Willingham, 2017) and are more efficient when learning new words during reading (Joseph & Nation, 2018). This interdependency of vocabulary and reading comprehension through development can thus perpetuate gaps relative to monolinguals for some EAL leaners, despite adequate decoding skills.

Vocabulary is an important part of EAL learners' reading comprehension. However, vocabulary knowledge is also multi-faceted, and some facets may be more relevant to understanding language group differences than others. Besides vocabulary size, or the number of words a child knows, vocabulary knowledge can also be operationalised as how well a word is known, a dimension which is known as vocabulary depth. There are many facets to vocabulary depth: a learner can know the

meaning, the form and/or the use of a word; for instance, they might be able to understand its meaning, but not be able to use it correctly, or indeed be able to produce correct sentences with specific words without knowing their meaning (Nation, 2001). Furthermore, some words are regularly combined to make up collocations such as *pay attention* or idioms such as *to kick the bucket* (Nation, 2001). However, drawing a distinction between vocabulary size and depth is not uncontroversial: while the two can be clearly distinguished in conceptualisations and measurements (Nation, 2001), they can be highly correlated in practice (Vermeer, 2001). Vermeer (2011) tested size through a picture-matching and a description task and vocabulary depth through an association task in a sample of 4- to 7-year-old Dutch bilinguals and monolinguals. She reported that vocabulary size and depth were strongly correlated both in monolinguals and bilinguals which she interpreted as them being indistinguishable (Vermeer, 2001). Vermeer's finding may, however, be due to methodological limitations, since performance on free association tasks like the one used in Vermeer (2011) is predicted by age, and may be less appropriate for use with younger children (Cremer, Dingshoff, de Beer, & Schoonen, 2011).

Ouellette (2006) suggests that vocabulary depth may be particularly important to explain comprehension of texts when reading: he tested 9- to 10-year-old children on various measures of reading ability, and predicted children's reading skills from their vocabulary size (as measured through tasks that required picture naming and matching pictures to words) as well as vocabulary depth (as measured through tasks that required producing definitions and choosing synonyms). He found that after controlling for other variables, size uniquely predicted word-level reading such as decoding and visual word recognition, while vocabulary depth was the strongest predictor of passage

comprehension (Ouellette, 2006). This indicates that measures of vocabulary depth may tap skills that are particularly important in children's reading comprehension.

Furthermore, a number of studies have reported that measures of vocabulary depth were particularly informative for comparing EAL and monolingual learners (Cameron & Besser, 2004; Cremer, Dingshoff, de Beer, & Schoonen, 2011; Cremer & Schoonen, 2013; Hessel & Murphy, 2018; Kan, 2014; Martinez & Murphy, 2011; McKendry, 2014; Schoonen & Verhallen, 2008; Smith & Murphy, 2015; Spätgens & Schoonen, 2017). These studies have relied on different measures of vocabulary depth, one of which has been used extensively with minority language learners of Dutch (children who, like EAL learners, are learning Dutch additionally to another language at home). This measure relies on assessing how many meaning associations a child can retrieve for a particular word (Cremer et al., 2011; Cremer & Schoonen, 2013; Schoonen & Verhallen, 2008; Spätgens & Schoonen, 2018). Children are presented with a target word and are asked to choose associated words from a list of options (Cremer & Schoonen, 2013; Schoonen & Verhallen, 2008; Spätgens & Schoonen, 2018) or to provide free associations (Cremer et al., 2011). Both task formats show differences in vocabulary depth between monolinguals and minority language learners; differences which run parallel to differences in reading comprehension (Cremer & Schoonen, 2013).

Another way in which vocabulary depth has been measured is by assessing children's understanding of words in the context of multi-word or figurative phrases (Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014; Smith & Murphy, 2015). For example, Kan (2014) compared EAL and monolingual primary school children on their understanding of high-frequency collocations such as *by and large*. Children read texts

containing collocations and then answered comprehension questions, and reported their self-perceived comprehension. EAL learners had more difficulties understanding collocations, as evidenced by lower accuracy on the comprehension questions. Yet these difficulties went unnoticed by EAL learners who judged their comprehension to be higher than their actual comprehension (Kan, 2014). Other research examined EAL learners' comprehension of collocations such as *pay attention* (Smith & Murphy, 2015), idioms such as *it's raining cats and dogs* (McKendry, 2014), and metaphors such as *hours fly by* (Hessel & Murphy, 2018). In a cross-sectional study with children aged 7–10 years, Smith and Murphy (2015) found that older EAL children did not produce many more collocations correctly compared to younger EAL learners, unlike monolinguals who experienced a "burst of multi-word vocabulary". In the later years of primary school, monolinguals' growth in multi-word vocabulary thus increased differences in vocabulary depth between EAL learners and monolinguals (Smith & Murphy, 2015).

Hessel and Murphy (2018) compared 5- to 7-year-old EAL learners and monolinguals on their comprehension of metaphors presented in story context. EAL learners had overall lower comprehension of metaphors; an effect that seemed driven most strongly by their lower accuracy in a task where children were asked to explain their metaphor interpretation, compared to multiple-choice questions or metaphor recall (Hessel & Murphy, 2018). McKendry (2014) tested 10- to 11-year-old EAL learners and monolinguals on a customised test of idiom comprehension. Monolinguals outperformed EAL learners on the idiom comprehension task, and performance on the idiom comprehension task predicted reading comprehension more strongly in EAL

learners than monolinguals, indicating that idiom comprehension captured an important part of the language group differences (McKendry, 2014).

Together, these studies indicate that vocabulary depth – as measured, for example, through the comprehension of multi-word and figurative phrases – taps an important facet of the group differences between EAL learners and monolinguals. In addition, these differences are predictive of reading comprehension, particularly in EAL learners (McKendry, 2014). None of the existing studies has, however, measured children's processing of these phrases, which makes it hard to know when difficulties arise, and why these difficulties are so predictive of EAL learners' reading comprehension. The following section will thus introduce theoretical and methodological questions on how to study language processing during children's reading, and suggest how it may be studied in EAL learners.

1.2 Reading processing and comprehension

Reading is complex, and only looking at shared variance may not be sufficient to understand some of these complexities. For example, the two component sets of skills assumed by the simple view – decoding and comprehension – are inter-related, since readers with strong word knowledge can use word knowledge to support decoding (Tunmer & Chapman, 2012). Furthermore, other skills contribute to reading beyond these two, such as background knowledge (Burgoyne et al., 2013), the ability to make inferences beyond what is stated in the text (Silva & Cain, 2015), and to keep an active memory of verbal information (Cain, Lemmon, & Oakhill, 2004). To complicate things further, these different skills are interrelated, since vocabulary may support inference making directly through its effect on verbal working memory (Cain & Oakhill, 2014; Currie & Cain, 2015), while inference making supports vocabulary development (Cain

et al., 2004). A way to understand these complexities is to consider the process of reading.

The reading systems framework (RSF) provides a fitting theory to think about the complexities of reading processing (Perfetti & Stafura, 2014). The RSF describes how a reader builds up a situation model – a full representation of the text – while reading: word meanings are woven into a situation model through ongoing integration processes, word meanings are shaped to fit the situation model, no longer relevant information is inhibited, and the situation model is continuously updated to fit the newly incoming information. The processes that follow lexical access and lead to the building of a situation model will from here on be summarised as post-lexical processing.

There is evidence of how skilled and less skilled comprehenders differ in post-lexical processing, for example showing that less skilled comprehenders are less successful than skilled comprehenders at suppressing no longer relevant verbal information (Pimperton & Nation, 2010) or noticing mismatched information (Connor et al., 2015; Eilers, Tiffin-Richards, & Schroeder, 2018a). The lexical quality hypothesis describes how these differences in processing can result from differences in word knowledge (Perfetti & Hart, 2002). Perfetti and Hart suggest that the cornerstone of successful comprehension processes is word knowledge of high lexical quality. High lexical quality is characterised by both *precise* and *flexible* representation of word meanings in a reader's lexicon. *Precise* means that a speaker has learned a word form well enough to reliably retrieve the correct meaning. For example, on reading the word *cat*, a precise word meaning would allow the reader to quickly exclude orthographic neighbours such as *cap* or *fat*. *Flexible*, on the other hand, means that a reader is able to activate many facets of a word. For example, reading the word *cat* would allow readers

with high lexical quality to activate different meaning facets, including the use of *cat* in phrases (*It's raining cats and dogs*) or its synonyms (*feline*). High lexical quality allows readers to process words efficiently and speedily – both during initial lexical access and during post-lexical processing (Perfetti, Yang, & Schmalhofer, 2008). For example, Perfetti et al. (2008) measured the reading processing of skilled and less skilled comprehenders by recording their brain responses using electroencephalogram (EEG) measurements. They found that skilled comprehenders were quick to access the word meaning, to update their text representation, and to integrate the meaning of the new word; in contrast, less skilled comprehenders were slower with initial lexical access which impacted their subsequent post-lexical processing (Perfetti et al., 2008).

A similar chain reaction from lexical access to post-lexical processing may also play a role in the lower reading comprehension of EAL learners compared to monolinguals. To date, however, there is no online evidence that would compare EAL learners' and monolinguals' post-lexical processing. One particularly compelling example of EAL learners' post-lexical processing may be provided by phrase processing, particularly of metaphors: understanding metaphors is harder for EAL learners than for monolinguals (Hessel & Murphy, 2018), and it requires two key components of general reading: integrating meanings (Coulson & Van Petten, 2002), and inhibiting interfering information that may lead to overly literal interpretations (Tartter, Gomes, Dubrovsky, Molholm, & Stewart, 2002). For example, Coulson and Van Petten (2002) found that processing sentences containing metaphors, such as *He knows that power is a strong intoxicant* triggered similar brain activation as processing sentences that require readers to integrate meanings literally, such as *He has used cough syrup as an intoxicant.* This indicates that similar processes are involved in metaphor

comprehension and in general post-lexical processing during reading (Coulson & Van Petten, 2002), making metaphor processing an interesting tool to study reading processing. Little is known, however, how these processes play out in children's metaphor comprehension more generally, or in EAL learners particularly. One purpose of this thesis was thus to establish the nature of metaphor processing in children's reading, and then to investigate any differences between EAL learners and monolinguals.

A second topic of interest is EAL learners' comprehension monitoring, or their ability to check and regulate their comprehension (Connor et al., 2015; Kim, Vorstius, & Radach, 2018; Oakhill, Hartt, & Samols, 2005). Comprehension monitoring seems to consist of two steps: first, readers evaluate the incoming information and notice a mismatch, and then, they re-analyse and repair the mismatch (Connor et al., 2015). These two processes are also involved in metaphor processing (Coulson & Van Petten, 2002; Levorato, Nesi, & Cacciari, 2004; Oakhill, Cain, & Nesi, 2016), and it has been suggested that EAL children may struggle with them more than monolinguals (Burgoyne et al., 2013; Kan, 2014; Martinez & Murphy, 2011). For example, EAL learners may struggle with figurative and multi-word phrases because they fail to notice difficulties with them (Kan, 2014; Martinez & Murphy, 2011) or because they struggle to fully integrate their meanings into the context (Burgoyne et al., 2013). A second purpose of this thesis was thus to assess children's comprehension monitoring during reading processing, and to again compare it between EAL learners and monolinguals.

One of the overarching goal of the thesis was thus to understand what processes link differences in vocabulary to lower reading comprehension. However, it is important to note that EAL learners may not necessarily be disadvantaged relative to

monolinguals, as they may be able to compensate vocabulary gaps in certain contexts. For example, Raudszus et al. (2017) found trends that inhibition skills were more strongly associated with syntactic skills and reading comprehension in minority language learners than in monolinguals. This finding indicates that minority language learners may have used increased executive control during reading to compensate weaker language skills. A similar interpretation was suggested by Shahar-Yamas and Prior (2018), who found that Russian minority language learners were able to infer as many words from context as monolinguals, despite gaps in vocabulary that could be expected to impact minority language learners' lexical inferencing. Such compensation behaviour may explain why bilingual readers are overall better in reading comprehension than monolinguals, relative to vocabulary knowledge (Melby-Lervåg & Lervåg, 2014). Studying reading processing in EAL learners is thus not only relevant to study how gaps in word knowledge impact reading comprehension, but also to see if EAL learners use different processing strategies to compensate for gaps in knowledge.

Importantly, both metaphor processing and online comprehension monitoring cannot be studied when reading has finished, but require measuring reading online – that is, as it happens. The two most sensitive online measures of reading are EEG measurements and eye-tracking (Perfetti & Stafura, 2014). Of both, eye-tracking is the only measure that allows readers to naturally move through a text, which makes it particularly advantageous to studying natural text reading behaviour (Rayner, Chace, Slattery, & Ashby, 2006). Eye-tracking offers high temporal precision and provides a millisecond record of where a reader fixates on text on a screen. This fixation record is traditionally processed by defining interest areas, which are areas in the text where relevant words were presented. All fixations and saccades in a region are then distilled

into different eye-movement measures that summarise different stages of reading processing.

Broadly, eye-movement measures can be divided into early and late measures (Pickering, Frisson, McElree, & Traxler, 2004). For example gaze duration is the amount of time a reader spends when first reading a word. Rereading time, in contrast, is the time spent when a reader returns to a word after having moved on. Early measures such as gaze duration are associated with early linguistic processes such as lexical access. Later measures such as rereading time or regression probabilities are argued to capture post-lexical processing, such as linking words to the context, and re-analysis of earlier interpretations (Pickering et al., 2004). Longer reading times across these different measures can index at what stage of processing a reader uses additional processing effort, thus providing additional information over other methods that do not record processing or only record overall reading time (Pellicer-Sanchez, 2015).

In recent years and with improving eye-tracking hardware, there has been a surge in developmental eye-tracking studies that have provided insights into children's reading processing (Blythe & Joseph, 2011). These studies have shown that children generally make longer fixations, shorter saccades and more regressions than adults (Blythe & Joseph, 2011). At the same time, children's reading behaviour also shows many similarities with that of adults: children, like adults, tend to first fixate a word left of the word centre, and their fixation durations are also influenced by linguistic factors such as word length (Joseph, Liversedge, Blythe, White, & Rayner, 2009) and word frequency (Joseph, Nation, & Liversedge, 2013). Age differences in reading behaviour seem largely due to the fact that children are still developing their language skills

compared to adults, rather than to differences in oculomotor control (Reichle et al., 2013).

Importantly, there have been a number of studies on children's post-lexical processing during reading (Connor et al., 2015; Eilers et al., 2018a; Eilers, Tiffin-Richards, & Schroeder, 2018b; Joseph et al., 2008; Joseph, Bremner, Liversedge, & Nation, 2015). This research has shown differences in the post-lexical processing of children and adult during their reading: for example, Joseph et al. (2008) reported that both children and adults spent longer in their initial reading of the target word *mouse* in anomalous sentences such as *Robert used a radio to play the horrible mouse that was very scared.* However, adults showed earlier reactions than children to implausible sentences such as *Robert used a hook to catch the horrible mouse that was very scared.* This indicates that children were slower than adults to integrate information in their situation model, as they took longer to notice incompatible information (Joseph et al., 2008).

Other studies have reported that children's ability to integrate information within texts was more strongly influenced by distance than it is in adults (Eilers et al., 2018b; Joseph et al., 2015). For example, Eilers et al. (2018b) compared 9-year-old children and adults in their sensitivity to an unusual repetition of character names, such as *Peter gets out of bed. Right away, Peter prepares breakfast.* This was compared to the reading of more commonly used referential pronouns as in *Right away, he prepares breakfast*. Both children and adults were sensitive to the unusual name repetition, as evidenced, for example, in more regressions out of the repeated name *Peter* than the pronoun *he*. However, only children were slower to read both pronouns and repeated names when they were further apart from the initial mention of the name, indicating that children's

processing was more strongly influenced by distance compared to adults (Eilers et al., 2018b). Similarly, Joseph et al. (2015) found that children spent more time on nominal anaphors that were close to the antecedent and typical (such as *vehicle* to refer back to *truck*), as opposed to more distant and atypical, indicating that they only resolved those easier references (Joseph et al., 2015). Together, these studies show similarities, but also differences in the post-lexical processing of children and adults.

On other hand, children vary greatly in their reading processing, and some 9-year-old children can already process texts similar to adults (Eilers et al., 2018a). A number of studies have studied these individual differences in children (Connor et al., 2015; Eilers et al., 2018a). For example, Eilers et al. (2018a) compared 9-year-old children and adults' eye movements when reading sentences that contained either matched pronouns (e.g. when *he* refers to a male character) or mismatched pronouns (e.g. when *he* refers to a female character). They found that all adults, but only half of all children reported noticing the pronoun mismatch. Children who reported a mismatch showed eye movements similar to adults, while those who failed to report it read more slowly overall, and made fewer regressions to the pronoun. These findings indicate that if children read fluently, they are able to notice and repair grammatical mismatches (Eilers et al., 2018a). This is in line with another study that reported that children with stronger language skills adapted their rereading more strongly to semantic mismatches than children with weaker language skills (Connor et al., 2015).

Taken together, these studies show that the eye-tracking methodology provides a sensitive measure of online reading, making it possible to examine children's post-lexical processing as it happens, and to uncover individual differences between children.

For all these reasons, eye-tracking has become a key methodology in developmental reading research, and was used in the current thesis.

1.3 Thesis Aims and Overview

The research reported in this thesis aimed to elucidate EAL and monolingual children's reading comprehension by studying how text is processed during reading. Based on previous offline evidence about EAL learners' comprehension skills, two areas were identified that merited further consideration: metaphor processing as an example of a multi-word phrase that promises to trigger post-lexical processing, and comprehension monitoring to probe children's sensitivity to mismatched information. Reading behaviour in these areas was compared between EAL and monolingual children, and was linked to children's individual differences in word knowledge. The following overarching questions were addressed:

1) How do children process metaphors during reading, and how does that differ by metaphor type and for children with varying word knowledge?
2) How do children monitor their comprehension during reading, and how is monitoring associated with word knowledge?
3) Are there differences between EAL learners and monolinguals in their online processing of metaphors and comprehension monitoring when reading, and if so, what are they?

Three experiments used eye-tracking methodology to measure EAL and monolingual children's processing while reading texts containing metaphors and internal inconsistencies. Chapter 2 introduces the eye-tracking method, and therefore provides more details about data collection and analyses than subsequent chapters. Experiment 1,

reported in Chapter 2, focuses on the effect of different metaphor types on children's reading and comprehension of verbal metaphors (such as *hours fly by*) and nominal metaphors (such as *grades are lotteries*) in relation to matched literal controls. The research aimed to explore at what point during reading children may struggle to process metaphors, and whether children have particular difficulties processing nominal compared to verbal metaphors. After reading stories that contained these metaphors, children's comprehension of the metaphors was tested. Children's reading and vocabulary were also assessed using standardised tests. This allowed comparison between the EAL and monolingual language groups, and provided data to link individual differences in online processing to differences in reading and language skill. Based on the results of this experiment, only nominal metaphors were used in the subsequent experiment.

Chapter 3 reports Experiment 2, which aimed to replicate Experiment 1 and to further probe how individual differences in reading fluency and vocabulary are related to EAL learners' and monolinguals' metaphor processing. To limit noise, metaphor stimuli were streamlined, and a slightly older and larger sample of children was tested. As in Experiment 1, children were tested in metaphor reading and on standardised tests of reading and language. A self-report measure was added to see if children differed in their awareness of metaphor comprehension. The first aim of Experiment 2 was to replicate the earlier findings and to extend the tested age range. This study examined whether the effects of metaphors on children's reading would prove reliable, and whether older children would be more skilful in their reading processing. The second aim was to conduct confirmatory analyses of how individual differences in reading fluency and vocabulary interact with language group in children's metaphor processing.

The purpose was to identify whether differences in language status and language skills were associated with metaphor reading, as they were in Experiment 1.

Chapter 4 reports Experiment 3, which focused on EAL and monolingual children's comprehension monitoring by recording their eye movements as they read texts containing internal inconsistencies, such as a *barking animal* that is later described as a *kitten*, as opposed to a *puppy*, as in the control condition (using a task adapted from Connor et al., 2015). Two self-report measures on self-perceived comprehension as well as on reading behaviour were combined with eye movement records (Hyönä & Nurminen, 2006). These measures examined whether EAL children would struggle more than monolinguals with monitoring their comprehension, and to what degree online comprehension monitoring, as manifest in eye movements, could be linked to strategic awareness of reading behaviour.

Finally, Chapter 5 summarises the results from the three experiments and links their findings to the literature, outlining wider theoretical, methodological and educational implications of the research, and suggesting future directions. This final chapter will discuss how the research presented in this thesis adds to our understanding of EAL learners' reading comprehension: the findings indicated the importance but also the complexities of children's rereading when processing texts, and indicate how reading processing is linked to word knowledge. Furthermore, they provide insights into which sub-groups of EAL learners may struggle in their reading comprehension and where. For teaching, these insights can help with understanding the processes underlying potential comprehension difficulties, and with targeting support accordingly.

Chapter 2: **The Processing of Verbal and Nominal Metaphors**

2.1 Introduction

Chapter 1 outlined the need for a better understanding of EAL readers' reading comprehension, and the value of studying it by measuring reading processing. Ample evidence shows that any differences between the reading comprehension of EAL learners and monolinguals are linked to differences in vocabulary, not reading fluency (Melby-Lervåg & Lervåg, 2014; Spencer & Wagner, 2017). Within vocabulary, language group differences seem widest on tests that assess the understanding of figurative and multi-word expressions (Kan, 2014; McKendry, 2014; Smith & Murphy, 2015; Verhallen & Schoonen, 1993, 1998). Experiment 1, reported in this chapter, will study these comprehension differences by focusing on EAL learners' metaphor processing as a compelling example of phrase-level processing in reading: metaphors require online processes such as meaning integrating (Coulson & Van Petten, 2002) and inhibition of incompatible information (here, of overly literal interpretations) (Tartter et al., 2002). These processes are relevant for reading comprehension more generally, for example when processing ambiguous words (Rodd, 2017) or processing meaning relationships within sentences (Coulson & Van Petten, 2002), and they are challenging for children (Connor et al., 2015; Levorato et al., 2004; Oakhill et al., 2016). Metaphor processing thus provides a window into important parts of children's reading processing where EAL learners likely differ from monolinguals.

2.1.1 Children's metaphor comprehension and processing

Children's metaphor comprehension has traditionally been investigated through the use of offline tasks such as asking children to match metaphors with a literal paraphrase, or

to explain metaphors verbally (Hessel & Murphy, 2018; Johnson & Pascual-Leone, 1989; Özçalışkan, 2007; Stites & Özçalişkan, 2013b; Waggoner & Palermo, 1989; Winner, Rosenstiel, & Gardner, 1976). Offline studies were able to show that metaphors are hard to understand for monolingual English children between 6 and 11 years of age (Cicone, Gardner, & Winner, 1981; Waggoner, Palermo, & Kirsh, 1997) and 5–7-year-old EAL learners (Hessel & Murphy, 2018). However, it has been suggested that this may be due to task effects as when presented in story context, metaphors do not cause any comprehension difficulties even to younger monolingual English children aged 5 years (Özçalışkan, 2007; Stites & Özçalişkan, 2013a). On the other hand, those latter studies may have failed to show metaphorness effects since they lacked literal controls against which to compare the comprehension of metaphors. Overall, offline studies suggest that metaphor comprehension may be a challenging task for children, particularly for EAL learners.

Existing offline studies of children's metaphor comprehension did, however, not directly tap children's processing, which makes it hard to know when the difficulties arise. There are a handful of eye movements studies showing that for adults, additional processing efforts are apparent in increased rereading of figurative phrases, for example on metaphors (Olkoniemi, Ranta, & Kaakinen, 2015) or metonymies (McElree, Frisson, & Pickering, 2006). In an eye movement study with Spanish-English bilingual adults, Heredia et al. (2016) replicated the effect of metaphors on rereading but also showed that for bilinguals with Spanish as their dominant language, metaphors also slowed down initial reading. This indicates that processing efforts on metaphors may be higher for some bilingual learners depending on language exposure, and possibly resulting differences in proficiency (Heredia & Cieslicka, 2016). Segal et al. (2018) compared

adult EAL learners to English monolinguals in their efficiency in making semantic judgements about metaphors when they were presented either to the right visual field/ left hemisphere, or the left visual field/ right hemisphere. They found that EAL learners' efficiency of processing metaphors in the left hemisphere was limited by smaller vocabulary size, indicating that vocabulary gaps can limit EAL learners' ability to retrieve meaning relationships from the left hemisphere up into adult years (Segal & Gollan, 2018). It remains to be seen how metaphor processing may differ between EAL and monolingual children. Based on self-report, explanations for EAL learners' difficulty with figurative and multi-word phrases included suggestions that they fail to notice them (Kan, 2014; Martinez & Murphy, 2011) or struggle to fully integrate their meanings into the context (Burgoyne et al., 2013). At the same time, EAL learners may also be able to compensate some difficulties through increased executive control or skilful inferencing (Raudszus et al., 2017; Shahar-Yames & Prior, 2018). Experiment 1 will be the first study to measure children's metaphor comprehension during processing, and to ask how it may differ between EAL learners and monolinguals.

2.1.2 Metaphor types

The time course and difficulty of processing varies for different metaphors. For example, a speaker's familiarity with a metaphor can make accessing either the metaphorical or literal meaning easier (Giora, 1999). Some metaphors are processed as fast as literal controls (McElree et al., 2006), while others are processed more slowly due to increased semantic processing demands (Coulson & Van Petten, 2002). For the present experiment, it was thus clear that the choice of metaphors would influence the kind of post-lexical processing that could be observed. To provide a fuller picture of how figurative language modulates children's reading, I thus decided to include two

metaphor types that I expected to pose different amounts of processing difficulty based on a review of the literature, as discussed below.

The first metaphor type are nominal metaphors, such as the phrase *grades are lotteries*. Nominal metaphors take the form (noun A is noun B) and compare two entities that are not usually linked. Understanding them is theorised to rely on integrating the meaning of the two nouns through active reasoning, such as analogical thinking (Gentner, Bowdle, Wolff, & Boronat, 2001; Kintsch, 2008). This account is supported by strong correlations between general reasoning skills and better comprehension as well as production of nominal metaphors (Beaty & Silvia, 2013; Billow, 1975; Johnson, 1991; Kogan, Connor, Gross, & Fava, 1980), and by similarities in brain activation in response to nominal metaphors and literal meaning integration (Coulson & Van Petten, 2002). An additional difficulty in understanding these metaphors may be literal interferences – for example, if the phrase *grades are lotteries* was interpreted overly literally as students buying lottery tickets to get grades, even if only momentarily (Tartter et al., 2002). Literal interference may be particularly common in children's comprehension of figurative language, as manifest in children's frequent literal misinterpretations of metaphors (Johnson & Pascual-Leone, 1989), and, amongst less skilled comprehenders, of idioms (Levorato et al., 2004). Due to both meaning integration and literal interference, nominal metaphors can thus be expected to place increased demands on children's reading processing compared to literal language.

Processing demands may, however, be much lower for the second metaphor type, i.e. verbal metaphors such as the phrase *hours fly by*, which take the form (verb + noun). Verbal metaphors are often cited as examples of *conceptual metaphors.* Conceptual metaphors are believed to be based on systematic associations between

abstract concepts and concrete experiences through which speakers make sense of the world (Lakoff & Johnson, 2003). For example, the metaphor *hours fly by* would be based on understanding the abstract idea of time through the concrete experience of moving through space (Lakoff & Johnson, 2003). The systematicity of these associations is supported by corpus (Semino, 2008) as well as behavioural studies (Boroditsky & Ramscar, 2002; Casasanto, 2008; Casasanto & Jasmin, 2012; Fuente, Santiago, Román, Dumitrache, & Casasanto, 2014). For example, people's understanding of references changed depending on their immediate spatial experience (Boroditsky & Ramscar, 2002). When talking about past and present, speakers of English and Mandarin systematically used gestures in line with the time-space metaphors of their language (Casasanto & Jasmin, 2012). Conceptual metaphors underlying verbal metaphors thus seem to habitually inform the speaking and thinking about abstract concepts. This systematicity can be expected to ease the processing of verbal metaphors in two ways: first, integrating meanings can be expected to require little to no extra processing effort if they are closely associated in a speaker's memory (Kintsch, 2008). Secondly, a close semantic overlap between meanings has been found to reduce the likelihood for initial misinterpretations (Pickering & Frisson, 2001), as seems also true for literal misinterpretations: verbal metaphors such *grasp an idea* are read faster when the literal meaning of *grasping* has been primed through doing or imagining the action, instead of slowed down as would be expected for literal interference (Wilson & Gibbs, 2007).

To review, nominal metaphors are theorised to be based on a singular comparisons, while verbal metaphors (often also referred to as conceptual metaphors) are theorised to be based on systematic, thus predictable, links between concepts, for example,

between time and motion (as in *hours fly by*) (Glucksberg & Keysar, 1990; Lakoff & Johnson, 2003). This difference can be expected to lead to increased efforts in post-lexical processing for nominal but not verbal metaphors compared to literal language, both in terms of literal interference and efforts in meaning integration. So far, however, the two metaphor types have only been examined in an offline study of children's listening comprehension task, without directly tapping processing (Hessel & Murphy, 2018). The present study is the first to directly compare verbal and nominal metaphors in reading processing.

2.1.3 The current experiment

The current experiment aimed to elucidate how children process metaphor during reading, and how that differs between metaphor types and language groups. Children's metaphor processing when reading them in texts was investigated through a combination of comprehension questions and eye movement recordings. Metaphor comprehension and processing are then compared across verbal and nominal metaphors to test if they indeed place different amounts of processing demands on children's reading. EAL learners and monolinguals were compared in their metaphor processing and comprehension as well as on tests of reading and language to see what language group differences are revealed by standardised and online reading tasks. All predictions for the current study were pre-registered prior to data analysis on the online platform *Open Science Framework* (see pre-registration document in Appendix A).

The following research questions were addressed:

1. Do EAL and monolingual children differ in their comprehension-related skills (vocabulary size and depth, reading comprehension), but not in their reading fluency?

EAL learners were expected to show weaker comprehension-related skills (vocabulary size and depth, reading comprehension), but not weaker reading fluency than monolinguals.

2. Do children read and comprehend metaphors differently from literal control phrases?

Children were expected to show more difficulty comprehending sentences containing metaphors than literal controls. The difficulty is expected to arise in eye-tracking measures that indicate post-lexical processing, that is, in increased and longer rereading of metaphors compared to their literal controls.

3. Does the type of metaphor (verbal vs. nominal) influence how children read and comprehend metaphors compared to literal controls?

Nominal metaphors were expected to place stronger demands on post-lexical processing compared to literal controls than verbal metaphors, manifesting in interactions between metaphorness and metaphor type that indicate increased and longer rereading in the nominal condition. This difficulty may also translate into lower comprehension of nominal metaphors.

4. Do children read and comprehend metaphors differently if they are learning EAL, or depending on their reading fluency and vocabulary size?

EAL readers were expected to show weaker comprehension overall, and particularly on metaphors. In online reading, this difficulty may either show in increased difficulties to process metaphors (a greater increase in rereading times for EAL readers on metaphors compared to monolinguals) (Burgoyne et al., 2013), or in a lack of noticing the need to process them (EAL readers showing no difference in their rereading between metaphors and literal controls) (Kan, 2014).

Alongside group differences, I also examined how individual differences in reading fluency and vocabulary were associated with metaphor processing and comprehension. These analyses were exploratory and were only conducted when the large heterogeneities within language groups became apparent. Results will be discussed briefly in this chapter and more fully in Chapter 3, alongside theoretical considerations and an additional dataset from Experiment 2.

2.2 Method

2.2.1 Design

This experiment has a mixed design with two within-subject manipulations: metaphorness (metaphor vs. literal) and metaphor type (nominal vs. verbal metaphors). The effect of these manipulations was compared between subjects of different language groups (EAL vs. monolinguals) and, in exploratory analyses, by considering associations with individual differences (reading fluency and vocabulary size). Dependent variables were children's eye movements on metaphors and literal controls, and their performance on subsequent comprehension questions. Differences in reading and language were assessed using standardised tests.

2.2.2 Participants

Forty-six children in Year 3 or 4 were recruited for this study. Recruitment was done locally through convenience sampling. The risk of fieldwork and ethical questions were assessed (see Appendix B). Following ethical approval from the CUREC (see Appendix C), primary schools in Oxford were contacted via email to solicit participation in the study. Three schools decided to take part.

To be eligible to participate, EAL children needed to have resided in the UK for at least 1 year. Two EAL children were excluded because they had only been in the UK for under a year. All children needed to have age-adequate word-reading skills, which was defined as not scoring more than 1 standard deviation below the age norm on the Test of Word Reading Efficiency Second Edition (TOWRE-2) (Torgesen, Rashotte, & Wagner, 2012). In the TOWRE-2, children read aloud lists of words or nonwords, and scores are calculated as the number of items read correctly within 45 seconds. All children in the sample were within the normal range of reading ability, both for non-word reading (range of 21-62, above cut-off of 10) and word reading (range 40-93, above cut-off of 35).

Within the three recruited schools, all parents were contacted who had eligible children in Year 4, plus children from one class which included both Year 3 and Year 4. Parents were informed about the study through personal conversations and flyers (see Appendix D for two flyer versions, including a simplified handout to accompany the main flyer). Informed consent was provided by the parents and children provided assent. Children received a sticker in appreciation of their participation.

The final sample comprised 44 children (EAL, n=21, 10 girls, and monolingual learners, n=23, 14 girls) with a mean age of 8;08 (range = 7;00-9;08). All children had normal or corrected-to-normal vision.

2.2.2.1 Validation of EAL status

Children were initially classified as EAL or monolingual English learners based on the school's record. School classifications are based on the Department for Education 2016 definition of an EAL learner as 'a child [that] was exposed to the language during early development and continues to be exposed to this language in the home or in the community'. These classifications are a useful initial indicator of EALness, but can also be misleading since the EAL label is not directly based on reported home language use, but only added later in the National Pupil Database (NPD) based on a child's first language as it is reported in the school census. However, there is no specific guidance on how the first language should be assessed for the school census (Department for Education, 2016), and a child may at least theoretically end up being classified as an EAL learner in the NPD without regularly being exposed to a home language other than English. To tap children's actual language use, it was thus decided to revisit the schools' classification using parent report. Parents were asked to fill out a Language Background Questionnaire (LBQ) and to answer questions on their children's language learning history, the languages used at home, and their child's proficiency in English and in the home language (see Appendix E for questionnaire). Based on the LBQ, two EAL children were reclassified as monolingual since their parents reported to use no home language other than English.

2.2.2.2 EAL learners and monolinguals

EAL learners and monolinguals were each other's peers: they all lived in the same school catchment area and attended the same classes at school. The two groups did not differ on age and non-verbal reasoning (see Table 1). Non-verbal reasoning was measured through the Matrices subtest from the Wechsler Abbreviated Scale of Intelligence (WASI II) (Wechsler, 2011), a task where children choose which picture out of four options a picture completes a sequence of four pictures. Matrices tasks such as the WASI II are a proxy of several cognitive control skills, such as updating (Friedman et al., 2006) and attention control (Unsworth & Engle, 2005), and are recognised as a reliable indicator of non-verbal reasoning (Ackerman, Beier, & Boyle, 2005). Both groups performed close to average on the task, as indicated by the standard *t*-scores.

Table 1: Language groups on age and non-verbal reasoning.

	Monolingual	EAL	Group comparisons	Cohen's d
Age	8;08 (0;08)	8;10 (0;06)	$t(40.7) = 0.67$, $p = 0.51$	0.16
Non-verbal reasoning				
Raw score	15.2 (4.0)	14.5 (5.2)		
			$t(39.4) = -0.52$, $p = .61$	0.20
Standard score	51.6 (7.4)	49.8 (11.5)		

*Notes: Round means. Standard deviations in brackets. * p<.05 ** p<.01 *** p<.001*
The non-verbal reasoning raw scores are the number of correctly completed sequences before three consecutive errors. The standard scores are t-scores with a mean of 50 and a standard deviation of 10.

2.2.2.3 EAL children's language background

The LBQ provided further information on EAL children's language background, beyond allowing to revisit their EAL classification. Gaining insights into children's language background is particularly important when studying EAL learners since the

term encompasses multiple different learning situations (Murphy, 2014). Unfortunately, only 47.8% of parents completed the questionnaire, which means the following cannot tell the whole story. The available data can, however, give an indication of the heterogeneity of the sample. EAL learners in the sample spoke many different home languages. Two children spoke Swahili, and two children spoke Spanish. Each of the other children spoke one of the following: Catalan, Malayalam, French, Italian, Hindi, a not further specified Ugandan language, and Polish. EAL learners generally started acquiring these home languages earlier (M = 0;04, SD = 1;03) than they started learning English (M = 1;0, SD = 1;07). Nevertheless, parents judged their children's language skills to be stronger in English (M = 3.6, SD = 0.5 on a scale of 4) than in their home language (M = 2.9, SD = 1.2 on a scale of 4). This makes sense given how much EAL children were reported to use each language: most of children's free-time activities took place in English (M = 80%, SD = 12%) rather than in the home language (M = 20%, SD = 12%). The home language was used to very different degrees at home (range 7-83% of time). That is, some of these children used their home language less than 10% of time. This, together with being schooled in an English-only school system, can explain why EAL children were generally rated to be more proficient in English than in the home language. On average, however, EAL children used their home language a little less than half the time (M = 46%, SD = 27%) compared to English (M = 55%, SD = 27%).

The language background of this sample is characteristic of minority language learners in the UK, and beyond. Many EAL learners also develop stronger skills in the majority language, due to comparatively limited input in their home language (Murphy, 2018). Just as in the current sample, UK EAL learners are known to speak many

different home languages (National Association of Language Development in the Curriculum (NALDIC, 2012), and vary greatly in their amount of language input (Murphy, 2018). As far as can be gathered from the available questionnaires, the current sample is a true reflection of the diversity of EAL learners in UK schools.

2.2.3 Materials

2.2.3.1 Assessment of metaphor reading and comprehension

Children read 40 short and child-appropriate stories while their eye movements were tracked (see Table 2 for examples, and Appendix F for a full list of stimuli). Each of the 40 stories contained either a metaphor or a literal control that differed by one word – the target word that created the metaphor. Half of the stories contained verbal and the other half contained nominal metaphors (or matched literal controls, for the literal version).

Stories were followed by two comprehension questions on separate screens: the first comprehension question asked for the reason of the final event in the story, thus asking about the metaphors or literal controls (for example: *Why does Sarah decide to go inside for lunch?*). Children could choose between two answer options, one being a paraphrase of the metaphor (for example: *Because it is soon lunchtime* for: *hours fly by*) and thus the correct answer for the metaphorical story version, and the other being a paraphrase of the literal control (for example: *Because the pigeons leave* for: *birds fly by*) and thus the correct answer for the literal story version. This way, the correct answer for one condition always served as a distractor in the other so that the same answer options were visible in both story conditions. This was done to reduce potential noise through varying distractor items. A second comprehension question was added to check attention on the first part of the story. Children had a high answer accuracy on the

second question (M = 86 %, SD = 9%) and no child performed below chance (range of subject means 62% - 100% correct), suggesting that children read stories attentively.

Stories were split into two lists following a Latin square design. Each child read all 20 verbal and 20 nominal metaphor stories, half in the metaphor condition and half in the literal control condition. Every child thus read each story once, and experienced 10 stories per condition. The stories were presented in fully randomized order.

Table 2: Example stories and questions.

Verbal	Sarah is a bit bored so she goes outside to feed the pigeons in the back garden. There are always a lot of pigeons in the garden. They have always really liked the seeds the feeds them. Sarah is happy to see all the pigeons! However, the *hours*/ *birds fly by* so she decides to go inside for lunch.	
	Comprehension question 1: Why does Sarah decide to go inside for lunch?	
	A: Because it is soon lunchtime.	B: Because the pigeons leave.
	Comprehension question 2: Why does Sarah go outside?	
	A: Because she is bored.	B: Because her mum tells her to.
Nominal	Jane is playing with her friends. Her friends like to draw but Jane thinks that playing the guitar is much more fun. Because of Jane's comments, everyone starts playing the guitar and singing along. Jane's mum sees what is happening and thinks that *Jane* is the *queen-bee*/ *music-lover* of the group and she smiles to herself.	
	Comprehension question 1: Why does Jane's mum smile?	
	A: Because Jane is the leader of her friend circle.	B: Because Jane likes playing the guitar so much.
	Comprehension question 2: What do Jane's friends like to do?	
	A: They like to draw.	B: They like to dance.

2.2.3.2 Story norming

Stories were sampled from previous studies (Cardillo, Schmidt, Kranjec, & Chatterjee, 2010; Hessel & Murphy, 2018; Özçalışkan, 2007) or newly written. All stories were adapted to integrate both the metaphor and its literal control in a plausible way. They were controlled closely in order to ensure age-appropriateness and minimise noise in the

data. Following existing procedures in the literature, stories were therefore matched in length, complexity, and episodic structure (Currie & Cain, 2015).

Length was controlled by keeping all stories to 55-60 words of length.

Story complexity was controlled using the online application Coh-Metrix Text Easability Assessor (Graesser, McNamara, & Kulikowich, 2011). The Text Easability Assessor examines texts across many factors that can make texts harder or easier to understand for children. The results showed that the stimuli have high narrativity (M = 89%, SD = 9%), high numbers of concrete and imageable words (M = 86%, SD = 17%), and high cohesion created through overlap of ideas (M = 82%, SD = 24%) and explicit causal connectives (M = 69%, SD = 36%). These numbers are similar to those reported for research looking at higher-level comprehension in 7–8-year-old children's listening comprehension (Currie & Cain, 2015). The only differences were the slightly more complex syntactic structures of the current stimuli (47%, SD = 26%, compared to M= 94% in Currie & Cain, 2015). However, syntactic structures are known to be more complex in children's written literature than in oral language (Montag & MacDonald, 2015). It thus seems appropriate to have more complex syntax in this study which focussed on reading rather than listening comprehension.

The episodic structure of stories was matched as follows: all stories were centred on a child character and a short series of events in their life. The penultimate event in the story was expressed through the metaphor or the literal control, and the last clause describes the final event - which is the outcome of the event expressed in the metaphor. This allowed for the comprehension question to ask for the cause behind the last event in the story, i.e. tapping comprehension of the metaphor or its literal control. This structure of story and questions follows a previously used paradigm (Hessel & Murphy,

2018; Özçalışkan, 2007). Basing the current stimuli on the existing paradigm afforded both an extension of existing knowledge and increased the comparability of the current study to the literature.

Several iterations were taken when stories were drafted. Texts and questions were repeatedly proofread and discussed with four speakers of British and Canadian English to make sure that the language was appropriate and acceptable.

The final set of stories and questions were piloted with EAL (n=5) and monolingual Year 4 students (n= 5) aged 8;07-9;05 (mean age 9;00). The children were drawn from the same population as the main study, but none of them participated in the final study. The pilot was run as a computer-based task and implemented in the freely available software *Psychopy* (Peirce, 2009). The children read all stories on a screen and answered the comprehension questions while their eye movements were tracked. Reading time and comprehension accuracy was monitored. The children were able to finish reading all texts without experiencing fatigue and answered comprehension questions to an appropriate level of accuracy (range 50-89 %, M = 69%), confirming the appropriateness of the task for the age group.

2.2.3.3 Metaphor norming

Metaphors were chosen to be decomposable but not lexicalised. This means that the overall meaning of metaphors can be understood from the individual words although the metaphors' exact wording is not conventionalised (Siyanova-Chanturia & Martinez, 2015). This excludes, for example, phrases such as *by and large*: if the figurative meaning "when everything about a situation is considered together" is unknown, the individual words *by* and *large* do not allow to infer it, and the figurative meaning can only be understood since it has been conventionalised (as manifest in its high phrase

frequency of 4.8 hits per million in the British National Corpus, 2007). The current experiment, however, was designed to tap post-lexical processing involved in linking individual words in a figurative way. Therefore, only metaphors that were not lexicalised but decomposable were sampled from previous studies (Cardillo et al., 2010; Glucksberg & Keysar, 1990; Özçalışkan, 2007) and from an idiom dictionary (*Cambridge Idioms Dictionary*, 2006). Metaphors were checked in the British National Corpus to ensure that their whole phrase frequency was sufficiently low (M = 0.22 hits per million, SD = 1.0). Furthermore, metaphor interpretability was checked through transparency ratings in an adult norming study (for procedures see Textbox 1). Metaphors were judged to be sufficiently transparent (M = 3.4 , SD = 1.3 out of possible 5), indicating their interpretability.

For each metaphor, a literal control was created by replacing one word in the metaphor. To review, the meaning of metaphors and literal controls needed to be distinct enough so that in comprehension questions, the paraphrase of the metaphor could serve as the distractor item for the literal condition and vice versa. Literal controls were therefore chosen so that they had a different meaning than the metaphors (such as *hours fly by* compared to *birds fly by*). The adult norming study confirmed that metaphors were rated as less transparent (M = 3.4, SD = 1.3) than literal controls (M = 4.2, SD = 1.0, $t(71.5) = 3.1$, $p < .01$[1]), which confirms that the link between words in metaphors was figurative and thus less direct than in literal controls. Furthermore, both metaphors and literal controls were rated at acceptable levels of naturalness (M = 3.6, SD = 1.3 for metaphors; M = 3.4, SD = 1.1 for literals, difference *ns*). Metaphors and

[1] All comparisons computed using Welch's *t*-test from the stats package in R (R Core Team, 2016) to adjust for unequal sample sizes and variances (Vasishth, 2014).

literal controls were also matched on word length within a metaphor-literal pair (≤ 2 letters difference) and across conditions (M = 5.9, SD = 1.6 for metaphors; M = 6.2, SD = 1.7 for literal controls, difference *ns*). Word frequency of replaced words did also not differ between metaphors and literal controls (M = 7.6, SD = 2.4 log frequency in the British National Corpus for metaphors; M = 8.1, SD = 2.2 for literal controls, difference *ns*).

Metaphors existed in two types: verbal and nominal metaphors. For both conditions, metaphors and literal controls were matched as described above, and differed neither in frequency (M = 9.4, SD = 1.4 log frequency for verbal metaphors compared to M = 8.9, SD = 1.2 for literal controls; and nominal metaphors M = 5.9, SD = 2.0 compared to literal controls M = 7.3, SD = 2.6, both differences *ns*), length (M = 5.9 letters, SD = 1.5 for metaphors; M = 6.1, SD = 1.8 for literal controls, difference *ns*), naturalness (M = 3.7, SD = 1.2 for verbal, and M = 3.8, SD = 1.2 for nominal metaphors, difference *ns*), or transparency (M = 4.0, SD = 1.1 for verbal, and M = 3.7, SD = 1.2 for nominal metaphors, difference *ns*).

Verbal and nominal metaphors were, however, expected to differ in the systematicity with which their ideas were mapped, since verbal metaphors are theorised to be systematic (Boroditsky & Ramscar, 2002; Casasanto, 2008; Casasanto & Jasmin, 2012; Fuente et al., 2014; Lakoff & Johnson, 2003; Semino, 2008). Since this mapping is assumed to be on the level of concepts and not words, it cannot be measured by simply testing the conventionality of a phrase. One way to measure it is to tag a corpus of texts for the linking of concepts (Koller, Hardie, Rayson, & Semino, 2008); however, the results of this method may be limited by the representativeness of the tagged corpus. An alternative is to test metaphors in terms of their predictability, since a reader's

knowledge of systematicities would arguably show in higher predictability of more systematic phrases, here in verbal metaphors (Siyanova-Chanturia & Martinez, 2015). In the current study, this was done using a cloze-test as part of the adult norming study (see Textbox 1), which confirmed that replaced words were more predictable for verbal metaphors than literal controls (M = .20, SD = .30 for metaphors; M = .04, SD = .09 for literal controls, $t(22.0) = -2.3$, $p < .05$), but not in nominal metaphors (M = .50, SD = .13 for metaphors; M = .60, SD = .15 for literal controls, difference *ns*).

Some variability was furthermore accepted to widen the pool of possible stimuli: replaced words in verbal phrases – metaphors and literal controls combined – were more frequent (M = 2.1, SD = 4.1 hits per million for verbal, and M = 0.4, SD = 0.9 for nominal metaphors, $t(40.4) = 2.57$, $p < .05$) and shorter (M = 9.1, SD = 1.3 log frequency for verbal, and M = 6.6, SD = 2.3 for nominal metaphors, $t(60.97) = -5.9$, $p < .001$) than those in nominal metaphors. Furthermore, replaced word changed the word class in ten cases (from noun to adjective, e.g. *the stores are jungles/ bustling*), and in three cases, articles preceding replaced words were replaced with personal pronouns (e.g. *the time/ his mum hurries*). Importantly, none one of these differences was significant between metaphors and literal controls, and was thus not expected to confound the metaphorness effects. Accepting the overall differences between conditions made it possible to create both nominal and verbal metaphors that were natural in the story context and matched with their literal controls, as discussed above. However, it also made it hard to directly compare reading times between verbal and nominal metaphors (Rayner, 1998). In the following analyses, verbal and nominal metaphors will thus not be compared directly, but only in terms of the metaphorness effect within each condition, for which metaphors and literal controls are closely matched.

Textbox 1: Procedures of adult norming study.

An initial pool of 50 metaphors was tested in a norming study. Adult native speakers (N=27) were recruited for participation through convenience sampling from personal and university acquaintances. The norming study was administered as an online questionnaire on the platform Qualtrics (https://www.qualtrics.com) (see Appendix G for a sample questionnaire). Test-takers rated metaphors and stories on naturalness, transparency and predictability.

Naturalness was chosen as a measure of how much the metaphors fit in the stories and in a usual story discourse (Cardillo et al., 2010). Test-takers were asked to rate how natural metaphors and literal controls sounded in the story context by choosing a number between 1 and 6 (from 1 = " Very unnatural" to 6 = " Very natural"). Naturalness was defined as "how normal an utterance sounds in the context it appears in, or how likely it is that a speaker might spontaneously express an idea in this particular manner".

Transparency was furthermore assessed to operationalise the degree to which metaphors required speakers to link distinct concepts, and to judge if they were interpretable. Transparency was defined as the degree to which "the individual words and the overall meaning of the phrase are related" (Skoufaki, 2006). Test-takers were asked to rate the link between a metaphor and a paraphrase of its meaning from 1 - 'Totally unrelated' to 6 - 'Very related'". The scale steps were explained to test-takers as ranging from no relationship (low transparency, e.g. *by and large* for 'when everything about a situation is considered together') to a figurative link (medium to high transparency, e.g. *reach the end of the road* for 'for a situation to reach its end') to a direct literal link (high transparency, , e.g. *do away with your fright* for: 'get rid of your fear').

Predictability was furthermore assessed since low predictability is known to slow down eye movements (Rayner, 2009). To assess predictability, test-takers were asked to fill out a cloze-test. In the cloze-test, one word had been deleted from the story. Deleted words were the first and the second word in the metaphor, for example _____*flies* or *time* _____. Survey takers were asked to insert the word that seemed appropriate to them.

2.2.3.4 Interest areas and eye-tracking measures

The eye-tracking record was split into the following three interest areas using the SR Research Data Viewer and Experiment Builder softwares (SR Research, 2009a): the first word of the metaphor, the second word of the metaphor, and the three words following the metaphor (see Table 3). Multiple interest areas are a common tool in eye movement studies that research post-lexical processing (cf. for example Wonnacott, Joseph, Adelman, & Nation, 2015), since effects can be delayed and only show on subsequent regions (Pickering et al., 2004), particularly in child readers (Joseph et al., 2008).

The length of target words ranged from 3-10 letters. To avoid skipping of shorter interest areas due to pre-processing in parafoveal preview (Rayner, 2009), neighbouring words were included in interest areas of smaller words to ensure that all areas included at least five letters, following procedures in the literature (cf. for example Eilers, Tiffin-Richards, & Schroeder, 2018).

Table 3: Example of interest areas.

		Word 1 of metaphor		Word 2 of metaphor	Post-metaphor region
Verbal	Metaphor	/ hours/		/ fly by/	/ so she decides/
	Literal control	/ birds/		/ fly by/	/ so she decides/
Nominal	Metaphor	/ grades/	/ are/	/ lotteries /	/ and she feels/
	Literal control	/ grades/	/ are/	/ misleading/	/ and she feels/

Based on those interest areas, eye movements were distilled into eye-tracking measures using the SR Research Dataviewer software (SR Research, 2009a). To review, eye-tracking measures can be divided into early and late measures (Pickering et al., 2004), with early measures indicating early linguistic processes and later measures indicating post-lexical processing (Rayner, 2009). Since the experiment focused is on post-lexical processing, I chose to concentrate on a larger number of late measures. The following eye-tracking measures were computed for each interest area:

Early:

- Gaze duration: The summed duration of all fixations on one word or region before first exiting that word or region.

Late:

- Rereading duration: The summed duration of all fixations on one word or region after exiting that word or region the first time.

- Rereading probability: The likelihood to reread a region.
- Regression in probability: The likelihood to go back into a region from a region to its right.
- Regression out probability: The likelihood to go back from a region into a region to its right.

Both early and late:

- Total time: The overall summed duration of all fixations on one word or region.

2.2.3.5 Apparatus

The stimuli were presented using Experiment Builder software (SR Research, 2009b) and eye movements were recorded using an Eyelink 1000 eye tracker at a rate of 1000 Hz. Sentences were presented in black, Courier New font, size 18, with double spacing, and on a grey background. Children read the sentences silently from a 15" Dell monitor, set at a refresh rate of 60 Hz with 1024 x 768 resolution, at a viewing distance of 60 cm. Although the children read with both eyes, only the movements of the right eye were monitored, as is standard procedure in reading eye-tracking. A chin and forehead rest was used to minimise head movements. To further minimise noise in the eye movement data, the camera was repeatedly calibrated on a five-point grid and every trial started with drift corrections to make sure calibration was sufficiently precise (< 0.5° of accuracy).

2.2.4 Assessments of reading and language

Each child in the study completed the following four standardised assessments.

Reading fluency was tested using the TOWRE-2 that was already introduced in the participants section (Torgesen et al., 2012). The TOWRE-2 was chosen as a quick and child-friendly test that is widely used in reading research, both with monolingual and bilingual learners (Bowyer-Crane et al., 2016; Kim, Wagner, Kim, & Wagner,

2015; Lervåg, Bråten, & Hulme, 2009; Mancilla-Martinez, 2016; Wonnacott et al., 2015).

Vocabulary size and depth were tested with two sub-tests of the Test of Word Knowledge (TOWK) (Wiig & Secord, 1992). For vocabulary size, children completed the Expressive Vocabulary subtest where children name objects and actions they see on pictures. Their score reflects the number of correctly named items before five consecutively misnamed and unnamed items. For vocabulary depth, children completed the Multiple Contexts subtest where children are given one word (e.g., *glasses*) and have to verbally or non-verbally provide two meanings or contexts for the word (e.g. *glasses you wear* and *glasses you drink out of*). The score represents the number of words for which two meanings were provided before the child gives five consecutive wrong or incomplete answers. The TOWK has been used in figurative language research with EAL primary school children (McKendry & Murphy, 2011; Smith & Murphy, 2015) which suggests its appropriateness for the population and research area.

Reading comprehension was assessed using an adapted version of the York Assessment of Reading for Comprehension (YARC) (Snowling et al., 2009). The YARC was adapted from a one-on-one interview into a pen-and-paper test that could be used in classroom testing. Children had 20 minutes to read up to three texts and answer the corresponding questions (Form A level 4, 3 and 5, see Appendix H). The YARC was chosen since it taps both literal and inferential reading comprehension (Snowling et al., 2009). Furthermore, the YARC has been used in similar research projects looking at EAL reading and vocabulary (Clancy, 2010; McKendry, 2014) and using it in the current study thus allows for better comparison to previous research.

2.2.5 Procedure

The equipment was set up in a quiet room for individual testing. In coordination with the class teachers, one child at a time was asked if they were willing to participate in the study. Having secured their assent, they walked to the experiment room with the researcher. The walking time and first minutes in the room were spent familiarising the child with the purpose and procedures of the experiment and the equipment. Testing started with the TOWRE-2. Children then did the eye-tracking experiment with a total of three breaks after reading approximately 10 stories each time. Breaks were filled with one of the other three standardised measures (WASI matrices, TOWK Expressive Vocabulary and Multiple Contexts). Each session lasted approximately 60-80 minutes. If necessary, children were tested in two sessions with a break in between to avoid fatigue if they showed early signs thereof.

2.3 Results and Discussion

2.3.1 Analytical strategy

Children's reading and comprehension of metaphors was assessed through a combination of eye-tracking and comprehension questions. Comprehension questions captured children's comprehension, and eye movement recordings documented children's on-line reading patterns. The initial eye movement record was read into the SR Research Data Viewer software (SR Research, 2009a). Following conventions in the literature, extremely short or long fixations (< 80ms and > 1200ms) were deleted or merged with adjacent fixations (short fixations at < 1°distance) using the Clean Data function in Data Viewer. This removed 1.7% of the data. Gaze drift (fixations recorded

outside of interest areas due to eye-tracking inaccuracies) was corrected manually by moving fixations upwards or downwards within a range of 3 pixels. The initial eye movement record was distilled into the following eye-tracking measures for the three interest areas (i.e., the first and second word of the metaphor, and the post-metaphor region) using the Interest Area Report function in Data Viewer.

The data were analysed using R (R Core Team, 2016) using *t*-tests to answer research question 1, and mixed effect models (MEM) to answer research question 2-5. MEM were chosen since they hold many advantages for psycholinguistic data (Baayen, 2008). MEM model both fixed effects (similar to independent or predictor variables in ANOVA or regressions) as well as random effects (thereby accounting for systematic variance) (Baayen, Davidson, & Bates, 2008). Random effects can be defined both for subjects and items, different from ANOVA and regression models that model only subject variance. MEM are thus better at controlling for item variance, which is particularly valuable when multiple observations are nested within both items and children, as in the current study. Furthermore, MEM can model both categorical and continuous predictors and are more robust than ANOVAs towards non-normal distributions, missing data, and unbalanced groups (Baayen, 2008). For all these reasons, MEM have become the gold standard in the analysis of eye-tracking data and were hence used in the following analyses.

Eye movement measures were both continuous (e.g. the total time spent on a word) and binary (e.g. regression in probability, the likelihood to go back into a word). Linear MEMs were computed for continuous measures using the lmer command, and binomial MEMs for binary measures using the glmer command, both in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Predictors were entered in the

following way: categorical predictors were defined using effect coding (.5 and -.5) so that the intercept represents the overall mean. Continuous predictors were tested for collinearity (Zuur, Ieno, & Elphick, 2010), and centred (by subtracting the mean) and rescaled (dividing by 2 standard deviations) to support convergence. Finally, reading times were converted from milliseconds to seconds, and log-transformed since this proved the best transformation to achieve normality, as confirmed through the BoxCox power transformation (Keshtiari & Vasishth, 2013).

One question that needed to be addressed was how to define random effects in the MEMs. Several options were available: A maximal random structure (including intercepts and slopes for all fixed effects), a minimal random structure (intercept only for each subject and item), or a data-based solution (for example by statistically testing for the appropriateness of each slope) (Bates, Kliegl, Vasishth, & Baayen, 2015). So far, there is no consensus in the field on which option is preferable. On the one hand, complex random effects frequently create convergence errors – in other words, the models are not computable (Eager & Roy, 2017). However, a maximal random structure also promises best protection from false positives through subject or item-specific error (Barr, Levy, Scheepers, & Tily, 2013). For example, a random slope for metaphorness by items can account for the particularly strong influence of one item on the overall reading time in the metaphor condition, and thereby avoid false positives based on single-item (or participant) effects. This is important given that recent simulation studies have indicated a high risk of false positives particularly in eye movement analyses (von der Malsburg & Angele, 2016). It was thus decided to aim for a maximal random structure (that is, containing both intercepts and slopes nested within subjects and items). When a maximal model failed to converge or showed bad fit to the data (in

form of non-normal residuals), singularity was checked to see if the random structure should be simplified. If singularity was a problem, the random structure was reduced, starting with the least explanatory[2] and most highly correlated higher-order term in the random structure[3]. If singularity was not the source of the problem, the package *numderiv* was used to improve the precision of the gradient estimate, and to restart the model from the more precise start value (Gilbert & Varadhan, 2016). If a more precise gradient estimation did not solve the problem, the model was rerun with different optimizers using the package *optimx* (Nash, 2014). Following this procedure, all models were computed with the most complex random structure that proved possible.

Another question was how best to decide if a result was significant. *P*-values are not directly outputted for linear lme4 models since the degrees of freedom for the *t*-statistic are difficult to determine (Baayen, 2008). In absence of directly accessible *p*-values, there are different options to check the significance of an effect (again, without a clear consensus within the field on what constitutes the best method). One option is to determine significance directly from the model *t*-values by checking if they lie beyond a two-tailed cut-off criterion at $|t| > 1.96$ (Vorstius, Radach, Mayer, & Lonigan, 2013). Another option is to use type-II model comparisons, as implemented in the Anova function of the car package (Fox, Friendly, & Weisberg, 2013). One problem with both approaches is that they can perform very poorly for a smaller sample size (Halekoh & Højsgaard, 2009). A more robust option is the Kenward-Rogers approximation as implemented in the *pbkrtest* package. The Kenward-Rogers provides *p*-values by

[2] For example, in a list of random effects that explain a variance of either "0.21" (random 1), "0.39" (random 2), or "0.003" (random 3), the random effect 3 would be considered to have the smallest explanatory power, and would be a first candidate to be taken out.
[3] For example, if there was a correlation of "1" between the random main effect for metaphorness and the metaphorness-metaphor type interaction nested within subject, the interaction effect would be taken out first, since it is the higher-order term of the two highly inter-correlated random effects.

approximating the mean and variance of the *F*-statistics and matching moments with an *F*-distribution, thereby obtaining the degrees of freedom of the error for each effect (Halekoh & Højsgaard, 2009). This makes it possible to address potential biases in the df estimation, for example when the covariance matrix is underestimated, and to be more conservative with small samples where necessary. Following the same logic as for the random structure, the latter, more conservative approach was chosen to compute all reported *p*-values for linear MEM. Alongside each *p*-value, an effect size is reported. Following reporting conventions in the published literature, different types of effect sizes were chosen for different models. For *t*-tests, Cohen's d is reported, for linear MEMs the beta coefficients, and odds ratios for binomial MEMs.

Once the final model had been established, model fit was checked. This was done visually by inspecting if residuals were normal and heteroscedastic (using base R for linear MEM (Baayen, 2008) and the *DHARMa* package for binomial models (Hartig, 2016). Finally, a bootstrap operation was used to check for outliers, using the package *influence.ME* (Nieuwenhuis & Grotenhuis, 2012). The *influence.ME* package takes an existing MEM and reruns it several times, removing either one subject or one item from the model calculations each time. The beta estimates and significance values are then compared across all of these leave-one-out simulations. Large differences of beta and/or significance values between the original model and its leave-one-out simulations would suggest that an item or child had a large influence on the model outcome. All these checks showed satisfactory model fit for all MEM reported below. The following presents the results by research questions.

2.3.2 Do EAL and monolingual children differ significantly in their comprehension-related skills (vocabulary size and depth, reading comprehension), but not in their reading fluency?

Children's language skills were assessed through standardised tests of reading fluency (TOWRE-II), vocabulary size and depth (subtests of the TOWK), and reading comprehension (YARC). Table 4 summarises the descriptives. Standard scores indicate that both language groups performed above average for reading fluency and that monolinguals were slightly above average on both vocabulary tests, while EAL learners were slightly below. To see if EAL and monolingual children differed on these measures, the groups were compared using Welch's *t*-tests (see Table 4). Welch's *t*-test was chosen over a traditional student's *t*-test since the former can correct for unequal variance across groups (Vasishth, 2014). As expected, EAL children had smaller and less deep vocabulary and weaker reading comprehension than their monolingual peers, but did not differ significantly from them on their reading fluency.

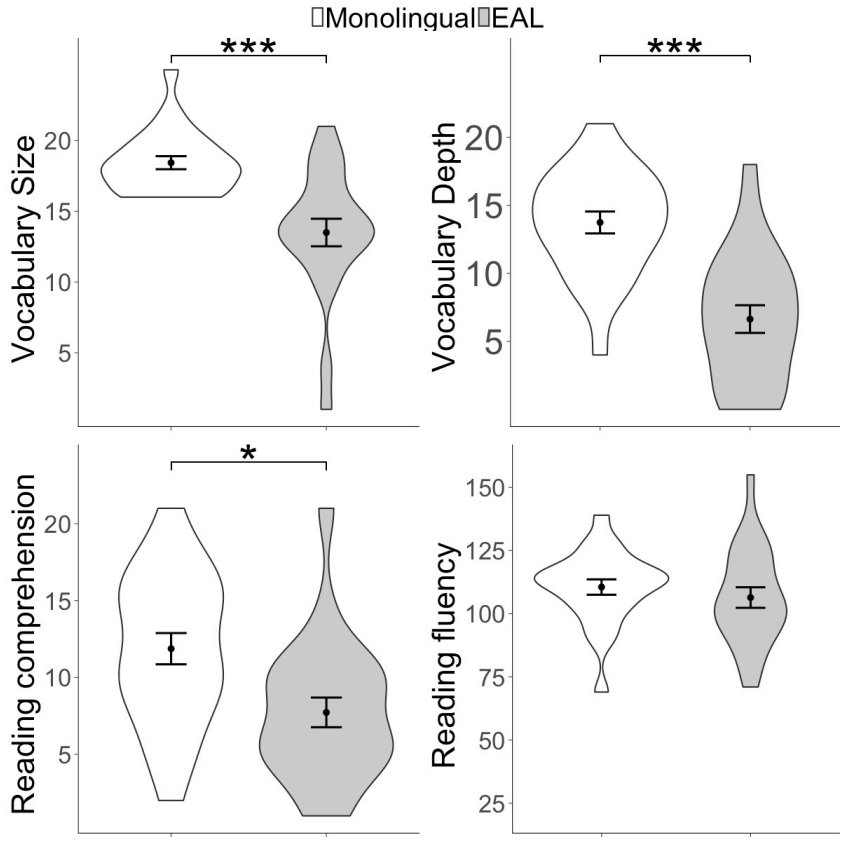
Table 4: Language groups compared on assessments of reading and language.

	Monolingual	EAL	Group comparisons	Cohen's d
Reading fluency				
Raw score	111 (14)	106 (19)	$t(39.3) = -0.82$, $p = 0.42$	0.3
Standard score	121 (11)	114 (14)		
Vocabulary size				
Raw score	18 (2)	14 (5)	$t(30.0) = -4.58$ ***	1.1
Standard score	106 (12)	90 (18)		
Vocabulary depth				
Raw score	14 (4)	7 (5)	$t(40.4) = -5.47$ ***	1.6
Standard score	104 (10)	90 (17)		
Reading comprehension *raw score*	12 (5)	8 (4)	$t(43.0) = -3.0$ *	0.9

*Notes: Round means. Standard deviations in brackets. All raw scores. * p<.05 ** p<.01 *** p<.001. Standard scores for vocabulary size and depth are based on the mean standardised score for the range of standardised scores equivalent to the sub-test raw score. No standard scores can be provided for reading comprehension since the test was administered in a group testing procedure that differs from the procedures underlying the YARC standard scores.*

The violin plots in Figure 1 illustrate not only the mean differences (indicated through error bars), but also the distribution of scores within each group (shown by the surface area). As is clear from Figure 1, the range of vocabulary scores is wider for EAL learners than monolinguals and extends further into the lower ranges. This observation will become relevant when discussing links between individual differences in vocabulary and reading patterns in research question 4.

Figure 1: Language group differences on assessments of reading and language.



Notes: ** p<.05 ** p<.01 *** p<.001. Error dots show the mean, error bars the standard error. Surface area shows probability density function mirrored along the midline. All raw scores.*

2.3.3 Do children read and comprehend metaphors differently from literal controls?

Metaphor comprehension was captured through comprehension questions. Metaphor processing was measured through eye movements that were summarised into eye-tracking measures. To review, these eye-tracking measures summarise children's first reading of a word before moving on (*gaze duration*), how long they take to reread it (*rereading time*), and how long they take to read it overall (*total time*). Other measures summarise how likely they are to reread a word (*rereading probability*) or go straight

back into it (*regression in probability*). Table 5 - Table 7 summarise the data on each measure.

Table 5: Means for children reading the first word of metaphors and literal controls.

	Word 1 of metaphor (*grades are lotteries / hours fly by)*					
	Gaze duration	Rereading time	Regression in probability	Regression out probability	Rereading probability	Total time
Monolingual						
Verbal						
Literal	352 (264)	569 (561)	0.37 (0.48)	0.47 (0.50)	0.21 (0.41)	621 (539)
Metaphor	412 (483)	683 (608)	0.42 (0.49)	0.52 (0.50)	0.20 (0.40)	765 (728)
Nominal						
Literal	305 (213)	643 (507)	0.41 (0.49)	0.49 (0.50)	0.23 (0.42)	622 (514)
Metaphor	311 (243)	617 (535)	0.27 (0.44)	0.46 (0.50)	0.20 (0.40)	593 (517)
EAL						
Verbal						
Literal	435 (309)	590 (639)	0.31 (0.46)	0.42 (0.49)	0.20 (0.40)	682 (570)
Metaphor	466 (386)	803 (783)	0.31 (0.46)	0.48 (0.50)	0.21 (0.41)	854 (736)
Nominal						
Literal	395 (277)	575 (493)	0.27 (0.44)	0.40 (0.49)	0.23 (0.42)	625 (472)
Metaphor	421 (377)	552 (459)	0.37 (0.48)	0.45 (0.50)	0.16 (0.36)	669 (510)

Notes: Reading times in milliseconds. Rounded means. Standard deviations in parentheses.

Table 6: Means for children reading the second word of metaphors and literal controls, and for answering comprehension questions.

	Word 2 of metaphor *(grades are lotteries / hours fly by)*						Offline
	Gaze duration	Rereading time	Regression in probability	Regression out probability	Rereading probability	Total time	Comprehension
Monolingual							
Verbal							
Literal	404 (352)	550 (448)	0.28 (0.45)	0.49 (0.50)	0.28 (0.45)	673 (545)	0.79 (0.41)
Metaphor	407 (329)	546 (355)	0.23 (0.42)	0.46 (0.50)	0.31 (0.46)	658 (432)	0.67 (0.47)
Nominal							
Literal	406 (359)	740 (710)	0.32 (0.47)	0.58 (0.49)	0.32 (0.47)	838 (803)	0.75 (0.44)
Metaphor	448 (416)	891 (771)	0.42 (0.50)	0.72 (0.45)	0.38 (0.49)	1,087 (824)	0.67 (0.47)
EAL							
Verbal							
Literal	452 (429)	714 (675)	0.25 (0.43)	0.42 (0.50)	0.22 (0.42)	755 (649)	0.71 (0.46)
Metaphor	472 (349)	554 (422)	0.22 (0.42)	0.45 (0.50)	0.25 (0.43)	719 (487)	0.61 (0.49)
Nominal							
Literal	466 (376)	832 (752)	0.32 (0.47)	0.51 (0.50)	0.24 (0.43)	891 (757)	0.62 (0.49)
Metaphor	522 (538)	933 (855)	0.35 (0.48)	0.62 (0.49)	0.31 (0.46)	1,104 (987)	0.57 (0.50)

Notes: Reading times in milliseconds. Rounded means. Standard deviations in parentheses.

Table 7: Means for children reading the post-metaphor region.

	Post-metaphor region *(and she decides / so she feels)*					
	Gaze duration	Rereading time	Regression in probablity	Regression out probablity	Rereading probablity	Total time
Monolingual						
Verbal						
Literal	619 (594)	815 (719)	0.30 (0.46)	0.46 (0.50)	0.26 (0.44)	992 (811)
Metaphor	568 (513)	673 (537)	0.24 (0.43)	0.41 (0.49)	0.26 (0.44)	841 (644)
Nominal						
Literal	575 (531)	959 (955)	0.33 (0.47)	0.51 (0.50)	0.31 (0.46)	1,066 (958)
Metaphor	577 (520)	860 (756)	0.36 (0.48)	0.65 (0.48)	0.41 (0.49)	1,135 (810)
EAL						
Verbal						
Literal	715 (646)	854 (846)	0.26 (0.44)	0.43 (0.50)	0.25 (0.43)	1,081 (896)
Metaphor	687 (607)	888 (763)	0.26 (0.44)	0.36 (0.48)	0.20 (0.40)	1,008 (847)
Nominal						
Literal	698 (609)	945 (742)	0.26 (0.44)	0.48 (0.50)	0.29 (0.45)	1,147 (908)
Metaphor	607 (535)	960 (805)	0.24 (0.43)	0.56 (0.50)	0.36 (0.48)	1,143 (880)

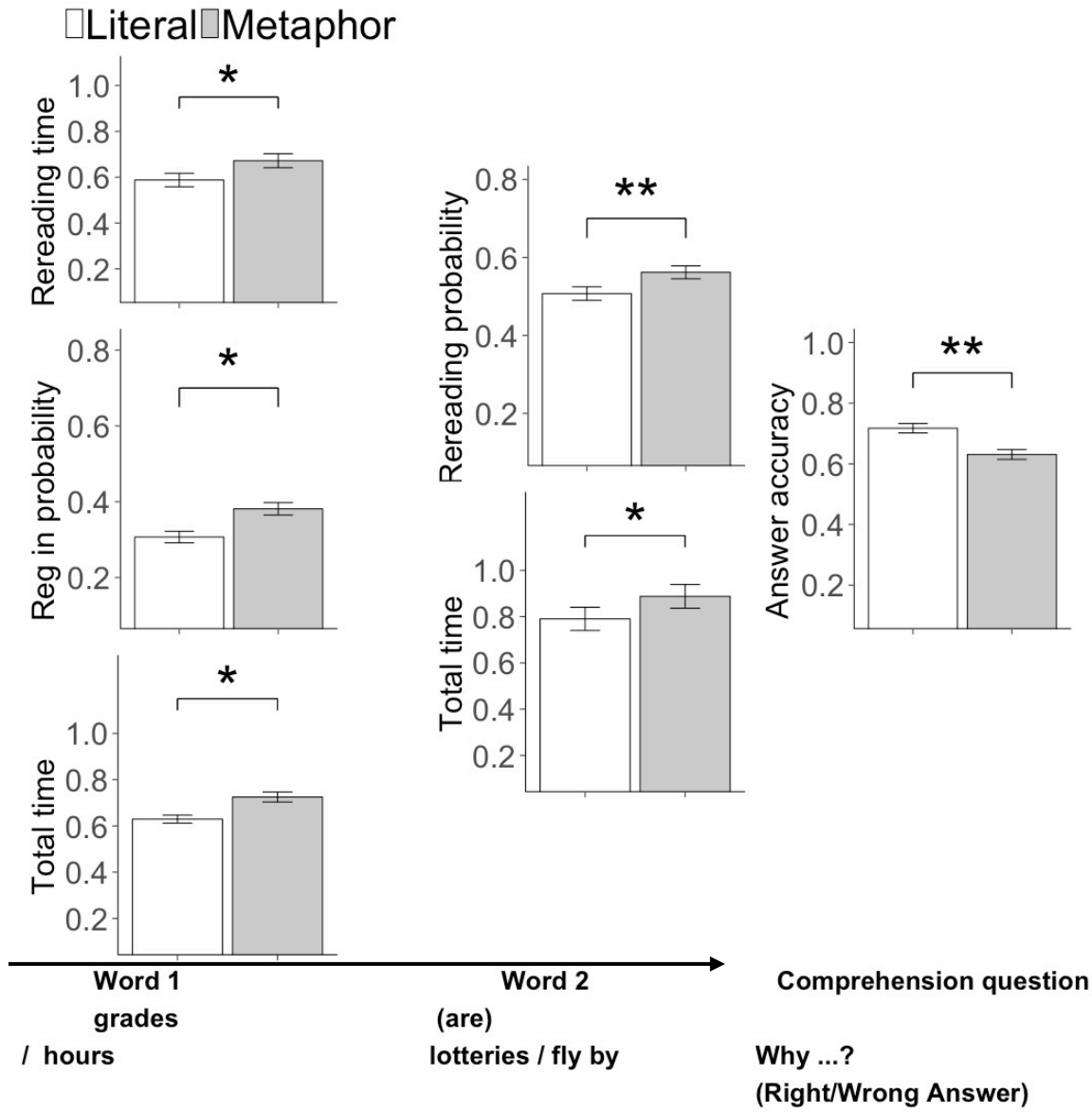
Notes: Reading times in milliseconds. Rounded means. Standard deviations in parentheses.

One model was run per measure and region, with language group, metaphorness, and metaphor type as predictors. Table 8, Table 9, and Table 10 show the results of these models. Figure 2 shows how children read metaphors or literal controls in three columns that represent the three interest areas which are organised from earlier to later position in the text, followed by children's performance on comprehension questions. As expected, children answered comprehension questions on metaphors less accurately than on literal language. Also as predicted, children took longer to read metaphors than literal controls in measures of rereading and total reading time, but not when first encountering them (*ns* on gaze duration). That is, children were more likely to regress back into the first word of the metaphor, and to reread the first and second word compared to literal controls. This indicates that metaphors required additional processing not during initial lexical access, but during post-lexical processing.

Given that the metaphorness effect largely showed in rereading patterns, I checked that there was enough data to support the analyses across children. Children reread on 755-852 trials in the three interest areas, which amounts to 46-54% of analysed trials. On the child level, the mean rereading probability was between 46-53% across interest areas, a high number relative to previous eye movements studies that focused predominantly on rereading (Meseguer, Carreiras, & Clifton, 2002; Murray & Kennedy, 1988; Schotter, Tran, & Rayner, 2014; von der Malsburg & Vasishth, 2011). Together, these numbers indicate that there were sufficient instances of rereading to afford reliable analyses, as was also supported by good results on the influence.ME bootstrap test. However, mean rereading probability varied across children, with 2-3 children rereading less than 20% of the time. This indicates that effects in rereading may be somewhat less representative across participants than measures based on first pass reading.

In answer to the research question, metaphors changed reading as expected: children understood metaphors less well than literal controls, a difficulty that show in frequent and slow rereading of metaphors during processing. Rereading effects were based on fewer instances than first-pass reading across children, but appeared reliable.

Figure 2: Children's processing and comprehension of metaphors and literal controls.



Notes: Reg in probability = Regression in probability. Error bars show standard errors.
** p<.05 ** p<.01 *** p<.001*

.

Table 8: Models using language group, metaphorness and metaphor type to predict reading the first word of the metaphor.

	Word 1 of metaphor (*grades are lotteries / hours fly by)*											
	Gaze duration		Rereading time		Regression in probability		Rereading probability		Regression out probability		Total time	
	B	*SE*	*B*	*SE*	*OR*	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*
Language group (LG)	0.13	0.08	-0.02	0.10	0.78	0.20	0.80	0.15	0.95	0.18	0.07	0.09
Metaphor Type	-0.13 *	0.06	-0.05	0.08	0.86	0.20	0.87	0.17	0.98	0.20	-0.15	0.08
Metaphorness (Met)	0.02	0.03	0.13 *	0.06	1.40 *	0.22	1.19	0.18	0.90	0.14	0.09 *	0.04
LG X Metaphor Type	0.07	0.06	-0.18	0.13	1.35	0.34	0.95	0.21	0.88	0.24	-0.02	0.07
LG X Met	-0.00	0.07	0.02	0.11	0.80	0.20	1.07	0.27	0.78	0.23	0.02	0.06
Met X Metaphor Type	0.02	0.06	-0.10	0.12	1.87 *	0.58	1.01	0.30	0.88	0.27	-0.04	0.08
LG X Met X Metaphor Type	0.18	0.13	-0.06	0.22	1.19	0.60	0.96	0.49	0.38	0.22	0.09	0.12

*Notes: Estimates are beta coefficients (B) for linear models on gaze duration and total time, and odds ratios (OR) for logit models on rereading probability and offline comprehension. * p<.05 ** p<.01 *** p<.001.*

Table 9: Models using language group, metaphorness and metaphor type to predict reading the second word of metaphors and answering questions.

	Word 2 of metaphor *(grades are lotteries / hours fly by)*												Offline	
	Gaze duration		Rereading time		Regression in probability		Rereading probability		Regression out probability		Total time		Comprehension	
	B	*SE*	*B*	*SE*	*OR*	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*	*OR*	*SE*
Language group (LG)	0.07	0.08	0.07	0.09	0.88	0.20	0.73	0.17	0.70	0.16	0.01	0.09	0.58 **	0.11
Metaphor Type	0.06	0.06	0.28 **	0.10	1.68 **	0.33	2.08 ***	0.37	1.28	0.23	0.30 **	0.09	0.78	0.14
Metaphorness (Met)	0.06	0.05	0.06	0.07	1.13	0.17	1.41 **	0.18	1.39	0.24	0.17 **	0.05	0.64 *	0.14
LG X Metaphor Type	-0.00	0.07	0.05	0.13	0.99	0.31	0.76	0.22	0.99	0.26	-0.06	0.09	0.83	0.23
LG X Met	0.06	0.07	-0.08	0.11	0.92	0.30	1.02	0.24	1.08	0.30	0.02	0.06	1.24	0.43
Met X Metaphor Type	0.02	0.09	0.21	0.15	1.46	0.44	1.67 *	0.43	1.08	0.35	0.19	0.10	1.28	0.54
LG X Met X Metaphor Type	-0.07	0.12	0.16	0.22	0.61	0.40	0.67	0.32	1.18	0.63	-0.10	0.12	0.97	0.61

*Notes: Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models. * p<.05 ** p<.01 *** p<.001.*

Table 10: Models using language group, metaphorness and metaphor type to predict reading the post-metaphor region.

	Post-metaphor region *(and she decides / so she feels)*											
	Gaze duration		Rereading time		Regression in probability		Rereading probability		Regression out probability		Total time	
	B	*SE*	*B*	*SE*	*OR*	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*
Language group (LG)	0.06	0.10	0.08	0.09	0.80	0.17	0.80	0.18	0.91	0.18	0.02	0.10
Metaphor Type	-0.03	0.09	0.18	0.10	0.95	0.33	1.83 *	0.45	1.66	0.52	0.14	0.11
Metaphorness (Met)	-0.05	0.04	-0.06	0.07	0.94	0.16	1.16	0.13	1.31	0.23	-0.03	0.04
LG X Metaphor Type	-0.00	0.07	-0.03	0.12	0.61	0.18	0.82	0.22	1.02	0.29	-0.07	0.08
LG X Met	-0.04	0.06	0.14	0.13	1.06	0.29	0.80	0.21	0.88	0.26	-0.03	0.06
Met X Metaphor Type	-0.03	0.07	0.02	0.15	1.23	0.42	2.04 **	0.47	1.93 *	0.62	0.16 *	0.08
LG X Met X Metaphor Type	-0.06	0.12	0.16	0.28	0.48	0.24	0.80	0.43	0.99	0.49	-0.07	0.12

*Notes: Estimates are beta coefficients (B) for linear models on gaze duration and total time, and odds ratios (OR) for logit models on rereading probability and offline comprehension * p<.05 ** p<.01 *** p<.001.*

2.3.4 Does the type of metaphor (verbal vs. nominal) influence how children read and comprehend metaphors compared to literal controls?

Children read two types of metaphors and literal controls: verbal metaphors such as *hours fly by* were compared to literal controls such as *birds fly by*, and nominal metaphors such as *grades are lotteries* were compared to literal controls such as *grades are misleading*. The relevant model results and descriptives are all included in the preceding Table 5-Table 10. As could be expected, metaphor type is a significant predictor of reading times. Since the main effect was added as a control variable for the stimuli heterogeneity between types only, it will not be discussed any further. Differences in the metaphorness effect across conditions was expected to show as an interaction between metaphorness and metaphor type, a comparison for which metaphors and controls had been tightly matched.

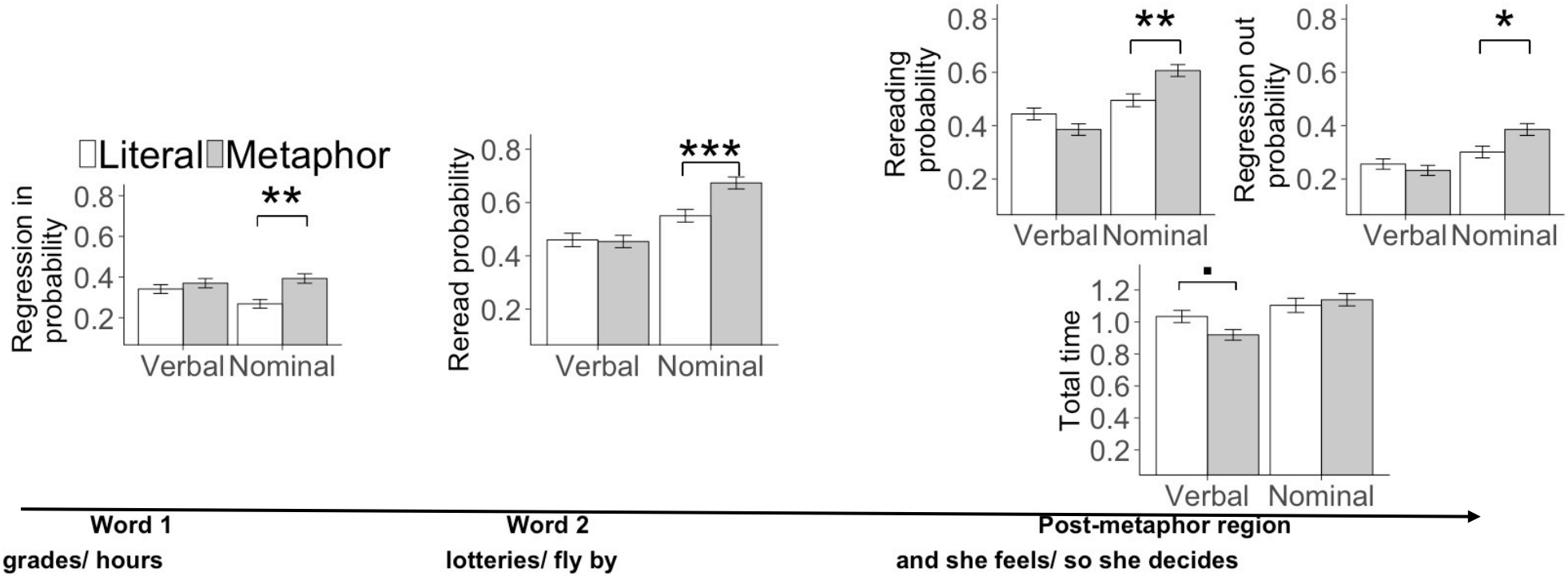
There were no interactions between metaphorness and metaphor type on offline comprehension, indicating that the metaphor type did not make a difference for how well children answered comprehension questions (difference *ns*). There were, however, significant interactions between metaphorness and metaphor type on children's reading of metaphors and the post-metaphor region. Post hoc tests were estimated using cell means coding and single degree of freedom contrasts as implemented in the *glht* function from the *multcomp* package (Hothorn, Bretz, & Westfall, 2008). These post hoc tests allowed the comparison of predicted means between metaphors and literal controls within the verbal and the nominal condition.

Figure 3 illustrates the results, again showing the time course of the effect in three columns that show children's reading from the first word of metaphors to the post-metaphor region. As predicted, the longer and increased rereading on metaphors seemed

largely driven by children reading nominal metaphors more than literal controls. That is, children were more likely to go back to read the first word of nominals metaphors than literal controls ($z = -3.0$, $p < .001$) and were more likely to reread the second word and the post-metaphor region of nominal metaphors than literal controls ($z = -3.4$, $p < .0001$ and $z = -3.0$, $p < .01$, respectively) and to return from the post-metaphor region to earlier regions ($z = -2.6$, $p < .05$). None of these effects was significant in verbal metaphors compared to literal controls. There was only one trend towards a faster reading of post-metaphor regions following verbal metaphors compared to literal controls ($z = 1.8$, $p = .07$).

To review, the metaphor type mattered for children's metaphor processing in the expected direction: reading nominal metaphors seemed required additional processing effort compared to matched literal controls - but not the reading of verbal metaphors. Children even showed a trend to speed up after having read verbal metaphors. However, these effects only show in online reading, but not in offline comprehension.

Figure 3: Significant interactions between metaphorness and metaphor type on reading metaphors and answering comprehension questions.



*Notes: Reg out = Regression out probability. * p<.05 ** p<.01 *** p<.001. Error bars show the standard error.*

2.3.5 Do children read and comprehend metaphors differently if they are learning EAL, or depending on their reading fluency and vocabulary size?

The following analyses focused on differences between children. Particularly, the question was how individual and group differences were associated with differences in children's reading, particularly on metaphors. Differences between language groups were already tested in the first models, summarised in in the preceding Table 5-Table 10. As predicted, EAL children were less accurate in answering comprehension questions – both on metaphors and literal controls – than monolinguals. Based on previous research (Hessel & Murphy, 2018; McKendry, 2014), EAL learners were furthermore expected to have particular difficulties with metaphors, which would manifest as an interaction between metaphorness and language group. However, there was no interaction between language group and metaphorness, and thus no indication that EAL learners struggled more with understanding metaphors than monolinguals. This discrepancy with previous studies may be due to older age or supporting story context in the current study (Hessel & Murphy, 2018; McKendry, 2014). Another possibility is that language group differences in online reading were masked by individual differences within language skills, as reported in previous studies (van den Bosch, Segers, & Verhoeven, 2018).

To see if language group differences would show if individual differences were considered, additional models were run with reading fluency and vocabulary added as predictors alongside metaphorness and language groups. I decided to enter only two individual difference measures since high collinearity made it impossible to add all four measures. Reading fluency and vocabulary were chosen since they are known to

uniquely contribute to reading comprehension (Gough & Tunmer, 1986), and since EAL children differed from monolinguals on vocabulary, but not on reading fluency. Table 11 - Table 13 summarise the data entered in the models.

Table 11: Model means for children reading the first word of the metaphor.

	Word 1 of metaphor (*grades are lotteries / hours fly by)*			
	Gaze duration	Rereading time	Rereading probability	Total time
Vocabulary				
Small				
Monolingual				
Literal	321 (160)	399 (184)	0.40 (0.46)	530 (159)
Metaphor	337 (168)	516 (192)	0.58 (0.52)	651 (193)
EAL				
Literal	333 (69)	406 (84)	0.41 (0.19)	510 (73)
Metaphor	336 (69)	485 (88)	0.38 (0.22)	523 (80)
Large				
Monolingual				
Literal	280 (66)	443 (80)	0.49 (0.21)	485 (73)
Metaphor	285 (72)	473 (86)	0.50 (0.23)	503 (86)
EAL				
Literal	321 (97)	433 (131)	0.38 (0.31)	499 (102)
Metaphor	321 (98)	441 (135)	0.67 (0.34)	649 (115)
Reading fluency				
Low				
Monolingual				
Literal	321 (160)	399 (184)	0.40 (0.46)	530 (159)
Metaphor	337 (168)	516 (192)	0.58 (0.52)	651 (193)
EAL				
Literal	333 (69)	406 (84)	0.41 (0.19)	510 (73)
Metaphor	336 (69)	485 (88)	0.38 (0.22)	523 (80)
High				
Monolingual				
Literal	280 (66)	443 (80)	0.49 (0.21)	485 (73)
Metaphor	285 (72)	473 (86)	0.50 (0.23)	503 (86)
EAL				
Literal	321 (97)	433 (131)	0.38 (0.31)	499 (102)
Metaphor	321 (98)	441 (135)	0.67 (0.34)	649 (115)

Notes: Back-transformed, rounded model means in milliseconds for the dependent measures on critical word of metaphor. Standard errors in parentheses.

Table 12: Model means for children reading the second word of the metaphor and answering comprehension questions.

	Word 2 of metaphor (*grades are lotteries / hours fly by)*				Offline
	Gaze duration	Rereading time	Rereading probability	Total time	Comprehension
Vocabulary					
Small					
Monolingual					
Literal	377 (219)	439 (171)	0.43 (0.58)	590 (236)	0.79 (0.49)
Metaphor	399 (175)	659 (148)	0.55 (0.52)	838 (189)	0.59 (0.51)
EAL					
Literal	332 (90)	555 (95)	0.39 (0.25)	535 (101)	0.64 (0.18)
Metaphor	366 (75)	529 (77)	0.48 (0.21)	622 (84)	0.61 (0.20)
Large					
Monolingual					
Literal	338 (92)	488 (84)	0.55 (0.28)	628 (104)	0.81 (0.25)
Metaphor	343 (75)	509 (75)	0.64 (0.26)	683 (87)	0.72 (0.26)
EAL					
Literal	355 (123)	430 (132)	0.55 (0.37)	659 (135)	0.76 (0.29)
Metaphor	386 (105)	569 (118)	0.67 (0.33)	874 (115)	0.58 (0.30)
Reading fluency					
Low					
Monolingual					
Literal	377 (219)	439 (171)	0.43 (0.58)	590 (236)	0.79 (0.49)
Metaphor	399 (175)	659 (148)	0.55 (0.52)	838 (189)	0.59 (0.51)
EAL					
Literal	332 (90)	555 (95)	0.39 (0.25)	535 (101)	0.64 (0.18)
Metaphor	366 (75)	529 (77)	0.48 (0.21)	622 (84)	0.61 (0.20)
High					
Monolingual					
Literal	338 (92)	488 (84)	0.55 (0.28)	628 (104)	0.81 (0.25)
Metaphor	343 (75)	509 (75)	0.64 (0.26)	683 (87)	0.72 (0.26)
EAL					
Literal	355 (123)	430 (132)	0.55 (0.37)	659 (135)	0.76 (0.29)
Metaphor	386 (105)	569 (118)	0.67 (0.33)	874 (115)	0.58 (0.30)

Notes: Back-transformed, rounded model means in milliseconds for the dependent measures on critical word of metaphor. Standard errors in parentheses.

Table 13: Model means for children reading the post-metaphor region.

	Post-metaphor region *(and she decides / so she feels)*			
	Gaze duration	Rereading time	Rereading probability	Total time
Vocabulary				
Small				
Monolingual				
Literal	489 (256)	501 (230)	0.28 (0.57)	641 (260)
Metaphor	460 (254)	580 (198)	0.56 (0.55)	852 (257)
EAL				
Literal	478 (102)	662 (93)	0.42 (0.21)	787 (107)
Metaphor	449 (97)	627 (98)	0.43 (0.22)	733 (104)
Large				
Monolingual				
Literal	469 (108)	691 (101)	0.55 (0.29)	889 (119)
Metaphor	457 (104)	557 (94)	0.55 (0.28)	815 (111)
EAL				
Literal	528 (135)	535 (157)	0.53 (0.36)	894 (139)
Metaphor	484 (133)	671 (152)	0.55 (0.37)	957 (138)
Reading fluency				
Low				
Monolingual				
Literal	489 (256)	501 (230)	0.28 (0.57)	641 (260)
Metaphor	460 (254)	580 (198)	0.56 (0.55)	852 (257)
EAL				
Literal	478 (102)	662 (93)	0.42 (0.21)	787 (107)
Metaphor	449 (97)	627 (98)	0.43 (0.22)	733 (104)
High				
Monolingual				
Literal	469 (108)	691 (101)	0.55 (0.29)	889 (119)
Metaphor	457 (104)	557 (94)	0.55 (0.28)	815 (111)
EAL				
Literal	528 (135)	535 (157)	0.53 (0.36)	894 (139)
Metaphor	484 (133)	671 (152)	0.55 (0.37)	957 (138)

Notes: Back-transformed, rounded model means in milliseconds for the dependent measures on critical word of metaphor. Standard errors in parentheses.

All models were rerun with vocabulary size and reading fluency as additional predictors added to language group and metaphorness. Table 14, Table 15 and Table 16 show the results of the models. Individual differences were modelled continuously, while interactions were investigated through post hoc comparisons that used point

estimates of children's reading at 1SD above or below the sample mean, here referred to as for example children with larger (+1SD) or smaller vocabulary size (-1SD). This affords clear group comparisons – e.g., between reading times of EAL and monolingual children with small vocabularies – while still benefitting from continuous predictors in the models. Considering continuous variation is preferable to a median split since it yields more accurate estimates and limits the risk of false positives (Preacher, MacCallum, Rucker, & Nicewander, 2005). Main effects showed that overall, children with higher reading fluency - as measured by the TOWRE - read more quickly, were less likely to regress back, and were more likely to answer comprehension questions correctly.

There were no main effects of vocabulary size (all effects *ns*), but vocabulary was associated with children's reading processing in interaction with metaphorness and language group. Figure 4 illustrates the three significant interactions between language group, vocabulary and metaphorness on the reading of metaphors and their context (again in three columns that show reading across the three interest areas). These interactions with vocabulary size were mirrored in a three-way interaction between language group, metaphorness and reading fluency on children's rereading time of the first word of metaphors (see Table 11 for predicted means). Like the main effects of metaphorness, these interactions showed on late eye-tracking measures that indicate post-lexical processing. Consider first the reading patterns of EAL readers. EAL readers with smaller vocabularies spent similarly long reading metaphors and literal controls across the first and second word. EAL readers with larger vocabularies, however, showed increased reading times on metaphors. This difference reached significance on EAL children's likelihood to reread the first word ($z = -3.2$, $p < .01$), while it remained

numerical on the rereading of the second word (*ns*). A similar reading pattern showed in monolinguals with smaller vocabularies who spent longer rereading metaphors than literal controls. This difference was marginal on monolingual children's rereading time of the second word ($z = 1.8$, $p = .07$), and remained numerical on the likelihood to reread the first word (*ns*). That is, both EAL children with *larger* vocabularies and monolinguals with *smaller* vocabularies adapted their rereading to metaphors. EAL readers with smaller vocabularies, on the other hand, did not modulate their rereading in response to metaphors. Longer and more frequent rereading generally indicates increased effort during post-lexical processing, such as when linking meanings or correcting interpretations. It seemed that monolinguals with smaller vocabularies and EAL readers with larger vocabularies invested such increased effort into the post-lexical processing of metaphors.

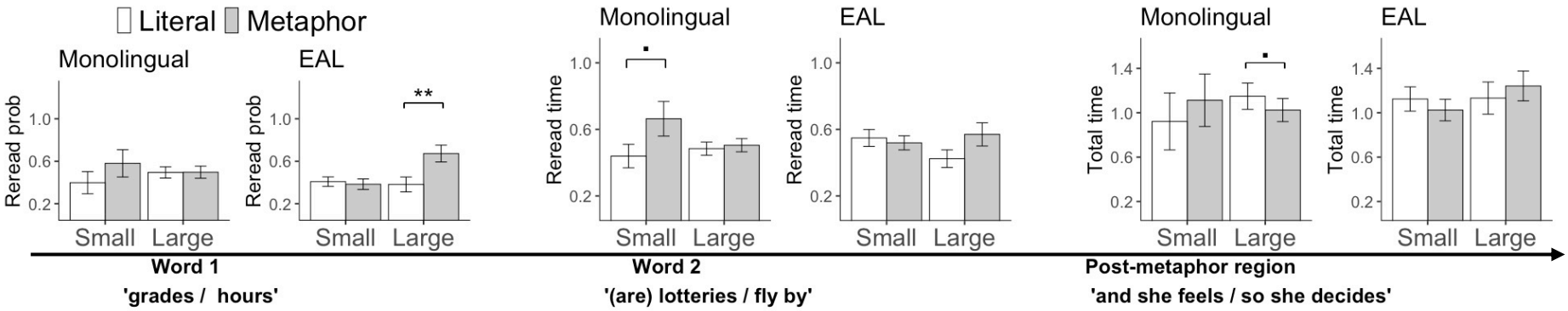
In contrast, monolinguals with larger vocabularies did not show any reliably difference in their rereading of metaphors and literal controls. This could suggest that the effect of vocabulary was simply reversed between the language groups, i.e. that reacting to metaphors was associated with larger vocabulary size in EAL children and with smaller vocabulary size in monolinguals. However, the reading patterns of the two groups were not exactly mirrored: monolinguals with larger vocabularies did not seem to reread the first word of metaphors more than literal controls, but they reread the second word metaphors longer than literal controls, at least numerically (difference *ns*). Reading time differences on higher-level manipulation may become smaller as children increase their efficiency of post-lexical processing (Tiffin-Richards & Schroeder, 2018). Perhaps the only numerical difference means that the difference between rereading metaphors and literal controls was too small to become significant in monolinguals with

large vocabularies. The idea that monolinguals wither larger vocabularies were efficient in their reading processing is supported by the finding that they even showed trends to read the post-metaphor region faster than following literal controls ($z = 1.93$, $p = .05$). However, this interpretation remains speculative at the current stage: post hoc tests seemed generally low powered, since even where significant three-way interactions gave reliable evidence of variation in reading time differences, post hoc comparisons remained non-significant in many cases. This indicates that individual differences were generally of low statistical power in current dataset, plausibly given the reasonably small sample size, combined with small effects and large variances. This makes smaller effects hard to interpret.

To review, there was no overall effect of language group on children's reading of metaphors, beyond EAL children's overall lower comprehension on the reading task compared to monolinguals. Since such large variation of language skills can mask language group differences (van den Bosch et al., 2018), additional models were run which included reading fluency and vocabulary as added predictors. Main effects showed that higher reading fluency – as measured by the TOWRE – was associated with more fluent reading and better comprehension. Interaction effects furthermore showed that monolinguals with smaller vocabularies adapted their rereading to metaphors, while EAL children did so only with larger vocabularies. Monolinguals thus seemed more sensitive to metaphors than EAL children, relative to vocabulary size. This difference may have been exacerbated by the fact that vocabulary scores within this group extended into much lower ranges, and that the estimation thus represents readers with particularly small vocabulary size (see research question 1). The pattern was mirrored in an interaction with reading fluency, suggesting that both facets of word

knowledge influenced children's metaphor processing in a similar way. Effects in monolinguals with larger vocabularies were harder to interpret due to low statistical power. It was tentatively suggested that these monolinguals were sensitive to metaphors, but that the effect was smaller and thus less detectable due to low power (Tiffin-Richards & Schroeder, 2018). This interpretation, however, remains speculative until tested in a stronger powered study.

Figure 4: Vocabulary, language group and metaphorness interactions on reading of metaphors and post-metaphor region.



Note: Error bars show the standard error. Small/ Large = Small vocabulary/ Large vocabulary, Reread prob = Rereading probability, Reread time = Rereading time.

Table 14: Models using individual differences, language group and metaphorness to predict reading the first word of the metaphor.

	Word 1 of metaphor (grades are lotteries or: hours fly by)							
	Gaze duration		Rereading time		Rereading probability		Total time	
	B	*SE*	*B*	*SE*	*OR*	*SE*	*B*	*SE*
Language group	-0.06	0.09	0.04	0.09	1.13	0.28	0.00	0.09
Metaphorness	-0.02	0.04	-0.13	0.08	0.63 **	0.11	-0.13 **	0.05
Metaphor Type	-0.13 *	0.05	-0.03	0.08	0.87	0.16	-0.15	0.08
Language Group X Metaphorness	-0.02	0.08	-0.06	0.14	1.19	0.37	0.02	0.08
Metaphorness X Metaphor Type	0.00	0.05	0.13	0.12	0.99	0.29	0.04	0.07
Vocabulary (Vocab)								
Vocab main	-0.09	0.12	-0.00	0.11	1.33	0.40	-0.04	0.12
Vocab X Language Group	-0.15	0.19	0.02	0.21	0.59	0.33	-0.27	0.20
Vocab X Metaphorness	0.02	0.09	0.17	0.17	0.75	0.28	-0.04	0.10
Vocab X Language Group X Metaphorness	0.03	0.17	0.03	0.34	7.58 **	5.66	0.42 *	0.20
Reading fluency (RF)								
RF Main	-0.44 ***	0.08	-0.45 ***	0.08	0.67	0.14	-0.59 ***	0.08
RF X Language Group	-0.09	0.14	0.08	0.15	0.83	0.34	-0.14	0.15
RF X Metaphorness	0.04	0.06	-0.02	0.13	1.23	0.33	0.06	0.07
RF X Language Group X Metaphorness	-0.13	0.12	-0.52 *	0.25	0.95	0.51	-0.24	0.14

Notes: ** p<.05 ** p<.01 *** p<.001.*

Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.

Regression in probability is not included since it showed no effects of reading fluency or vocabulary.

Table 15: Models using individual differences, language group and metaphorness to predict reading the second word of metaphors and answering comprehension questions.

	Word 2 of metaphor (grades are lotteries or: hours fly by)								Offline	
	Gaze duration		Rereading time		Rereading probability		Total time		Comprehension	
	B	*SE*	*B*	*SE*	*OR*	*SE*	*B*	*SE*	*OR*	*SE*
Language group	0.01	0.11	0.01	0.08	1.03	0.30	0.02	0.12	1.45	0.32
Metaphorness (Met)	-0.07	0.06	-0.16	0.09	0.68 *	0.11	-0.23 ***	0.06	1.85 *	0.46
Metaphor Type	0.06	0.06	0.30 **	0.10	2.03 ***	0.37	0.31 ***	0.09	0.76	0.13
Language Group X Met	0.06	0.09	-0.11	0.14	1.02	0.31	-0.00	0.09	1.29	0.54
Met X Metaphor Type	-0.03	0.09	-0.20	0.15	0.62	0.16	-0.19	0.10	0.77	0.32
Vocabulary (Vocab)										
Vocab	-0.04	0.13	-0.09	0.09	1.76	0.64	0.10	0.14	1.32	0.34
Vocab X Language Group	-0.19	0.22	-0.02	0.18	0.74	0.47	-0.34	0.24	1.15	0.56
Vocab X Met	0.03	0.11	0.01	0.16	0.97	0.37	0.07	0.10	1.12	0.49
Vocab X Met X Language Group	0.02	0.21	0.74 *	0.33	1.33	0.99	0.40	0.20	0.28	0.24
Reading fluency (RF)										
RF Main	-0.32***	0.09	-0.49 ***	0.07	0.85	0.21	-0.54***	0.10	1.59 **	0.28
RF X Language Group	0.13	0.16	-0.22	0.13	1.17	0.56	-0.20	0.17	1.36	0.47
RF X Met	-0.03	0.08	0.11	0.12	1.17	0.31	0.09	0.07	0.59	0.18
RF X Met X Language Group	0.11	0.15	0.12	0.24	1.25	0.67	0.04	0.15	2.65	1.59

Notes: ** p<.05 ** p<.01 *** p<.001.*
Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
Regression in probability is not included since it showed no effects of reading fluency or vocabulary.

Table 16: Models using individual differences, language group and metaphorness to predict reading the post-metaphor region.

	Post-metaphor region *(and she decides / so she feels)*							
	Gaze duration		Rereading time		Rereading probability		Total time	
	B	*SE*	*B*	*SE*	*OR*	*SE*	*B*	*SE*
Language group	-0.03	0.14	-0.07	0.10	0.93	0.27	-0.03	0.06
Metaphorness (Met)	0.06	0.04	-0.02	0.09	0.73 *	0.11	-0.01	0.02
Metaphor Type	-0.03	0.08	0.18	0.10	1.70 *	0.37	0.06	0.05
Language Group X Met	-0.04	0.08	0.12	0.17	0.59	0.19	-0.03	0.04
Met X Metaphor Type	0.02	0.07	-0.04	0.15	0.49 **	0.12	-0.08	0.04
Vocabulary (Vocab)								
Vocab	0.03	0.18	0.04	0.12	1.66	0.58	0.06	0.08
Vocab X Language Group	-0.11	0.28	0.22	0.23	1.12	0.70	-0.03	0.13
Vocabulary X Met	0.00	0.09	0.04	0.20	1.80	0.64	0.04	0.04
Vocab X Language Group X Met	-0.06	0.19	0.63	0.39	3.39	2.38	0.25 **	0.09
Reading fluency (RF)								
RF	-0.46***	0.12	-0.36*	0.09	1.08	0.26	-0.23 ***	0.06
RF X Language Group	-0.10	0.21	-0.15	0.17	1.64	0.75	-0.05	0.10
RF X Met	-0.12	0.07	0.16	0.14	0.89	0.22	0.02	0.03
RF X Language Group X Met	0.02	0.13	-0.43	0.27	0.92	0.46	-0.07	0.06

Notes: ** p<.05 ** p<.01 *** p<.001.*
Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
Regression in probability is not included since it showed no effects of reading fluency or vocabulary.

2.4 General Discussion

This experiment focused on EAL and monolingual children's reading and comprehension of metaphors and literal controls, and compared it across metaphor types. Children's eye movements were recorded as they read stories containing verbal and nominal metaphors (or literal controls), and they also answered comprehension questions about the story. Children were also assessed on their reading and language skills. Using these measures, the experiment compared children's processing and comprehension of verbal and nominal metaphors, and linked them to individual and group differences.

2.4.1 Metaphor reading and comprehension

Metaphors were harder to understand than literal controls, as manifest in children's lower accuracy on comprehension questions following metaphors. The finding chimes with previous findings that metaphors were hard to understand for monolingual English children between 6 and 11 years of age (Cicone et al., 1981; Waggoner et al., 1997) and 5–7-year-old EAL learners (Hessel & Murphy, 2018). However, the results are in contrast with findings that metaphors presented in story context did not cause any comprehension difficulties even to younger monolingual English children aged 5 years (Özçalışkan, 2007; Stites & Özçalişkan, 2013a). However, those studies may have failed to show a metaphorness effect since they lacked literal controls against which to compare the comprehension of metaphors. The current experiment is the first to show that understanding metaphors remains a challenge for EAL children later in development, aged 8 years.

Previous studies have studied children's understanding of metaphors as part of listening comprehension, and the research on children's reading processing of figurative phrases is sparse. This experiment is the first to document an effect of metaphors on children's reading comprehension. The distinction is not trivial: while both listening and reading comprehension recruit many of the same language skills, they also differ in important ways. Compared to listening, reading comprehension involves the added difficulties of decoding letters into sounds and processing more complicated syntax and demanding vocabulary (Castles et al., 2018). At the same time, texts also offer possibilities to support comprehension that are not available in every listening context: readers can move around in a text to selectively resample parts of the linguistic input, by rereading an earlier passage they have misunderstood. Developing all of these skills with texts is an important part of children's development at primary school and has consequences for their academic achievement (García-Madruga, Vila, Gómez-Veiga, Duque, & Elosúa, 2014). This is particularly important for populations that are known to struggle with reading comprehension, as in the case of EAL learners. The current study thus adds to the growing body of EAL research by linking metaphor comprehension to text processing.

Different explanations have been suggested to account for EAL children's comprehension difficulties with figurative phrases - such as a failure to integrate meanings (Burgoyne et al., 2013), or to notice a figurative language use (Kan, 2014; Martinez & Murphy, 2011). So far, however, there have been no online studies aiming to identify when in processing metaphors become difficult during children's reading. Experiment 1 showed that children's additional processing effort was manifest in increased rereading, as previously reported with adults' reading of metaphors

(Olkoniemi et al., 2015). This is similar to reading on other post-lexical processing difficulties, for example when readers process words that are semantically or syntactically ambiguous (Pickering & Frisson, 2001; von der Malsburg & Vasishth, 2012) or anomalous (Joseph et al., 2008). However, metaphors did not change children's initial reading of words, which would have indicated difficulties at the word level, such as when readers access the meaning of less frequent words (Joseph et al., 2013). This indicates that metaphors placed additional demands on children's post-lexical processing, rather than on lexical access. The findings are in line with adult eye movement studies on figurative processing that reported increased rereading of, for example, one word metaphors (Olkoniemi et al., 2015) or metonymies (McElree et al., 2006). This experiment is the first to demonstrate an effect of metaphorness on post-lexical processing in primary school children.

What then, makes post-lexical processing more difficult for metaphors than for literal language? One explanation is interferences from overly literal interpretations – for example, if the phrase *grades are lotteries* was interpreted as students buying lottery tickets to get grades. Overly literal interpretations are common in children's comprehension of figurative phrase, as manifest in children's frequent literal misinterpretations of metaphors (Johnson & Pascual-Leone, 1989), and, amongst less skilled comprehenders, of idioms (Levorato et al., 2004). Literal interference could cause increased rereading due to the time required to suppress an overly literal interpretation, or to re-analyse metaphors after an initial literal misinterpretation. Rereading has been found to indicate such re-analysis of misinterpretations in studies where readers reread semantic and syntactic ambiguities after a later context had disambiguated them (Kambe, Rayner, & Duffy, 2001; von der Malsburg & Vasishth,

2012). Whether readers simply took longer to process metaphors or needed to re-analyse initial misinterpretations may differ depending on their processing skills (Tiffin-Richards & Schroeder, 2018; von der Malsburg & Vasishth, 2012). Both processes, however, are likely to result in increased rereading.

A slightly different explanation is that the difficulty with metaphors stems not from literal interference, but from the additional effort needed to integrate the meanings of its two words. It has been suggested that metaphors are harder to process primarily because the meanings of their parts are less strongly associated in the readers' lexicon than for literal phrases (Kintsch, 2008). The account is supported by ERP evidence showing similar activations for metaphors and linking inferences (Coulson & Van Petten, 2002; De Grauwe, Swain, Holcomb, Ditman, & Kuperberg, 2010) as well eye movement evidence showing more rereading on figurative phrases that require building new meaning relationships (McElree et al., 2006). In the current experiment, this meaning distance was operationalised in the transparency rating where the overall meaning of metaphors was rated to be less strongly associated with the meanings of individual words than literal controls. Children's rereading on the metaphors may thus be a result of this semantic distance, and of the resulting effort needed in meaning integration. Plausibly, both literal interference and semantic distance may together have placed the additional demands on children's metaphor processing and thus have led to the observed increased rereading. While checks confirmed that there were overall sufficient instances of rereading to afford reliable analyses, rereading was also less pervasive across participants than measures of initial reading, which may somewhat limit the representativeness of rereading measures of children's reading more generally.

2.4.2 Differences between metaphor types

The effect of metaphors on children's processing was compared between two metaphor types and their matched literal controls: verbal metaphors of the type *hours fly by*, and nominal metaphors such as *grades are lotteries*. Metaphor types did not change the effect of metaphorness in children's offline comprehensions. However, metaphorness effects in children's online processing seemed largely driven by nominal metaphors, as manifest in interactions across all regions: children reread nominal metaphors and their following region longer and more often than literal controls. However, this was not the case for verbal metaphors compared to literal controls, and children even spent less time in the regions following a verbal metaphor than a literal control. The reason for this difference plausibly lies in the higher systematicity of verbal metaphors (Lakoff & Johnson, 2003) which was expected to reduce processing efforts in two ways: first, efforts involved in meaning integration are theorised to be largely influenced by semantic distance (Kintsch, 2008). In nominal metaphors, active reasoning seems required to bridge the gap between the two word meanings (Beaty & Silvia, 2013; Billow, 1975; Coulson & Van Petten, 2002; Johnson, 1991; Kogan et al., 1980). For verbal metaphors, however, the meanings, for example of time and space, are closely related in speakers' memories (Boroditsky & Ramscar, 2002; Casasanto, 2008; Casasanto & Jasmin, 2012; Fuente et al., 2014; Semino, 2008), and integrating them may thus require little effort, as manifest in children's lack of rereading. Secondly, verbal metaphors may also trigger less to no literal interference: reading times on verbal metaphors have been found to be faster when linked to a literal interpretation (Wilson & Gibbs, 2007), and close semantic overlaps seem to reduce the likelihood for initial

misinterpretations (Pickering & Frisson, 2001) – both of which could further explain children's lack of rereading on verbal metaphors.

The fact that metaphor types affected eye movements but not answers on comprehension questions may indicate that while nominal metaphors required more processing effort, children eventually reached a sufficiently good interpretation. Alternatively, the absence of a metaphor type effect in offline comprehension may also indicate potential limitations of the forced-choice questions with which it was measured, as some differences in metaphor comprehension may only show in different task formats, such as explaining metaphors (Hessel & Murphy, 2018). Verbal and nominal metaphors have so far only been researched in separate studies, with the exception of one offline study of listening comprehension (Hessel & Murphy, 2018). The present experiment is the first to document how the different processing demands of verbal and nominal metaphors resulted in higher rereading of nominal – but not verbal – metaphors compared to literal language.

Arguably, Experiment 1 offered only a broad comparison of metaphor types, as some lexical differences remained between the two types. For example, words in the verbal condition were overall more frequent and longer than in the nominal conditions, which may have further supported quick lexical access and thereby facilitated post-lexical processing (Perfetti & Stafura, 2014). However, differences in length and frequency only exist between metaphor types generally, but not between metaphors and literal controls, which means that they cannot fully account for differences in the metaphorness effect. Furthermore, predictability was used as an indicator of a more general systematicity in verbal metaphors (Siyanova-Chanturia & Martinez, 2015) and may have contributed to their processing ease (Fernández, Shalom, Kliegl, & Sigman,

2014). There is a wealth of literature documenting wider differences between metaphor types that can be expected to influence processing (Boroditsky & Ramscar, 2002; Casasanto, 2008; Casasanto & Jasmin, 2012; Coulson & Van Petten, 2002; Fuente et al., 2014; Kintsch, 2008; Semino, 2008). Further research could explore the exact mechanisms driving the metaphor type effect, for example, by unpacking the effect of predictability through context manipulations distinguishing between effects of semantic overlap and sentence context (Davenport & Coulson, 2011). While this experiment can only be a first step towards understanding differences between metaphor types, it is still the first to demonstrate how children's reading is influenced by different metaphors types.

2.4.3 Individual and group differences

Children were compared in two ways. First, language groups were compared on the standardised tests of vocabulary and reading, and then, children were compared on their metaphor reading and comprehension.

In the standardised tests, EAL and monolingual children did not differ in terms of reading fluency, but they differed on vocabulary and reading comprehension. This finding chimes with previous reports that if EAL and monolinguals differ in their language skills, they do so in comprehension, not decoding (Bialystok et al., 2009; Melby-Lervåg & Lervåg, 2014; Murphy, 2014; Oller et al., 2007; Pearson et al., 1993; Spencer & Wagner, 2017). Qualitative and quantitative differences in exposure between EAL learners and monolinguals (Murphy, 2018; Oller et al., 2007; Paradis, 2016) likely affect the development of vocabulary or decoding to different degrees: while the English letter-sound relationships are less affected since they are finite and generally

acquired early in primary school (Castles et al., 2018), vocabulary is more strongly affected since its development continues throughout people's lifetime (Brysbaert et al., 2016) and children with stronger vocabulary quickly become better and more avid readers and better word learners, thus widening rather than narrowing the gap (Joseph & Nation, 2018; Verhoeven & van Leeuwe, 2012; Willingham, 2017).

EAL and monolingual children's metaphor reading and comprehension was compared in two different types of models: first models compared language groups only, and then individual differences in reading fluency and vocabulary size were added. Main effects showed lower comprehension for EAL learners compared to monolinguals, and for less fluent compared to fluent readers, as manifest in fewer correct answers on comprehension questions across both metaphors and literal controls. EAL children can be expected to struggle more with the comprehension questions, particularly given the large differences on standardised tests of comprehension and vocabulary. For the less fluent readers, decoding may have limited their comprehension, as is often the case for younger readers that are still acquiring letter-sound relationships (Verhoeven & van Leeuwe, 2012).

Based on previous research (Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014; Smith & Murphy, 2015), a prediction was that EAL children would struggle particularly with metaphors, as would manifest in an interaction effect between language group and metaphorness. Unexpectedly, however, there was no such interaction on comprehension, giving no indication that EAL children struggled particularly with reading or understanding compared to literal controls in the current study. In line with previous research (van den Bosch et al., 2018), further analyses did, however, reveal differences in EAL and monolingual children's metaphor processing in

an interaction with reading fluency and vocabulary. It seems that language groups were too heterogeneous to find group differences without considering important individual differences. However, the current study had not been designed with individual difference analyses in mind and may have lacked the statistical power to fully analyse them. The following discussion will thus focus on the clearest effects in the data, while postponing a fuller discussion of all effects until the next chapter, where they will be considered in combination with a second, larger experiment.

Interactions between vocabulary size, language group and metaphorness indicated differences between EAL and monolingual children's metaphors processing, particularly in late processing. Previous research has suggested that EAL readers may struggle with understanding figurative phrases since they may fail to notice that a phrase carries a non-literal meaning (Kan, 2014; Martinez & Murphy, 2011), or that a literal interpretation does not fit the context (Burgoyne et al., 2013). In line with idea of "failing to notice", EAL children with small vocabulary did not differentiate in their rereading between metaphors and literals. This effect may have been exacerbated by the many EAL learners with particularly small vocabularies who likely were at an earlier stage of reading development, and were thus more likely to fail to adapt their reading selectively. This is similar to previous eye movement evidence which showed that less strong readers are less sensitive to higher-level manipulations (Connor et al., 2015; Joseph et al., 2008; Reichle et al., 2013). For example, lower ability readers are less likely to reread in order to undertake a selective reanalysis of semantic inconsistencies such as *barking kittens* or *chopping carrots* with an *axe* (Connor et al., 2015; Joseph et al., 2008). This was different, however, for EAL readers with larger vocabularies who reread metaphors more frequently and more often than literal controls. This suggests

that whether EAL children struggle to notice metaphors depends on sufficient vocabulary knowledge, further supporting the idea that EAL learners in particular could benefit from explicit vocabulary instruction (Murphy, 2018).

In contrast to EAL children with smaller vocabularies, monolinguals with smaller vocabularies were found to be sensitive to metaphors, as manifest in increased rereading of metaphors compared to literal controls. Monolinguals with larger vocabularies reread metaphors more only numerically than literal controls. While this could be interpreted as showing that the effect of vocabulary was simply reversed between EAL learners and monolinguals, this seems unlikely given the exact reading patterns: monolinguals differed numerically, if not significantly, in their rereading time of metaphors and literal controls. A more plausible explanation seems to be that after an initial large reading time difference on higher-level manipulations, reading time differences become smaller as children increase their efficiency of post-lexical processing, as suggested based on longitudinal evidence (Tiffin-Richards & Schroeder, 2018). Perhaps monolinguals with large vocabularies were simply more efficient in their processing of metaphors so that reading time difference between metaphors and literal controls was too small to become significant in the current, relatively low-powered study. This suggestion will thus be revisited in the following chapter, together with additional data from an older and larger sample.

2.5 Conclusion

To conclude this chapter, the results have provided a first insight into EAL children's processing of metaphors when they read texts. Four key questions were posed. The first was whether EAL and monolingual children would differ in specific areas of their language skills. The results showed the predicted pattern, with EAL children lagging behind monolinguals on comprehension-related skills, but not reading fluency as measured by the TOWRE-2. The second question was whether children would differ in their reading and comprehension of metaphors compared to literal controls. Metaphors were challenging to understand for children, as manifest in lower comprehension and increased post-lexical processing during rereading. This links to the third question, whether metaphor type would influence this metaphorness effect. As predicted, only nominal metaphor placed increased demands on children's post-lexical processing, while children did not slow down on verbal metaphors and even speeded up after reading them. In line with metaphor theory, this was interpreted as a result of reduced efforts required in terms of meaning integration and literal interference due to the systematicity of verbal metaphors.

The final question was how metaphor reading and comprehension differed between EAL and monolingual children. EAL children had overall lower comprehension on the reading task, but against predictions, not particularly so for metaphors compared to literal controls. However, language groups differed in their metaphor processing when vocabulary size was considered. The fact that none of these differences showed in offline comprehension may be due to methodological differences: offline comprehension had been measured using a single comprehension question only, and provided less information than the continuous eye movement record collected on

the reading process. Interactions in reading processing showed that only EAL children with larger vocabularies, but not those with smaller vocabularies, adapted their rereading to metaphors, in line with the suggestion that EAL readers may fail to notice figurative phrase (Kan, 2014; Martinez & Murphy, 2011). Monolinguals with small vocabularies, in contrast, adapted their reading to metaphors, which may indicate that they were less limited in their metaphor processing by vocabulary gaps. It was suggested that monolinguals with larger vocabulary were more efficient in their metaphor processing, which may have decreased their reading time differences between metaphors and literal controls to a numerical one (Tiffin-Richards & Schroeder, 2018). However, the current study had not been designed to tap such small individual differences. Chapter 3 thus moves on investigate metaphor processing with a particular focus on individual differences.

The results have implications for both research and practice. For metaphor research, this study provides first insights into children's processing of metaphor types, focusing on a broad comparison of verbal and nominal metaphors. Future studies wishing to further tease apart the mechanisms underlying the metaphor type difference may wish to test processing on a larger stimuli set manipulated on multiple factors, such as semantic overlap and sentence contexts. For EAL research, the results indicate that whether EAL children struggle to notice figurative phrases may depend on vocabulary. Future research studying this may benefit from considering vocabulary as a covariate. Such individual differences will be considered in the following chapter.

For practice, the results support the importance of teaching word knowledge to support reading comprehension, particularly to EAL learners. Teaching vocabulary may support children's ability to adapt their reading to challenging phrases, which promises

to support reading comprehension and through that, overall educational attainment (Demie & Strand, 2006; Strand & Demie, 2005; Strand et al., 2015; Whiteside et al., 2017).

An open question is whether the observed differences in metaphor processing can be replicated reliably in a study designed to test them. Chapter 3 thus moves on to investigate individual differences in metaphor processing in a larger and older sample, using more tightly controlled stimuli.

Chapter 3: **Individual Differences in the Processing of Nominal Metaphors**

3.1 Introduction

Chapter 2 showed first insights on how children process metaphors during reading, and how that differs between metaphors types. Metaphors, in particular nominal metaphors, placed large demands on children's post-lexical processing, as was apparent in more frequent and longer rereading of metaphors than literal controls. Exploratory analyses furthermore indicated that EAL learners and monolinguals differed in their metaphor processing, but that these differences only showed when individual differences in reading fluency and vocabulary size were considered. Only EAL learners with larger vocabulary spent more time rereading metaphors than literal controls, while monolinguals did this already with smaller vocabulary size. These findings were also mirrored in interactions with reading fluency. They may indicate that EAL learners had more difficulties adapting their reading to metaphors than monolinguals. Findings were less clear for monolinguals with larger vocabulary, where reading time difference between metaphors and literal controls failed to reach significance. These children still read metaphors numerically longer than literal controls, but reread metaphors faster than monolinguals with smaller vocabularies, possibly thanks to more efficient processing. It is an open question whether these findings would replicate, both in terms of children's metaphor processing, but also in terms of how EAL learners and monolinguals differed in their metaphor processing in relation to reading fluency and vocabulary.

Experiment 2, reported in this chapter, thus aimed to replicate these key findings from Experiment 1, while also amending the materials and procedures. All predictions

and analyses were again pre-registered prior to data analysis on the online platform *Open Science Framework* (see pre-registration document in Appendix I). This includes the individual differences analyses, which, unlike in Experiment 1, were confirmatory instead of exploratory. Furthermore, Experiment 2 was amended by adapting stimuli, and by testing an older and larger sample. Stimuli were updated in two ways: first, only nominal metaphors were kept since they had triggered post-lexical processing most clearly in Experiment 1. Secondly, stimuli were more tightly controlled in overall story layout and interest areas to decrease noise and increase power. The sample was larger to increase statistical power for the planned individual difference analyses. Participants were also slightly older, which holds two advantages: first, readers tend to become less heterogeneous in their reading behaviour as they become fluent decoders (Blythe & Joseph, 2011), which limits variance and further increases statistical power. Second, testing first younger and now older children extends the data to include more advanced stages of children's metaphor processing where effects came out less clearly in Experiment 1. Furthermore, age differences in children's eye movements are indicative of differences in lexical and post-lexical processing (Reichle et al., 2013). With appropriate language skills, young children can often read as skilfully as older children, or even adults (Aaronson & Ferres, 1984; Eilers et al., 2018a). Juxtaposing metaphor processing in Experiment 1 and 2 can thus indicate how different levels of word knowledge are linked to increasingly skilful post-lexical processing in children's reading.

3.1.1 Word knowledge in children' reading

Readers need to know the words in a text to understand what they read. This statement seems almost too simple to need stating. However, the quality of word knowledge also

has additional, subtle influences on the efficiency of moment-to-moment processing during reading. These influences are the research focus of Experiment 2, reported in this chapter.

Chapter 1 first introduced the lexical quality hypothesis which provides a valuable framework to understand the subtle influences of word knowledge on reading processing (Perfetti & Hart, 2002). To review, Perfetti and Hart suggest that the cornerstone of successful comprehension processes is word knowledge of high lexical quality, that is both, both *precise* and *flexible* representation of word meanings in a reader's lexicon. High lexical quality allows readers to process words efficiently and speedily- both during initial lexical access and during post-lexical processing (Perfetti et al., 2008). In a comparison of skilled and less skilled comprehenders, Perfetti and colleagues were able to demonstrate how strong vs. weak lexical access shaped the readers' post-lexical processing. Participants read short stories that contained target words which placed different processing demands on readers' meaning integration, while the participants' brain responses were measured as event-related potentials (ERP). The measured ERPs allow for the investigation of the difficulty associated with lexical access and post-lexical processing. Of particular interest is the N400 component – a brain response that indicates meaning integration (Kutas & Van Petten, 2006). A bigger N400 indicates greater difficulty with integrating meanings. As expected, Perfetti and colleagues found that short stories across conditions differed in the size of the N400 they triggered, indicating varying difficulty of meaning integration.

Intriguingly, skilled and less skilled comprehenders differed in the time course and size of all ERPs up to and including the N400- from lexical access to post-lexical processing. Skilled comprehenders showed an early ERP indicating lexical access, and

then they integrated word meanings into the text in two phases: after around 300ms, a positive peak indicated that they updated their text memory (for example when encountering the paraphrase). After around 400ms, they then integrated the new word meaning, as indicated by a strong N400. This indicates that skilled comprehenders were quick to access the word meaning, to update their text representation, and to integrate the meaning of the new word. In contrast, less skilled comprehenders were slower with initial lexical access which impacted their subsequent post lexical processing. They did not adapt their processing to the different conditions and did not show any signs of updating their text memory. As a result, their N400 was delayed and smaller than that of skilled comprehenders, indicating slow and less efficient meaning integration. These findings illustrate how slow lexical processing can trigger a chain reaction that leads to delayed and even incomplete post-lexical processing. A similar chain reaction from lexical access to post-lexical processing may also play a role in children's individual differences in metaphor processing: the processes indicated by the N400 are also the cornerstone of readers' processing of metaphors (Coulson & Van Petten, 2002; De Grauwe et al., 2010). The interesting question is how these mechanisms influence natural text reading. In particular, how may word knowledge shape children's eye movements when reading metaphors?

A reaction time study indicates that vocabulary size can limit even adult learners' efficiency of metaphor processing (Segal & Gollan, 2018), but there is little direct evidence on how individual differences shape metaphor processing in text reading (Olkoniemi et al., 2015), and none focusing on word knowledge. There are, however, eye movement studies that link individual differences in reading and language to other post-lexical processes (Connor et al., 2015; Eilers et al., 2018a; Hopp, 2016; Joseph &

Liversedge, 2013; Luke, Henderson, & Ferreira, 2015; Tiffin-Richards & Schroeder, 2018). These studies have tapped individual differences through different assessments of reading and language, as well as through group comparisons between children and adults, or first and second language readers. The way that differences in reading and vocabulary were operationalised in those experiments is in many case different from this thesis, where children were assessed on reading fluency and vocabulary size. However, individual differences in reading and vocabulary have similar effects on post-lexical processing across studies, which indicates the relevance of these studies also for the current study. Individual differences can be broadly categorised as (1) differences in the likelihood to engage in post-lexical processing, and (2) different levels of the effort involved in the processing.

Readers with stronger language skills are more likely to notice when structures require additional processing (Connor et al., 2015; Eilers et al., 2018a; Oakhill et al., 2016). For example, only children with strong comprehension skills reported to have noticed a gender mismatch of pronouns (Eilers et al., 2018a), or adapted their rereading times to semantic mismatches (Connor et al., 2015). This indicates that a certain level of word knowledge is necessary to adapt rereading to challenging language features. How much word knowledge is needed seems to vary with the difficulty of a language feature. For example, children with poor reading comprehension only took longer to read familiar, but not novel idioms compared to literal controls (Oakhill et al., 2016), indicating that they did not appreciate that also novel idioms required particular processing effort. Similarly, children spend more time on nominal anaphors that were close to the antecedent and typical (such as *vehicle* to refer back to *truck*), as opposed to more distant and atypical, indicating that they only resolved those easier references

(Joseph et al., 2015). What may account for this lack of noticing? Plausibly, post-lexical processing was so slow for the various reader groups with weaker reading and vocabulary – that is, for poor vs. good comprehenders, children vs. adults, children with weaker vs. stronger academic language (Connor et al., 2015; Joseph et al., 2015; Oakhill et al., 2016) - that they had moved on in the text before they could make the decision to reread. The newly fixated words may then have replaced all memories of unresolved processes so that readers never noticed that more processing would have been needed on earlier text parts. If a reader's attention is limited in such a way, this may also limit the children's own awareness of their lack of comprehension - as has been reported for EAL children who did to not notice their difficulty with figurative phrases (Kan, 2014; Martinez & Murphy, 2011). To test this idea, I decided to include a question about self-perceived comprehension to compare to children's actual comprehension, and to test the variability of this link between EAL and monolingual children.

Stronger word knowledge is not only linked to more frequent, but also to fast and efficient post-lexical processing (Hopp, 2016; Joseph & Liversedge, 2013; Luke et al., 2015; Tiffin-Richards & Schroeder, 2018). For example, children take longer to resolve temporally ambiguous sentences such as "The boy poked the elephant with the long trunk from outside the cage" compared to adult readers (Joseph & Liversedge, 2013). The delayed response suggests that children are slower to update their interpretation. Similarly, both adult readers with lower lexical quality and younger children needed more time to read and reread the ends of a sentence or clause, indicating that they are less efficient in finishing sentence final processing, called sentence wrap up (Luke et al., 2015; Tiffin-Richards & Schroeder, 2018). The link

between lexical access and post-lexical processing was tested directly by Hopp (2014) by inserting either high or low-frequency words in easier subject clefts and harder object clefts. Low-frequency words make lexical access harder and high-frequency words make it easier. When sentences contained only high-frequency words (thus posing little demands on lexical access), native speakers did not take significantly longer to read object compared to subject clefts, but they did when sentences contained low-frequency words. This indicates that the faster lexical access in the high-frequency condition increased the efficiency of post-lexical processing so that even challenging sentence structures could be processed fluently.

The two quite distinct types of post-lexical processing- to fail to notice, or to take long to process- may even be two steps within the same developmental trajectory. Based on longitudinal data, it has been suggested that children's sentence wrap up develops in an upside down U-shape: from no reaction, to strong processing demands, to eventually fast processing (Tiffin-Richards & Schroeder, 2018). It is possible that children's metaphor processing follows a similar trajectory. The current experiment can test the plausibility of this account by gathering data on metaphor processing in an older sample and comparing this to the pattern of findings seen in younger children in Experiment 1.

To review, readers' word knowledge influences how likely they are to engage in post-lexical processing as well as their efficiency of processing. Based on the lexical quality hypothesis, this relationship between individual differences in reading and vocabulary on the one hand and post-lexical processing on the other hand can be interpreted as a chain reaction where slow lexical access impacts post-lexical

processing. Difficulties with post-lexical processing may show as a failure to reread or slow processing, with the two phenomena being linked developmentally.

3.1.2 The current study

Experiment 1 provided first insights into how the presence of nominal metaphors influences children's reading, particularly their post-lexical processing. Exploratory analyses furthermore indicated that EAL learners and monolinguals differed in their metaphor processing when individual differences in reading fluency and vocabulary were considered. Only EAL learners with larger vocabulary read longer on metaphors than literal controls, while monolinguals did so already with smaller vocabulary size. There were trends that monolinguals with larger vocabularies were faster in processing metaphors, possibly because their processing became more efficient. It was suggested that these findings are part of a U-shaped trajectory of how children's post-lexical processing develops with increasing language skills- from not reacting to challenges, to effortfully processing them, to selective and fluent rereading on text difficulties (Tiffin-Richards & Schroeder, 2018).

It is an open question, however, whether the findings of Experiment 1 would replicate in a higher-powered study, and how individual differences would play out in slightly older children who can be expected to have stronger language skills and thus be more efficient in their metaphor processing (Reichle et al., 2013; Segal & Gollan, 2018). Experiment 2 focused on individual differences in a large sample of EAL and monolingual children reading nominal metaphors only. Metaphor processing was tested in older children and linked to their individual differences in reading fluency and vocabulary, which extends the data gathered on younger children in Experiment 1. Furthermore, a self-report measure on children's self-perceived comprehension was

added to see if EAL learners and monolinguals also differed in their awareness of comprehension difficulties on metaphors, as previously suggested for their comprehension of collocations (Kan, 2014).

The following research questions were addressed:

1. Do EAL and monolingual children differ significantly in their vocabulary size, but not in their reading fluency?

Based on Experiment 1 and previous findings in the literature, EAL learners were expected to have smaller vocabularies than monolinguals, but be comparable on reading fluency.

2. Do children read and comprehend metaphors differently from literal control phrases?

Children were expected to again show more difficulty comprehending metaphors than literal controls, as evidenced by lower accuracy on comprehension questions following metaphors. This difficulty was expected to arise in eye-tracking measures that indicate post-lexical processing, particularly in more frequent and longer rereading of metaphors compared to literal controls.

3. Do children read and comprehend metaphors differently if they are learning EAL, or depending on their reading fluency and vocabulary size?

Both EAL children and children with lower reading fluency and smaller vocabulary size were expected to be less accurate on comprehension questions. In online reading as revealed by the eye movement record, higher reading fluency and larger vocabulary size was expected to be associated with more fluent reading overall, and with either a higher likelihood to reread metaphors, or shorter time spent rereading metaphors compared to

literal controls. However, the current sample was also older and it was hard to predict the exact stage at which they would be in their metaphor processing.

4. Are children aware of their actual level of comprehension, and does this vary between children?

If children were aware of their comprehension, this was expected to manifest as a significant and positive relationship between actual and self-perceived comprehension. That is, a child with lower comprehension on the task would also be expected to rate their comprehension as lower, and vice versa.

If EAL and monolingual children or children with varying word knowledge differed in their ability to accurately judge their comprehension, this was expected to manifest as an interaction of language group or word knowledge with self-perceived comprehension when predicting actual comprehension.

3.2 Method

3.2.1 Design

This experiment has a mixed design with one within-subject manipulation (metaphor vs. literal). The effect of this manipulation was compared between subjects across different language groups (EAL vs. monolinguals) and by considering associations with individual differences (in reading fluency and vocabulary size). Dependent variables were children's eye movements on metaphors and literal controls, and their performance on subsequent comprehension questions. Individual differences in reading fluency and vocabulary were assessed using standardised tests. In additional exploratory analyses, children were also compared on their ability to accurately judge their comprehension on the metaphor task.

3.2.2 Participants

Sixty-four children aged 9;01-10;10 were recruited for this study. As in Experiment 1, recruitment was done locally through convenience sampling. Ethical procedures were re-assessed to ask for permission to recruit following an opt-out (in lieu of the previously used opt-in) procedure (see Appendix J). Following ethical approval from the CUREC (see Appendix K), primary schools in Oxford were contacted at research into practice events to solicit participation in the study. Two schools decided to take part.

Within the recruited schools, parents with children in Year 5 were contacted. The first school requested opt-in procedures be used; thus, the same procedure was followed as in Experiment 1. In the second school, an opt-out procedure was followed: information leaflets were sent out to parents twice with a one-week break in between, and at least two weeks before the start of testing. Each information leaflet was accompanied by a letter explaining that if parents did not want their children to participate in the experiment, they had two weeks to opt-out (see Appendix L for all flyers and accompanying letters). Two parents decided to opt-out of the experiment and their children were not included in the study. All other parents either did not choose to opt out, or in the case of the first school, gave their written consent. All children in the final sample provided assent and received a sticker in appreciation of their participation.

Eligibility criteria were the same as in Experiment 1. All EAL children had resided in the UK for at least one year and were thus eligible to participate. Children were also assessed on age-adequate word-reading skills on the TOWRE-II. One EAL child performed below the <1SD age cut-offs of 34 for word-reading and 14 for non-word reading and was thus excluded. All other children were within the normal range of

reading ability (range = 55-91 for word reading, and range = 19-63 for non-word reading).

The final sample comprised 63 children (EAL learners, n=23, 11 girls; monolinguals, n=40, 22 girls) with a mean age of 10;02 (range = 9;01-10;08). All children had normal or corrected-to-normal vision.

3.2.2.1 Validation of EAL status

Different from Experiment 1, children could not be classified as EAL learners based on the school's classification since the opt-out procedure prevents asking for information about sub-groups of children. Children were thus classified as EAL or monolingual based on a Language Background Questionnaire. In the questionnaire, children were asked if they spoke any other language at home or in their community, and how often they spoke it. If they reported to regularly be speaking or hearing a language other than English, they were classified as EAL, if they answered in the negative, they were classified as monolingual.

3.2.2.2 EAL learners and monolinguals

As in Experiment 1, EAL and monolingual children were each other's peers: they all lived in the same school catchment area and attended the same classes at school. The two groups did not differ on non-verbal reasoning (as measured through the WASI II matrices subtest) or age (see Table 17). Both groups performed slightly below average, as indicated by the standard *t*-scores.

Table 17: Language groups on age and non-verbal reasoning.

	Monolingual	EAL	Group comparisons	Cohen's d
Age	10;02 (0;04)	10;03 (0;04)	$t(52.7) = -1.10$, $p = .28$	0.28
Non-verbal reasoning				
Raw score	15.2 (4.0)	14.9 (4.9)	$t(38.0) = 0.30$, $p = .77$	0.08
Standard score	46.9 (8.2)	46.2 (10.5)		

Notes: Round means. Standard deviations in brackets. ** p<.05 ** p<.01 *** p<.001*
The non-verbal reasoning raw scores are the number of correctly completed sequences before three consecutive errors. The standard scores are t-scores with a mean of 50 and a standard deviation of 10.

3.2.2.3 EAL children's language background

Parent's low response rate on the language background questionnaire had been a problem in Experiment 1. As the children were older in Experiment 2, it was decided to administer the questionnaire by interviewing the children themselves. To make the questionnaire as child-friendly as possible, the number of questions was reduced, and all questions were anchored in a recent and concrete time period (such as "[*Over the last month,*] how often did your parents speak in your home language to you?"). Additionally, answer scales were presented visually and introduced with a practice item (see Appendix M for a sample questionnaire and scales). The first question always asked if children spoke another home language than English. Children who did not report another home language were classified as monolingual English, and the LBQ was discontinued with these children. Among the EAL children who reported another home language, many different languages were spoken: six children spoke Arabic, and two spoke Kiswahili. One of the Arabic speakers also spoke French. The other children each spoke one of the following: Italian, Japanese, Malay, Malayalam, Filipino, Serbian, Shona, German, Urdu, Welsh, Zulu, and Hawaiian pidgin.

The remainder of the LBQ focused on EAL children's current and past English language exposure. Two EAL children reported to have moved to the UK at age 2 and 8, respectively. Both these children had already started learning English at age three. All other EAL children had been born in the UK and had started learning English from birth or, in the case of four children, when entering nursery at age 3-4. The questionnaire also assessed the quantity and quality of children's current English language input (Paradis, 2016). Input quantity was assessed by asking children about the proportion of time they spent speaking English at home. Answers were summarised in a composite score (range 0-5, higher numbers indicate more English language input). One child did not speak any English at home. The other children varied greatly in their amount of English language exposure, from little English to English being used most of the time (range 1- 4, M = 2.8, SD = 0.7). Input quality was assessed by asking children about different sources of English language exposure at home and in their free-time, for example when speaking with friends, or when reading and using other media. Answers were again summarised in a composite score, with scores ranging from 0-5. Children's input quality varied greatly, from using English in many to only few different contexts and media (range = 2-4, M = 2.9, SD = 0.6).

3.2.3 Materials

3.2.3.1 Assessment of metaphor reading and comprehension

Children read 20 single paragraph stories that were followed by a forced choice comprehension question (see Table 18 for an example, and Appendix N for a full list of stimuli). The stories followed the same logic as in Experiment 1: each story existed in a metaphorical and a literal version that differed by one word – the target word that

created the metaphor. Story contexts were written to ensure that both versions were plausible. Plausibility of contexts was confirmed when stories were proofread and discussed with five British Native speakers (three when preparing Experiment 1 and another two when preparing Experiment 2), and by norming metaphors on naturalness as part of the adult norming study (see Chapter 2).

Stories were split up into lists following a Latin square design. Each child read all 20 stories, half in the metaphor condition and half in the literal control condition. Every child thus read each story once, and experienced 10 stories per condition. The stories were presented in fully randomized order.

Table 18: Example story and question.

Libby always thought she was good with numbers. But today, she got an odd grade in maths. When she comes home, Libby tells her mother that she now thinks she's useless at maths. Her mother says that sometimes, grades are *lotteries* / *misleading* and then Libby feels much better.	
Comprehension question: Why does Libby feel better?	
A: She sees that grades can be random.	B: She sees that she has probably misunderstood the meaning of her grade.

3.2.3.2 Story norming

Stories were adapted from Experiment 1. To review, stories were already controlled for story complexity and episodic structure. For Experiment 2, stories were furthermore shortened and the second comprehension question was removed. This reduced children's reading time, freeing up testing time to include additional measures, such as the language background questionnaire. Furthermore, the screen layout was streamlined by matching stories on the number of words and characters preceding a target region (40 words, between 220-228 letters), and on words and characters following the target word (6 words, between 30-35 letters). Thus, all stories had the same number of lines

preceding a target word and words following a target word within the last line, which streamlined the location of target words on the screen. This was done to reduce noise created by text length or screen layout (Radach, Huestegge, & Reilly, 2008; Rayner, 1977), and to make the texts amenable to scanpath analyses. The shortened texts were proofread and discussed with two speakers of British English to make sure that the language was appropriate and acceptable. Furthermore, stories and questions were piloted with a 7-year-old child for readability.

3.2.3.3 Metaphor norming

Metaphors were adapted from Experiment 1. To review, those metaphors were not lexicalised but interpretable, and had been normed on transparency and naturalness. For the current experiment, I only kept nominal metaphors (e.g. *Jane is the queen-bee*) and their literal controls (e.g. *Jane is the music-lover*). Nominal metaphors had shown reliable effects in Experiment 1. Focusing on them further reduced testing time and any variance caused by stimulus heterogeneity.

In Experiment 1, literal controls were created by replacing one word that was matched on length, frequency and predictability with its counterpart. For Experiment 2, it was decided to furthermore match the two words on the number of syllables and morphemes, while the match in frequency and letter length was maintained (see Table 19). The words had a maximum age of acquisition of 10 years (the average age of the participants) (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Welch's *t*-tests confirmed that metaphors and literal controls did not differ on any of these dimensions.

Table 19: Matching of replaced word in metaphors and literal controls.

	Metaphor	Literal	*t*-test results
Log frequency	2.2 (2.2)	2.3 (2.4)	$t(35.8) = -0.1$, $p = .93$
Number of letters	6.1 (1.8)	6.8 (1.9)	$t(38.0) = 1.2$, $p = .24$
Number of syllables	1.6 (0.6)	2.0 (0.8)	$t(36.0) = 1.6$, $p = .11$
Number of morphemes	1.7 (0.7)	1.7 (0.7)	$t(37.7) = 0.2$, $p = .82$
Age of acquisition	6.5 (2.2)	7.2 (1.6)	$t(32.6) = 1.2$, $p = .25$
	Maximum: 10.7	Maximum: 9.7	

Note: Frequency based British National Corpus.

3.2.3.4 Interest areas and eye-tracking measures

As in Experiment 1, the raw data from the eye movement record was analysed into the same three interest areas using SR Research software. Table 20 shows examples. The only difference to Experiment 1 is that for interest areas in the current experiment, all words were at least 5 letters long so that no neighbouring words were added. Eye-tracking measures were again calculated for each region, with a focus on post-lexical processing. Based on findings from Experiment 1, the number of eye-tracking measures was limited to reduce the risk of false positives created by multiple comparisons (von der Malsburg & Angele, 2017). The following measures were analysed: for the second word of the metaphor and the post-metaphor region, I analysed gaze duration, rereading time, rereading probability, regression in probability, and for the first word of the metaphor, regression in probability.

Table 20: Example of interest areas.

	Word 1 of metaphor		Word 2 of metaphor	Post-metaphor region
Metaphor	/ grades/	/ are/	/ lotteries /	/ and she feels/
Literal control	/ grades/	/ are/	/ misleading/	/ and she feels/

3.2.3.5 Apparatus

The same equipment and set-up were used as in Experiment 1.

3.2.4 Assessments of reading and language

Each child in the study completed the following two standardised assessments.

Reading fluency was tested using the (TOWRE-2) (Torgesen et al., 2012).

Vocabulary size was tested using the British Picture Vocabulary Scale, Version 2 (BPVS) (Dunn, Dunn, Whetton, & Burley, 1997). Only one assessment of vocabulary was included since having three assessments in Experiment 1 had not added more value: all three measures were highly correlated, they showed the same difference between language groups, and only vocabulary size could be used as a predictor of metaphor processing.

The BPVS is a quick test that is frequently used in research on bilingual primary school children and figurative language (Rundblad & Annaz, 2010; Stites & Özçalişkan, 2013b). Its predictive validity (Cunningham & Stanovich, 1997; Hart & Risley, 2003) and concurrent validity (Dunn et al., 1997; Howlin & Cross, 1994) have been established in a wealth of correlational and longitudinal studies. In the BPVS, children see four pictures on a page and are asked to choose the picture that fits the word read out to them. The test score represents the number of correctly chosen pictures until a child makes 8 mistakes within a set of 12 items.

3.2.5 Self-perceived comprehension

Children's self-perceived comprehension was measured by asking how much children believed they had understood in the stories. This was done to see if children were able to accurately describe their actual comprehension, as would be reflected in a link between self-perceived and actual comprehension, and how children might vary in this ability, as would be reflected in interactions with language group or individual differences. They were asked the question "How much do you think you understood from the stories you just read?" and were prompted to answer using a 4-point scale (*nothing- something- most of it- all of it*, see Appendix O). Children heard the questions and both saw and heard the scale. The question was modelled on previous studies (Kan, 2014; Martinez & Murphy, 2011), with the difference that only one question was asked after all stories had been read, instead of after each story. This was necessary since questions during a reading task are known to influence reading behaviour (Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, & Staub, 2017), an influence which I wanted to avoid.

3.2.6 Procedure

Overall, the same procedure was followed as in Experiment 1, the only difference being that testing was done in two sessions to allow sufficient time for all assessments. The two sessions occurred on separate days and took place 1-3 school days apart from each other. The first testing session started with the language background questionnaire. Children then read the metaphor and control stories with one break after approximately 10 stories. The break was filled with either the TOWRE-II, the BPVS, or the WASI. In a second session, children completed the remaining standardised assessments and a

second eye-tracking experiment (discussed in Chapter 4). The two sessions lasted approximately 20-30 minutes each.

3.3 Results and Discussion

3.3.1 Analytical strategy

As in Experiment 1, children's reading and comprehension of metaphors was assessed through a combination of eye-movement recordings and comprehension questions. Eye movement data was processed the same way as in Experiment 1. Deleting extremely short or long fixations (<80ms and >1200ms) removed less than 0.01% of the data. The data were then analysed in R (R Core Team, 2016) using Welch's *t*-tests for group comparisons in research question 1, as well as linear and binomial mixed effect models (MEMs) for research question 2 and 3. Research question 4 was analysed using a regression model that had self-perceived comprehension, language group, and individual differences as predictors of actual comprehension. I chose a regression model rather than a MEM since there was only one observation of self-perceived comprehension per child, making it impossible to model random effects for stories or subjects. A regression model was needed rather than a simple correlation since actual and self-perceived comprehension were assumed to be differently linked for groups of children, making it necessary to include an interaction term that could account for such moderating effects.

The following presents the results by research questions.

3.3.2 Do EAL and monolingual children differ significantly in their vocabulary size, but not in their reading fluency?

Table 21 summarises the descriptives and group comparisons for vocabulary and reading fluency. The standard scores show that both EAL and monolingual children were slightly above the population mean, particularly in their reading fluency. As expected, EAL and monolingual children did not differ in their reading fluency. Contrary to predictions, however, the two groups did not differ in vocabulary size either. Numerically, EAL learners showed slightly better reading fluency and smaller vocabularies, but neither of these differences was significant.

Table 21: Language groups compared on reading fluency and vocabulary size.

	Monolingual	EAL	Group comparisons	Cohen's d
Reading fluency				
Raw score	119 (18)	124 (17)	$t(48.0) = -1.0$, $p = 0.31$	0.3
Standard score	116 (14)	119 (14)		
Vocabulary size				
Raw score	108 (14)	103 (13)	$t(49.3) = 1.5$, $p = 0.13$	0.4
Standard score	109 (13)	104 (12)		

Notes: Round means. Standard deviations in brackets. Standard scores with mean of 100 and standard deviation of 15.

3.3.3 Do children read and comprehend metaphors differently from literal control phrases?

Comprehension of metaphors was measured by performance on the comprehension questions, while metaphor reading was captured through eye movements. Eye movements were distilled into five dependent measures: children's first reading of a word or region before moving on (*gaze duration*), how long they reread a word (*rereading time*) or read it overall (*total time*), and how likely they were to reread the

word (*rereading probability*) or to go straight back into it (*regression in probability*). The Table 22 and Table 23 summarise the data.

Table 22: Means for children reading the second word of metaphors and answering comprehension questions.

	Second word of metaphor (*grades are lotteries*)				Offline
	Gaze duration	Regression in probability	Rereading probability	Total time	Comprehension
Monolingual					
Literal	353 (276)	0.13 (0.34)	0.34 (0.48)	538 (492)	0.74 (0.44)
Metaphor	411 (362)	0.27 (0.44)	0.46 (0.50)	691 (607)	0.69 (0.46)
EAL					
Literal	346 (198)	0.22 (0.42)	0.41 (0.49)	638 (591)	0.74 (0.44)
Metaphor	377 (264)	0.32 (0.47)	0.58 (0.49)	824 (847)	0.61 (0.49)

Notes: Reading times in milliseconds. Rounded means. Standard deviations in parentheses.

Table 23: Means for children reading the first word of metaphors and the post-metaphor region.

	First word of metaphor (*grades are lotteries*)	Post-metaphor region (*so she feels*)			
	Regression in Probability	Gaze duration	Regression in probability	Rereading probability	Total time
Monolingual					
Literal	0.20 (0.40)	586 (433)	0.38 (0.49)	0.45 (0.50)	859 (586)
Metaphor	0.31 (0.46)	576 (430)	0.35 (0.48)	0.54 (0.50)	970 (644)
EAL					
Literal	0.21 (0.41)	553 (420)	0.39 (0.49)	0.51 (0.50)	970 (838)
Metaphor	0.32 (0.47)	577 (424)	0.41 (0.49)	0.61 (0.49)	1002 (722)

Notes: Reading times in milliseconds. Rounded means. Standard deviations in parentheses.

As in Experiment 1, one MEM was run per measure and region, with language groups and metaphorness as predictors. The Table 24 and Table 25 summarise the results of these models. Figure 5 shows that as expected, children answered comprehension questions on metaphors less accurately than on literal language. Also as predicted, children read metaphors for longer and fixated them more often, relative to literal controls, but not when first encountering them (*ns* on gaze duration). They were more likely to reread the first and second word of metaphors and the post-metaphor

region, and to go back directly into the first or second word, resulting in longer total times on metaphors than literal controls. In other words, metaphors slowed down children's rereading of texts and impacted their comprehension, which indicates that metaphors posed increased demands on post-lexical processing.

Since as in Experiment 1, effects showed mainly in rereading, I checked again if there was enough data to support the analyses. Children reread on 412 trials in the first and 523 trials in the second word of metaphors, which amounts to 37% and 42% of analysed trials, respectively. On the child level, their mean rereading probability was between 37-44% on the two interest areas, which, even if lower than in Experiment 1, is still a high number relative to previous eye movements studies that focused strongly on rereading (Meseguer et al., 2002; Murray & Kennedy, 1988; Schotter et al., 2014; von der Malsburg & Vasishth, 2011). Together, these numbers indicate that there were again sufficient instances of rereading to afford reliable analyses, as was also supported by good results on the influence.ME bootstrap test. However, the mean rereading probability on the two interest areas varied across children, with 5 and even 12 children rereading less than 20% of the time on the first and second word, respectively. This indicates that effects in rereading were somewhat less representative across participants than measures on first pass reading times.

In answer to the research question, metaphors changed reading as expected: children understood metaphors less well than literal controls, a difficulty that during reading processing showed in terms of frequent and slow rereading of metaphors. These findings largely resonate with those of Experiment 1. Different from the younger children in Experiment 1, however, the 9–10-year-olds in this experiment regressed directly into the first and second word of the metaphor and reread the post-metaphor

region more often. This indicates that older children may reread metaphors more selectively and spent more time linking metaphors to the text (but see Section 3.4.3. for limitations of the comparison between the two experiments).

Table 24: Results of models using language group and metaphorness to predict reading the second word in the metaphor and answering comprehension questions.

	Word 2 of metaphor (*grades are lotteries*)								Offline	
	Gaze duration		Regression in Probability		Rereading probability		Total time		Comprehension	
	B	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*	*OR*	*SE*
Language group	0.01	0.07	1.59 *	0.35	1.61	0.40	0.15	0.10	0.84	0.13
Metaphorness	-0.04	0.07	2.10 ***	0.45	2.04 ***	0.36	0.25 **	0.07	0.64 **	0.09
Language Group X Metaphorness	-0.00	0.06	0.68	0.22	1.29	0.43	0.06	0.08	0.66	0.19

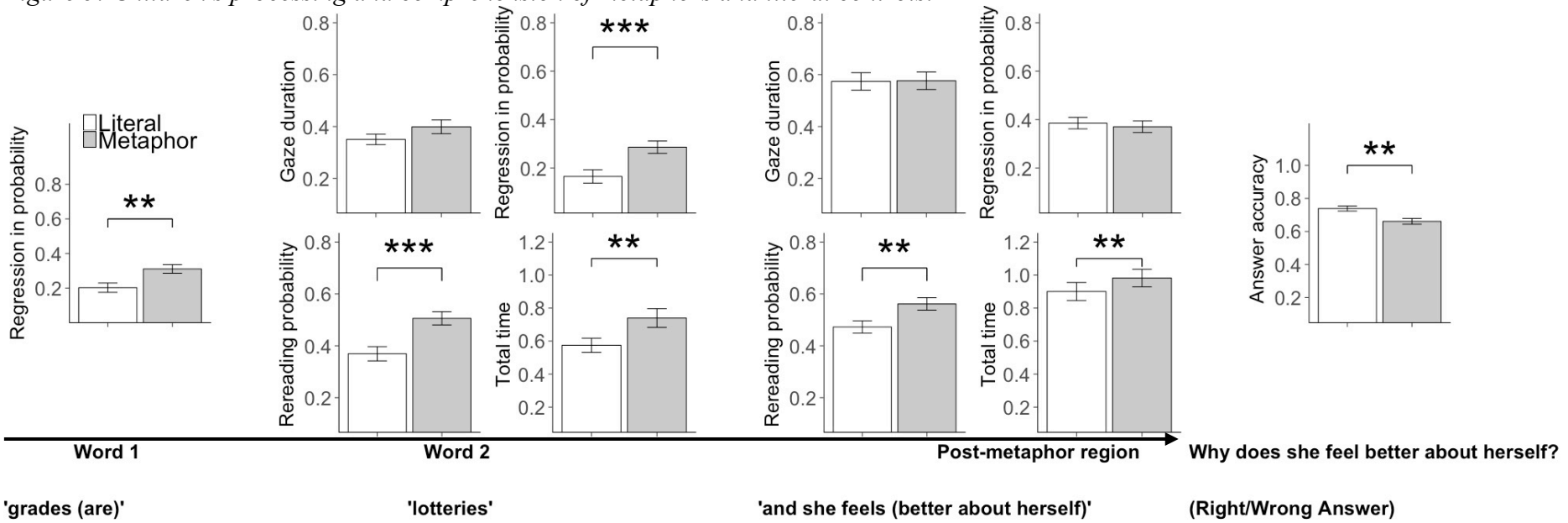
Notes: Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
** p<.05 ** p<.01 *** p<.001.*

Table 25: Models using language group and metaphorness to predict reading the first word of the metaphor and the post-metaphor region.

	First word of metaphor (*grades are lotteries*)		Post-metaphor region (*so she feels*)							
	Regression in Probability		Gaze duration		Regression in Probability		Rereading probability		Total time	
	OR	*SE*	*B*	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*
Language group	1.01	0.27	-0.01	0.12	1.23	0.32	1.43	0.37	0.07	0.11
Metaphorness	1.89 **	0.40	0.01	0.04	0.96	0.14	1.53 **	0.20	0.11 **	0.04
Language Group X Metaphorness	1.10	0.48	0.05	0.07	1.34	0.37	1.00	0.26	-0.12	0.07

Notes: Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
** p<.05 ** p<.01 *** p<.001.*

Figure 5: Children's processing and comprehension of metaphors and literal controls.



Note: Reading times in seconds. Rereading duration was analysed for the second word and the post metaphor region but is not included in this figure because metaphorness had no significant effect on this measure in any region. Error bars show the standard error.

3.3.4 Do children read and comprehend metaphors differently if they are learning EAL, or depending on their reading fluency and vocabulary size?

As in Experiment 1, differences between children were investigated by comparing language groups, and by relating performance on the experiment to individual differences in reading fluency and vocabulary. Differences between language groups are already summarised in Table 22-Table 25. Table 26 and Table 27 summarise the descriptives of individual differences which are based on predicted means for reading times at +/- 1 SD of reading fluency or vocabulary size, respectively.

Table 26: Predicted means for children reading the second word of the metaphor and answering comprehension questions.

	Word 2 of metaphor (*grades are lotteries*)				Offline
	Gaze duration	Rereading time	Rereading probability	Total time	Comprehension
Vocabulary					
Small					
Monolingual					
Literal	509 (118)	331 (148)	0.30 (0.31)	403 (114)	0.73 (0.34)
Metaphor	509 (118)	402 (130)	0.47 (0.29)	537 (118)	0.65 (0.3)
EAL					
Literal	473 (100)	490 (103)	0.48 (0.28)	546 (101)	0.75 (0.30)
Metaphor	465 (103)	515 (128)	0.64 (0.34)	701 (120)	0.65 (0.30)
Large					
Monolingual					
Literal	410 (94)	335 (110)	0.32 (0.26)	405 (92)	0.82 (0.32)
Metaphor	403 (91)	407 (106)	0.44 (0.24)	489 (100)	0.77 (0.27)
EAL					
Literal	447 (143)	406 (181)	0.26 (0.39)	381 (140)	0.82 (0.39)
Metaphor	492 (149)	397 (161)	0.58 (0.42)	551 (147)	0.60 (0.35)
Reading fluency					
Low					
Monolingual					
Literal	564 (95)	426 (103)	0.39 (0.26)	526 (92)	0.75 (0.31)
Metaphor	540 (92)	438 (104)	0.45 (0.25)	596 (101)	0.69 (0.27)
EAL					
Literal	623 (139)	652 (145)	0.43 (0.34)	573 (134)	0.78 (0.37)
Metaphor	598 (145)	623 (148)	0.75 (0.43)	842 (144)	0.57 (0.34)
High					
Monolingual					
Literal	370 (117)	260 (150)	0.23 (0.31)	310 (113)	0.80 (0.35)
Metaphor	380 (117)	374 (129)	0.46 (0.29)	441 (117)	0.73 (0.31)
EAL					
Literal	340 (107)	305 (124)	0.30 (0.30)	363 (106)	0.80 (0.31)
Metaphor	383 (110)	328 (135)	0.45 (0.35)	459 (123)	0.67 (0.30)

Notes: Reading times in milliseconds. Back-transformed, rounded model means. Standard errors in parentheses.

Table 27: Predicted means for children reading the first word of the metaphor and the post-metaphor region.

	First word of metaphor (*grades are lotteries*)	Post-metaphor region (*so she feels*)			
	Regression in Probability	Gaze duration	Rereading time	Rereading probability	Total time
Vocabulary					
Small					
Monolingual					
Literal	0.23 (0.08)	509 (118)	481 (131)	0.44 (0.31)	763 (112)
Metaphor	0.26 (0.09)	509 (118)	523 (112)	0.53 (0.32)	852 (107)
EAL					
Literal	0.25 (0.07)	473 (100)	609 (110)	0.56 (0.29)	875 (100)
Metaphor	0.26 (0.08)	465 (103)	531 (97)	0.64 (0.31)	867 (93)
Large					
Monolingual					
Literal	0.23 (0.06)	410 (94)	420 (105)	0.44 (0.26)	609 (91)
Metaphor	0.23 (0.08)	403 (91)	584 (86)	0.56 (0.27)	751 (84)
EAL					
Literal	0.21 (0.10)	447 (143)	468 (157)	0.49 (0.39)	717 (138)
Metaphor	0.22 (0.11)	492 (149)	468 (132)	0.61 (0.41)	810 (132)
Reading fluency					
Low					
Monolingual					
Literal	0.25 (0.06)	564 (95)	461 (104)	0.41 (0.27)	791 (92)
Metaphor	0.27 (0.08)	540 (92)	684 (88)	0.50 (0.27)	946 (85)
EAL					
Literal	0.23 (0.10)	623 (139)	735 (146)	0.61 (0.36)	1196 (134)
Metaphor	0.25 (0.10)	598 (145)	699 (127)	0.64 (0.38)	1165 (128)
High					
Monolingual					
Literal	0.21 (0.08)	370 (117)	438 (127)	0.46 (0.31)	588 (111)
Metaphor	0.22 (0.09)	380 (117)	447 (109)	0.59 (0.32)	676 (106)
EAL					
Literal	0.22 (0.08)	340 (107)	388 (120)	0.43 (0.31)	525 (106)
Metaphor	0.23 (0.08)	383 (110)	356 (99)	0.61 (0.33)	603 (99)

Notes: Reading times in milliseconds. Back-transformed, rounded model means. Standard errors in parentheses.

As in Experiment 1, further models were run to see how individual differences in vocabulary and reading fluency are associated with differences in metaphor processing and comprehension. Continuous predictors were centred and scaled before entering them into the models. Collinearity checks confirmed that reading fluency and vocabulary size correlated with each other, but not too highly to not be entered together (r = 0.5). Table 28 and Table 29 show the results of the models.

Different from Experiment 1, children did not differ on their offline comprehension in the eye-tracking task as measured by comprehension questions – neither by language group, nor by individual differences (all *ns*). However, there were clear online differences. Overall, children with higher reading fluency on the TOWRE read faster and reread less often than less fluent readers. With increasing vocabulary size, children showed shorter gaze durations on the second word. These effects replicate the findings from Experiment 1, where reading fluency was the strongest predictor of fast reading.

Reading fluency also interacted with language group. As in Experiment 1, post hoc tests were estimated using cell means models and single degree of freedom contrasts. This means that individual differences in vocabulary and reading fluency were modelled continuously (Preacher et al., 2005), while interactions were investigated through post hoc comparisons that used point estimates of children's reading at 1SD above or below the sample mean, here referred to as less fluent (-1SD) or fluent readers (+1SD). Post hoc results showed that less fluent, but not fluent EAL readers differed from monolinguals in total reading time on the post-metaphor region (z = 2.3, p < .05 for less fluent readers, z = -0.9, difference *ns* for fluent readers). Numerical differences in both gaze duration and rereading time further suggest that the significant effect on

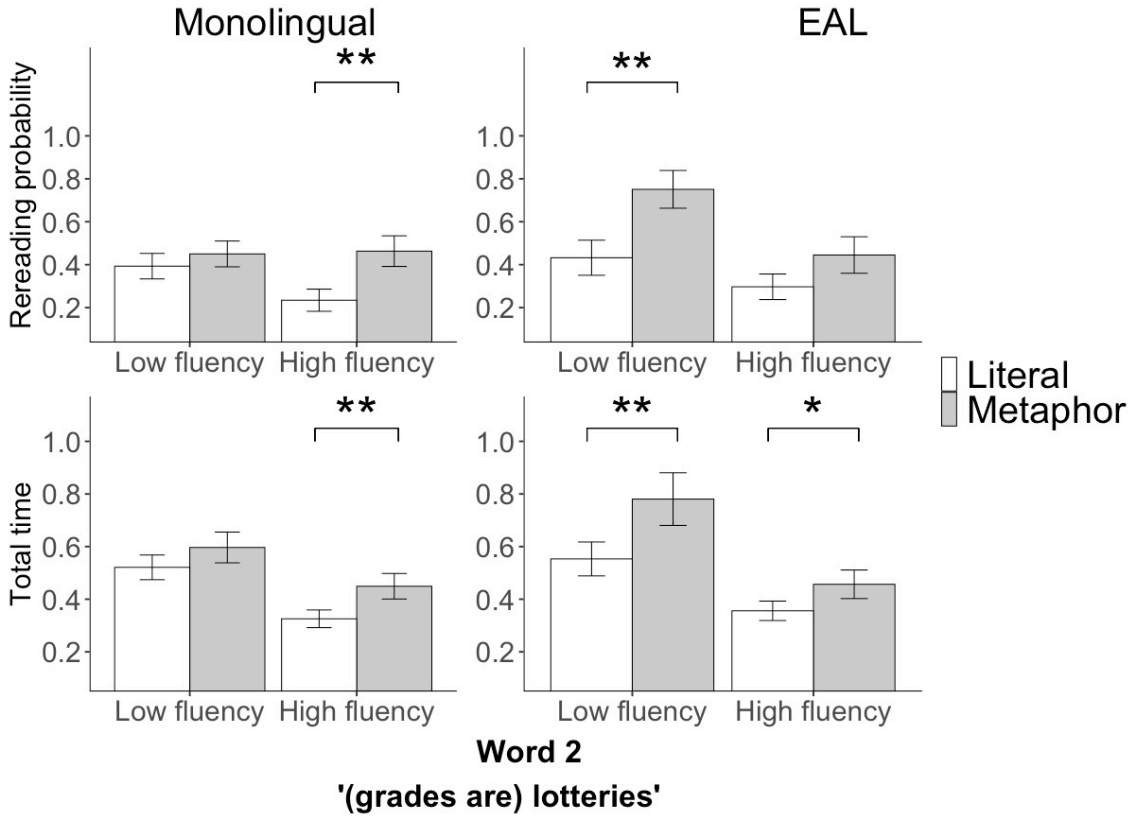
total time added up from both early and later reading (see Table 27 for predicted means). The effect possibly became significant on the post-metaphor region, but not other regions because reading times added up across the multiple words in the post-metaphor region. The effect suggests that EAL readers were more limited by lower reading fluency than monolinguals. This idea is also supported by further interactions with metaphorness that will be discussed next.

There were two three-way interactions between language group, reading fluency and metaphorness on how often and how long children read and reread metaphors (see Figure 6). As in Experiment 1, these effects varied by language group. Consider first EAL readers. Less fluent EAL learners reread metaphors more often ($z = 3.3$, $p < .01$) and for longer ($z = 3.0$, $p < .01$) than literal controls, indicating that metaphors posed particularly high processing demands to these readers. This effect was so large that it also showed in a main effect of language group ($p < .05$ on regression in probability on second word). This was different for fluent EAL readers who reread metaphors considerably faster and less often than less fluent peers, indicating a decrease in processing demands on metaphors. Like the fluent monolingual readers in Experiment 1, fluent EAL readers read metaphors only numerically, but not significantly slower than literal controls. This indicates that some EAL readers in Experiment 2 may be at an earlier stage of metaphor processing than monolingual age peers, as older EAL children read metaphors similarly to younger rather than same-aged monolinguals.

Monolinguals, in contrast, seemed more advanced in their metaphor processing. Less fluent monolinguals read metaphors like fluent EAL readers. That is, they reread metaphors only numerically longer and more often ($z = 1.3$, $p = 0.19$ for total time, and $z = 0.9$, $p = 0.38$ for rereading probability). Fluent monolinguals, however, reread literal

controls less often and more quickly - leading to significant differences between the short reading on literal controls and long reading of metaphors ($z = 3.1$, $p < .01$ on total time, and $z = 3.2$, $p < .01$ on rereading probability). In other words, fluent monolingual readers read faster on literal language, while more fluent EAL read faster on metaphors.

Figure 6: Monolingual and EAL children with low and high reading fluency reading the second word of metaphors or literal controls.



Note: Reading times in seconds. Error bars show the standard error.

To review, main effects indicated that both vocabulary size and reading fluency predicted fast reading, while reading fluency was the stronger predictor of both initial and later reading. Interactions with language group indicated that low reading fluency had a stronger effect on EAL than monolingual readers. Unlike in Experiment 1, further interactions with metaphorness indicated that all children adapted their rereading to metaphors. Metaphor processing was influenced by both language group and word

knowledge. For EAL learners, initially very long and frequent rereading on metaphors decreased with higher reading fluency (similar to younger monolinguals in Experiment 1), while monolinguals read the literal controls more quickly.

Table 28: Models using individual differences, language group and metaphorness to predict reading the second word of metaphors and answering comprehension questions.

	Second word of metaphor (*grades are lotteries*)								Offline	
	Gaze duration		Rereading time		Rereading probability		Total time		Comprehension	
	B	*SE*	*B*	*SE*	*OR*	*SE*	*B*	*SE*	*OR*	*SE*
Language group	-0.01	0.06	0.20	0.10	1.65 *	0.41	0.16	0.08	0.85	0.14
Metaphorness (Met)	0.07	0.06	0.11	0.09	2.34***	0.44	0.27***	0.07	0.56	0.19
Language Group X Metaphorness	-0.01	0.07	-0.19	0.15	1.43	0.52	0.07	0.08	0.61	0.20
Reading fluency (RF)										
RF Main	-0.16*	0.07	-0.51***	0.10	0.50 *	0.14	-0.47***	0.09	1.30	0.23
RF X Language Group	0.19	0.14	-0.38	0.20	0.61	0.34	-0.12	0.18	1.06	0.37
RF X Met	-0.03	0.08	0.23	0.15	1.16	0.40	0.04	0.09	1.10	0.37
RF X Language Group X Met	-0.02	0.15	-0.21	0.29	0.21 *	0.14	-0.38*	0.18	1.44	0.98
Vocabulary (Vocab)										
Vocab Main	-0.16*	0.07	-0.11	0.11	0.81	0.22	-0.17	0.09	1.39	0.25
Vocab X Language Group	-0.13	0.14	-0.24	0.21	0.55	0.31	-0.26	0.18	0.64	0.24
Vocabulary X Met	-0.07	0.08	-0.03	0.17	1.23	0.42	0.01	0.09	0.73	0.26
Vocabulary X Language Group X Met	-0.00	0.16	-0.07	0.33	3.00	2.07	0.22	0.19	0.53	0.38

Notes: Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
** p<.05 ** p<.01 *** p<.001.*

Table 29: Models using individual differences, language group and metaphorness to predict reading the first word of metaphors and the post-metaphor region.

	First word of metaphor (*grades are lotteries*)				Post-metaphor region (*so she feels*)					
	Regression in Probability		Gaze duration		Regression in Probability		Rereading probability		Total time	
	OR	*SE*	*B*	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*
Language group	1.09	0.30	0.03	0.10	1.36	0.37	1.40	0.38	0.10	0.09
Metaphornes s (Met)	1.92 **	0.41	0.02	0.04	0.89	0.14	1.48 **	0.21	0.11 **	0.04
Language Group X Met	1.22	0.54	0.05	0.07	1.22	0.35	0.96	0.27	-0.10	0.08
Reading fluency (RF)										
RF Main	0.80	0.24	-0.46 ***	0.11	0.85	0.26	0.98	0.30	-0.53 ***	0.10
RF X Language Group	1.22	0.73	-0.14	0.22	0.54	0.33	0.48	0.30	-0.42 *	0.20
RF X Met	0.50	0.18	0.11	0.08	1.57	0.50	1.43	0.45	0.06	0.08
RF X Language Group X Met	0.62	0.46	0.09	0.16	3.17	2.03	1.91	1.22	0.21	0.17
Vocabulary (Vocab)										
Vocab Main	1.35	0.42	-0.11	0.11	1.44	0.43	0.90	0.28	-0.15	0.10
Vocab X Language Group	0.86	0.53	0.23	0.22	0.92	0.56	0.86	0.53	0.04	0.19
Vocab X Met	0.96	0.37	0.05	0.08	0.97	0.31	1.15	0.37	0.11	0.09
Vocab X Language Group X Met	0.98	0.77	0.13	0.16	0.67	0.44	0.87	0.56	0.03	0.17

Notes: Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
** p<.05 ** p<.01 *** p<.001.*

3.3.5 Are children aware of their actual level of comprehension, and does this vary between children?

Self-perceived comprehension was measured by a self-report question where children rated their comprehension using a scale from 1 ("I understood nothing") to 4 ("I understood everything"). Their self-perceived comprehension on the experiment was transformed into proportions (so that 0 represents "Nothing" and 1 represents "Everything"), making it the same scale as actual comprehension, which is a child's proportion of correct answers on comprehension questions. Table 30 summarises children's actual and self-perceived comprehension by language groups.

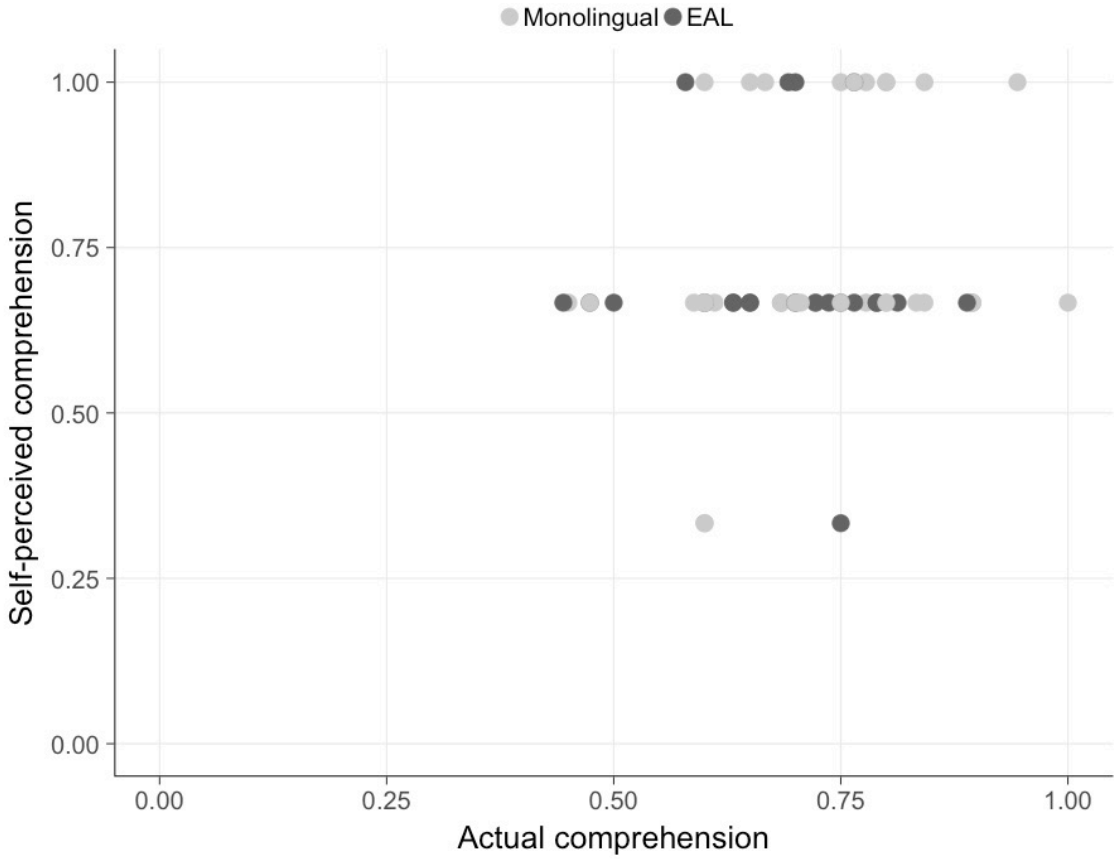
Table 30: Language groups compared on their actual and self-perceived comprehension.

	Actual comprehension	Self-perceived comprehension
Monolingual	0.72 (0.12)	0.73 (0.17)
EAL	0.68 (0.11)	0.71 (0.15)

Notes: Rounded means. Standard deviations in brackets.

Figure 7 shows that children varied in their ability to accurately estimate their comprehension, both monolingual (indicated by circles) and EAL children (indicated by triangles). To explore if self-perceived and actual comprehension were linked and if the linked was moderated by language group, a regression model was run with children's actual comprehension as the outcome variable, and self-perceived comprehension and language group as predictors. Individual differences in reading fluency and vocabulary were added in a second step. However, all predictors were *ns*. In other words, there was no evidence that children's actual comprehension was linked to their self-reported comprehension or that it was moderated by group or individual differences.

Figure 7: Children's actual and self-perceived comprehension by language groups.



3.4 General Discussion

This experiment examined the reading of nominal metaphors in a sample of 9–10-year-old EAL learners and monolinguals. Children's eye movements were recorded as they read stories containing nominal metaphors or literal controls, and children also answered comprehension questions about the story. Reading fluency, vocabulary size and self-perceived comprehension were also measured. Using these measures, the experiment compared children's processing of metaphors and literal language, and linked them to individual and group differences.

3.4.1 Metaphor reading and comprehension

Based on findings in Experiment 1, I predicted that children would struggle more to understand and process metaphors than literal controls. Both online and offline findings support this prediction: metaphors were again harder to understand than literal controls, as reflected in children's lower accuracy on comprehension questions following stories that contained metaphors. This finding is in line with the research discussed in Chapter 2 (Cicone, Gardner, & Winner, 1981; Hessel & Murphy, 2018; Waggoner, Palermo, & Kirsh, 1997), but see (Özçalışkan, 2007; Stites & Özçalişkan, 2013a). In other words, the current experiment supports the idea that metaphors make comprehension more difficult. Importantly, the current findings replicate the difficulty of metaphors in an older sample, thereby extending the observations made in Experiment 1 to older children. These findings further underline the importance of investigating the development of multi-word and figurative vocabulary throughout primary school years, in line with previous research (Kan, 2014; McKendry, 2014; Smith & Murphy, 2015).

As predicted, metaphors also placed higher demands on children's post-lexical processing. This was reflected in children's longer reading of metaphors. As before, children did not modulate their reading when first reading metaphors, but did so when rereading. This is in line with the literature discussed in Chapter 2 (McElree et al., 2006; Olkoniemi et al., 2015), but see (Heredia & Cieslicka, 2016) who also report longer first fixations for Spanish dominant bilinguals reading English metaphors. These effects in first fixation may, however, be due to difficulties in lexical access rather than metaphor processing as such. In the current experiment, metaphor processing showed in increased rereading which may stem from an increased need to suppress or reanalyse a literal interpretation (Johnson & Pascual-Leone, 1989; Kambe et al., 2001; Levorato et al., 2004; Oakhill et al., 2016; von der Malsburg & Vasishth, 2012), or to establish a figurative relationship between two distinct word meanings (Coulson & Van Petten, 2002; De Grauwe et al., 2010; McElree et al., 2006). Effects of metaphorness are plausibly due to a combination of both difficulties. Taken together, the findings support the conclusions drawn on metaphor processing from Experiment 1.

Older children also seemed to reread metaphors more selectively and to link them more closely to the context. While the 7–9-year-olds in Experiment 1 only regressed directly into the first word of a metaphor and did not reread the post-metaphor region, the 9–10-year-olds in the current experiment went directly back into both the first and the second word of the metaphor and also reread the post-metaphor region more often and for longer. Thus, the older children made targeted regressions to metaphors and spent time rereading the context of metaphors, indicating that they were integrating metaphors into the text. This effect may be partially due to differences between stimuli in Experiment 1 and 2, such as the focus on nominal metaphors in

Experiment 2, and the slightly shorter text lengths: the absence of verbal metaphors may have made the metaphorness effects come out more clearly. Furthermore, shorter text length can lead to overall reduced rereading (Radach et al., 2008), which may in Experiment 2 have reduced children's general rereading probability on metaphors in favour of targeted regressions. However, effects of text length are probably unlikely to fully explain differences between experiments since they have to date only been documented for sentences compared to paragraphs (Radach et al., 2008), while both Experiment 1 and 2 tested reading on paragraphs. It is likely that at least partially, reading differences between Experiment 1 and 2 were influenced by the age difference between the two samples. Eye movements during reading develop with the increased lexical and post-lexical processing skills that come with older age (Aaronson & Ferres, 1984; Reichle et al., 2013). Such developments are likely to improve text memory and free up processing capacities for children in Experiment 2 – both of which are important to target selective regressions (Murray & Kennedy, 1988). In other words, children in Experiment 2 were likely better at targeting their regressions directly at metaphors thanks to more efficient language processing.

To review, the current experiment lends further support to the idea that metaphors are difficult to understand, and that they pose additional demands during post-lexical processing. Furthermore, the 9–10-year-olds in this experiment seemed better able to selectively target their regressions and link metaphors to the context, relative to the younger children who participated in Experiment 1, indicating stronger ability with post-lexical processing.

3.4.2 Individual and group differences in Experiment 2

As in Experiment 1, children were compared in two ways. First, language groups were compared using standardised tests of vocabulary size and reading fluency, then children were compared on their metaphor reading and comprehension.

In the standardised tests, EAL and monolingual children did not differ in terms of vocabulary or reading fluency, that is, not only were children in both language groups similarly fluent readers (as in Experiment 1), but their vocabulary sizes also did not differ significantly. This finding is different from Experiment 1 and many previous studies, where EAL learners had smaller vocabularies but comparable reading fluency (Murphy, 2018). While unexpected, this finding is not implausible: EAL learners have highly heterogeneous language skills due to child-internal factors, such as their language aptitude, or environmental factors, such as richness of language input (Paradis, 2011) and different language learning contexts (Murphy, 2014). This thesis sampled EAL and monolingual learners from the whole population, which differs from many previous studies that specifically sampled EAL learners with poor comprehension skills (Spencer & Wagner, 2017). The current study likely differs from other UK results because the sample represents a snapshot of a different part of the EAL population. Similar differences across samples have also been reported from other countries. In the Dutch school context, for example, some studies found significant differences in vocabulary between Dutch monolinguals and minority language learners (Droop & Verhoeven, 2003), while these differences remained numerical in others (Spätgens & Schoonen, 2018).

Beyond sampling effects, an alternative explanation is the measure of vocabulary used: the BPVS, a measure of receptive vocabulary size – may not measure

the full extent of differences between EAL learners and monolinguals, which is also indicated by further language group differences on metaphor processing (see below). One possibility is that the BPVS may only reveal part of the actual difference between EAL learners and monolinguals, which may be larger in tasks that require depth of vocabulary, as reported for activities that tap into semantic associations (Cremer et al., 2011; Cremer & Schoonen, 2013; Schoonen & Verhallen, 2008; Spätgens & Schoonen, 2018) and understanding of multi-word and figurative phrases (Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014; Smith & Murphy, 2015). In addition, it has been suggested that the psychometric properties of the BPVS differ between bilingual and monolingual participants (Chan, 2007), which may further limit its usefulness in identifying differences between language groups.

Overall, children with higher reading fluency and larger vocabularies showed faster reading times in the experiment. This pattern is in line with predictions based on Experiment 1 and the lexical quality hypothesis in indicating that stronger word knowledge predicts efficient lexical access with (Perfetti & Hart, 2002). In the metaphor condition, the effect of reading fluency differed between EAL learners and monolinguals. Individual differences were modelled continuously, while interactions were investigated through post hoc comparisons that used point estimates of children's reading at 1SD above or below the sample mean, here referred to as fluent (+1SD) or less fluent (-1SD) readers. Less fluent EAL readers took noticeably longer to read metaphors compared to literal controls. More fluent EAL readers, in contrast, spend less additional time reading metaphors compared to literal controls, indicating that EAL learners' processing demands on metaphors decreased with increased reading fluency. Turning to the monolingual group, fluent readers maintained their frequent rereading on

metaphors, but they took less time to read literal controls. This means that fluent monolingual readers spent less time rereading on easy language (i.e., literal controls), but reread frequently when needed, i.e., on metaphors. This pattern is consistent with the idea that overall reading processing becomes more fluent with increasing ease of lexical access and, thereby, supports more efficient post-lexical processing (Perfetti et al., 2008), as has been found across populations and language features (Hopp, 2016; Joseph & Liversedge, 2013; Luke et al., 2015; Tiffin-Richards & Schroeder, 2018).

Unexpectedly, however, differences between the language groups in processing were not mirrored by differences in offline comprehension for the experimental stories: offline comprehension was unrelated to the language group or individual differences in reading fluency and vocabulary. This may mean that differences in rereading were unrelated to text comprehension. This is, however, unlikely given the difficulties posed by metaphors and the link between rereading and comprehension, particularly on difficult texts (Schotter et al., 2014). Possibly, the absence of differences was due to limitations in the way comprehension had been measured: to not interrupt children's reading during the task, comprehension questions were limited to forced-choice questions that could be answered by clicking a button. While this made the questions highly compatible with a classical eye-tracking task, the forced-choice questions may have lacked the necessary sensitivity to identify different levels of metaphor comprehension. This idea is supported by a study with younger monolingual and EAL children which used similar forced-choice questions following metaphors: differences in metaphor comprehension between language groups did not show on the forced-choice question itself, only on an additional task where children explained their chosen answer (Hessel & Murphy, 2018). Future research may, thus, wish to add more fine-grained

measures of comprehension after each text, i.e. explaining metaphors or written text summaries (Hyönä, Lorch, Robert F., & Kaakinen, 2002). However, it remains to be seen how the presence of more extensive questions during an experiment itself may influence children's eye movements (for effects of questions on eye movements, see for example Weiss et al., 2017).

There were also no differences between children (neither by language group, nor reading fluency or vocabulary) identifiable in the measure of self-report, which had been included to assess children's ability to accurately describe their comprehension (Eilers et al., 2018a; Kan, 2014; Martinez & Murphy, 2011). However, actual comprehension did not appear linked to self-perceived comprehension among children, nor was it moderated by differences between children. This could be seen as evidence that both monolinguals and EAL learners at age 10 are equally unable to accurately judge their comprehension, contradicting previous suggestions that EAL children may have particular difficulty with noticing misinterpretations of figurative language (Burgoyne et al., 2013; Kan, 2014).

However, the absence of an effect may also be due to limitations in the measures used: as with the comprehension questions, the self-report questions had been adapted to fit with the eye-tracking experiment. This means that the self-report was limited to a single question following the whole experiment, compared to previous research that measured self-perceived comprehension after each experimental text (Martinez & Murphy, 2011). While this may be sufficient to identify awareness in a simpler situation, for example, to see if children notice mismatches or not (Eilers et al., 2018a), it may not be sufficient to discern comprehension differences for metaphors. Future research may wish to present questions around comprehension monitoring in more

sensitive manipulations, one possibility of which will be presented in the following chapter.

To conclude, Experiment 2 contributes further insight to understanding metaphor processing in older children and replicates previous findings on how children read metaphors. It also replicated the finding that EAL learners' and monolinguals' metaphor processing is associated with individual differences; in this case this was primarily reading fluency. At the same time, Experiment 2 raises further questions on comprehension monitoring which will be re-considered in the coming chapter.

3.4.3 Individual differences across Experiments 1 and 2

Experiments 1 and 2 together collected data from 107 children, data which speak to the roles of language groups and individual differences in reading fluency and vocabulary in children's metaphor processing. It is important to note here that Experiments 1 and 2 were not designed as cross-sectional studies: stimuli were adapted in Experiment 2 – by shortening texts and focusing on nominal metaphors – in response to what had been learned from Experiment 1. All following discussions are, thus, limited by the general caveat that they would need to be replicated in a cross-sectional or longitudinal design to test their reliability. However, a comparison of metaphor reading across Experiments 1 and 2 also affords a valuable first indication of how metaphor processing may develop with increased lexical and post-lexical processing skills that come with older age (Aaronson & Ferres, 1984; Reichle et al., 2013), and links metaphor processing to previous suggestions of an inverted U-shaped trajectory in children's post-lexical processing (Tiffin-Richards & Schroeder, 2018). The stronger lexical skills of older children in Experiment 2 compared to children in Experiment 1 is illustrated in Table 31 using their scores on the TOWRE. Both groups of children are from the same

population of EAL learners and monolinguals in Oxford primary schools, and read comparable stories containing metaphors and literal controls. The following discussion will, thus, suggest an interpretation that considers evidence from both experiments and relates them to the wider literature on children's eye movements during reading.

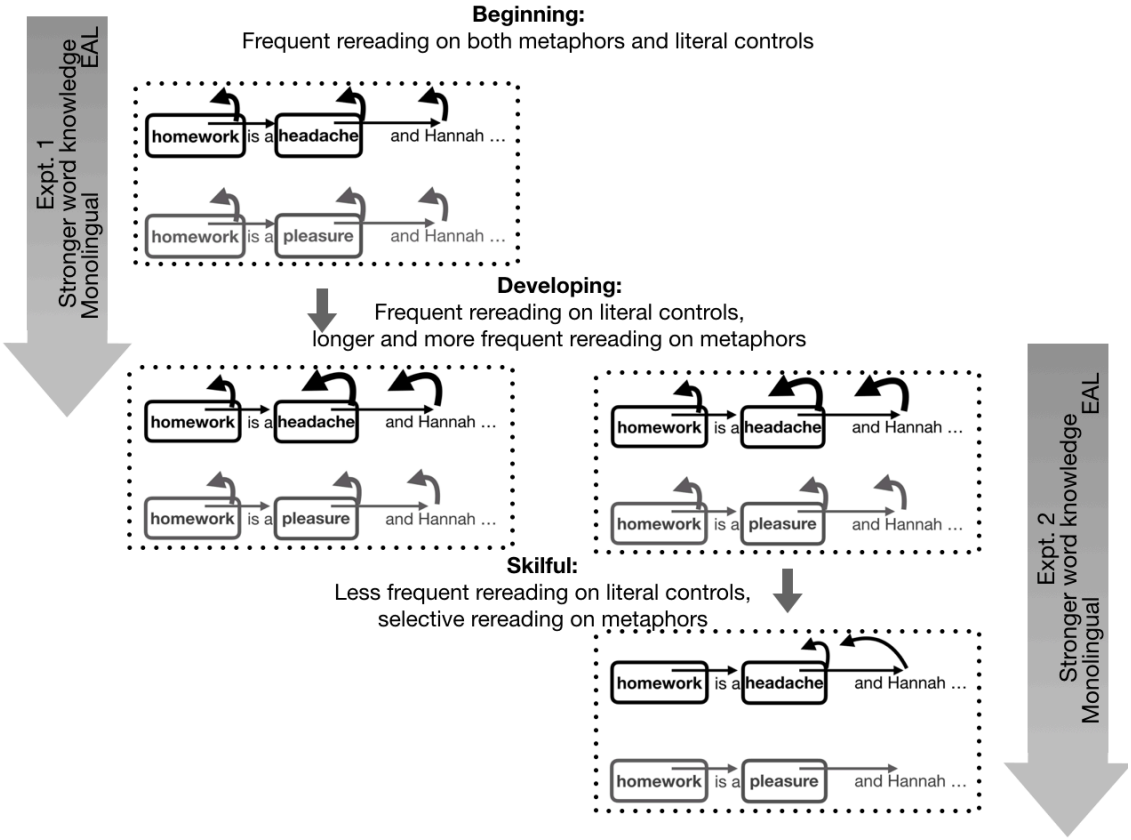
Table 31: Reading fluency of children in Experiments 1 and 2.

	Experiment 1	Experiment 2	Cohen's d
Reading fluency (TOWRE raw scores)	108 (17)	121 (18)	0.7

Notes: Round means. Standard deviations in brackets.

Figure 8 presents a way to reflect on the data from both experiments along a continuum of increasing metaphor processing abilities. The continuums within age group are juxtaposed to highlight parallels with longitudinal trends as summarised in the inverted U-shaped model, with children's post-lexical processing developing from no reaction to metaphors (in Figure 8: "Beginning"), to strong processing demands ("Developing), to eventual fast processing ("Skilful") (Tiffin-Richards & Schroeder, 2018). Figure 8 summarises differences in word knowledge as reading fluency and vocabulary size. Considering them together is appropriate, given that reading fluency and vocabulary size seem to have similar effects on metaphor reading, both within Experiment 1, and between effects predicted by vocabulary in Experiment 1 and those predicted by reading fluency in Experiment 2. Plausibly, both vocabulary size and reading fluency interact to make reading efficient, thus supporting post-lexical processing (Perfetti & Stafura, 2014; Perfetti et al., 2008), as also indicated by the fact that different individual difference measures show similar effects in online reading (Connor et al., 2015; Luke et al., 2015; Perfetti et al., 2008).

Figure 8: Schema of increasingly skilful metaphor processing in children's reading.



Notes: the upper row of each box shows reading on metaphors, while the lower row shows reading on literal controls. Forward arrows show progressive fixations. Backwards arrows show regressions. Thicker arrows indicate more frequent rereading, thinner arrows indicate less frequent rereading.

What, then, does the juxtaposition of Experiments 1 and 2 suggest about EAL and monolingual children's metaphor processing? As shown in Figure 8, differences in metaphor processing were strongly linked to word knowledge within both age groups, with the exception that EAL children were one step behind monolinguals with comparable word knowledge. At the "Beginning" stage, children reread frequently on literal language but did not adapt their rereading to metaphors. This stage only showed in EAL readers with less developed word knowledge in Experiment 1, possibly because this sub-group comprised children with particularly small vocabulary sizes. Their

rereading pattern is similar to children with weaker comprehension skills who failed to adapt their rereading to the presence of novel idioms (Oakhill et al., 2016), sentence endings (Tiffin-Richards & Schroeder, 2018), or noun anaphors at greater distances from their referent (Joseph et al., 2015). In an upside-down U-shaped curve, this stage would represent the initial low starting point, where children do not adapt their reading when faced with difficulties in the text (Tiffin-Richards & Schroeder, 2018). This reading pattern may be due to slower lexical access, which could have impacted post-lexical processing to the point where children did not decide to reread metaphors more than literal controls (Perfetti et al., 2008). An additional influence may be that EAL children have learned to accept a low level of comprehension (Oakhill et al., 2005), which could have lowered the likelihood that they will reread on metaphors, despite comprehension difficulties (Bicknell & Levy, 2011). Questions regarding EAL learners' comprehension monitoring will be considered further in Chapter 4 to add to current evidence on how word knowledge is linked to reading processing on metaphors.

In contrast to the beginning stage, children with more developed word knowledge adapted their rereading to metaphors and spent a long time rereading them (see reading pattern "Developing" in Figure 8). This was suggested, for example, for monolinguals with less developed word knowledge in Experiment 1, and it was also clearly enacted by EAL readers with less developed word knowledge in Experiment 2. These children took particularly long to read metaphors compared to literal language, leading to as much as 269ms longer mean total time when reading the second word of metaphors compared to literal controls. They clearly did not fail to adapt their reading to metaphors, but neither was their metaphor processing efficient. It seems that their post-lexical processing – processes such as integrating the meaning of both words in a

metaphor – was slowed down considerably by effortful lexical access, as has been reported for adult EAL learners interpreting metaphors (Segal & Gollan, 2018). Slower post-lexical processing is likely to necessitate frequent rereading, for example, to resample linguistic information that was hard to remember long enough to finish processing it (Nation, Marshall, & Altmann, 2003). Similarly high processing demands have been reported for children on other language features, for example syntactic ambiguities (Joseph & Liversedge, 2013) and clause-final words (Tiffin-Richards & Schroeder, 2018). The developing stage is equivalent to the highest point of the upside-down U-shaped curve (Tiffin-Richards & Schroeder, 2018).

Finally, children at the upper end of word knowledge skills took considerably less time to reread both metaphors and literal controls, rereading primarily on metaphors (see reading pattern "Skilful" in Figure 8). Trends towards this were visible in, for example, monolinguals with more developed word knowledge in Experiment 1 and EAL children with more developed word knowledge in Experiment 2, but was most prominent in monolinguals with developed word knowledge in Experiment 2. This indicates that some children can reach a high stage of selective rereading by age 10. The decrease in processing effort on metaphors is in line with previous findings that readers became more efficient in their post-lexical processing with increased lexical quality (Luke et al., 2015; Tiffin-Richards & Schroeder, 2018). In an upside-down U-shaped curve, this stage would represent the decrease of processing demands after the initial peak. Furthermore, the overall, more fluent reading is reminiscent of the selective rereading that has in the past been described as the "explorer" style and likened to strategic reading and skilful comprehension monitoring (Connor et al., 2015; Ehrlich, Remond, & Tardieu, 1999; Olson, Kliegl, Davidson, & Foltz, 1985).

The trends seen in Experiments 1 and 2 illustrate how individual differences in children's metaphor reading follow a U-shaped development curve from first failing to notice, to cumbersome processing, to efficient processing (Tiffin-Richards & Schroeder, 2018). The patterns observed resonate with findings for other language features and populations, which potentially indicate that children's reading may develop in similar ways across different language features (Eilers et al., 2018a; Hopp, 2016; Luke et al., 2015). The current interpretation sees word knowledge as central to becoming a skilful reader, as well as for EAL children reading metaphors and beyond. To directly test this assumption, future research could observe how metaphor processing and word knowledge develop longitudinally or, better still, train children's word knowledge to observe the effect on their metaphor reading.

Another key point illustrated in Figure 8 is that EAL children were one step behind monolinguals in their metaphor processing relative to word knowledge as measured in standardised tests. The way that metaphor processing differed between EAL learners and monolinguals coincides with different accounts in the literature, depending on their stage in the suggested developmental trajectory. At the beginning stage, EAL learners with small vocabulary sizes failed to adapt their rereading to metaphors, in line with previous accounts of readers lacking awareness of comprehension difficulties with figurative language (Kan, 2014; Martinez & Murphy, 2011). Additionally, for more advanced readers, monolinguals with relatively strong word knowledge – but not EAL learners – reached the stage of skilful metaphor processing where they reread fluently and selectively on metaphors, a finding that resonates with the suggestion that EAL learners particularly struggle to fully process figurative phrases (Burgoyne et al., 2013).

Wider language group differences in metaphor processing may stem from the degree to which understanding metaphors requires efficient meaning processing, which, even for highly proficient adult EAL learners, is limited by smaller vocabulary sizes (Segal & Gollan, 2018). Metaphor processing in Experiments 1 and 2 may tap a dimension of language processing – establishing extensive meaning relationships – that is particularly challenging for EAL learners. This may be due to particular gaps in vocabulary knowledge, as reported for EAL learners' limited meaning associations with words they know (Cremer et al., 2011; Cremer & Schoonen, 2013; Schoonen & Verhallen, 2008; Spätgens & Schoonen, 2018) and difficulties understanding multi-word and figurative phrases (Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014; Smith & Murphy, 2015). For example, it would be possible to score highly on the TOWRE by only knowing how words are pronounced but not what they mean, and to score highly on the BPVS and the TOWK with very shallow knowledge of meaning, i.e., to know enough to link a word to a picture. Furthermore, picture based vocabulary tests, such as the TOWK and BPVS, seem best suited to assess concrete words, which may limit their validity in assessing the actual extent of children's vocabulary knowledge (Monster, 2018). If EAL learners have, for example, fewer meaning associations for each word in a metaphor, this could result in less efficient metaphor processing – even relative to monolinguals with comparable vocabulary sizes.

However, the distinction between vocabulary depth and size is not uncontroversial: while the two can be clearly distinguished in the way they are conceptualised and measured (Nation, 2001), vocabulary size and depth development can be strongly correlated in both monolingual and bilingual children (Vermeer, 2001). Furthermore, comparisons between experimental tasks of depth and standardised tasks

(as in Cremer, 2013; Kan, 2014; McKendry, 2014; Schoonen & Verhallen, 2008; Smith & Murphy, 2015) may be confounded by methodological limitations of using standardised tests with bilinguals. The tests used in this study were normalised using a monolingual population (Dunn et al., 1997; Torgesen et al., 2012; Wiig & Secord, 1992). Additionally, these tests may have different psychometric properties in bilingual compared to monolingual test-takers (Chan, 2007), which may limit a test's reliability in estimating bilingual children's language skills. Further research may wish to compare bilingual and monolingual children across different standardised and experimental measures of vocabulary depth and size, in order to gain a more reliable estimate of the extent of differences in vocabulary depth beyond size.

3.5 Conclusion

Experiment 2 has replicated earlier findings in children's metaphor processing and presented new insights into individual differences. In the introduction to this chapter, three main questions were posed. The first was whether Experiment 2 would replicate the metaphorness effects. As expected, children showed clear signs that metaphors were challenging, both through lower comprehension and their increased rereading of them. The older sample in Experiment 2 showed further differences to their younger peers: they made more targeted regressions into the metaphor and read the following context more often, indicating the older children were better able to selectively direct their attention during reading.

The second question explored language group differences in standardised tests. As expected, EAL and monolingual children did not differ in reading fluency. In addition, and unexpectedly, EAL children did not differ significantly from

monolinguals in their vocabulary, albeit EAL learners had numerically smaller vocabularies. This may be due to sampling effects, or a failure on the part of the BPVS to capture all differences between language groups, such as differences in vocabulary depths.

This links to the third question: whether EAL and monolingual children differed in their metaphor processing and comprehension. Results showed that language groups differed in their metaphor processing when reading fluency was considered. All children reread metaphors more frequently and for longer than literal controls. With increased reading fluency, children became more efficient in their metaphor processing. EAL learners were, again, behind monolingual as a result of their word knowledge, and only fluent monolingual readers reread metaphor quickly and selectively. Together with the findings from Experiment 1, children's metaphor processing was interpreted as developing in an inverted U-shaped trajectory with increasing word knowledge: from failing to react to metaphors (EAL learners with small vocabularies in Experiment 1), to frequent rereading on metaphors, to efficient and selective rereading (Tiffin-Richards & Schroeder, 2018). This suggestion remains to be tested, however, in cross-sectional designs or training studies.

The results have implications for both research and practice. For EAL research, they indicate that word knowledge influences reading processing in a complex way, beyond a simple correlation of more vocabulary with greater reading comprehension. This indicates the value of online processing studies to shed light on these complexities, possibly in conjunction with training studies that directly test causal relationships between word knowledge and skilful post-lexical processing. Furthermore, EAL learners in Experiments 1 and 2 varied greatly in their English proficiencies. An open

and highly relevant question is what can explain this heterogeneity in abilities? There are some studies in Canadian contexts which investigate this by linking EAL learners' individual differences and their variable English proficiencies (Blom, Paradis, & Duncan, 2012; Paradis, 2011; Paradis & Kirova, 2014). Similar studies would be needed in UK contexts to gauge the extent of variability in English abilities and to see how differences in children's abilities or learning contexts are associated with differences in English proficiency.

For practice, the results support the importance of teaching word knowledge to support reading comprehension, particularly to EAL learners. Teaching vocabulary may have wider influences on comprehension by facilitating skilful post-lexical processing. Together with studies highlighting the importance of English fluency for EAL children's educational achievement (Demie & Strand, 2006; Strand & Demie, 2005; Strand et al., 2015; Whiteside et al., 2017), this suggests that language support can be a powerful tool to support all children in their school careers, particularly EAL learners.

An additional question concerning EAL learners' skilful reading processing is whether EAL learners and monolinguals differ also in comprehension monitoring. Comprehension monitoring has classically been tested through (1) targeted manipulations where children need to notice mismatches in texts, combined with (2) self-report questions that examine the readers' awareness of the mismatch and their reading behaviour. Chapter 4, thus, moves on to investigate EAL learners' comprehension monitoring by combining a semantic mismatch paradigm with self-reporting.

Chapter 4: **Individual Differences in Comprehension Monitoring**

4.1 Introduction

Chapters 2 and 3 outlined how metaphors are processed and understood by monolingual and EAL learners, and how this varies by children and metaphor types. The results indicated that metaphors, particularly nominal metaphors, place high demands on children's post-lexical processing. With growing word knowledge, both EAL and monolingual children started to selectively reread metaphors. EAL children were, however, one step behind monolinguals in their metaphor processing relative to word knowledge: they showed less selective rereading on metaphors given their level of word knowledge, relative to the monolingual children. This may be due to differences in vocabulary depth that go beyond the differences reflected in estimates based on standardised tests of vocabulary (Cameron & Besser, 2004; Cremer, Dingshoff, de Beer, & Schoonen, 2011; Cremer & Schoonen, 2013; Hessel & Murphy, 2018; Kan, 2014; Martinez & Murphy, 2011; McKendry, 2014; Schoonen & Verhallen, 2008; Smith & Murphy, 2015; Spätgens & Schoonen, 2017).

Another possibility is that EAL learners struggle with comprehension monitoring, as has been suggested based on self-report collected alongside comprehension measures on idiomatic and figurative phrases (Burgoyne et al., 2013; Kan, 2014; Martinez & Murphy, 2011). As introduced in Chapter 1, the term comprehension monitoring describes the ways in which readers check and regulate their comprehension (Connor et al., 2015; Kim et al., 2018; Oakhill et al., 2005). Two separate processes seem involved: first, readers evaluate the incoming information and notice a mismatch, and then, they re-analyse and repair the mismatch (Connor et al.,

2015). Both processes are also involved in figurative language processing (Coulson & Van Petten, 2002; Levorato et al., 2004; Oakhill et al., 2016) and EAL readers may struggle particularly with them (Burgoyne et al., 2013; Kan, 2014; Martinez & Murphy, 2011). To date, however, there have been no studies that used online measurement to directly tap EAL children's comprehension monitoring during reading itself. Experiment 3, reported in this chapter, will address this by measuring EAL and monolingual children's comprehension monitoring as they read text containing semantic mismatches, such as when a *kitten* suddenly *barks*. Reading will be measured through eye-movement recordings followed by self-report measures. This procedure allows a targeted investigation of EAL and monolingual children's comprehension monitoring when reading, promising to shed further light on the way their reading processing might differ.

4.1.1 Children's comprehension monitoring during reading

Comprehension monitoring has traditionally been studied by inserting mismatched information in a text – for example a syntactic mismatch between pronoun and verb, or a semantic mismatch between two pieces of information in the story, and by asking readers to underline or verbalise the mismatch (Baker, 1984; Kim & Phillips, 2014; Meyers, Lytle, Palladino, Devenpeck, & Green, 1990; Oakhill et al., 2005). Such offline studies have shown that comprehension monitoring is closely linked to overall comprehension skill (Kim & Phillips, 2014) and is stronger in older than young children (Baker, 1984), and in skilled than less skilled comprehenders (Oakhill et al., 2005). A limitation of these studies, however, is that children need to be instructed to look for mismatches and to point them out. This differs from actual comprehension monitoring

during natural reading, which needs to happen spontaneously. An alternative methodology is to test comprehension monitoring by recording online processing.

A number of studies have tested comprehension monitoring by recording children's eye movements while they read mismatches (Connor et al., 2015; Kim et al., 2018; Kinnunen & Vauras, 1995; van der Schoot, Reijntjes, & van Lieshout, 2012; Vorstius et al., 2013). When children fail to adapt their reading to the mismatch, this is seen as a sign of comprehension monitoring failure, while longer reading times are seen as indicative of increased attention on the mismatch, indicating successful monitoring. This method allows the measurement of children's reactions to mismatches without needing explicit instructions, secondary tasks or verbalisations. For example, Connor and colleagues asked 10–12-year-old students to read short two-sentence stories where the first sentence introduced a scenario (such as an animal that was *barking*) and the second continued in a way that was either plausible (i.e. consistent with the first sentence, such as the animal being a *puppy*), or implausible (i.e. inconsistent with the first sentence, such as the animal being a *kitten*). They found that all children spent longer reading implausible words initially and in rereading, but that children with stronger academic language skills (a composite of vocabulary, story recall, and academic knowledge) showed a bigger difference in their rereading of implausible words. This indicates that children with stronger word knowledge are more skilful in their rereading. These results indicate important advantages of the task: it involves monitoring of semantic information, just like the metaphor processing task used in Experiments 1 and 2, and it differentiates between children of differing word knowledge. I thus adapted Connor et al.'s task to measure EAL children's online comprehension monitoring in this experiment.

Connor et al. suggested that the two stages of comprehension monitoring – detecting a mismatch, and repairing it – are reflected in different eye movement measures, with longer gaze durations indicating detection, and longer rereading times indicating repair (2015). They further suggested that detection was an automatic process (since all children did it), while repair was a deliberate process during rereading that only children with stronger academic language fully engaged in. There are several reasons to believe that rereading is indeed important for re-analysis and repair: skilled readers tend to reread more when texts contain difficult content (Rayner et al., 2006) or implausible and anomalous words and structures (Connor et al., 2015; Joseph et al., 2008; Rayner, Warren, Juhasz, & Liversedge, 2004; von der Malsburg & Vasishth, 2011; Weiss et al., 2017; Zabrucky & Ratner, 1989). Furthermore, comprehension suffers if readers are unable to reread (Schotter et al., 2014), possibly because rereading allows readers to identify or repair misinterpretations (Metzner, von der Malsburg, Vasishth, & Rösler, 2016; van der Schoot et al., 2012). It is thus plausible that increased rereading on implausible words indicates the re-analysis part of comprehension monitoring.

However, this interpretation of rereading is not shared unanimously. Regressions have also been interpreted as signs of confusion (Murray & Kennedy, 1988), or as means to pause the intake of new information, for example to finish lexical processing (Mitchell, Shen, Green, & Hodgson, 2008). One way to distinguish the purpose of a regression may be to analyse their spatial pattern: regressions directed at re-analysis seem longer and more targeted, while regressions that only serve to pause seem shorter and more diffuse (Murray & Kennedy, 1988). This distinction is particularly relevant when studying children's rereading, where the latter type of

refixations and regressions is more frequent (Blythe & Joseph, 2011). For the current experiment, I thus decided to analyse not only rereading time (Connor et al., 2015), but also rereading probability and regression in probability. In particular, regression in probability has a strong spatial component in that it only includes direct regressions into a word. Regression in probability may thus be a good indicator of the kind of selective rereading that indicates re-analysis as part of successful comprehension monitoring.

Another question is whether rereading is indeed a deliberate strategy, as suggested by Connor et al. Earlier offline studies share this view (Meyers et al., 1990; Oakhill et al., 2005). It is possible, however, that comprehension monitoring is only deliberate when measured in metacognitive tasks, such as when children are asked to identify and verbalise mismatches and describe how they can be solved. During natural reading, however, comprehension monitoring as reflected in selective rereading may be much less deliberate. Perhaps selective rereading looks like a deliberate strategy, but in actual fact is the result of efficient reading processing (Olson et al., 1985; Perfetti & Hart, 2002). For example, efficient reading processing may enable readers to engage in more advanced post-lexical processing (Perfetti et al., 2008) or remember word locations precisely enough to launch regressions (Murray & Kennedy, 1988). This link between fluent reading overall and selective rereading has been described as an *explorer* style of reading, in contrast to a slow and diffuse *plodder* style of reading (Olson et al., 1985). Based on the eye movement record alone, it is thus not possible to establish conclusively if rereading indicates a deliberate strategy.

Hyönä and Nurminen (2006) addressed this by combining eye movement recording with self-report on rereading behaviour. Hyönä and Nurminen found that adult readers were able to accurately describe their rereading patterns, and that

increased regressions were also linked to better comprehension, indicating that their rereading was a deliberate strategy that supported comprehension (Hyönä & Nurminen, 2006). By linking a reader's awareness to their actual rereading behaviour, their study indicates that regressive eye movements can indeed be strategic, in line with other strategy research that investigates awareness as a criterion for strategic action (Macaro, 2006). In the Hyönä and Nurminen study, however, rereading was investigated only in adults and only in terms of global rereading, such as a reader's likelihood to return to headings or topic sentences. There is some evidence to suggest that such global rereading may be more amenable to conscious control than local rereading of mismatches (Weiss et al., 2017). It is an open question if readers, and in particular children, are equally aware of their local rereading. Thus, an adapted version of Hyönä and Nurminen's question was added to Experiment 3, alongside the same question about self-perceived comprehension used in Experiment 2. This way, Experiment 3 combines both online and offline data on children's comprehension monitoring.

4.1.2 The current experiment

Experiment 1 and 2 indicated that EAL children are behind monolinguals in their metaphor processing, relative to their word knowledge. It has been suggested that such difficulties are due to a lack of comprehension monitoring. For this experiment, I thus adapted an existing task to measure EAL and monolingual children's comprehension monitoring when reading text containing implausible words. This allows for the investigation of whether EAL children will struggle to notice or fully process semantic difficulties more generally, which may explain difficulties in metaphor processing. Online rereading patterns were analysed, and combined with self-report measures on self-perceived comprehension and rereading behaviour. All predictions and analyses

were pre-registered prior to data analysis on the online platform *Open Science Framework* (see pre-registration document in Appendix I).

The following research questions are addressed:

1. Do children adapt their reading to implausible words, indicating online comprehension monitoring?

Based on previous findings with the same stimulus set, it was predicted that children would spend longer reading implausible than plausible words both during gaze duration and rereading (Connor et al., 2015).

2. Does children's comprehension monitoring differ by language group (monolingual vs. EAL) or individual differences in reading fluency and vocabulary size?

Based on previous research (Connor et al., 2015), it was predicted that children with stronger word knowledge (larger vocabularies, higher reading fluency) would adapt their rereading (but not their gaze duration) more to implausible words than children with weaker word knowledge.

3. Are children aware of their reading behaviour and comprehension, indicating deliberate comprehension monitoring?

A self-report measure tapped children's self-perceived rereading behaviour and this was compared to their actual reading (Hyönä & Nurminen, 2006). A close overlap between self-perceived and actual rereading on implausible words would indicate awareness of local inconsistences and support the idea that this is an aspect of deliberate comprehension monitoring (Connor et al., 2015). A second self-report question probed children's self-perceived comprehension. It was expected that if children actively

monitored their comprehension and noticed the presence of implausible words, this would show in lower self-perceived comprehension.

4.2 Method

4.2.1 Design

This experiment has a mixed design with one within-subject manipulation of internal consistency (plausible vs. implausible words). The effect of this manipulation was compared between subjects of different language groups (EAL vs. monolinguals) and related to individual differences in reading fluency and vocabulary size. Dependent variables were children's eye movements on plausible and implausible words. Individual differences in reading and vocabulary size were assessed using standardised tests. Children were also compared on their self-reported comprehension and the link between their reported and actual rereading behaviour.

4.2.2 Participants

This experiment was completed by the same participants as in Experiment 2, as described in Chapter 3.

4.2.3 Materials

4.2.3.1 Comprehension monitoring task

Children read 20 two-sentence stories (see Table 32 for an example, and Appendix P for a full list of stimuli). The first sentence of the paragraphs describes a simple event or action, such as an animal barking, and the second sentence continued this action in a way that was either consistent or inconsistent with the first sentence. Internal consistency varied by inserting a target word that continued the story either in a

plausible (e.g. *puppy*) or in an implausible way (e.g. *kitten*). One third of all stories was followed by forced-choice comprehension questions that tapped memory of the story contexts to encourage attentive reading.

As in Experiments 1 and 2, stories were split into lists following a Latin square design. This means that each child read all 20 stories, with half in the implausible condition and half in the plausible condition. Every child thus read each story once, and there were 10 stories per condition, presented in fully randomized order.

Table 32: Example story and question.

Rover barks at all passing animals on the street. He's the most alert *kitten / puppy* in the whole neighbourhood in fact.
Comprehension question: Does Rover make a lot of noise?
A: Yes, he does. B: No, he is very quiet.

4.2.3.2 Stimulus development

The task was adapted from Connor et al. where plausible targets and implausible words were matched for number of letters, syllables and morphemes, as well as age of acquisition and frequency. The stimuli set was thus already tightly controlled. The following additional changes were made: since the original study had been conducted with American speakers of English, words specific to American English were replaced with closely matched British English alternatives (e.g. *nail* for *bolt*). The adapted words were checked again on all dimensions to make sure that they did not differ across plausibility conditions (see Table 33). The layout of all stories was furthermore streamlined by matching stories on the number of words and characters preceding a target region (13 words, or 69-77 letters), and on words and characters following the target word (6 words, or 27-42 letters). This was done to reduce noise (Rayner, 1977),

and to make the texts amenable to scanpath analyses. The adapted texts were proofread and discussed with two speakers of British English to make sure that the language was appropriate and acceptable.

Table 33: Matching of implausible words and plausible controls.

	Implausible words	Plausible controls	*t*-test results
Frequency	3.2 (1.2)	2.2 (2.2)	$t(29.4) = 1.67$, $p = .11$
Number of letters	5.5 (1.1)	5.4 (1.2)	$t(37.7) = 0.14$, $p = .89$
Number of syllables	1.2 (0.4)	1.3 (0.5)	$t(35.9) = -1.13$, $p = .27$
Number of morphemes	1.6 (0.6)	1.6 (0.6)	$t(38.0) = 0.27$, $p = .80$
Age of acquisition	5.4 (1.7)	5.7 (2.7)	$t(32.5) = -0.42$, $p = .68$
	Maximum: 9.5	Maximum: 12.5	

Note: Log frequency based British National Corpus.

4.2.3.3 Interest area and eye-tracking measures

In line with the procedure used by Connor et al., eye movements were analysed for one interest area to summarise all fixation and saccades on the implausible word or its plausible control (e.g. / kitten/ or / puppy/). The following reading measures were analysed: gaze duration, rereading time, rereading probability, and regression in probability.

4.2.3.4 Apparatus

The same equipment and set-up were used as in Experiment 1 and 2.

4.2.3.5 Self-perceived comprehension

As in Experiment 2, children were asked to rate how well they had understood the stories after they had finished reading them, using the same question format and procedure describe in Chapter 3.

4.2.3.6 Self-perceived rereading behaviour

Children also answered four questions on their rereading behaviour. They were asked to rate four statements about their own rereading, such as “I reread words when I don't understand them”. They could rate on a 3-point scale if the statement sounded: “Not like me”, “A little like me”, or “A lot like me”. They were also given the option to answer “Don’t know”. Children heard the statements and both saw and heard the scale steps they could choose from (see Appendix Q for questions and scales). The questions were modelled on a previous study (Hyönä & Nurminen, 2006), but adapted for use with children: the number of questions was reduced from 8 to 4 by removing questions that tapped global reading patterns (and thus were not directly relevant to the current investigation of local comprehension monitoring). Furthermore, the number of scale steps was reduced from 6 to 3, and the phrasing of questions was simplified. A pilot interview with a 7-year-old child confirmed that the questions were not too difficult for use with primary school children.

4.3 Procedure

The current experiment was administered in the second testing session of Experiment 2 following the procedures described in Chapter 3. Children were told that some stories would be followed by comprehension questions.

4.4 Results and Discussion

4.4.1 Analytical strategy

As in Experiments 1 and 2, children’s reading of internal inconsistencies was assessed by analysing eye-movement patterns. Eye movement data were processed the same way

as in Experiments 1 and 2. Extremely short or long fixations (<80ms and >1200ms) were deleted, amounting to 3.1% of the data. The data were then analysed in R (R Core Team, 2016) using linear and binomial MEMs for eye movement data, and regressions for self-report.

The following presents the results by research questions.

4.4.2 Do children adapt their reading to implausible words, indicating online comprehension monitoring?

Children's reading of plausible and implausible words was captured through eye movements which were analysed using five measures: gaze duration, rereading time, total time, rereading probability, and regression in probability. Table 34 summarises the data.

Table 34: Means for children reading the target word. All reading times in milliseconds.

	Target word (*puppy* / *kitten*)				
	Gaze duration	Regression in probability	Rereading probability	Rereading time	Total time
Monolingual					
Plausible	324 (197)	0.15 (0.36)	0.26 (0.44)	104 (236)	428 (318)
Implausible	344 (235)	0.28 (0.45)	0.46 (0.50)	271 (449)	615 (510)
EAL					
Plausible	300 (157)	0.14 (0.35)	0.26 (0.44)	152 (666)	451 (681)
Implausible	408 (400)	0.29 (0.45)	0.49 (0.50)	259 (472)	667 (609)

Notes: Reading times in milliseconds. Rounded means. Standard deviations in parentheses.

One MEM was run per measure with language groups and consistency as predictors. Table 35 summarises the results of these models and Figure 9 presents the time course of the implausibility effect in three columns: the first column shows children's initial reading on the target word, the second their rereading, and the final column the overall reading time. As expected, children read implausible words more slowly both in initial reading and in rereading. These findings replicate those from

Connor et al. (2015). Furthermore, the current experiment also measured regression in probability. Children showed increased regression in probability in the implausible condition, consistent with them re-reading more selectively.

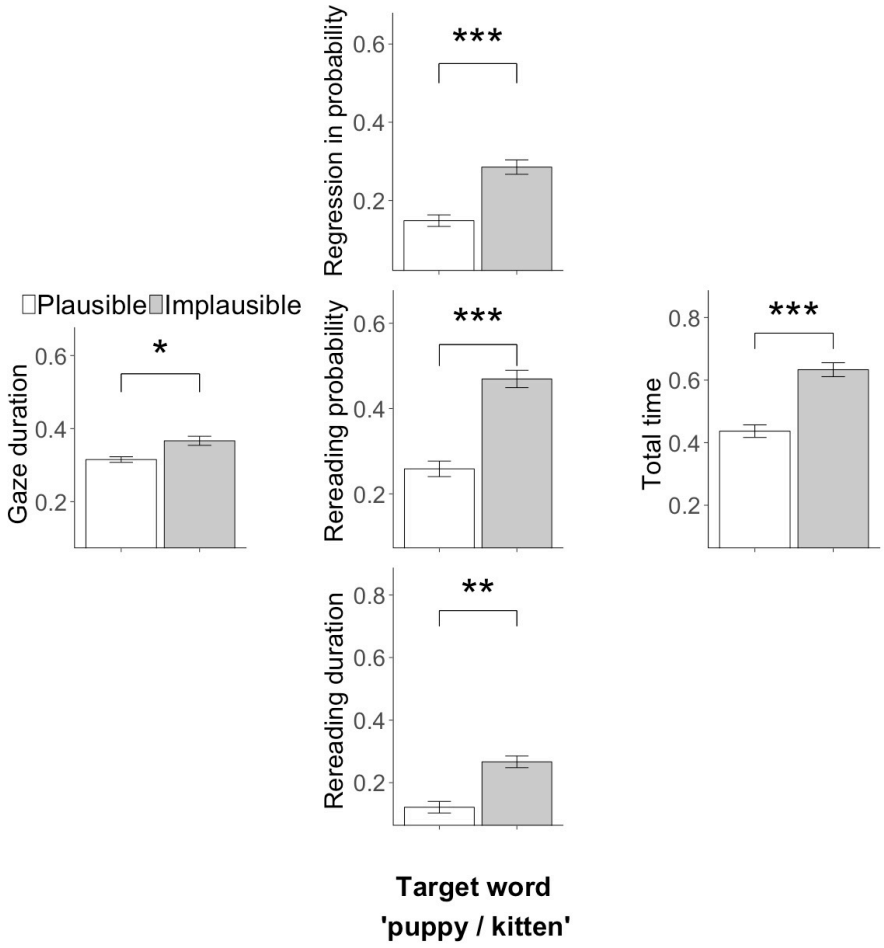
Table 35: Models using language group and implausibility to predict reading the target word.

	Target word (*puppy* / *kitten*)									
	Gaze duration		Regression in Probability		Rereading probability		Rereading time		Total time	
	B	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*	*B*	*SE*
Language group	0.01	0.04	1.03	0.23	1.12	0.23	0.04	0.06	0.01	0.08
Consistency	-0.05*	0.02	2.16***	0.41	2.85 ***	0.46	-0.11**	0.04	0.35 ***	0.06
Language Group X Consistency	0.03	0.02	1.11	0.38	1.14	0.36	-0.04	0.04	0.06	0.08

Notes: Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
** p<.05 ** p<.01 *** p<.001.*

As in Experiments 1 and 2, further checks were run to see if there was enough data to support the analyses of rereading. Children reread on 433 trials on target words, which amounts to 37% of analysed trials. On the child level, their probability mean rereading probability was 36% on target words, which is similar to Experiment 2, indicating that there were sufficient instances of rereading to afford reliable analyses, as was also supported by good results on the influence.ME bootstrap test. However, mean rereading probability varied again across children, with 11 children rereading less than 20% of the time, indicating that, as in Experiments 1 and 2, rereading patterns were somewhat less representative across participants than first pass reading times.

Figure 9: Children's reading processing of plausible and implausible words.



*Note: Reading times in seconds. Error bars show the standard error. * p<.05 ** p<.01 *** p<.001.*

4.4.3 Does children's comprehension monitoring differ by language group (monolingual vs. EAL) or individual differences in reading fluency and vocabulary size?

As in Experiments 1 and 2, differences between children were investigated by comparing the two language groups and assessing associations with individual differences in reading fluency and vocabulary. Differences between language groups are summarised in Table 34 and Table 35, presented above. As in Experiments 1 and 2, there was no main effect of language group, indicating, in answer to the first part of the research question, that EAL and monolingual readers do not differ in general in their comprehension monitoring.

Table 36 summarises individual differences based on predicted means for reading times at +/- 1 SD of reading fluency or vocabulary size, respectively. To analyse these data, further models were run to test the impact of the individual differences measures. As discussed in Experiment 2, reading fluency and vocabulary size were centred and scaled, and collinearity checks had confirmed that they could be entered together for the current sample. Table 37 shows the results of the models.

Table 36: Predicted means for children reading the target word by language group and consistency condition.

	Target word (*puppy* / *kitten*)				
	Gaze duration	Regression in Probability	Rereading probability	Rereading time	Total time
Vocabulary					
Small					
Monolingual					
Plausible	309 (19)	0.18 (0.04)	0.22 (0.05)	396 (53)	387 (30)
Implausible	309 (23)	0.22 (0.05)	0.44 (0.07)	388 (46)	478 (46)
EAL					
Plausible	280 (16)	0.13 (0.03)	0.24 (0.04)	399 (47)	368 (25)
Implausible	373 (26)	0.21 (0.04)	0.46 (0.06)	404 (43)	567 (53)
Large					
Monolingual					
Plausible	255 (13)	0.11 (0.02)	0.24 (0.04)	273 (29)	313 (20)
Implausible	277 (17)	0.29 (0.05)	0.45 (0.06)	459 (43)	467 (37)
EAL					
Plausible	261 (20)	0.15 (0.04)	0.24 (0.06)	266 (43)	332 (31)
Implausible	258 (24)	0.4 (0.08)	0.55 (0.09)	386 (56)	445 (54)
Reading fluency					
Low					
Monolingual					
Plausible	319 (16)	0.14 (0.03)	0.28 (0.05)	342 (35)	403 (26)
Implausible	338 (21)	0.26 (0.05)	0.46 (0.06)	515 (48)	564 (46)
EAL					
Plausible	319 (24)	0.12 (0.03)	0.25 (0.05)	483 (74)	441 (39)
Implausible	366 (32)	0.35 (0.07)	0.53 (0.08)	508 (69)	626 (73)
High					
Monolingual					
Plausible	248 (16)	0.15 (0.03)	0.18 (0.04)	316 (44)	301 (24)
Implausible	253 (19)	0.25 (0.05)	0.44 (0.07)	346 (41)	397 (39)
EAL					
Plausible	229 (13)	0.16 (0.03)	0.23 (0.05)	220 (28)	277 (19)
Implausible	263 (19)	0.25 (0.05)	0.48 (0.07)	307 (33)	403 (38)

Notes: Back-transformed, rounded model means. Reading times in milliseconds. Standard errors in parentheses.

Table 37: Results of models using individual differences, language group and consistency to predict reading of the target word.

	Target word (*puppy* / *kitten*)									
	Gaze duration		Regression in Probability		Rereading probability		Rereading time		Total time	
	B	*SE*	*OR*	*SE*	*OR*	*SE*	*B*	*SE*	*B*	*SE*
Language group	0.01	0.03	1.12	0.26	1.18	0.25	0.04	0.05	0.03	0.06
Consistency (Cons)	-0.05	0.02	2.34 ***	0.46	2.92 ***	0.50	-0.11*	0.05	0.33 ***	0.06
Language Group X Cons	0.03	0.02	1.34	0.48	1.14	0.39	-0.02	0.04	0.06	0.09
Reading fluency (RF)										
RF Main	-0.29***	0.06	0.90	0.21	0.81	0.19	-0.43***	0.10	-0.38 ***	0.07
RF X Language Group	0.02	0.06	1.27	0.59	1.34	0.62	0.18	0.10	-0.13	0.15
RF X Imp	0.01	0.04	0.65	0.25	1.23	0.45	0.01	0.08	-0.02	0.10
RF X Language Group X Cons	0.01	0.04	0.54	0.42	0.85	0.63	0.15	0.08	0.09	0.19
Vocabulary size (Vocab)										
Vocab	-0.19 **	0.06	1.32	0.32	1.14	0.28	-0.17	0.11	-0.15	0.08
Vocab X Language Group	0.05	0.06	1.57	0.76	1.03	0.50	0.09	0.11	-0.06	0.15
Vocab X Cons	0.04	0.04	2.43 *	1.00	1.14	0.44	-0.23**	0.08	0.03	0.10
Vocab X Language Group X Cons	-0.08 *	0.04	0.99	0.81	1.26	0.98	-0.05	0.08	-0.33	0.20

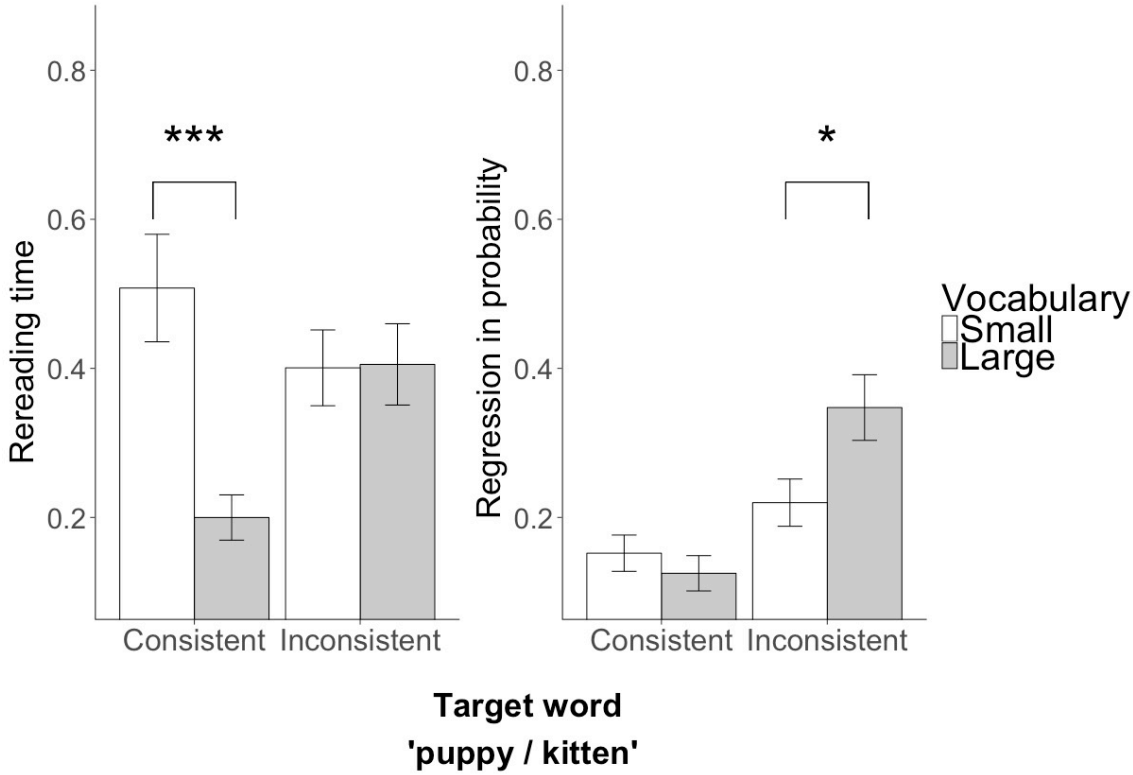
Notes: Estimates are beta coefficients (B) for linear models and odds ratios (OR) for logit models.
**p<.05 **p<.01 ***p<.001.*

As in Experiment 1 and 2, children with higher reading fluency and larger vocabularies generally read faster in this experiment. As before, vocabulary was associated with shorter gaze durations, while reading fluency predicted faster rereading. Together, this indicates that fluent reading in the experiment is associated with strong word knowledge, in particularly reading fluency as measured by the TOWRE-2.

Vocabulary size also interacted with children's rereading of the target words, as summarised in Figure 10. As in Experiments 1 and 2, post hoc tests were estimated using cell means models and single degree of freedom contrasts. These post hoc tests compared the predicted means for reading times at +/- 1 SD of vocabulary size and reading fluency, respectively. Results showed that only children with large vocabularies were significantly more likely to directly go back to target words in the implausible condition; they also spent significantly longer reading implausible than plausible words. Note that at least numerically, the same trends were visible in reading fluency, for example in a larger probability to reread implausible compared to plausible words particularly amongst more fluent readers (see Table 26). These effects largely replicate rereading differences shown by the children with strong academic language in Connor et al.'s experiment (2015). What is different from those findings, however, is the way in which exactly children with larger vocabularies differed from children with smaller vocabularies. In Connor et al. (2015), children with strong academic language showed a bigger difference in rereading time between plausible and implausible words than children with weaker language skills. In the current experiment, there were two distinct differences. First, children with larger vocabularies took less time to reread plausible words compared to those with smaller vocabularies (z = -3.4, p < .001), and second, they were more likely to regress directly back into implausible words (z = 2.3, p < .05).

This indicates that readers with large vocabularies not only reread implausible words for longer, they also became faster in their processing of plausible controls and at the same time more selective in their rereading on implausible targets.

Figure 10: Children with small and large vocabulary reading plausible and implausible words, plotted for predicted means at +/- 1 SD of vocabulary size .

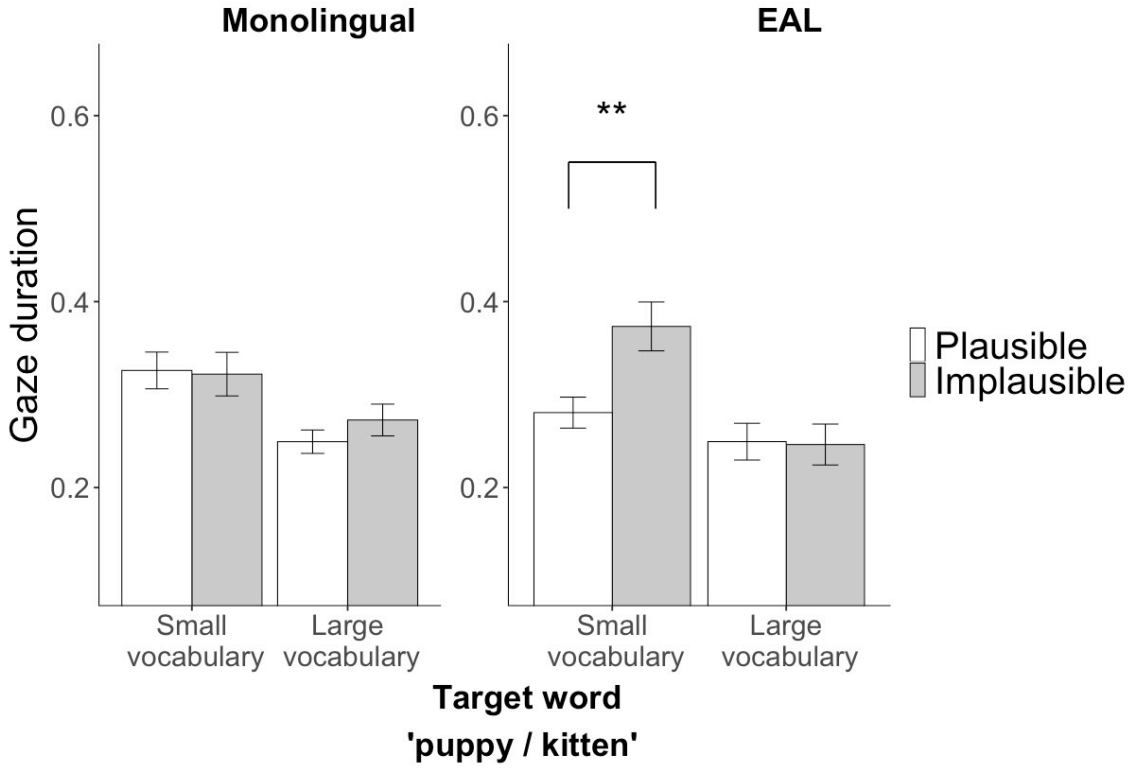


*Note: Predicted means for reading times at +/- 1 SD of vocabulary size. Reading times in seconds. Error bars show the standard error. * p<.05 ** p<.01 *** p<.001.*

Furthermore, there was an unexpected three-way interaction between language groups, vocabulary size, and consistency on initial reading, shown in Figure 11. Only EAL children with low vocabulary showed significantly longer gaze durations when reading implausible words than plausible words. That is, longer gaze durations on implausible compared to plausible words were carried by EAL children with small vocabularies, which is a different group of children than those children with large

vocabularies who spent more time on targets in their rereading. This differs from previous findings where individual differences only interacted with consistency in rereading, but not in gaze duration (Connor et al., 2015).

Figure 11: EAL and monolingual children with small and large vocabulary reading plausible and implausible words, plotted for predicted means at +/- 1 SD of vocabulary size.



*Note: Reading times in seconds. Error bars show the standard error. * p<.05 ** p<.01 *** p<.001.*

In summary, large vocabulary size was linked to increased rereading of implausible words, as predicted (Connor et al., 2015). This was mirrored by a numerical trend associated with reading fluency. There were, however, also two new findings: children with larger vocabularies not only reread implausible words for longer, they also reread plausible words more fluently, and reread implausible words more selectively. Furthermore, individual differences also influenced initial reading, where only EAL

learners with smaller vocabularies spent longer reading implausible words than plausible words.

4.4.4 Are children aware of their reading behaviour and comprehension, indicating deliberate comprehension monitoring?

Children were asked four questions that assessed their self-perceived rereading behaviour. Answers were added up to create a total score that captured the degree to which children judged their own rereading to be selective. Answers to one were was reversed before being added since the question assessed non-selective rereading (i.e. children rated the statement "I reread each sentence before moving to the next"). The total score of self-perceived selective rereading was compared to children's actual rereading to assess children's awareness of their rereading patterns. Additionally, children's awareness of implausible words was measured by asking a question about how much children believed to have understood in the stories, where active monitoring was expected to show in low self-perceived comprehension. The total score of selective rereading and the answer on the self-perceived comprehension question were both rescaled from 0 to 1 for readability. The data are summarised by language group in Table 38.

Table 38: Language groups compared on their self-perceived comprehension and rereading.

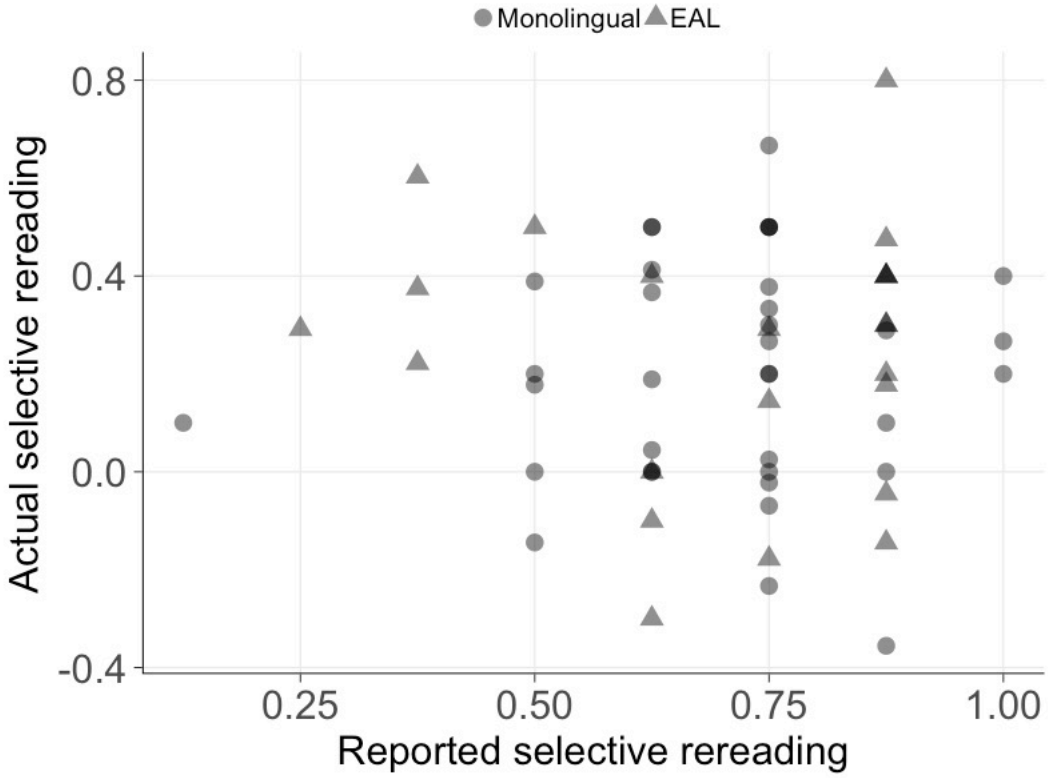
	Self-perceived comprehension	Self-perceived rereading
Monolingual	0.80 (0.21)	0.70 (0.17)
EAL	0.87 (0.19)	0.71 (0.21)

Notes: Round means. Standard deviations in brackets.

Consider first self-perceived rereading. On average, both EAL and monolingual children perceived their rereading to be fairly selective. This means that children reported, for example, to reread only words they found difficult. To investigate if

children's self-report answers indicated actual awareness of their rereading patterns, their reported selective reading was linked to their actual selective rereading (Hyönä & Nurminen, 2006). A score for actual selective rereading was obtained for each child by subtracting their probability to reread on plausible controls from their probability to reread implausible targets. Awareness of rereading was expected to manifest as a strong correlation between a child's actual and self-perceived selective rereading. Figure 12 plots the two scores against each other. This shows that children varied greatly in their ability to accurately describe their rereading, which was true for both monolingual (indicated by circles) and EAL children (indicated by triangles). Similarly, a regression model did not show any significant link between children's reported and actual degree of selective rereading, or any influence of language group or vocabulary size (all predictors *ns*).

Figure 12: EAL and monolingual children's actual and reported selective rereading.



A lack of awareness was furthermore supported by children's high self-perceived comprehension. On average, both EAL and monolingual children rated their comprehension to be at least at 0.80 (out of 1). For comparison, this is higher than children's self-perceived comprehension of metaphors (which was 0.68 and 0.72 for EAL and monolingual children, respectively). Given that 50% of stories contained implausible words, this may indicate that children were not aware of the difficulty caused by these implausible elements in the stories.

Taken together, there was no strong evidence that children were aware of the difficulties posed by implausible words or the degree to which they reread them, thus yielding no support to the interpretation that more frequent rereading on implausible words is strategic and deliberate.

4.5 General Discussion

This experiment investigated the online comprehension monitoring of 9–10-year-old children learning EAL or English only. Children's eye movements were recorded when reading stories containing implausible words that were mismatched with earlier story information, or when reading the same stories with plausible words. Their reading fluency and vocabulary size were also assessed, using standardised measures. The experiment investigated children's comprehension monitoring of implausible words, and linked it to individual and group differences in reading fluency and vocabulary. Additionally, children's awareness and strategic use of comprehension monitoring was investigated through measures of self-perceived comprehension and rereading patterns.

4.5.1 Online comprehension monitoring when reading implausible words

Based on previous findings (Connor et al., 2015; Zabrucky & Ratner, 1989), children were predicted to take longer to read implausible than plausible words both during initial reading and during rereading, with the difference in rereading time being larger in children with strong word knowledge. The findings fulfilled these predictions. Overall, the 9–10-year-old EAL and monolingual children were sensitive to the implausible words during rereading. This is not a trivial finding given that the words were made implausible based on internal inconsistencies within the story, which is a mismatch that is more difficult to process for children than, for example, mismatches with world knowledge (Joseph et al., 2008; Vorstius et al., 2013). So far, comprehension monitoring in the form of increased rereading on internal inconsistencies has only been found in slightly older and monolingual children (Connor et al., 2015; Zabrucky & Ratner, 1989). Connor et al. suggested that this longer rereading reflects deliberate re-analysis. To test the suggestions that rereading reflected (1) re-analysis and (2) a deliberate process, the current experiment analysed eye-tracking measures that tap targeted rereading and combined this with self-report measures. The results partially support, but also extend previous accounts (Connor et al., 2015).

Consider first the question if rereading indicates re-analysis. Children reread implausible words more than plausible controls, just as has been found for children and adults reading semantic or syntactic mismatches (Connor et al., 2015; Joseph et al., 2008; Rayner et al., 2004; von der Malsburg & Vasishth, 2011; Weiss et al., 2017; Zabrucky & Ratner, 1989). As in Connor et al., the rereading on implausible words was also more pronounced in children with larger vocabularies. From longer rereading times alone, however, it is not clear if they indicate re-analysis (Metzner et al., 2016; van der

Schoot et al., 2012), or a simple pause, or more general rereading (Mitchell et al., 2008; Murray & Kennedy, 1988). The spatial analyses of rereading showed that the longer rereading of implausible words in children with larger vocabularies was driven both by (1) faster rereading on plausible controls and (2) targeted rereading on implausible words. In other words, children with large vocabularies reread less often overall, but when they reread, they did so for longer and more selectively in the implausible condition. Such selective rereading patterns make it conceivable that children's rereading on implausible words indeed reflects selective re-analysis rather than general rereading (Murray & Kennedy, 1988). The findings support the idea that children with stronger language skills are more likely to engage in comprehension monitoring during reading, in line with previous online (Connor et al., 2015; Zabrucky & Ratner, 1989) and offline evidence (Oakhill et al., 2005).

The next question is whether children's rereading (as captured in the eye movement record) indicates a strategic choice (Connor et al., 2015). This was tested through measures of self-report. Awareness is an often cited criterion for an action to qualify as a strategy (Hyönä & Nurminen, 2006; Macaro, 2006). In this experiment, however, children showed no awareness of their rereading or of the presence of internal inconsistencies. In other words, children's self-report gave no indication that their rereading on implausible words was deliberate. One possibility is that the absence of an effect is again due to methodological limitations, as suggested for the self-report measure used in Experiment 2. The questions on selective rereading may, for example, have been harder to answer for children than they were for the adults in the original study, despite attempts to make them child-friendly (Hyönä & Nurminen, 2006). Children may still be used to varying reading contexts that can influence reading

behaviour, such as shared compared to silent reading (Evans & Saint-Aubin, 2005). It is possible that a more explicit anchoring of the questions in the present reading task would have made it easier for the children to accurately describe their reading, for example by changing "I reread words when I don't understand them" to "When reading the stories on the screen, I reread the words that I didn't understand". The idea that anchoring might have helped is supported by the questions raised by some children during testing, such as when they asked if they were meant to report on how they read at home or at school. Thus, the absence of an effect may be due to methodological limitations.

Another possibility is that the null result represents a true absence of an effect, and that selective rereading may appear like a strategic choice, but actually largely results from efficient reading processing. This is further supported by how exactly the rereading differed between children with smaller and larger vocabularies: targeted regressions co-occurred with fluent reading on plausible control words, indicating that children became more selective in their reading at the same time as they became fluent readers. That is, children's online comprehension monitoring on implausible words- as evidenced in selective rereading – may be the result of efficient word processing (Perfetti & Hart, 2002). This would be different from earlier findings that adult readers were able to accurately describe their global rereading, which indicated that they used it as a strategy (Hyönä & Nurminen, 2006). Perhaps the local comprehension monitoring tapped in the current study is more automatic than global monitoring of comprehension, such as when rereading headings to remember content. Such a distinction is supported by the finding that readers adapt their rereading to comprehension demands on the global, but not on the local level of rereading mismatches (Weiss et al., 2017). This may

mean that rereading on mismatches is less strategic and deliberate than other, more global types of comprehension monitoring. However, this suggestion remains speculative until local and global comprehension monitoring can be compared directly in children, and linked to more robust measures of self-report.

Based on their finding that all children showed longer gaze durations, Connor et al. (2015) had furthermore suggested that gaze durations reflect the automatic initial noticing stage of comprehension monitoring, which would then be followed by rereading. In the current study, although longer gaze durations on implausible words were observed, the effect was driven by EAL children with smaller vocabularies. Not only did this effect not manifest in all children, it was a different group within the sample who showed the reactions to implausibility in gaze duration than those who did in rereading, as discussed above. This makes it unlikely that two eye movement measures represent two successive stages of comprehension monitoring. The distinction between gaze duration and rereading is also supported by a principal component analysis that confirmed that longer gaze duration on inconsistent target words loads on a different factor than longer rereading on the same targets (Kim et al., 2018). Longer gaze durations may thus reflect differences in lexical access, a possibility that will be considered further in relation to EAL children's comprehension monitoring. The reason that, different from the Connor et al. study, the current study was able to pick up on individual differences in gaze duration may be a combination of: a younger sample that also included EAL readers, differences in used predictors, and different ways of modelling individual differences. The results of the present study may reflect a wider range of abilities so that the differences in gaze duration could come out more strongly.

Not interpreting gaze duration as the detection stage of comprehension monitoring means, however, that there is only one eye movement record (increased rereading) that subsumes two processes (detecting as well as repairing a mismatch). This may seem unsatisfactory at first glance, but is plausible given general limitations of eye movement data. Unlike for example ERP data, eye movements do not allow for the disentanglement of two processes if they overlap in time (Perfetti et al., 2008). Combined recording of eye movements and EEG showed that regressions on syntactic and semantic mismatches were accompanied by two distinct ERPs – the P600 and the N400 – which seemed to indicate detecting and re-analysing, respectively (Metzner et al., 2016). In eye movements, however, both processes showed in regressions only. In the current study, children's rereading of implausible words may equally indicate both detection and repair. Both stages of comprehension monitoring may thus be reflected in children's rereading.

4.5.2 EAL children's comprehension monitoring

A central focus of the experiment was to uncover whether and how EAL learners and monolinguals differ in their comprehension monitoring (Burgoyne et al., 2013; Kan, 2014). The only difference between EAL learners and monolinguals was that EAL learners with smaller vocabularies initially spent longer reading implausible words. Longer gaze durations are commonly considered signs of slower lexical access, for example on longer or less frequent words (Rayner, 1998). The longer gaze durations on implausible words may indicate that their meaning was harder to retrieve for EAL children with small vocabularies, possibly because they are less predictable from context than plausible controls. Low predictability can have a stronger effect in less skilled comprehenders, arguably because they rely more strongly on context to predict

upcoming words and support lexical access than skilled comprehenders (Ashby, Rayner, & Clifton, 2005). This may be similar in EAL children with smaller vocabularies. These EAL children may have deficits in vocabulary depth or lexical quality that go beyond what is reflected in vocabulary size relative to monolinguals (see discussion in Chapter 3). Such gaps can slow down their lexical access (Tannenbaum, Torgesen, & Wagner, 2006), making predicting of upcoming words a useful tool to maintain fluent reading. This is not the same as saying that EAL readers would infer new word meanings from context, which is generally harder for readers with small comprehension skills (Cain et al., 2004; Joseph & Nation, 2018). Rather, EAL children, like less skilled adult comprehenders, may predict upcoming words simply to speed up their lexical access (Ashby et al., 2005). This idea, while remaining speculative for now, would be in line with the finding that minority language learners can efficiently make use of context during reading despite vocabulary gaps, for example to even infer new word meanings (Shahar-Yames & Prior, 2018). Such use of context and knowledge sources can be a useful way for minority language learners to make up for vocabulary gaps, and may be considered as comprehension monitoring on the level of lexical access (Kinnunen & Vauras, 1995). Predicting upcoming words would fail, however, when a word does not fit the context, as with implausible words in the current experiment. It may thus be the failure of an otherwise useful processing style that drives some EAL children's longer gaze durations on implausible words. To probe this idea directly, future research could adopt the procedures used in Ashby et al. (2005) to compare EAL and monolingual children on their reading of words that vary in predictability and lexical difficulty.

Different from EAL learners with smaller vocabularies, EAL children with large vocabularies read implausible words fast initially and reread them selectively, just like monolinguals. This is surprising given that in Experiments 1 and 2, EAL children processed metaphors differently from monolinguals with comparable levels of vocabulary. Consider first the absence of an effect in gaze duration. Possibly, slower speed in lexical access mattered only for children with small vocabularies since overall, all target words were easy to process for all children with large vocabularies: both implausible and plausible words were on average learnt early (age of acquisition was 5-6), fairly short (5-6 letters), and moderately frequent (43-44 hits per million in the BNC). All these factors are known to ease lexical access (Rayner, 1998). For children with large vocabularies, the words may thus have been so easy to access that none of them needed to rely on context.

EAL and monolingual children were furthermore similar in their comprehension monitoring, as reflected in their comparable rereading behaviour. This is different from Experiments 1 and 2, where even EAL children with large vocabularies seemed behind monolinguals in metaphor processing. This suggests that given sufficient vocabulary knowledge, EAL readers do not differ from monolinguals in comprehension monitoring, but possibly only when processing metaphors and similar phrases. Crucially, this seems to be the case despite the similarity of the two tasks, since both rely on noticing mismatches and re-analysis of semantic information. In other words, EAL readers do not seem to struggle with these processes per se, but rather only when processing metaphors, or other multi-word and figurative phrases (Burgoyne et al., 2013; Kan, 2014; Martinez & Murphy, 2011; McKendry, 2014; Smith & Murphy, 2015). Arguably, this indicates that the difficulty lies not in the processes, but in the

kind of knowledge that needs to be processed, in this case vocabulary depth. The experiment thus lends further support to the idea that when EAL learners struggle with their reading processing, this is associated with vocabulary gaps.

At the same time, there seemed to be no particular disadvantage for EAL children in the current experiment: with sufficient vocabulary knowledge, their comprehension monitoring on internal inconsistencies did not differ from that of monolinguals. And EAL children with smaller vocabularies may have been particularly skilful in counterbalancing slow lexical access by predicting upcoming words from context. This may, at a smaller scale, be akin to of minority language learners' stronger propensity to infer new word meanings (Shahar-Yames & Prior, 2018). Such use of inferencing may also explain why overall, EAL children have tended to have higher reading comprehension than what would be expected on the basis of their vocabulary size (Melby-Lervåg & Lervåg, 2014; Spencer & Wagner, 2017). Taken together, EAL children thus seem to struggle with metaphor processing, but not comprehension monitoring per se.

4.6 Conclusion

Experiment 3 has partially replicated but also extended previous findings on children's comprehension monitoring. Three key questions were posed. The first was whether Experiment 3 would replicate the finding that children adapt their reading to implausible words, indicating online comprehension monitoring (Connor et al., 2015). As expected, children took longer to read implausible words than plausible controls, both in initial reading and in rereading. A closer analysis of rereading patterns also showed that children were more likely to regress directly into implausible words. This

finding supports the idea that increased rereading on implausible words indicates selective re-analysis.

The next question asked about individual and group differences between EAL and monolingual children. As expected, children with larger vocabularies reread implausible words more frequently and for longer, indicating a link between language skills and online comprehension monitoring. This was mirrored numerically in reading times associated with higher reading fluency. Different from previous findings, however, the effect was not only driven by longer rereading on implausible words, but by both fluent reading on the plausible control words and targeted regressions into implausible words. Furthermore, there were also unexpected individual differences in initial reading: longer gaze durations on implausible words were driven by EAL children with smaller vocabularies. This may be due to EAL learners counterbalancing slow lexical access by predicting upcoming words, a strategy that would fail on unpredictable – since implausible - words. Importantly, the fact that different sub-groups of children carried the effects in gaze duration and rereading is at odds with the suggestion that the two reflect two successive stages of comprehension monitoring (Connor et al., 2015). Longer gaze durations on implausible words may not indicate mismatch detection, but difficulties with lexical access when context is missing.

This links to the third question: whether children were aware of implausible words and their rereading on them, which would indicate strategic comprehension monitoring (Hyönä & Nurminen, 2006). However, there was no link between children's self-reported and actual selective rereading, nor did their judgement of their own level of comprehension indicate any awareness of implausible words. On the one hand, this may be a reflection of methodological limitations, since for example the self-report

questions would have needed to be anchored in the present reading task to make them more appropriate for children. On the other hand, the absence of a relationship may reflect that children were truly not aware of their rereading, which would make it unlikely that children's rereading on local mismatches was the result of a strategic decision. The co-occurrence of targeted regressions with fluent reading may indicate that comprehension monitoring is instead the result of fluent reading. This suggestion is in line with the lexical quality hypothesis which sees efficient word processing at the heart of successful post-lexical processing (Perfetti & Hart, 2002; Perfetti et al., 2008). Crucially, this may be very different for more global comprehension monitoring such as targeting attention to headings (Hyönä & Nurminen, 2006) or rereading texts before answering difficult comprehension questions (Weiss et al., 2017).

Further research is needed to see if the suggested distinction between a fairly automatic local monitoring and a more explicit global comprehension monitoring would hold when both manipulations are tested alongside each other, and combined with more robust self-report measures. Furthermore, it seems that eye movements alone may not allow for the distinction between the two stages of comprehension monitoring, detection and re-analysis. Future studies wishing to study the two processes separately may thus need to co-register eye movements with ERP data.

The current discussion also leads to several implications for practice: teachers may want to pay particular attention to whether the texts they use in the classroom pose strong demands on children's monitoring skills. For example, teachers could be instructed about the potential challenges involved in establishing text coherence for children with small vocabularies so that they can offer targeted support. Furthermore, the link between word knowledge and selective rereading could be explained to

teachers, and they could be encouraged to include vocabulary instruction when aiming to support strategic reading.

Chapter 5: **General Discussion**

5.1 Overview of Main Findings and Conclusions

The three experiments reported in this thesis aimed to elucidate EAL children's reading comprehension, in particular their metaphor processing and comprehension monitoring. This final chapter summarises the major findings and sets them in the broader theoretical and educational context. Methodological issues and limitations are discussed, and throughout, issues for future research are identified.

Broadly, the present research assessed EAL and monolingual children's reading processing in two areas that were identified as meriting further research, that is, in metaphor processing and comprehension monitoring (Burgoyne, Whiteley, & Hutchinson, 2013; Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014). Existing research on these areas has relied on offline measures, such as comprehension questions on collocations (Kan, 2014), metaphors (Hessel & Murphy, 2018) or idioms (McKendry, 2014). The experiments in this thesis are the first studies to directly test EAL children's metaphor processing and comprehension monitoring in online reading. In Experiments 1 and 2, metaphor processing was compared across metaphor types and between children with varying word knowledge. In Experiment 3, comprehension monitoring was associated with word knowledge and self-report measures. Across all three experiments, language group differences were identified. The following will present the main findings and resulting conclusions from the three experiments, presented under the heading of the three questions posed in Chapter 1.

5.1.1 How do children process metaphors during reading, and how does that differ by metaphor type and for children with varying word knowledge?

The present research established that metaphors were challenging to understand for children not only when listening (Hessel & Murphy, 2018), but also when reading. In both Experiment 1 and 2, this was evident in lower accuracy on comprehension questions following stories that contained metaphors compared to literal controls. Eye movements furthermore revealed that children's difficulty with metaphors arose during post-lexical processing, as manifested by more frequent and longer rereading of metaphors than literal controls. This finding is consistent with online evidence from ERP and eye movement studies with adults as they read metaphors (Coulson & Van Petten, 2002; Olkoniemi et al., 2015) and indicates that in metaphor processing, the same processes are difficult for children as for adults. The experiments link metaphor processing to questions about children's reading more generally, and add to a growing body of research on the role of vocabulary depth in EAL reading comprehension (Cremer, 2013; Cremer et al., 2011; Cremer & Schoonen, 2013; Kan, 2014; McKendry, 2014; Schoonen & Verhallen, 2008; Smith & Murphy, 2015; Spätgens & Schoonen, 2018).

Experiment 1 further compared processing across metaphor types. Metaphor type did not change the effect of metaphorness in children's offline reading comprehension. Metaphor type mattered, however, for children's reading processing: interaction effects showed that children's increased processing of metaphors was largely driven by nominal metaphors such as *grades are lotteries*, as manifested in longer and more frequent rereading of nominal metaphors than literal controls. These findings

indicate that, as predicted, only nominal metaphors placed increased demands on children's post-lexical processing, arguably because these metaphors require integrating meanings and suppressing meaning facets that would lead to overly literal interpretations (Coulson & Van Petten, 2002; Tartter et al., 2002). Verbal metaphors such as *hours fly by*, in contrast, were not reread more often or longer than literal controls, and there were even trends that children read faster following verbal metaphors. This indicates that, as expected, verbal metaphors were as easy to process as literal controls, arguably since their systematicity in language increased their predictability, and reduced processing demands (Boroditsky & Ramscar, 2002; Casasanto, 2008; Casasanto & Jasmin, 2012; Fuente et al., 2014; Lakoff & Johnson, 2003). The fact that metaphor types affected eye movements but not answers on comprehension questions may indicate that while nominal metaphors required more effort to be processed, children eventually reached a sufficiently good interpretation of them. Another candidate explanation is that the absence of a metaphor type effect in the comprehension question may be linked to potential limitations of the way comprehension was measured (see Section 5.3.3 for a more detailed discussion). To review, Experiment 1 showed how different types of metaphors affect children's reading. Further research could probe which exact dimensions drive the metaphorness effect of nominal, but not verbal metaphors.

Experiment 2 investigated children's processing of nominal metaphors only, with a focus on how differences in processing were associated with word knowledge, as measured by reading fluency and vocabulary size. Within both experiments, stronger word knowledge was associated with reading patterns that indicated skilful metaphor processing. In Experiment 1, larger vocabulary and higher reading fluency were

associated with children adapting their reading to metaphors by rereading more frequently on metaphors than literal controls, similar to more skilled readers who reread increasingly on other challenging elements in texts (Connor et al., 2015; Eilers et al., 2018a; Oakhill et al., 2016) . In Experiment 2, higher reading fluency was associated with increasingly shorter and more targeted rereadings of metaphors, similar to other studies where skilled readers became more efficient in their processing of difficult structures and thus decreased their rereading (Hopp, 2016; Joseph & Liversedge, 2013; Luke et al., 2015; Tiffin-Richards & Schroeder, 2018). To summarise, while stronger word knowledge was generally correlated with reading patterns that seem to indicate skilful processing, these rereading patterns differed across experiments. These differences are in line with the idea of an upside-down U-shaped developmental trajectory: from not reacting to challenges, to effortfully processing them, to selective and fluent rereading (Tiffin-Richards & Schroeder, 2018). Suggestions on how to test the idea of an inverted U-shape development in metaphor processing will be considered in Section 5.2.2. Furthermore, the variability in skilful reading across experiments indicates challenges of mapping eye movements to linguistic processes that will be discussed in Section 5.2.1.

5.1.2 How do children monitor their comprehension during reading, and how is monitoring associated with word knowledge?

Experiment 3 showed that 9–10-year-old children were sensitive to internal inconsistencies within stories caused by the inclusion of implausible words. This was manifested in longer gaze durations during the initial reading, as well as longer and more frequent rereading on implausible compared to plausible words. These findings

replicate those from an earlier study where slightly older children were tested on the same stimulus set (Connor et al., 2015). Unlike previous findings, however, effects in initial reading and rereading were carried by different sub-groups of children. This finding is at odds with the idea that there are two successive stages of comprehension monitoring, with gaze duration reflecting automatic detection followed by re-analysis of inconsistencies during rereading (Connor et al., 2015). Rather, gaze duration may indicate differences in lexical access, while rereading alone may indicate comprehension monitoring (Kim et al., 2018) and capture both detection and re-analysis of mismatched information (Metzner et al., 2016).

Experiment 3 furthermore showed that stronger word knowledge was associated with fluent reading on easy (i.e. plausible) words and selective rereading on difficult (i.e. implausible) words. This co-occurrence makes it plausible that word knowledge drives first efficient word processing and through that, successful post-lexical processing (Perfetti & Hart, 2002; Perfetti et al., 2008). This idea is also supported by the absence of a link between children's self-reported and their actual selective rereading. While this may be a reflection of methodological limitations of the self-report measures, it may also indicate that local comprehension monitoring, such as rereading on mismatches, is more implicit than more global comprehension monitoring, such as rereading headings or trying to memorise texts (Hyönä & Nurminen, 2006; Weiss et al., 2017).

5.1.3 Are there differences between EAL learners and monolinguals in their online processing of metaphors and comprehension monitoring when reading, and if so, what are they?

EAL and monolingual children differed in their online reading across all three experiments, even when this was not mirrored in language group differences in offline comprehension or standardised tests. In Experiment 1, EAL and monolingual children differed on standardised tests of comprehension-related skills (reading comprehension, vocabulary size and depth) and in their overall comprehension in the metaphor experiment, but not particularly on metaphors. In Experiments 2 and 3, EAL children did not differ on standardised tests nor in their overall comprehension in the metaphor experiment. In all of these instances, however, language groups differed in their online processing when word knowledge was considered.

Consider first language group differences in metaphor processing. In Experiment 1, only EAL learners with a relatively large vocabulary size reread more on metaphors than literal controls; in contrast, monolinguals with relatively smaller vocabularies showed increased rereading. This indicates that monolinguals were sensitive to metaphors in their processing, even when they had smaller vocabularies. These effects were mirrored in reading patterns associated with reading fluency as measured by the TOWRE in Experiments 1 and 2. In Experiment 2, for example, EAL learners but not monolinguals with low reading fluency took particularly long to read metaphors, indicating that EAL learners were less efficient in their metaphor processing than monolinguals. Fluent monolingual readers, in turn, processed literal language fluently while frequently rereading metaphors, resulting in efficient and selective rereading of metaphors. It seems that reading metaphors taps a difference between EAL

learners' and monolinguals' language skills that are not fully assessed on standardised tests such as the BPVS or the TOWRE, or by answers to the metaphor questions in my experiments.

These stronger effects in metaphor processing may be due to the degree to which understanding metaphors requires efficient meaning processing: even for highly proficient adult EAL learners, efficiency in meaning retrieval when reading metaphors such as *He is a late bloomer* seems limited by smaller vocabulary size (Segal & Gollan, 2018). Similarly, metaphor processing in Experiments 1 and 2 may have tapped a dimension of language processing – establishing extensive meaning relationships – that becomes more efficient into adult years for EAL learners. Perhaps this is a result of EAL learners' larger gaps relative to monolinguals in vocabulary depth than vocabulary size, as reported for studies investigating meaning relations (Cremer et al., 2011; Cremer & Schoonen, 2013; Schoonen & Verhallen, 2008; Spätgens & Schoonen, 2018) and understanding of multi-word and figurative phrases (Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014; Smith & Murphy, 2015). If EAL learners had fewer meaning associations for each word in a metaphor, this could lead to a similarly lower efficiency as reported in Segal et al. (2018), even relative to monolinguals with comparable vocabulary size.

Differences between EAL learners' and monolinguals' reading processing were also evident in comprehension monitoring in Experiment 3. Different from metaphor processing, however, only EAL children with small vocabularies differed in their comprehension monitoring from monolinguals, since they were the sole sub-group who took longer to read inconsistent words initially. The longer gaze durations of EAL learners with smaller vocabulary sizes may stem from their stronger reliance on context

to predict upcoming words compared to other readers, which they may use to support lexical access (Ashby et al., 2005). A strong reliance on context would however fail on unpredictable (since in the context implausible) words, and could thus lead to the observed longer reading times. With larger vocabularies, however, EAL learners were as likely as their monolingual peers to notice and re-analyse internal inconsistencies. This indicates that there are similar skills in comprehension monitoring across the two language groups, given sufficient word knowledge.

To summarise, EAL learners may struggle with retrieving and establishing meaning relationships, as manifested in less efficient metaphor processing than monolinguals, but not necessarily with comprehension monitoring. EAL learners may even rely more on context than other readers to compensate otherwise slower lexical access, in line with compensation accounts in other studies (Raudszus et al., 2017; Shahar-Yames & Prior, 2018). Furthermore, the findings indicate that language group differences are most pronounced for children with weaker word knowledge and that group differences can vanish with sufficient word knowledge. Overall, these findings support the view that EAL learners' reading processing may differ from that of monolinguals in complex ways – both in terms of compensation behaviour (Raudszus et al., 2017) and in terms of how gaps in word knowledge may show in delayed post-lexical processing on metaphors (Perfetti et al., 2008). The findings indicate the usefulness of studying online processing to understand minority language learners' reading comprehension.

5.2 Theoretical Implications

5.2.1 Regressive eye movements and children's post-lexical processing

Throughout this thesis, eye movements were used as indicators of children's reading processing, both on metaphors and in comprehension monitoring. Eye-tracking has become a valued method to study reading, and there is a wealth of evidence that links different eye-movement measures to different components of linguistic processing (Rayner, 1998). Using different measures, this thesis was able to show where children's effort arises during post-lexical processing (Experiment 1 and 2) and where different groups of children struggle at different time points of processing (Experiment 3). However, three issues around interpreting eye movements in children's reading have emerged in the research underlying this thesis.

First, rereading is widely interpreted as an indicator of post-lexical processing, for example when readers re-analyse ambiguous or mismatched information (Connor et al., 2015; Joseph et al., 2008; Rayner et al., 2004; von der Malsburg & Vasishth, 2011; Weiss et al., 2017; Zabrucky & Ratner, 1989). In the current study, increased rereading was thus interpreted as indicating meaning integration and literal suppression in metaphors (Experiment 1 and 2), and detection and re-analysis of mismatches in comprehension monitoring (Experiment 3). However, regressions may also indicate confusion (Murray & Kennedy, 1988) or a simple means to pause reading (Mitchell et al., 2008). This ambiguity of regressions is particularly prominent in children's reading since regressions are frequent in young children's reading more generally (Blythe & Joseph, 2011). It has been suggested that different kinds of rereading can be distinguished by considering both their spatial and temporal pattern (Murray & Kennedy, 1988). For example, scanpath analyses that combine both spatial and

temporal information afford distinctions between different types of rereading on ambiguous sentences, and were shown to differentiate between adults with different working memory capacities (von der Malsburg & Vasishth, 2012). While scanpath analyses were beyond the scope of the current research, future research on children's reading may employ them to tease apart different types of rereading. For example, scanpaths could be used to distinguish between general pausing or rereading of the last word, as could be predicted to show in short regressions of 1–2 words length, and selective re-analysis, as manifested in longer regressions targeted at text difficulties such as metaphors or implausible words (Murray & Kennedy, 1988).

A second problem for interpretation is the ambiguity in the ways in which children's regressions were linked to the metaphorness effect in Experiment 1 and 2: while in Experiment 1, more skilled post-lexical processing was linked to *increasing probability* to reread, in Experiment 2 it was manifest as *decreasing time* spent rereading. That is, in younger children's skilful metaphor processing meant *more* rereading, as evident in the higher rereading probability, while in the older children, skilful metaphor processing meant *less* rereading, as indexed by shorter rereading times. This seeming contradiction is not uncommon in the literature around eye movements and post-lexical processing: in some studies, increased rereading is interpreted as a sign of skilful processing (Connor et al., 2015; Eilers et al., 2018a; Oakhill et al., 2016), and in others, faster rereading is seen as an indicator of skilful processing (Hopp, 2016; Joseph & Liversedge, 2013; Luke et al., 2015; Tiffin-Richards & Schroeder, 2018). A developmental account has been suggested to link the two: the type of post-lexical processing may depend on a reader's skills, and progress in an inverted U-shape from no reaction, to increased rereading, to finally fast rereading (Tiffin-Richards &

Schroeder, 2018). The trends associated with word knowledge in Experiments 1 and 2 are in line with this suggestion. The experiments in the current thesis were however not designed to test development directly, and comparisons between Experiments 1 and 2 are not cross-sectional, nor within items or within participants. Future research could test the suggested inverted U-development of different rereading patterns in metaphor processing in both cross-sectional and longitudinal designs. At the same time, it is important to note here that any age-related development is plausibly not driven by age itself, but by developing language skills (Reichle et al., 2013). Another design to study developmental trajectories in post-lexical processing may thus include training studies that strengthen language skills and observe if reading processing subsequently moves along the suggested U-shaped developmental trajectory, as outlined further in the following section.

A final issue that emerged in the current research relates to two limitations concerning what regressive eye movements can reveal about reading processing. Firstly, analyses that rely on rereading are typically based on a smaller number of trials since not all trials are reread. For the three experiments in this thesis, the proportion of reread trials was sufficiently high for analyses (between one third and half of all trials per child and interest area, which is similar or higher to other studies focusing on rereading, e.g. Meseguer, Carreiras, & Clifton, 2002; Murray & Kennedy, 1988; Schotter, Tran, & Rayner, 2014; von der Malsburg & Vasishth, 2011). However, not all children reread on every interest area, indicating that eye-tracking measures relying on rereading are not as valid across readers as first-pass measures, such as gaze duration. This indicates a limitation for studies focusing on effects that are evident in rereading, such as studies of metaphor processing (Olkoniemi et al., 2015), comprehension monitoring (Connor et

al., 2015; Kim et al., 2018), or anaphoric relationships (Eilers et al., 2018a; Joseph et al., 2015).

The second limitation relates to the degree to which different linguistic processes can or cannot be distinguished based on eye movement records. Connor et al. (2015) had proposed that a two-stage monitoring process was reflected in different eye movement measures, with gaze duration indicating noticing a mismatch, and rereading reflecting re-analysis to update the initial interpretation. Experiment 3 showed, however, that effects in gaze duration and rereading were carried by different sub-groups of children, making it unlikely that the two measures indicated subsequent stages. Rather, the two stages of comprehension monitoring – noticing a mismatch, and updating one's interpretation – both seemed associated with rereading. This idea is supported by a study that collected EEG and eye movement data simultaneously: Metzner et al. (2016) found that rereading in the eye movement record was accompanied by two distinct brain potentials in the EEG record, indicating that rereading captured both noticing and re-analysis. Future research may wish to co-register EEG and eye movements to distinguish specific processes involved in comprehension monitoring, and to further understand how children differ in them.

5.2.2 Word knowledge in children's reading processing

Each of the three experiments in this thesis has provided data on how children's word knowledge is associated with reading metaphors (Experiments 1 and 2) and comprehension monitoring (Experiment 3). Word knowledge was measured through standardised tests of word reading fluency (i.e. the TOWRE-II) and vocabulary size (the TOWK expressive vocabulary sub-test in Experiment 1 and the BPVS in Experiment 2

and 3). Both reading fluency and vocabulary size were similarly associated with metaphor processing, and numerical trends suggested the same for comprehension monitoring. This is somewhat surprising given that in children's standardised measures of reading comprehension, vocabulary becomes more important beyond the first years of learning to read, while the relative influence of reading fluency declines with development (Protopapas, Sideridis, Mouzaki, & Simos, 2007). The findings in the current thesis may indicate how both vocabulary size and reading fluency interact in determining reading efficiency, thus jointly facilitating or hindering post-lexical processing (Perfetti & Stafura, 2014; Perfetti et al., 2008).

At the same time, EAL learners were less efficient in their metaphor processing relative to monolinguals who performed similarly on standardised tests. This may indicate that EAL learners are less efficient than monolinguals in retrieving and establishing meaning relations when striving to understand metaphors, relative to vocabulary size (Segal & Gollan, 2018). This may be linked to EAL learners' particular difficulties with tasks that require depth of vocabulary, as reported on tasks that tap semantic associations (Cremer et al., 2011; Cremer & Schoonen, 2013; Schoonen & Verhallen, 2008; Spätgens & Schoonen, 2018) and understanding of multi-word and figurative phrases (Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014; Smith & Murphy, 2015). That is to say, EAL learners' less efficient metaphor processing may be indicative of differences in vocabulary depth. However, establishing specific difficulties in vocabulary depth is not uncontroversial: while vocabulary size and depth can be clearly distinguished in conceptualisations and measurements (P. Nation, 2001), the two can be highly correlated in practice, both in monolingual and bilingual children (Vermeer, 2001). Establishing language group differences that are specific to

vocabulary depth is further complicated by questions around task design, such as what constitutes an appropriate measure of depth (Vermeer, 2001), and what methodological limitations may arise when using standardised tests with bilinguals (see Section 5.3.4). To further establish the role of different dimensions of word knowledge in EAL learners' reading comprehension, future research could combine several tests of vocabulary size and depth, distil the variance of all size and depth measure into orthogonal latent scores, and compare their predictive power of reading processing between EAL learners and monolingual. This way, effects of individual tests could be minimised and collinearity reduced, affording more reliable analyses than when predictors are based on individual tests (James, Fraundorf, Lee, & Watson, 2018).

Some theoretical issues would however remain unresolved in any correlational study. For example, effects of word knowledge may be confounded with individual differences in comprehension standards: if children with lower lexical quality become used to accepting low levels of comprehension, they may in turn set lower standards for comprehension during reading, which could result in differences in reading behaviour beyond direct effects of vocabulary on reading efficiency (Bicknell & Levy, 2011). A way to test the relationship between lexical quality and reading processing directly would be a training study: lexical quality could be manipulated through a vocabulary instruction programme that has been shown to improve lexical access and reading comprehension beyond taught words (cf. for example Beck, Perfetti, & McKeown, 1982). The effect of such an intervention on reading processing could then be tested on online reading tasks, for example in metaphor processing or comprehension monitoring to see if gains in lexical quality leads to more skilful post-lexical processing and selective rereading. To tease apart interactions between word knowledge and strategic

choices during reading, the intervention could be crossed with strategy instruction, such as encouraging active engagement with texts in shared reading exercises (McNamara, 2017). This approach could test previous suggestions that strategies may only become accessible to readers when a certain efficiency in reading processing is reached (Kaakinen & Hyönä, 2007).

5.3 Methodological Considerations

5.3.1 Sampling effects on language group differences

EAL learners vary greatly in their language skills, and in how far they differ from monolinguals (Paradis & Jia, 2017). The current research aimed to represent this heterogeneity in UK classrooms by recruiting any EAL and monolingual children that fulfilled baseline criteria necessary to participate in the experiment (such as age-appropriate decoding and at least two years of exposure to English). Alternatively, I could have only recruited EAL learners with specific home languages (Droop & Verhoeven, 2003) or low English proficiency (Whiteside et al., 2017). While this may have allowed a focus on sub-groups that can be expected to struggle (Demie, 2018; Strand et al., 2015), it would have limited what can be learned about the reality of EAL learners in the classroom, where it is clear that there is considerable variation in language backgrounds, achievement levels, and English proficiency (Strand & Hessel, 2018; Strand et al., 2015). Any sampling decision has implications for the pattern of results seen. For example, if only EAL children with weak comprehension skills had been recruited, overall language group differences may have come out more strongly. Furthermore, children were recruited through convenience sampling, which lead to slightly diverging results between the two samples in the thesis, and may have

contributed to the U-shaped trajectory in metaphor processing: for example, the many EAL children with particularly small vocabulary sizes in Experiment 1 may have exacerbated the contrast between EAL readers with small vocabulary who did not adapt their rereading to metaphors at all, compared to monolinguals who did. The differences between the two samples indicate that while tested at the whole class level, each sample remains a snapshot of a part of the EAL population. Further research would be needed to see if the results of this thesis would replicate in other, possibly larger samples to test their generalisability to the wider EAL population.

Within such larger samples, differences in EAL learners' environment could be used to predict variance in EAL learners' English proficiency, which may facilitate the understanding of the origins of the large heterogeneity between EAL learners (Blom et al., 2012; Paradis, 2011; Paradis & Kirova, 2014). My experiments always included language background questionnaires to be able to understand EAL learners' language backgrounds descriptively, but their use for inferential analyses was limited by large numbers of missing values (Experiment 1) and generally shorter length (Experiment 2 and 3). Further research with a focus on the role of language backgrounds could employ extensive and normed questionnaires, such as the Bilingualism Profile Index which distils information from an extensive questionnaire about children's language exposure into a latent variable that is predictive of both English proficiency and performance on inhibition tasks (de Cat, Gusnanto, & Serratrice, 2016; de Cat & Serratrice, *under review*).

5.3.2 Stimulus development for studies of post-lexical reading processing

Several choices were made during stimulus development that influenced which insights could be gained. For example, metaphor processing was measured by comparing children's processing of metaphors to that of matched literal controls in the same story context. Alternatively, metaphors could have been tested by manipulating the context so that the same phrase takes either a literal or a figurative meaning (Olkoniemi et al., 2015). Such a design would make streamlining stimuli easier, since reading times could be compared on the same phrase. However, this design would also make it impossible to compare children's comprehension of the same texts with and without metaphors, and does thus not afford educational implications of how metaphors impact children's comprehension. In the present research, it was thus decided to present metaphors and literal controls in the same story context. This led to somewhat reduced experimental control over lexical variables: items were particularly constrained by the requirements of finding pairs of metaphors (e.g. *hours fly by*) and literal controls (e.g. *birds fly by*) that were matched on length and frequency, while expressing two different but equally plausible events given the same story context. The resulting stimulus set was relatively heterogeneous regarding word class, but care was taken to match the literal and metaphor condition on length, frequency and naturalness. Furthermore, all statistical analyses were run with maximally possible random structure to further account for item effects. This ensured that the observed reading time differences between metaphors and literal controls were reliable despite item heterogeneity, as confirmed through model checks.

Another choice was made regarding how stimuli would be normed. While the implausible and plausible conditions in Experiment 3 were largely kept the same as

previous research (Connor et al., 2015), metaphors in Experiments 1 and 2 were normed on several criteria: for example, metaphors and literal controls were compared on transparency, or the link between individual words and overall meaning, and metaphor types were compared on predictability, as measured in a cloze test. With a different research focus, it would have also been possible to norm metaphors on other criteria that are more or less associated with those tested for Experiments 1 and 2 (Cardillo et al., 2010; Davenport & Coulson, 2011). The fact that phrases can be normed and compared on so many criteria may raise the question of what ultimately drives effects of manipulations of metaphorness, metaphor types, or inconsistency. To answer this question, future research could compare manipulations of metaphorness and implausibility, or different metaphor types across degrees of underlying criteria such as predictability and semantic overlap to see what drives reading time differences across manipulations (Wlotko & Federmeier, 2012).

5.3.3 Comprehension and self-report measures in eye-tracking studies

In eye-tracking studies, eye movement records are often studied without additional questions which tap offline comprehension, or self-report measures to tap decision-making during reading. However, this may limit the information that can be gained about reading since it is not possible to probe if eye movement records reflect conscious decisions (Hyönä & Nurminen, 2006), or if differences in reading patterns are mirrored in differences in comprehension (Schotter et al., 2014). Both types of information are informative for both theoretical and educational questions regarding what is difficult to understand, and for what reasons. In the present experiments, eye movement recordings were thus combined with comprehension questions designed to assess whether the

metaphors were understood, and self-report measures tapping the child's own awareness of comprehension and rereading behaviour (Hyönä & Nurminen, 2006; Martinez & Murphy, 2011).

To not interrupt children's reading during the task, comprehension questions were limited to forced-choice questions that could be answered using a button press, and self-report questions were administered after the experiment had finished. While this made the questions highly compatible with a classical eye-tracking task, the lack of overlap between children's eye movements on the one hand, and their self-reported reading and offline comprehension on the other hand may indicate weaknesses in this approach. For example, the forced-choice questions may have lacked the necessary sensitivity to tap different levels of metaphor comprehension, as also suggested by offline studies using different task designs (Hessel & Murphy, 2018). Possibly, a task that requires explaining one's metaphor interpretation or writing a text summary would have shown more differences in comprehension that would have mirrored the differences in eye movements (Hessel & Murphy, 2018; Hyönä & Nurminen, 2006). An additional limitation may have been introduced by the constraint that the correct answer for the metaphor condition were used as distractors in the literal condition and vice versa; possibly, a less constrained choice of distractors would have made it possible to construct questions that would have been more appropriate to tap comprehension at the right level of difficulty. The self-report measure on awareness of comprehension was limited by being based on a single question, compared to previous research that measured self-perceived comprehension after each text (Martinez & Murphy, 2011). Future research could add more fine-grained measures of comprehension after each text, followed by a rating of self-perceived comprehension. It remains to be seen, however,

how the presence of such extensive questions during an experiment itself may influence children's eye movements (for effects of questions on eye movements, see for example Weiss et al., 2017).

Finally, children's self-report on rereading may have been limited by the fact that it was not explicitly linked to the testing situation, following procedures used with adults (Hyönä & Nurminen, 2006). For use with children, it may be beneficial to link the self-report more directly to the testing situation since children may still be used to varying reading contexts that can make for different reading behaviours, such as shared compared to silent reading (Evans & Saint-Aubin, 2005). A clear anchoring of the self-report questions to the testing situation may make it easier for children to give accurate descriptions of their reading record in the experiment.

5.3.4 Use of standardised tests with EAL learners

EAL and monolingual children's metaphor processing were differently associated with reading fluency and vocabulary size, as assessed on standardised tests. On the one hand, this may reflect weaknesses in vocabulary depth in EAL learners that goes beyond differences in vocabulary size (Cremer, 2013; Kan, 2014; McKendry, 2014; Schoonen & Verhallen, 2008; Smith & Murphy, 2015). However, this effect may partly be due to limitations of using standardised tests in bilingual samples. For example, the standardised tests used in this study were normed on a monolingual population (Dunn et al., 1997; Torgesen et al., 2012; Wiig & Secord, 1992). This can be an issue since item difficulty may vary between bilingual and monolingual learners (Chan, 2007). If standardised tests have only limited validity when used with EAL learners, this can lead to different associations of word knowledge with experimental measures between

language groups. Such differences can make it harder to identify areas of difficulty for EAL learners, since a different effect on an experimental measure may either mean that it measures a particularly important language skill, or simply that it is more reliable in testing EAL learners' language skills. Future research is needed to test the reliability of standardised tests for the use with EAL learners, for example by comparing the psychometric properties of commonly used standardised tests between EAL learners and monolinguals.

5.4 Educational Implications

The research presented in this thesis holds several implications for educational practice. EAL children with strong vocabulary knowledge seemed equally proficient as monolingual learners in comprehension monitoring, supporting the idea that with sufficient vocabulary knowledge, EAL learners can develop high proficiency in English while reaping the benefits of learning a second language. Across experiments, differences between EAL learners and monolinguals seemed widest amongst children with relatively weaker word knowledge. This indicates that EAL support would be best targeted at EAL learners with lower levels of language proficiency. Together with studies highlighting the importance of English proficiency for EAL children's educational achievement (Demie & Strand, 2006; Strand & Demie, 2005; Strand et al., 2015; Whiteside et al., 2017), this suggests that language instruction for EAL learners at the lower end of the ability spectrum can be a powerful tool to support their school career.

The results also hold implications for how interventions may be targeted. Results on the metaphor processing task indicated that EAL readers may struggle

specifically with processing metaphors, possibly driven by lower efficiency in retrieving and integrating meanings due to weaker word knowledge (Segal & Gollan, 2018). This finding may be a further indicator of previously reported larger differences between EAL learners and monolinguals in tests of meaning associations (Cremer et al., 2011; Cremer & Schoonen, 2013; Schoonen & Verhallen, 2008; Spätgens & Schoonen, 2018) and understanding of multi-word and figurative phrases (Hessel & Murphy, 2018; Kan, 2014; McKendry, 2014; Smith & Murphy, 2015). Vocabulary interventions could be targeted at supporting these areas of depth of vocabulary, for example through shared reading of stories designed to teach connected words in the context of phrases and semantically related words (Hadley, Dickinson, Hirsh-Pasek, & Golinkoff, 2018). Furthermore, different dimensions of word knowledge, e.g. reading fluency and vocabulary size, were related to efficient reading processing across children. This indicates that vocabulary instruction that trains different dimensions of knowledge could be used to support reading processing, for example by combining definition and classification tasks to train vocabulary depth, and timed production tasks to train fluency (Beck et al., 1982).

Finally, the results were in line with the idea that word knowledge may facilitate skilful post-lexical processing, and result in selective and targeted rereading. While this idea remains to be probed in direct training studies, teachers could for now be informed about the well-established links between word knowledge and skilful higher-level processing (Perfetti, 2007; Perfetti et al., 2008). Teachers could also be instructed about the potential challenges involved in establishing text coherence during reading, particularly for children with small vocabularies, and they could be encouraged to include vocabulary instruction when aiming to support strategic reading.

5.5 Conclusion

The experiments in this thesis investigated the effect of nominal metaphors on children's reading, and results indicated how EAL learners may struggle with metaphor processing relative to monolinguals with comparable word knowledge. In comprehension monitoring, however, only EAL children with small vocabularies differed from monolinguals, while all children with larger vocabularies monitored their comprehension proficiently. This indicates a prime role of word knowledge in children's skilful reading processing. A close examination of online reading of metaphors revealed different rereading patterns across experiments, indicating potential developmental trends in children's post-lexical processing that merit further research.

Two main implications are clear. The first is that online tasks of reading processing hold valuable information about EAL learners' reading comprehension, such as potential difficulties with processing metaphors, but also, given sufficient vocabulary knowledge, strong skills in comprehension monitoring. The second implication is the importance of word knowledge in understanding children's post-lexical processing during reading, which may be tested in further longitudinal or training studies.

Research projects such as the one presented in this thesis can provide new insights into the reading learning of children who are learning more than one language – insights that are increasingly valuable in this increasingly bilingual world.

Returning to the early literacy classroom at the outset of this thesis, we now better appreciate the significance of the different activities they are engaged in, and better understand which of the children – both monolingual and EAL – are likely to struggle in reading comprehension. We have a better idea of the importance of vocabulary and its impact on reading comprehension and the processing of metaphors.

With this knowledge in hand, we are in a better position to help both EAL and monolingual learners.

References

Aaronson, D., & Ferres, S. (1984). Reading strategies for children and adults: some empirical evidence. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 189–220. http://doi.org/10.1016/S0022-5371(84)90137-3

Ashby, J., Rayner, K., & Clifton, C. J. (2005). Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Human Experimental Psychology*, *58*(6), 1065–1086. http://doi.org/10.1080/02724980443000476

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

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. http://doi.org/10.1007/s11145-014-9536-x

Babayigit, S., & Stainthorp, R. (2014). Correlates of early reading comprehension skills: a componential analysis. *Educational Psychology*, *34*(2), 185–207. http://doi.org/10.1080/01443410.2013.785045

Baker, L. (1984). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology*, *76*(4), 588–597. http://doi.org/10.1037/0022-0663.76.4.588

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing : keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. *ArXiv Preprint ArXiv:1506.04967*, 1–27. http://doi.org/arXiv:1506.04967

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: cognitive abilities and the production of figurative language. *Memory & Cognition*, *41*(2), 255–267. http://doi.org/10.3758/s13421-012-0258-5

Beck, I., Perfetti, C. A., & McKeown, M. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension . *Journal of Educational Psychology*, *74*(4), 506–521.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2009). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. http://doi.org/10.1017/S1366728909990423

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: is there a bilingual advantage? *Applied Psycholinguistics*, *24*(1), 27–44. http://doi.org/10.1017/S014271640300002X

Bialystok, E., Shenfield, T., & Codd, J. (2000). Languages, scripts, and the environment: factors in developing concepts of print. *Developmental Psychology*, *36*(1), 66–76. http://doi.org/10.1037/0012-1649.36.1.66

Bicknell, K., & Levy, R. (2011). Why readers regress to previous words: a statistical analysis. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 931–936).

Billow, R. M. (1975). Cognitive developmental study of metaphor comprehension. *Developmental Psychology*, *11*(4), 415–423.
Blom, E., Paradis, J., & Duncan, T. S. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular -s in child L2 English. *Language Learning*, 1–30. http://doi.org/10.1111/j.1467-9922.2012.00715.x
Blythe, H. I., & Joseph, H. S. S. L. (2011). Children's eye movements during reading. *The Oxford Handbook of Eye Movements*, 643–662. http://doi.org/10.1093/oxfordhb/9780199539789.013.0036
Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*(2), 185–189. http://doi.org/10.1111/j.1467-9639.1991.tb00167.x
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. http://doi.org/10.1007/s11145-016-9699-8
Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1–11. http://doi.org/10.3389/fpsyg.2016.01116
Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2011). The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual, English-speaking peers. *The British Journal of Educational Psychology*, *81*, 344–54. http://doi.org/10.1348/000709910X504122
Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2013). The role of background knowledge in text comprehension for children learning English as an additional language. *Journal of Research in Reading*, *36*(2), 132–148. http://doi.org/10.1111/j.1467-9817.2011.01493.x
Cain, K., Lemmon, K., & Oakhill, J. (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), 671–681. http://doi.org/10.1037/0022-0663.96.4.671
Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: is vocabulary more important for some aspects of comprehension? *L'année Psychologique*, *114*, 647–662.
Callahan, R. M., & Gandara, P. C. (2014). Contextualising bilingualism in the labor market: new destinations, established enclaves and the information age. In R. M. Callahan & P. C. Gandara (Eds.), *The bilingual advantage: language, literacy and the US labor market* (pp. 3–15). Bristol: Bilingual Education & Bilingualism.
Cameron, L., & Besser, S. (2004). *Writing in English as an additional language at key stage 2*.
Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*(3), 651–64. http://doi.org/10.3758/BRM.42.3.651
Casasanto, D. (2008). Similarity and proximity: when does close in space mean close in mind? *Memory & Cognition*, *36*(6), 1047–1056. http://doi.org/10.3758/MC.36.6.1047

Casasanto, D., & Jasmin, K. (2012). The hands of time: temporal gestures in English speakers. *Cognitive Linguistics*, *23*(4), 643–674. http://doi.org/10.1515/cog-2012-0020

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

Chan, L. L. . (2007). An exploration of assessment of early English language development for kindergarten children in Hong Kong. In *Talk presented at the British Educational Research Association Annual Conference*.

Cicone, M., Gardner, H., & Winner, E. (1981). Understanding the psychology in psychological metaphors. *Journal of Child Language*, *8*(1), 213–216. http://doi.org/10.1017/S0305000900003123

Connor, C. M., Radach, R., Vorstius, C., Day, S. L., Mclean, L., Morrison, F. J., & Mcdonald, C. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: an eye-movement study. *Scientific Studies of Reading*, *19*(2), 114–134. http://doi.org/10.1080/10888438.2014.943905

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: an event-related potential study. *Memory & Cognition*, *30*(6), 958–968. http://doi.org/10.3758/BF03195780

Cremer, M. (2013). *Accessing word meaning: Semantic word knowledge and reading comprehension in Dutch monolingual and bilingual fifth-graders*. University of Amsterdam.

Cremer, M., Dingshoff, D., de Beer, M., & Schoonen, R. (2011). Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *International Journal of Bilingualism*, *15*(2), 187–204. http://doi.org/10.1177/1367006910381189

Cremer, M., & Schoonen, R. (2013). The role of accessibility of semantic word knowledge in monolingual and bilingual fifth-grade reading. *Applied Psycholinguistics*, *34*(6), 1195–1217. http://doi.org/10.1017/S0142716412000203

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945.

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. http://doi.org/10.1016/j.jecp.2015.03.005

D'Angiulli, A., Siegel, L. S., & Serra, E. (2001). The development of reading in English and Italian in bilingual children. *Applied Psycholinguistics*, *22*(4), 479–507. http://doi.org/10.1017/S0142716401004015

Davenport, T., & Coulson, S. (2011). Predictability and novelty in literal language comprehension: an ERP study. *Brain Research*, *1418*, 70–82. http://doi.org/10.1016/j.brainres.2011.07.039

de Cat, C., Gusnanto, A., & Serratrice, L. (n.d.). Identifying a threshold for the executive function advantage in bilingual children.

de Cat, C., & Serratrice, L. (n.d.). Predicting language proficiency in bilingual children. *Under Review*.

De Grauwe, S., Swain, A., Holcomb, P. J., Ditman, T., & Kuperberg, G. R. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, *48*(7), 1965–84. http://doi.org/10.1016/j.neuropsychologia.2010.03.017

Demie, F. (2018). English as an additional language and attainment in primary schools in England. *Journal of Multilingual and Multicultural Development*, *39*(3), 210–223. http://doi.org/10.1080/01434632.2017.1348508

Demie, F., & Strand, S. (2006). English language acquisition and educational attainment at the end of secondary school. *Educational Studies*, *32*(2), 215–231. http://doi.org/10.1080/03055690600631119

Department for Education. (2016). *School census 2016 to 2017*.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second- language learners. *Reading Research Quarterly*, *38*(1), 78–103.

Dunn, L. M., Dunn, D. M., Whetton, C., & Burley, J. (1997). *The British picture vocabulary scale II*. *BPVS II*. Windsor: NFER-Nelson.

Eager, C., & Roy, J. (2017). *Mixed Effects Models are Sometimes Terrible*.

Ehrlich, M. F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: differences in metacognitive monitoring. *Reading and Writing*, *11*(1), 29–63. http://doi.org/10.1023/A:1007996502372

Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2018a). Individual differences in children's pronoun processing during reading: detection of incongruence is associated with higher reading fluency and more regressions. *Journal of Experimental Child Psychology*, *173*, 250–267. http://doi.org/S0022096517304770

Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2018b). The repeated name penalty effect in children's natural reading: evidence from eye tracking. *Quarterly Journal of Experimental Psychology*, 174702181875771. http://doi.org/10.1177/1747021818757712

Evans, M. A., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading: evidence from eye movement monitoring. *Psychological Science*, *16*(11), 913–920. http://doi.org/10.1111/j.1467-9280.2005.01636.x

Fernández, G., Shalom, D. E., Kliegl, R., & Sigman, M. (2014). Eye movements during reading proverbs and regular sentences: the incoming word predictability effect. *Language, Cognition and Neuroscience*, *29*(3), 260–273. http://doi.org/10.1080/01690965.2012.760745

Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. *The R Journal*, *5*(1), 39–52.

Fuente, J. de la, Santiago, J., Román, A., Dumitrache, C., & Casasanto, D. (2014). When you think about it, your past is in front of you: how culture shapes spatial conceptions of time. *Psychological Science*, *25*, 1682–1690.

Gao, F. (2012). Imagined identity of ethnic Koreans and its implication for bilingual education in China. *International Journal of Bilingual Education and Bilingualism*, *15*(3), 343–353. http://doi.org/10.1080/13670050.2011.636146

García-Madruga, J. A., Vila, J. O., Gómez-Veiga, I., Duque, G., & Elosúa, M. R. (2014). Executive processes, reading comprehension and academic achievement in 3th grade primary students. *Learning and Individual Differences*, *35*, 41–48. http://doi.org/10.1016/j.lindif.2014.07.013

García, O., & Baetens Beardsmore, H. (2009). *Bilingual education in the 21st century:*

A global perspective. Malden, MA ; Oxford: Wiley-Blackwell.
Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 199–253). Boston, Mass: MIT Press.
Gilbert, P., & Varadhan, R. (2016). numDeriv: Accurate Numerical Derivatives.
Giora, R. (1999). On the priority of salient meanings: studies of literal and figurative language. *Journal of Pragmatics*, *31*(7), 919–929. http://doi.org/10.1016/S0378-2166(98)00100-3
Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*(1), 3–18.
Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. http://doi.org/10.1177/074193258600700104
Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. http://doi.org/10.3102/0013189X11413260
Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15.
Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). Building semantic networks: the impact of a vocabulary intervention on preschoolers' depth of word knowledge. *Reading Research Quarterly*. http://doi.org/10.1002/RRQ.225
Halekoh, U., & Højsgaard, S. (2009). A Kenward-Roger approximation and parametric bootstrap Mmethods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software*, *10*(2), 1–24. http://doi.org/10.18637/jss.v069.i12
Hart, B., & Risley, T. R. (2003). The early catastrophe: the 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.
Hartig, F. (2016). DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.2.0.
Heller, M. (2010). Globalization and the commodification of bilingualism in Canada. In *The Handbook of Language and Globalization* (pp. 287–304). http://doi.org/10.1002/9781444324068.ch12
Heredia, R. R., & Cieslicka, A. B. (2016). Metaphoric reference: an eye movement analysis of Spanish–English and English–Spanish bilingual readers. *Frontiers in Psychology*, *7*, 1–10. http://doi.org/10.3389/fpsyg.2016.00439
Hessel, A. K., & Murphy, V. A. (2018). Understanding how time flies and what it means to be on cloud nine: English as an Additional Language (EAL) learners' metaphor comprehension. *Journal of Child Language*, 1–27. http://doi.org/10.1017/S0305000918000399
Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, *37*(5), 1253–1280. http://doi.org/10.1017/S0142716415000569
Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.
Howlin, P., & Cross, P. (1994). The variability of language test scores in 3- and 4-year-old children of normal non-verbal intelligence: a brief research report. *European Journal of Disorders of Communication : The Journal of the College of Speech and Language Therapists, London*, *29*(3), 279–288.
Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The

developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading*, *26*(1), 141–423. http://doi.org/10.1111/1467-9817.261003

Hyönä, J., Lorch, Robert F., J., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*(1), 44–55. http://doi.org/10.1037/0022-0663.94.1.44

Hyönä, J., & Nurminen, A.-M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, *97*, 31–50. http://doi.org/10.1348/000712605X53678

James, A. N., Fraundorf, S. H., Lee, E., & Watson, D. G. (2018). Individual differences in syntactic processing: is there evidence for reader-text interactions ? *Journal of Memory and Language*, *102*, 155–181. http://doi.org/10.1016/j.jml.2018.05.006

Johnson, J. (1991). Developmental versus language-based factors in metaphor interpretation. *Journal of Educational Psychology*, *83*(4), 470–483. http://doi.org/10.1037//0022-0663.83.4.470

Johnson, J., & Pascual-Leone, J. (1989). Developmental levels of processing in metaphor interpretation. *Journal of Experimental Child Psychology*, *48*(1), 1–31.

Joseph, H. S. S. L., Bremner, G., Liversedge, S. P., & Nation, K. (2015). Working memory, reading ability and the effects of distance and typicality on anaphor resolution in children. *Journal of Cognitive Psychology*, *27*(5), 622–639. http://doi.org/10.1080/20445911.2015.1005095

Joseph, H. S. S. L., & Liversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PloS One*, *8*(1), 1–13. http://doi.org/10.1371/journal.pone.0054141

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, *61*(5), 708–723. http://doi.org/10.1080/17470210701400657

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, *49*(16), 2078–2086. http://doi.org/10.1016/j.visres.2009.05.015

Joseph, H. S. S. L., & Nation, K. (2018). Examining incidental word learning during reading in children: the role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. http://doi.org/10.1016/j.jecp.2017.08.010

Joseph, H. S. S. L., Nation, K., & Liversedge, S. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, (June).

Joshi, R. M., Ji, X. R., Breznitz, Z., Amiel, M., & Yulia, A. (2016). Validation of the simple view of reading in Hebrew – a semitic language. *Scientific Studies of Reading*, *19*, 243–252. http://doi.org/10.1080/10888438.2015.1010117

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: an analysis of eye movements and reported encoding strategies. *Memory*, *15*(6), 634–646. http://doi.org/10.1080/09658210701457096

Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: evidence from eye fixations. *Memory & Cognition*, *29*(2), 363–72.

Kan, R. (2014). *Effects of frequency and idiomaticity on second language reading comprehension in children*. University of Oxford.
Keshtiari, N., & Vasishth, S. (2013). Reactivation of antecedents by overt versus null pronouns: evidence from Persian. *Journal of Language Modelling*, *1*(2), 243–266.
Kim, Y. S., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly*, *49*(3), 269–281. http://doi.org/10.1002/rrq.74
Kim, Y. S., Vorstius, C., & Radach, R. (2018). Does online comprehension monitoring make a unique contribution to reading comprehension in beginning readers? Evidence from eye movements. *Scientific Studies of Reading*, 1–17. http://doi.org/10.1080/10888438.2018.1457680
Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction*, *5*(2), 143–165. http://doi.org/10.1016/0959-4752(95)00009-R
Kintsch, W. (2008). How the mind computes the meaning of metaphor: a simulation based on LSA. In *The Cambridge handbook of metaphor and thought*.
Kogan, N., Connor, K., Gross, A., & Fava, D. (1980). Understanding visual metaphor: developmental and individual differences. *Monographs of the Society for Research in Child Development*, *45*(1), 1–90. http://doi.org/10.2307/1165832
Koller, V., Hardie, A., Rayson, P., & Semino, E. (2008). Using a semantic annotation tool for the analysis of metaphor in discourse. *Metaphorik.De*, *15*, 141–160. http://doi.org/10.1136/bmjspcare-2014-000785
Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990. http://doi.org/10.3758/s13428-012-0210-4
Kutas, M., & Van Petten, C. (2006). Psycholinguistics electrified II: event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego: Academic Press.
Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago: The University of Chicago Press.
Lervåg, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *51*(5), 612–620. http://doi.org/10.1111/j.1469-7610.2009.02185.x
Lesaux, N. K., Rupp, A. a., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology*, *99*(4), 821–834. http://doi.org/10.1037/0022-0663.99.4.821
Levorato, M. C., Nesi, B., & Cacciari, C. (2004). Reading comprehension and understanding idiomatic expressions: a developmental study. *Brain and Language*, *91*(3), 303–314. http://doi.org/10.1016/j.bandl.2004.04.002
Luke, S. G., Henderson, J. M., & Ferreira, F. (2015). Children's eye-movements during reading reflect the quality of lexical representations: an individual differences approach. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(6), 1675–1683. http://doi.org/10.1037/xlm0000133
Macaro, E. (2006). Strategies for language learning and for language use: revising the theoretical framework. *The Modern Language Journal*, *90*(3), 320–337. http://doi.org/10.1111/j.1540-4781.2006.00425.x

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, *34*(1), 9. http://doi.org/10.2307/3588095

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*(2), 267–290. http://doi.org/10.5054/tq.2011.247708

McCarthy, M. (2006). *Cambridge idioms dictionary*. Cambridge: Cambridge University Press.

McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science*, *30*, 181–192. http://doi.org/10.1207/s15516709cog0000_49

McKendry, M. (2014). *Investigating the relationship between reading comprehension and semantic skill in children with English as an Additional Language: A focus on idiom comprehension*. University of Oxford.

McNamara, D. S. (2017). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes*, *54*(7), 479–492. http://doi.org/10.1080/0163853X.2015.1101328

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, *140*(2), 409–433. http://doi.org/10.1037/a0033890

Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*(4), 551–561. http://doi.org/10.3758/BF03194956

Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2016). The importance of reading naturally: evidence from combined recordings of eye movements and electric brain potentials. *Cognitive Science*, 1–32. http://doi.org/10.1111/cogs.12384

Meyers, J., Lytle, S., Palladino, D., Devenpeck, G., & Green, M. (1990). Think-aloud protocol analysis: an investigation of reading comprehension strategies in fourth- and fifth-grade students. *Journal of Psychoeducational Assessment*, *8*(2), 112–127. http://doi.org/10.1177/073428299000800201

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: a reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*(3), 266–293. http://doi.org/10.1016/j.jml.2008.06.002

Monster, I. (2018). Dimensions in children's word knowledge across the upper primary grades. In *Talk presented at the 8th International Summer School on Literacy Research*. Egmond aan Zee.

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8-and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*(2), 447–468. http://doi.org/10.1037/xge0000054

Murphy, V. A. (2014). *Second language learning in the early school years: Trends and contexts*. Oxford: Oxford University Press.

Murphy, V. A. (2018). Literacy development in linguistically diverse pupils. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language: The state of the science across its subfields.* (Studies in, Vol. 54, pp.

315–323). Amsterdam: John Benjamins. http://doi.org/10.1075/btl.128.02gar
Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: evidence from eye movement analyses. *The Quarterly Journal of Experimental Psychology*, *40*(4), 693–718. http://doi.org/10.1080/14640748808402294
Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing*, *20*(7), 691–719. http://doi.org/10.1007/s11145-006-9045-7
NALDIC. (2016). The new EAL "stages of proficiency." *EAL Journal*, *Autumn*, 25–28.
Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*(2), 1–14.
Nation, K., Marshall, C. M., & Altmann, G. T. . (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314–329. http://doi.org/10.1016/j.jecp.2003.09.001
Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
National Association of Language Development in the Curriculum (NALDIC). (2012). Languages in school. Retrieved July 14, 2017, from http://www.naldic.org.uk/research-and-information/eal-statistics/lang
Nieuwenhuis, R., & Grotenhuis, M. (2012). influence.ME : tools for detecting influential data in mixed effects models. *R Journal*, *4*(2), 38–47.
Oakhill, J., Cain, K., & Nesi, B. (2016). Understanding of idiomatic expressions in context in skilled and less skilled comprehenders: online processing and interpretation. *Scientific Studies of Reading*, *20*(2), 124–139. http://doi.org/10.1080/10888438.2015.1092973
Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*(7–9), 657–686. http://doi.org/10.1007/s11145-005-3355-z
OECD. (2006). Where immigrant students succeed: a comparative review of performance and engagement in PISA 2003. *Intercultural Education*, *17*(5), 507–516. http://doi.org/10.1080/14675980601063900
OECD. (2016). *Programme for International Student Assessment Results from PISA 2015: Germany*.
Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2015). Individual differences in the processing of written sarcasm and metaphor: evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 433–450.
Oller, K. D., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, *28*(2), 191–230. http://doi.org/10.1017/S0142716407070117
Olson, R. K., Kliegl, R., Davidson, B. J., & Foltz, G. (1985). Individual and developmental differences in reading disability. *Reading Research: Advances in Theory and Practice, Vol. 4.*, 1–64.
Ouellette, G. P. (2006). What's meaning got to do with it: the role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554–566. http://doi.org/10.1037/0022-0663.98.3.554
Özçalışkan, Ş. (2007). Metaphors we move by: children's developing understanding of

metaphorical motion in typologically distinct languages. *Metaphor and Symbol*, *22*(2), 147–168. http://doi.org/10.1080/10926480701235429
Paradis, J. (2011). Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. http://doi.org/10.1075/lab.1.3.01par
Paradis, J. (2016). Parent report data on input and experience reliably predict bilingual development and this is not trivial. *Bilingualism: Language and Cognition*, 1–2. http://doi.org/10.1017/S136672891600033X
Paradis, J., & Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, *20*. http://doi.org/10.1111/desc.12433
Paradis, J., & Kirova, A. (2014). English second-language learners in preschool: profile effects in their English abilities and the role of home language environment. *International Journal of Behavioral Development*, *38*(4), 342–349. http://doi.org/10.1177/0165025414530630
Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*(3), 399–410. http://doi.org/10.1017/S014271640707021X
Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Language Learning*, *43*(1), 93–120. http://doi.org/10.1111/j.1467-1770.1993.tb00174.x
Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 1–8. http://doi.org/10.3389/neuro.11.010.2008
Pellicer-Sanchez, A. (2015). Using eye-tracking in incidental L2 vocabulary acquisition research. In *Applied Linguistics and Seocnd Language Acquisition Seminars*. Oxford: Department of Education, University of Oxford.
Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. http://doi.org/10.1080/10888430701530730
Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy*. Amsterdam: John Benjamins.
Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. http://doi.org/10.1080/10888438.2013.827687
Perfetti, C., Yang, C.-L., & Schmalhofer, F. (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology*, *22*, 303–318. http://doi.org/10.1002/acp
Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 556–573. http://doi.org/10.1037//0278-7393.27.2.556
Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In *The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond* (Vol. 44, pp. 1–27). http://doi.org/10.4324/9780203509050
Pimperton, H., & Nation, K. (2010). Suppressing irrelevant information from working memory: evidence for domain-specific deficits in poor comprehenders. *Journal of Memory and Language*, *62*(4), 380–391. http://doi.org/10.1016/j.jml.2010.02.005

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en espagnol: toward a typoloy of code-switching. *Linguistics*, *18*(7–8), 581–618.
Preacher, K. J., MacCallum, R. C., Rucker, D. D., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, *10*(2), 178–192. http://doi.org/10.1037/1082-989X.10.2.178
Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2007). Development of lexical mediation in the relation between reading comprehension and word reading skills in Greek. *Scientific Studies of Reading*, *11*(3), 165–197. http://doi.org/10.1080/10888430701344322
R Core Team. (2016). R: a language and environment for statistical computing. Retrieved from http://www.r-project.org/
Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–688. http://doi.org/10.1007/s00426-008-0173-3
Raudszus, H., Segers, E., & Verhoeven, L. (2017). Lexical quality and executive control predict children's first and second language reading comprehension. *Reading and Writing*, *31*(2), 1–20. http://doi.org/10.1007/s11145-017-9791-8
Rayner, K. (1977). Visual attention in reading: eye movements reflect cognitive processes. *Memory & Cognition*, *5*(4), 443–448. http://doi.org/10.3758/BF03197383
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. http://doi.org/10.1037/0033-2909.124.3.372
Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. http://doi.org/10.1080/17470210902816461
Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3). http://doi.org/10.1207/s1532799xssr1003
Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290–1301. http://doi.org/10.1037/0278-7393.30.6.1290
Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*(2), 110–149. http://doi.org/10.1016/j.dr.2013.03.001
Rodd, J. (2017). Lexical Ambiguity. In *Oxford Handbook of Psycholinguistics.* (pp. 1–43). http://doi.org/10.17605/OSF.IO/YEZC6
Rundblad, G., & Annaz, D. (2010). Development of metaphor and metonymy comprehension: receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*(3), 547–563.
Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, *25*(2), 211–236. http://doi.org/10.1177/0265532207086782
Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): comprehension is supported by regressions during reading. *Psychological*

Science, *25*(6), 1218–1226. http://doi.org/10.1177/0956797614531148
Segal, D., & Gollan, T. H. (2018). What's left for balanced bilinguals? Language proficiency and item familiarity affect left-hemisphere specialization in metaphor processing. *Neuropsychology*, 1–14.
Semino, E. (2008). *Metaphor in discourse*. Cambridge: Cambridge University Press.
Shahar-Yames, D., & Prior, A. (2018). The challenge and the opportunity of lexical inferencing in language minority students. *Reading and Writing*. http://doi.org/10.1007/s11145-018-9830-0
Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *American Psychological Association*, *107*(2), 321–331.
Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549–569. http://doi.org/10.1093/applin/amt054
Skoufaki, S. (2006). *Investigating idiom instruction methods*. University of Cambridge.
Smith, S. A., & Murphy, V. A. (2015). Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, *28*(3), 347–369. http://doi.org/10.1007/s11145-014-9527-y
Snowling, M., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., … Hulme, C. (2009). *York assessment of reading for comprehension*. London: GL Assessment.
Spätgens, T., & Schoonen, R. (2018). The semantic network, lexical access, and reading comprehension in monolingual and bilingual children: an individual differences study. *Applied Psycholinguistics*, *39*(1), 225–256.
Spencer, M., & Wagner, R. K. (2017). The comprehension problems for second-language learners with poor reading comprehension despite adequate decoding: a meta-analysis. *Journal of Research in Reading*, *40*(2), 199–217. http://doi.org/10.1111/1467-9817.12080
SR Research. (2009a). Data Viewer.
SR Research. (2009b). Experiment Builder.
Stites, L. J., & Özçalişkan, Ş. (2013a). Developmental changes in children's comprehension and explanation of spatial metaphors for time. *Journal of Child Language*, *40*(5), 1123–1137. http://doi.org/10.1017/S0305000912000384
Stites, L. J., & Özçalişkan, Ş. (2013b). Teasing apart the role of cognitive and verbal factors in children's early metaphorical abilities. *Metaphor and Symbolic Activity*, *28*(2), 116–129. http://doi.org/10.1080/10926488.2013.768511
Strand, S., & Demie, F. (2005). English language acquisition and educational attainment at the end of primary school. *Educational Studies*, *31*(3), 275–291. http://doi.org/10.1080/03055690500236613
Strand, S., & Hessel, A. K. (2018). *English as an additional language (EAL), proficiency in English and pupil's educational achievement: an analysis of local authority data*.
Strand, S., Malmberg, L., & Hall, J. (2015). English as an Aditional Language (EAL) and educational achievement in England: an analysis of the national pupil database, 1–94.
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. http://doi.org/10.1207/s1532799xssr1004

Tartter, V. C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R. V. (2002). Novel metaphors appear anomalous at least momentarily: evidence from N400. *Brain and Language*, *80*(3), 488–509. http://doi.org/10.1006/brln.2001.2610

The British National Corpus. (2007). Retrieved from http://www.natcorp.ox.ac.uk/

Tiffin-Richards, S. P., & Schroeder, S. (2018). The development of wrap-up processes in text reading: a study of children's eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1–43. http://doi.org/10.1037/xlm0000506

Tilstra, J., McMaster, K., Van Den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*(4), 383–401. http://doi.org/10.1111/j.1467-9817.2009.01401.x

Tomasello, M. (2008). The usage-based theory of language acquisition. In *The Cambridge handbook of child language* (pp. 69–88). Cambridge: Cambridge University Press.

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *TOWRE 2: test of word reading efficiency*. Austin, Texas : Pro-Ed.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, *45*(5), 453–466. http://doi.org/10.1177/0022219411432685

van den Bosch, L. J., Segers, E., & Verhoeven, L. (2018). Online processing of causal relations in beginning first and second language readers. *Learning and Individual Differences*, *61*, 59–67. http://doi.org/10.1016/j.lindif.2017.11.007

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*(7), 1665–1690. http://doi.org/10.1007/s11145-011-9337-4

Vasishth, S. (2014). *An introduction to statistical data analysis (Summer 2014) Lecture notes*.

Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied Linguistics*, *14*(4), 344–363.

Verhallen, M., & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of third and fifth graders. *Applied Linguistics*, *19*(4), 452–470. http://doi.org/10.1093/applin/19.4.452

Verhoeven, L. (2000). Components in Early Second Language Reading and Spelling. *Scientific Studies Of Reading*, *4*(4), 313–330. http://doi.org/10.1207/S1532799XSSR0404

Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*(8), 1805–1818. http://doi.org/10.1007/s11145-011-9346-3

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217–234. http://doi.org/10.1017/S0142716401002041

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, (94), 119–133. http://doi.org/10.1016/j.jml.2016.10.003

von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*(2), 109–127.

http://doi.org/10.1016/j.jml.2011.02.004
von der Malsburg, T., & Vasishth, S. (2012). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1–34. http://doi.org/10.1080/01690965.2012.728232
Vorstius, C., Radach, R., Mayer, M. B., & Lonigan, C. J. (2013). Monitoring local comprehension monitoring in sentence reading. *School Psychology Review*, *42*(2), 191–206.
Waggoner, J. E., & Palermo, D. S. (1989). Betty is a bouncing bubble: children's comprehension of emotion-descriptive metaphors. *Developmental Psychology*, *25*(1), 152–163.
Waggoner, J. E., Palermo, D. S., & Kirsh, S. J. (1997). Bouncing bubbles can pop: contextual sensitivity in children's metaphor comprehension. *Metaphor and Symbol*, *12*(4), 37–41. http://doi.org/10.1207/s15327868ms1204
Wechsler, D. (2011). *WASI -II: Wechsler abbreviated scale of intelligence- second edition*. San Antonio, Tex.: Psychological Corporation.
Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2017). Comprehension demands modulate re-reading, but not first pass reading behavior. *The Quarterly Journal of Experimental Psychology*, 1–37. http://doi.org/10.1080/17470218.2017.1307862
Westman, M., Korkman, M., Mickos, A., & Byring, R. (2008). Language profiles of monolingual and bilingual Finnish preschool children at risk for language impairment. *International Journal of Language & Communication Disorders*, *43*(6), 699–711. http://doi.org/10.1080/13682820701839200
Whiteside, K. E., Gooch, D., & Norbury, C. F. (2017). English language proficiency and early school attainment among children learning English as an additional language. *Child Development*, *88*(3), 812–827. http://doi.org/10.1111/cdev.12615
Wiig, E. H., & Secord, W. (1992). *Test of word knowledge (TOWK): Complete kit*. Cambridge: Pearson.
Willingham, D. (2017). *The reading mind: A cognitive approach to understanding how the mind reads*. Oxford: Wiley.
Wilson, N. L., & Gibbs, R. W. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science*, *31*(4), 721–731. http://doi.org/10.1080/15326900701399962
Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, *12*(4), 289–297. http://doi.org/10.1037//0012-1649.12.4.289
Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356–366. http://doi.org/10.1016/j.neuroimage.2012.04.054
Wonnacott, E., Joseph, H. S. S. L., Adelman, J. S., & Nation, K. (2015). Is children's reading "good enough"? Links between online processing and comprehension as children read syntactically ambiguous sentences. *Quarterly Journal of Experimental Psychology (2006)*, 1–25. http://doi.org/10.1080/17470218.2015.1011176
Zabrucky, K., & Ratner, H. H. (1989). Effects of reading ability on children's compreehsion evaluation and regulation. *Journal of Reading Behaviour*.
Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to

avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. http://doi.org/10.1111/j.2041-210X.2009.00001.x

Appendices

Appendix A. Pre-registration document Experiment 1

(Uploaded on OpenScienceFramework on 7/9/16, 11:24AM)

EAL reading metaphors project.
Annina Hessel, Victoria Murphy, Kate Nation
University of Oxford

Pre-registration Document

Interest areas:
There are three target areas (see figure below):
1) First word of metaphor
2) Second word of metaphor
3) Post metaphor region (3 words)[a]

hours	fly by	so she decides

(1) (2) (3)

Target areas begin with space before and end with last character of each word. There are some cases where target areas span more than one word:
- Phrasal verbs are coded as one area, ex. '[fly by]'.
- Target area that would be less than five characters long are extended to include another word until they had reached a minimum of five characters, e.g. '[*the* time]'. Whenever possible, the word to be added was chosen so that it reduced the space between the first and the second word of the metaphor, e.g. for 'film was a flop' choose 'film was [*a* flop] and' rather than 'film was a [flop *and*]'.[b]

[a] The post metaphor region is set around three rather than one word in order to capture delayed spill over effects as they occur in children (Joseph et al., 2008).
[b] Target areas are extended to be at least five characters in order to allow for analysing later measures in particular which arguably will only show on longer words, as they have in previous studies (Joseph et al., 2008).

Eye tracking measures:
- look at all eye tracking measures for every interest area, only report gaze duration among all early measures as indicator of lexical processing
- Measures to look at: (underlined= the ones that I plan on reporting)
Early:
IA_FIRST_RUN_DWELL_TIME = Gaze duration
IA_FIRST_FIXATION_DURATION= First Fixation Duration

Late:
IA_REGRESSION_PATH_DURATION = Go-past time
IA_REGRESSION_IN_COUNT - Returns the number of regressions in
IA_REGRESSION_OUT_COUNT - Returns the number of regressions out
IA_DWELL_TIME - Total time

Exclusions:
It is common in eye tracking research to exclude participants with low answer accuracy on comprehension questions since the assumption is that these participants have not fully read the material. While this may be true, these participants might have attempted to comprehend the material and failed, a process that the current study would be particularly interested in, given the vast heterogeneity in the target population (Strand, Malmberg, & Hall, 2015). In order to give justice to both approaches, the following was decided:
As a default, eye movements of all participants will be analysed. When analyses make it necessary to limit the sample to a subset of participants, the following procedure will be adopted: We will establish a cut-off point of exclusion based on descriptive data of answer accuracy. The cut-off will be chosen in order to not exclude too many participants while excluding cases with extremely low answer accuracy.

Coding of variables
Group: 1= English as an additional language (EAL) learners, 2= Monolingual

1

EAL reading metaphors project.
Annina Hessel, Victoria Murphy, Kate Nation
University of Oxford
Condition = 1=Metaphor, 2= Literal
Type= 1= Verbal, 2 = Nominal
Cond_Type = 1 = Met-Verb, 2 = Lit-Verb; 3= Met-Nom, 4 = Lit-Nom
IA: 0= Practice, 1= Pre_Context, 2=Met1, 3=Mid_Context, 4= Met2, 5=Post_Met, 6=Post_Context

Met_Type and Condition are going to be used either together or separately depending on LME model to be calculated.

Model building in general:
1. The starting point of every model will be defined first and the full random structure will be trimmed until the model converges.
2. Trim random structure following the standard approach in reading eye tracking research.
3. Centre continuous variables
4. Log-transform ET data
5. Before running the LME model, we will use the "contr.sdif"(MASS package) to specify the two categorical factors (Met_Type and Story).

1) Do EAL and non-EAL children differ in their comprehension and processing of metaphors (of different types)?

Starting point of the model:
Eye tracking measure ~ Condition*Group*Type + (Condition*Group*Type |ID) + (condition*group*Met_Type |story)
(Or for offline answer accuracy:
Answer accuracy ~ condition*group*Met_Type + (condition*group*Met_Type|ID) + (condition*group*Met_Type|story) family = "binomial")

Predictions:
1 Condition
a) No longer RT on metaphors itself in early measures + lower accuracy offline measures
(even if in some studies that study ambiguity, but not here, evidence against (Heredia & Cieslicka, 2016; Inhoff, Lima, & Carroll, 1984)
1 b) Probably see differences between metaphor and literal rather later (go past, total time) or even in later region
- based on research on low comprehenders that show delayed effects in children, as well as research on ambiguous figurative language that finds effects in later measures (Frisson & Pickering, 1999; H. Joseph, 2016)
(resolving "ambiguity" when notice later that first interpretation of metaphor might not make sense, or just ignored unknown phrase- next event in storyline building up on metaphor)

2) Group
Overall, longer reading times for EAL (ex. Number of fixations, shorter saccade amplitude, sentence reading times) + lower accuracy offline measures can be expected based on previous research showing lower reading comprehension abilities in EAL children (Spencer & Wagner, 2016).

3) Metaphor type
On manipulation (nominal + verbal):
Nominal less transparent (transparency meaning here the link between the single words of the metaphor and its literal meaning), low transparency of figurative phrases has been found to cause comprehension difficulties in children (Nippold & Taylor, 2002)→ longer reading times on nominal metaphors + lower accuracy offline measures

4) Group * metaphor type:
Note: Beyond my MSc research in which I looked at offline comprehension of metaphors in oral stories (Hessel, 2014), I am aware of no study that would compare the comprehension of verbal and nominal metaphors, in particular not in an eye tracking study. The following predictions are based on my MSc research, other offline projects on EAL figurative language comprehension, as well as eye tracking studies comparing other types of figurative language that can be hypothesized to relate to the comparison verbal and nominal metaphors.

Based on the type of evidence described above, we expect shorter total time and higher answer accuracy in comprehension questions on verbal metaphors for monolinguals, but not EAL (bilinguals generally and EAL more specifically have been found to not show strong differences between different types of familiar vs. unfamiliar or novel vs. real figurative phrases- both in comprehension tasks and in eye tracking) (Hessel, 2014; McKendry, 2014; Siyanova-Chanturia, Conklin, & Schmitt, 2011). However, we are aware of the fact that our target population have extremely heterogeneous profiles and that, given the convenience sampling adopted in the cited studies as well as in the current study, we find different patterns depending on the nature of our current sample.

5) Condition*Group:
Find effects of resolving the metaphor in later regions in EAL kids than in monolinguals in late measures (Monolinguals on metaphor, EAL on later context (after "and") (H. Joseph, 2016)

Possible second step:
If we found longer RT on later measures for metaphors (signs of reintegration), then we will do the following extra analyses:

6) Differentiate between trials that were answered correctly and those that weren't
→ See what reading pattern is linked to giving correct answer (either knowing metaphor and going through or integrating the meaning in late measures)
→ what is linked to getting it wrong (ignoring it or trying to figure it out and not being able to)
→ cf.(H. Joseph, 2016) -second paper: accuracy and RT not always directly linked, sometimes see signs of seemingly successful reading in online measures but no difference in offline (difficult to differentiate between successful comprehension monitoring and unsuccessful one based on online measures only)

Further Analyses
Based on the results from analysing the experimental manipulation as described above, we will continue exploring links of the experimental measure to external measures further. In these more exploratory analyses, we will address the following questions:

2) Do reading comprehension and/or vocabulary knowledge explain differences in processing literal and figurative language (better than the group comparison between EAL and monolingual children)?
Analyses of national student databases have that the common grouping as EAL and monolingual students becomes much more informative when combined with measures of language proficiency (Strand & Demie, 2005). It also has been pointed out that there are distinct similarities between monolingual low comprehenders and EAL students with low comprehension skills (Spencer & Wagner, 2016). We therefore plan to compare the grouping EAL- monolingual to our continuous variables tapping vocabulary and reading comprehension in order to see what we can learn about the role that either play for online reading and offline comprehension. Results will be compared to previous correlational studies of EAL comprehension.

The current study has tested two type of vocabulary, namely expressive size (through a naming pictures task) and receptive depth of vocabulary (through a giving multiple meanings for polysemous words tasks). McKendry reported that in her correlational study of reading and idiom comprehension, differences between groups disappeared when partialling our depth of vocabulary, while when controlling for vocabulary size, differences across idiom types disappeared only for EAL children, but not for their monolingual peers (2014). This suggests that vocabulary depth captures most of the variance between EAL and monolingual students. It further suggests that children approach figurative phrases differently, possibly by relying more on their expressive vocabulary. It remains to be seen if the same influences of different types of vocabulary knowledge can be found in this study, using the same vocabulary tests, but different phrases (metaphors not idioms) and a different task (reading comprehension not multiple-choice).

Regarding reading comprehension, a meta-analysis of 16 studies showed that EAL students have lower oral comprehension than their monolingual peers even when matched on reading comprehension (Spencer & Wagner, 2016). This suggests again that EAL children approach comprehension tasks differently than their monolingual peers since they manage to reach the same level of comprehension having lower oral comprehension. We want to investigate if EAL and monolingual readers differ in online reading times when matched on reading comprehension. If we found differences in online reading, this would lend more information for what exactly is different in the way EAL children approach higher level comprehension tasks.

3) Does the answer accuracy in my test (on inferential and metaphor-related questions especially) predict reading comprehension as measured by the YARC (beyond what is tested by the TOWK and better in EAL learners)?

Existing oral comprehension and vocabulary tests do not tap all the variance of EAL comprehension skills reflected in their written comprehension, possibly because they mainly focus on single-word comprehension (Spencer & Wagner, 2016). Given the previously reported strong correlation between idiom knowledge and reading comprehension in EAL students (McKendry, 2014), we would like to see how much the present test predicts in EAL reading comprehension.

Seeing a stronger predictive power between the here presented higher level comprehension test and standardised measures of vocabulary would lend further support to the idea that when testing EAL students' language skills, tapping higher level skills as well as multi-word and figurative vocabulary are most informative.

4) Are performances on different metaphor types (verbal vs. nominal) correlated with different abilities?

Different types of metaphor have been researched in previous studies. These metaphor types are comparable to different kinds of vocabulary whose comprehension seems to be related to different skills. Previous research has found different correlations between individual differences in vocabulary and executive control or reasoning suggest that 1) EAL use different abilities to solve figurative phrases (Hessel, 2014; McKendry, 2014) and 2) nominal metaphors more linked to executive control/ reasoning than verbal metaphors which might be more linked to vocabulary knowledge (Beaty & Silvia, 2013; Hessel, 2014). Looking at the links between individual differences and metaphor comprehension allows for extending our understanding of what skills are crucial when processing different metaphors. The resulting improved understanding can inform teaching and theory-building alike.

Based on the cited evidence, we can expect stronger link between processing and comprehension of nominal metaphors and nonverbal reasoning on the one hand, and verbal metaphors and vocabulary on the other hand.

References

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: cognitive abilities and the production of figurative language. *Memory & Cognition*, *41*(2), 255–267. http://doi.org/10.3758/s13421-012-0258-5

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: evidence from eye movements. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(6), 1366–1383. http://doi.org/10.1037/0278-7393.25.6.1366

Heredia, R. R., & Cieslicka, A. B. (2016). Metaphoric Reference : An Eye Movement Analysis of Spanish – English and English – Spanish Bilingual Readers. *Frontiers in Psychology*, *7*(March), 1–10. http://doi.org/10.3389/fpsyg.2016.00439

Hessel, A. K. (2014, July). *Exploring Metaphor Comprehension in UK Primary School Children with English as Their First or Additional Language*. University of Oxford.

Inhoff, A. W., Lima, S. D., & Carroll, P. J. (1984). Contextual effects on metaphor comprehension in reading. *Memory & Cognition*, *12*(6), 558–567.

Joseph, H. (2016). What can eye movements tell us about children's reading comprehension (difficulties)? In *Institute of Education Research into Practice Event*. Reading.

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, *61*(5), 708–723. http://doi.org/10.1080/17470210701400657

McKendry, M. (2014). *Investigating the relationship between reading comprehension and semantic skill in children with English as an Additional Language: A focus on idiom comprehension*. University of Oxford.

Nippold, M. A., & Taylor, C. L. (2002). Judgments of Idiom Familiarity and Transparency : A Comparison of Children and Adolescents. *Journal of Speech, Language, and Hearing Research*, *45*(April), 384–391.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272. http://doi.org/10.1177/0267658310382068

Spencer, M., & Wagner, R. K. (2016). The comprehension problems for second-language learners with poor reading comprehension despite adequate decoding : A meta-analysis. *Journal of Research in Reading*, *0*(0), 1–19. http://doi.org/10.1111/1467-9817.12080

Strand, S., & Demie, F. (2005). English language acquisition and educational attainment at the end of primary school. *Educational Studies*, *31*(3), 275–291. http://doi.org/10.1080/03055690500236613

Strand, S., Malmberg, L., & Hall, J. (2015). English as an Aditional Language (EAL) and educational achievement in England: An analysis of the national pupil database, (January), 1–94.

Appendix B. Ethics Applications for Experiment 1, Including Risk Assessment

Form CUREC/1A

University of Oxford

CENTRAL UNIVERSITY RESEARCH ETHICS COMMITTEE (CUREC)

CUREC/1A Checklist for the Social Sciences and Humanities

The University of Oxford places a high value on the knowledge, expertise, and integrity of its members and their ability to conduct research to high standards of scholarship and ethics. The research ethics clearance procedures have been established to ensure that the University is meeting its obligations as a responsible institution. They start from the presumption that all members of the University will take their responsibilities and obligations seriously and will ensure that their research on human subjects is conducted according to the established principles and good practice in their fields and in accordance, where appropriate, with legal requirements. Since the requirements of research ethics review will vary from field to field and from project to project, the University accepts that different guidelines and procedures will be appropriate. Please check the CUREC website to ensure that you have the correct form for your project.

This form does not cover research governance, satisfactory methodology, compliance with the requirements of publishers when administering their tests or questionnaires, or the health and safety of employees and students. As principal investigator, it is your responsibility to ensure that requirements in these areas are met. Please carry out a risk assessment of the project, in consultation with all researchers involved, using the checklist and CUREC's other documentation.

The use of an asterisk in this form indicates a phrase defined in the glossary. The glossary and further information on the University's research ethics procedures are available from the CUREC website:

www.admin.ox.ac.uk/curec

This form is designed largely for research that falls within the Divisions of Social Sciences and Humanities and which does not involve a high level of risk to the subjects. Elite interviews, field work and oral history are included in the CUREC process. Please take a moment to read through it and if you have any questions or doubts as to whether it is the appropriate form, please review Section A or consult the CUREC website.

Note on anonymised data and audit: you do not need to obtain ethical approval for your study if:

- you are using previously collected **anonymised data** about people which neither you nor anyone else involved in your study can trace back to the individuals who provided them (e.g. census data, administrative data, secondary analysis). Please refer to the definition of *personal data in the glossary and FAQ A4 for further guidance; or
- you are conducting research on behalf of or at the request of a service provider that matches the definition of *audit in the glossary.

If your research is audit or uses prior-anonymised data, please check this box: ☐

You do not need to seek ethical approval from CUREC, and you do not need to complete any more of this form. However, please check with your department's requirements, as some departments require you to lodge this form with them.

Office use only: IDREC Ref. No. _____

Date of confirmation that checklist accepted on behalf of IDREC: // //

SECTION A	Yes	No
1) Are you using research methodologies commonly used in biomedical or behavioural laboratory sciences?		x
2) Is there a significant risk that the research will induce anxiety, stress or other harmful psychological states in participants that might persist beyond the duration of any test or interview in which they are participating?		X
3) Will the research involve human participants recruited by means of their status as present or past NHS patients or their relatives or carers?		X
*4) Does the research involve *human participants aged 16 and over who do not have *capacity to consent for themselves? See the Mental Capacity Act 2005*		X
5) Is the study to be funded by the US National Institutes of Health or another US federal funding agency?		x

If you answered 'yes', please **stop** work on this checklist and

- **for questions 1 and 2**, complete CUREC/1 instead (available from www.admin.ox.ac.uk/curec);
- **for questions 3 and 4**, submit your proposal to the appropriate NHS ethics committee (see www.nres.nhs.uk and www.admin.ox.ac.uk/researchsupport/=ctrg for further information);
- **for question 5**, or **if you answered 'yes' to questions 1, 2 or 4 and your research will take place outside the EU and is a biomedical study**, submit your proposal to OXTREC, which uses separate documentation. **Applications to OXTREC using this form will not be accepted.** If your research is not a biomedical study and does not have US funding, but will take place outside the EU, you may use this form to submit your application for approval to the Social Sciences and Humanities IDREC.

If you have answered 'no' to all questions in Section A, please complete Sections B-E. This form and any supporting materials should be typewritten.

SECTION B

Principal investigator/ supervisor/student researcher (title and name):	***Supervisor: Professor Victoria Murphy, Department of Education & Professor Kate Nation, Department of Experimental Psychology Student researcher: Annina Hessel, MSc& MEd***
Name of supervisor (STUDENT RESEARCH PROJECTS ONLY):	***Professor Victoria Murphy & Professor Kate Nation***
Degree programme, e.g. DPhil, MPhil, MSc (STUDENT RESEARCH PROJECTS ONLY):	***DPhil in Education***
Department or institute:	***Department of Education***
Address for correspondence (if different):	
Email and phone contact:	***Annina.Hessel@education.ox.ac.uk***
Title of research project:	***Metaphor comprehension in English as an additional language (EAL) learners***

Title and brief lay description of *research (about 150 words), plus description (about 200 words) of the nature of participants (including the criteria for inclusion/exclusion & method of recruitment, how professional guidelines are being applied (if applicable) and use to which the results/data will be put.

Please describe how you will obtain informed consent, citing and attaching, where applicable, documentation produced in support of your application such as generic recruitment and advertisement material, participant information sheet(s), consent form(s) and debriefing document(s).

My research focuses on the language learning of students learning English as an additional language (EAL). I am researching the comprehension processes that EAL children engage in when reading metaphor. Some EAL children experience comprehension difficulties that impede their literacy development, which in turn can affect their overall academic attainment (Hutchinson, Whiteley, Smith, & Connors, 2003; OECD, 2006). Furthermore, EAL learners' comprehension problems can be linked to difficulties understanding metaphors (Bountrogianni, 1984; Johnson, 1991). Importantly, different metaphors seem to cause different problems (Hessel, 2014; Kan, 2014). Measuring reading comprehension in EAL learners helps to understand these difficulties and develop targeted teaching. I thus will measure the processing of different metaphors online to see whether and how EAL children differ from monolinguals in their processing and comprehension of metaphor.

Participants will include typically-developing primary school children in year 3 (8 years of age), including both monolingual and EAL children (≅30 per group). They will be tested for vocabulary knowledge through the Test of Word Knowledge (Wiig & Secord, 1992). In order to make sure they are able to do the reading metaphor test, their literacy skills will be tested through the Test of Word Reading Efficiency 2nd Edition (Torgesen, 2012). These tests are quick to complete and have been developed and validated for this age range. For inclusion into the EAL group, language background questionnaires adapted from previous studies (McKendry & Murphy, 2011) and (Paradis, Emmerzael, & Duncan, 2010) will be administered to both the parents and children to establish the degree of exposure to home language (see Appendix A). If necessary, the questionnaire will be translated into the parents' first languages. To test metaphor comprehension, the children will read stories on a computer screen (see Appendix B). The stories will be 55-63 words to avoid tiring the children. Children will read 30 literal stories and 30 stories that contain metaphors. After every story, they will answer a comprehension question (see Appendix B). During the story reading, children's eye movements will be filmed through a table mounted Eyelink eye tracker.

All testing will be done in the children's schools. Children will be tested individually in a calm area of the school across two sessions of no more than 30-40 minutes each. Piloting with monolingual and EAL children (n ≅10 per group) will help to make sure that the tasks are age-appropriate. In case the tasks appear too difficult for the children or if any other problem is detected during the testing, the tasks will be considered age-inappropriate and will be adapted to meet the children's needs and abilities. During the first testing session, the standardised tests and the language background questionnaire will be administered on one day together with approximately 20 stories from the metaphor test, with the remaining 40 stories of the metaphor comprehension test being administered on another day. All data will be saved under a code assigned to the participants from the first data collection session onwards so that no personal information will be retained in the data sheets. I will keep the list of children's names and numbers in a password-protected computer file, and I will erase this list at the end of the project.

Metaphors will be pretested for ecological validity and frequency in the Oxford Children's Corpus (Banerji, Gupta, Kilgarriff, & Tugwell, 2013). Stories will be piloted with adult native speakers (N≅150) in order to make sure they sound natural and that literal and metaphorical stories don't vary in predictability (see Appendix C). There will be five pre-tests that investigate the literal and the metaphorical version of the story respectively or different aspects of the phrases (for cloze tests). Each survey will take approximately 15-20 minutes and will be filled out by approximately 20-30 individuals. The first participants taking these questionnaires will be asked if they are happy with being contacted for a one-one-one conversation to comment on potential difficulties connected to stimuli and tasks (n=2-3 per pre-test). In recruitment, I will rely on a convenience sample of students and members of societies and groups I am a member of, such as the Oxford University Ballet Society. The data will be treated as described above for children's data.

For recruitment of monolingual and EAL children, I am cooperating with a local charity called Jacari providing tutoring for EAL children. As tutors, Jacari recruits University of Oxford and Oxford Brookes students who volunteer to teach an EAL child for one year. My language testing will be done as part of a project that Jacari has asked me to implement to help them examine their work with EAL students. To that end, I will send out online questionnaires via email to Jacari tutors (see Appendix F) to explore the tutors' development, their evaluation of Jacari support, and their relationship with parents and students. The tutors who Jacari will approach for this online questionnaire will be both the tutors of the children I will be testing as well as other a few other tutors working with children from other year groups. The same questionnaire will be sent out at two time points, once in the beginning of the year and then at the end of the tutoring. The questionnaire will be distributed through the online platform Qualtrics. The tutors' answers will be analysed for the three areas named above. Furthermore, the questionnaire responses of the tutors whose Jacari students I will have tested in the above described study will be linked to their students' language abilities. Linking the tutors' experiences to the students' current state of knowledge will provide an overview of the current state of tutoring to Jacari. The data will be presented in a clear and professional way that helps Jacari to evaluate and improve their own services without compromising any participant's anonymity (either tutor or pupil). Jacari tutors will be given the option to opt in for discussing their tutoring experience within a focus group of several tutors organised by Jacari. For those tutors opting in to the focus group, a feedback will be provided on their development between the two testing points, for example on how their self-efficacy in teaching has developed from the beginning of the year to the end of their tutoring. Jacari will help establish contacts to the schools and families they work with in Oxford which will help me to reach the intended population of my study in the ways outlined below. Additionally, I will recruit other local schools by first emailing the head teachers and, upon obtaining their consent to do my study in their school, I asking the parents for consent.

I have received a CRB check in December 2013 that discloses no convictions, cautions, final warnings or reprimands. The described research project may best be conducted by following the protocol P25 provided by the central university research ethics committee. In order to comply with this protocol, the following procedures in recruiting participants and securing consent will be adopted.
All participants will receive information on the project together with an opt-in form to take part in the online questionnaire for tutors and NS in the pilots (integrated in the questionnaires, see Appendix C and F) and to take part in the language testing for teachers, children and their parents, respectively (see Appendix D and E). The child participants will be recruited and tested in local primary schools. The schools' head teachers will be contacted by myself together with Jacari and will be asked for permission to conduct the study in their school (see Appendix D). With the head teacher's consent, the aforementioned information letters will be given out to the parents. In case the EAL children's guardians are not in the position to understand the information provided in English, translations will be provided. The need for such additional material will be established in consultation with head and class teachers. If a child shows signs of being unwell or distressed during testing, the session will be stopped, and the event reported to the child's teacher.

List actual or probable location(s) where project will be conducted, if known:	Testing in local primary schools in and around Oxford in combination with online questionnaires
If your research involves overseas travel or fieldwork, have you completed and returned a travel risk assessment form? (Bear in mind that this may be necessary to ensure that the travel or fieldwork is covered by the University's travel insurance – see http://www.admin.ox.ac.uk/finance/insurance/travel/.)	**No overseas travel planned.**
Anticipated duration of project:	__32__ months
Anticipated start date:	20 / 01 / 16
Anticipated end date:	30 / 9 /18
Name and status (e.g. 3rd year undergraduate; post-doctoral research assistant) of others taking part in the project:	No other researchers involved.
Please indicate what training on research ethics the researchers involved with this study have received, e.g. the title of the online or in-person course, and date completed (online training available at www.admin.ox.ac.uk/researchsupport/integrity/human/):	Philosophy of Education course as part of PRS training in Department of Education

SECTION C

Methods to be used in the study (**tick** as many as apply: this information will help the committee understand the nature of your research and may be used for audit).

	Please tick
Interview	
Questionnaire	X
Analysis of existing records	
Participant performs verbal/paper and pencil/computer based task	X
Measurement/recording of motor behaviour	X
Audio recording of participant	
Video recording or photography of participant	
Physiological recording from participant	
Participant observation	
Covert observation	
Systematic observation	
Observation of specific organisational practices	

Other (please specify)	

SECTION D

Have you read one or more of the following professional guidelines and do you undertake to use the principles listed there as a guide for your own work? Please note that this is not intended to be an exhaustive list. Links to the guidelines listed below are included on the CUREC website.

	Please tick
British Society of Criminology: Code of Ethics for Researchers in the Field of Criminology [www.britsoccrim.org/codeofethics.htm]	
British Educational Research Association Ethical Guidelines for Educational Research [www.bera.ac.uk/guidelines]	x
Academy of Management's Code of Ethics [www.aomonline.org/aom.asp?ID=&page_ID=242]	
Association of American Geographers Statement on Professional Ethics [www.aag.org/cs/resolutions/ethics]	
Oral History Society of the UK Ethical Guidelines [http://www.oralhistory.org.uk/index.php]	
American Political Science Association (APSA) Guide to Professional Ethics in Political Science (Section H) [www.apsanet.org/content_9350.cfm]	
Political Studies Association Guide to Good Professional Conduct [www.psa.ac.uk/psa]	
British Psychological Society Code of Ethics and Conduct [www.bps.org.uk/what-we-do/ethics-standards/ethics-standards]	
Ethics Guidelines of the Association of Social Anthropologists of the UK and Commonwealth [www.theasa.org/ethics/guidelines.shtml]	
Social Research Association: Ethical Guidelines [http://the-sra.org.uk/sra_resources/research-ethics/ethics-guidelines/]	
Statement of Principles of Ethical Research Practice from the Socio-Legal Studies Association [www.slsa.ac.uk/content/view/247/270/]	
Statement of Ethical Practice for the British Sociological Association [www.britsoc.co.uk/about/equality/statement-of-ethical-practice.aspx]	
Other professional guidelines (please specify):	

SECTION E

Please put a tick in the yes/no column as appropriate to indicate your response.

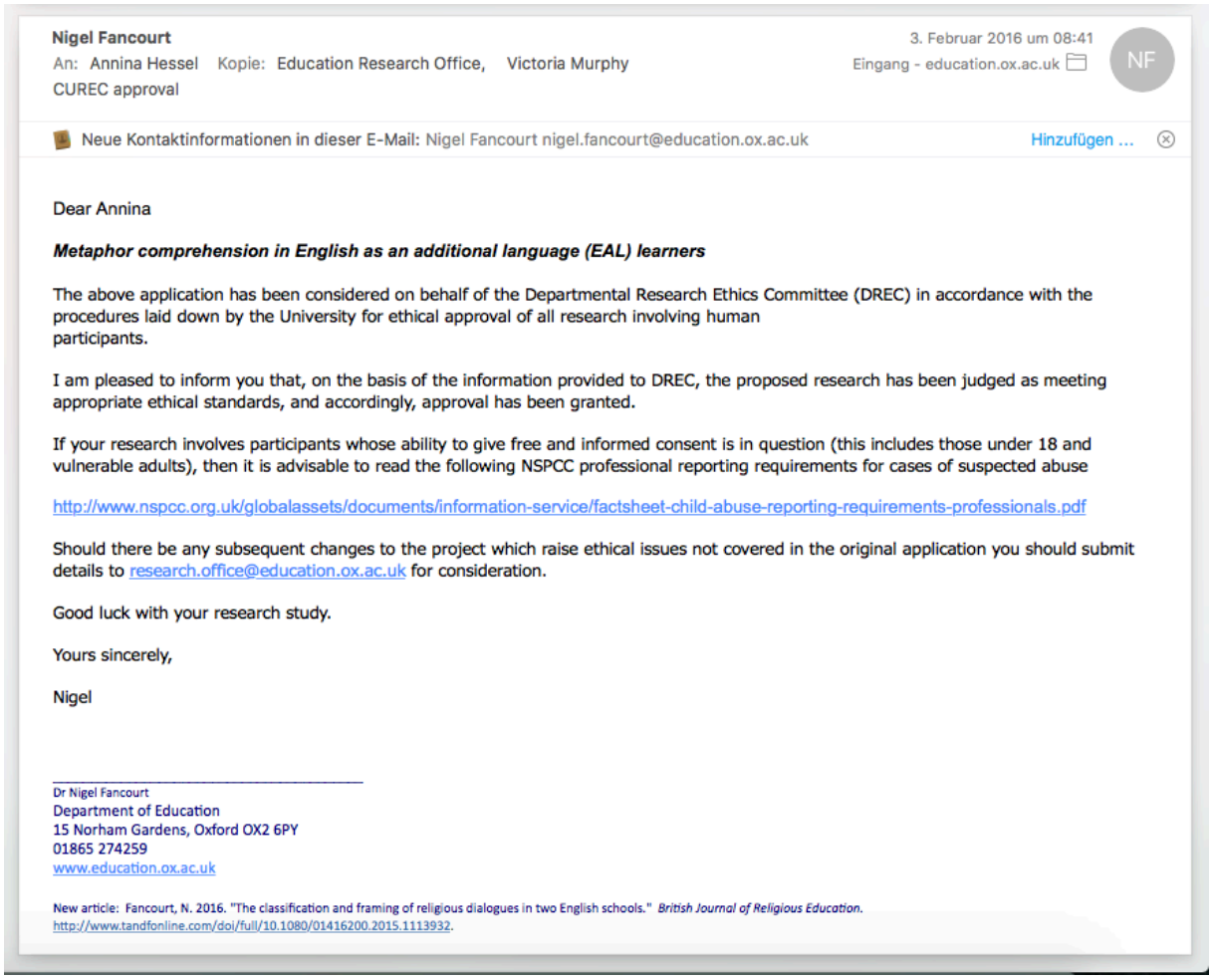
1) Will you obtain informed consent according to good practice in your discipline before participation?	**Yes**	**No**
	X	
2) Will you ensure that *personal data collected directly from participants or via a *third party is held and processed in accordance with the provisions of the Data Protection Act?	**Yes**	**No**
	X	
3) Does the research involve as participants *people whose ability to give free and	**Yes**	**No**

informed consent is in question? ***Note: participants aged under 16 are generally considered to require consent of a parent or guardian (only answer 'no' to this question if you can cite one of the protocols listed under 'children'). For participants aged 16-17, consult FAQ C13***	**x**	
4) As a consequence of taking part in the research, will participants be at serious risk of rendering themselves liable to criminal prosecution (e.g. by providing information on drug abuse or child abuse)?	**Yes**	**No**
		x
5) Does the research involve the *deception of participants, as part of the investigation/experiment?	**Yes**	**No**
		x
If any of your answers above are in a shaded box, please indicate whether those aspects of your project are fully covered by the following.		
6) Research protocol(s) which has/ve received IDREC/CUREC approval?	**Yes**	**No**
If yes, please give protocol number(s): **DREC_2014_P25**	**x**	
7) Professional guidelines that you will be following, as noted under Section D?	**Yes**	**No**
I will be following the *British Educational Research Association Ethical Guidelines for Educational Research.*	**x**	

If any of your answers in Section E are in a shaded box and are not covered by a protocol or by professional guidelines, please complete CUREC/2, available to download from the CUREC website. Then submit both this form (you need not complete section F) and the CUREC/2 to the Social Sciences and Humanities IDREC.

If all your answers in Section E are in the unshaded boxes or your answers in shaded boxes are covered by a protocol or professional guidelines, complete Section F and submit this form and any accompanying documents to the Social Sciences and Humanities IDREC or to the relevant officer/committee at departmental level (see notes and address below).

Appendix C. Ethical Clearance from Central University Research Ethics Committee



Nigel Fancourt
An: Annina Hessel Kopie: Education Research Office, Victoria Murphy
CUREC approval

3. Februar 2016 um 08:41
Eingang - education.ox.ac.uk

NF

Neue Kontaktinformationen in dieser E-Mail: Nigel Fancourt nigel.fancourt@education.ox.ac.uk Hinzufügen ...

Dear Annina

Metaphor comprehension in English as an additional language (EAL) learners

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse

http://www.nspcc.org.uk/globalassets/documents/information-service/factsheet-child-abuse-reporting-requirements-professionals.pdf

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Yours sincerely,

Nigel

Dr Nigel Fancourt
Department of Education
15 Norham Gardens, Oxford OX2 6PY
01865 274259
www.education.ox.ac.uk

New article: Fancourt, N. 2016. "The classification and framing of religious dialogues in two English schools." *British Journal of Religious Education.*
http://www.tandfonline.com/doi/full/10.1080/01416200.2015.1113932.

Appendix D. Recruitment letters Experiment 1

Main Flyer

WHAT SHOULD YOU DO NEXT?

If you would like to discuss the project, please do not hesitate to contact me:

Annina Hessel
Department of Education
University of Oxford
15 Norham Gardens
OX2 6PY
annina.hessel @education.ox.ac.uk

Phone: 07802883835

If now or at any time you have any concerns about this research and you are not satisfied with the answers I have given you, you can contact my supervisor, Professor Victoria Murphy: victoria.murphy@education.ox.ac.uk

If you are happy for your child to take part, please tick the box on the slip at the bottom of this page, fill in your child's name and return the slip to your child's teacher or their Jacari tutor. If you would like a general summary of the results of the project, please put your email or postal address on the slip at the bottom of this page and return it to your child's teacher or their Jacari tutor.

Yours sincerely,
Annina Hessel

☐----------

I would like to have a summary of the results of this project. My email or postal address is:

==========

==========

☐ I am happy for my child to participate in this project. My child's name is

---

Signed by parent or guardian:

_____

Name of parent/ guardian: ----------

Date: ----------

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024
Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk
www.education.ox.ac.uk

Director: Professor Jo Anne Baird

To learn to read is to light a fire



A research project by

University of Oxford in collaboration with Jacari



LEARNING TO READ PROJECT

My name is Annina Hessel. I am investigating how children learning English at school and an additional language at home. In the past, I have successfully worked with a number of schools in studies in Germany and the UK. The current Learning to Read Project is part of my PhD research I am conducting at the University of Oxford Departments of Education and Experimental Psychology together with Professor Victoria Murphy and Professor Kate Nation.

In partnership with the University of Oxford and Jacari, your child's school has agreed to take part in my project on children's language learning. I would like to invite your child to participate in my project. Before you decide whether you are happy with this, please read this letter carefully, and ask me any questions you would like.

WHAT IS THIS PROJECT TRYING TO FIND OUT?

We think that children learn to read at different paces, and that this learning is influenced by additional languages they speak.

For example, eight-year-old children might learn to read well in general, but they might have problems with reading some types of figurative phrases (such as *to learn to read is to light a fire*) because they do not speak English at home. While a lot of research has been carried out on reading figurative language in adults with one language, not much research has been done on how young children learning an additional language read figurative phrases.

WHY AM I INVITING YOUR CHILD TO TAKE PART?

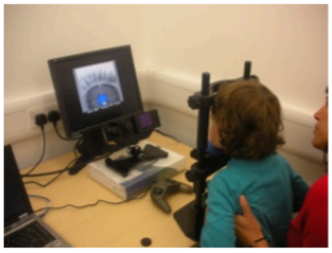
I am inviting two groups of eight-year-old children to take part in this research: one group of children who speak English at school and at home with their parents, family and friends, and another group of children who speak English at school and an additional language at home. This is so I can see whether there are differences between the groups in their reading.

I am inviting your child to take part because they are 8 years old, and because they speak either English at school and at home, or English at school and an additional language at home.

DOES YOUR CHILD HAVE TO TAKE PART?

Your child does not have to take part in this project. If you or your child decide against taking part, they will not be at any educational disadvantage.

You can even change your minds later: your child can stop participating at any time, without having to give me a reason – you just need to let me know.



Holly and her daughter, using a camera that films her daughter's eye movements.

WHAT WILL HAPPEN IF YOUR CHILD DOES TAKE PART?

If you choose for your child to take part, I will come to the school and do some individual activities with them, in a quiet room. Before working with your child, I will ask them whether they are happy to do the activities. I will let them know that they can stop the activities at any time.

I and a research assistant will meet with your child for less than 60 minutes. Your child will do five short tasks with us: a vocabulary task, a word reading task, a reasoning task, a memory task, and a story reading task. During their reading of the stories, their eye movements will be filmed with a little camera like the one in the picture to the left. This is to observe how they read in real time.

If at any time your child gets tired or loses interest, I will stop the activities and arrange to meet them again at another time.

WHAT WILL HAPPEN TO THE RESULTS OF THE PROJECT?

All the information I collect will be kept strictly confidential. I will give each child a code number, and only this code number will be written on the task papers. I will keep the list of children's names and numbers in a password-protected computer file, and I will erase this file at the end of the project. When I have finished the project, I will give a general summary of findings about all children to the school, and you are welcome to have this summary too. However, I cannot give you or anyone else any information about your child's individual results.

Who is supervising this project?
The project is supervised by Professor Victoria Murphy. The project has received ethics clearance through the University of Oxford's ethical approval process (Central University Research Ethics Committee).

.

Simplified Handout

Learning to Read Study



I am Annina Hessel. I am a student at the University of Oxford. I am doing a project together with your child's school & Jacari.

The study looks at how children learn to read.

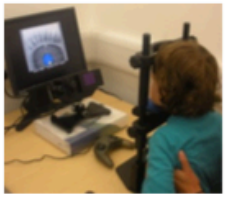


I am inviting your child to take part in my project.

What will your child do?

They will...

- Name pictures
- Put pictures in the right order
- Read words
- Memorise words
- Read stories & answer questions



When they read stories, their eye movements will be filmed using a little camera.

How long will it take?

Less than 60 minutes





Where and when?

In their school school during school hours.



All information that I collect will be kept confidential.

What if my child wants to stop?

Your child can stop at any time they want.



Questions?

Please feel free to call me and ask questions:
07802883835

What to do now?

If you are happy for your child to be in my project, please ....

1. Sign sheet below
2. Cut it out
3. give it to your Jacari tutor or child's teacher.

1  2 3 



Your signature: _____

Your name: _____

Your child's full name: _____

Date: _____
(Day/Month/Year)

Appendix E. Language Background Questionnaire Experiment 1



To learn to read is to light a fire

QUESTIONNAIRE FOR PARENTS

UNIVERSITY OF OXFORD

Jacari time to teach

LEARNING TO READ PROJECT

My name is Annina Hessel. In partnership with the University of Oxford and Jacari, your child's school has agreed to take part in my project on children's language learning.

I am trained as a school teacher and language researcher, and I am carrying out this project as a part of my PhD research in Education.

WHAT IS THE PROJECT TRYING TO FIND OUT?

We think that children learn to read at different paces, and that this learning is influenced by additional languages they speak.

To understand how reading and learning an additional language are linked we also ask families about the languages they speak.

WHY AM I INVITING YOU TO TAKE PART?

You as a parent know best how often you and your child speak English and how often you speak an additional language.

Your answers will be invaluable for us to understand what languages your child learns.

WHAT WILL HAPPEN IF YOU TAKE PART?

On the following pages, you will find questions such as "How often do you speak English with your child?" and you will choose your answer between 0- never and 4 – always.

It will take you 15 minutes to answer all questions. Your answers will be anonymised and will only be used for the *Learning to Read* project.

WHO CAN I ASK IF I HAVE QUESTIONS?

If you would like to discuss the project, please do not hesitate to contact me (annina.hessel@education.ox.ac.uk, 07802883835). If now or at any time you have any concerns about this research and you are not satisfied with the answers I have given you, you can contact my supervisor, Professor Victoria Murphy: victoria.murphy@education.ox.ac.uk

LEARNING TO READ- QUESTIONNNAIRE FOR PARENTS

☐ I agree to take part in the questionnaire for parents for the study *To learn to read is to light a fire.*

Name of your child: _____

You are the : ☐ Mother ☐ Father

Date: _____

Name of interpreter (if present): _____

I would like to know how often you and your child speak English and how often you speak your home language, that is, the other language you speak most often at home besides English. For example, if you speak Bengali at home and your children learn Arabic to read the Quran, your home language is Bengali. If you speak Urdu at home and you and your child sometimes watch Punjabi movies or communicate in Urdu with Punjabi speakers, your home language is Urdu but you also count the time spent watching Punjabi movies when asked: 'How often does your child watch movies in your home language?' For each question, please tick the appropriate answer, for example by choosing an answer from "never" (0) to "always (4), or "yes" or "no", and provide numbers (for example number of siblings) where needed.

English						**Your Home Language**					
When did your child start learning English?	At age _____.					When did your child start learning your home language?	At age _____.				
When did you child first live in the UK?	At age _____.					What is you home language, i.e. what language do you speak most at home besides English?	We speak _____.				
When did you child first go to a school in the UK?	At age _____.					Does your child learn to read and write in your home language?	Yes.				
Language use at home *Please choose the appropriate answer from "never" (0) to "always" (4).*	NEVER				ALWAYS		NEVER				ALWAYS
How often do you speak English with your child?	0	1	2	3	4	How often do you speak Your home language with your child?	0	1	2	3	4
How often does your child speak English with you?	0	1	2	3	4	How often does your child speak your home language with you?	0	1	2	3	4
How often does their father or mother speak English with your child?	0	1	2	3	4	How often does their father or mother speak your home language with your child?	0	1	2	3	4
How often does your child speak English with their father or mother?	0	1	2	3	4	How often does your child speak your home language with their father or mother?	0	1	2	3	4

Are there other adult relatives in the home? For example, a grandmother?	Yes. *Please answer questions in grey below.*				No. *Please go to questions marked with an *.*						
How many adult relatives live at home?											
Are one of these adults your child's primary caregiver?	Yes.				No.						
	NEVER				ALWAYS		NEVER				ALWAYS
How often do(es) the adult relative(s) speak English with your child?	0	1	2	3	4	How often do(es) the adult relative(s) speak your home language with your child?	0	1	2	3	4
How often does your child speak English with the adult relative(s)?	0	1	2	3	4	How often does your child speak your home language with the adult relative(s)?	0	1	2	3	4
* *Continue here.*						* *Continue here.*					
Does your child have brothers or sisters?	Yes. *Please answer questions in grey below.*				No. *Please go to questions marked with an *.*						
Their first sibling is ...	Older				Younger	First sibling					
	NEVER				ALWAYS		NEVER				ALWAYS
How often does their first sibling speak English with your child?	0	1	2	3	4	How often does their first sibling speak your home language with your child?	0	1	2	3	4
How often does your child speak English with their first sibling?	0	1	2	3	4	How often does your child speak your home language with their first sibling?	0	1	2	3	4
How often do you speak English with their first sibling?	0	1	2	3	4	How often do you speak your home language with their first sibling?	0	1	2	3	4
*(In case your child has only one sibling, please continue at the *.)*											
Their second sibling is ...	Older				Younger	Second sibling					
How often does their second sibling speak English with your child?	0	1	2	3	4	How often does their second sibling speak your home language with your child?	0	1	2	3	4

Your child's abilities *Please rate the following statements from "completely disagree" (0) to "fully agree" (4).* *If you're struggling to decide, please circle "I don't know."*	COMPLETELY DISAGREE				COMPLETELY AGREE		COMPLETELY DISAGREE				COMPLETELY AGREE
Compared with other children of the same age, your child expresses themselves very well in English.	0	1	2	3	4	Compared with other children of the same age, your child expresses themselves very well in your home language.	0	1	2	3	4
	I don't know.						I don't know.				
Compared with other children of the same age, your child pronounces words very well in English.	0	1	2	3	4	Compared with other children of the same age, your child pronounces words very well in your home language.	0	1	2	3	4
	I don't know.						I don't know.				
Is it easy for your family or friends to have a conversation with your child in English.	0	1	2	3	4	Is it easy for your family or friends to have a conversation with your child in your home language.	0	1	2	3	4
	I don't know.						I don't know.				
Compared with other children of the same age, your child has difficulty producing correct sentences in English.	0	1	2	3	4	Compared with other children of the same age, your child has difficulty producing correct sentences in your home language.	0	1	2	3	4
	I don't know.						I don't know.				
Your child speaks English like the children in England.	0	1	2	3	4	Your child speaks your home language like the children in the country where your home language is spoken.	0	1	2	3	4
	I don't know.						I don't know.				

Additional sheet

If your child has more than two siblings, please answer the questions below for each additional brother or sister:

Their third sibling is ...	Older			Younger		Third sibling					
How often does their third sibling speak English with your child?	0	1	2	3	4	How often does their third sibling speak your home language with your child?	0	1	2	3	4
How often does your child speak English with their third sibling?	0	1	2	3	4	How often does your child speak your home language with their third sibling?	0	1	2	3	4
How often do you speak English with their third sibling?	0	1	2	3	4	How often do you speak your home language with their third sibling?	0	1	2	3	4
Their fourth sibling is ...	Older			Younger		Fourth sibling					
How often does their fourth sibling speak English with your child?	0	1	2	3	4	How often does their fourth sibling speak your home language with your child?	0	1	2	3	4
How often does your child speak English with their fourth sibling?	0	1	2	3	4	How often does your child speak your home language with their fourth sibling?	0	1	2	3	4
How often do you speak English with their fourth sibling?	0	1	2	3	4	How often do you speak your home language with their fourth sibling?	0	1	2	3	4
Their fifth sibling is ...	Older			Younger		Fifth sibling					
How often does their fifth sibling speak English with your child?	0	1	2	3	4	How often does their fifth sibling speak your home language with your child?	0	1	2	3	4
How often does your child speak English with their fifth sibling?	0	1	2	3	4	How often does your child speak your home language with their fifth sibling?	0	1	2	3	4
How often do you speak English with their fifth sibling?	0	1	2	3	4	How often do you speak your home language with their fifth sibling?	0	1	2	3	4
Their sixth sibling is ...	Older			Younger		Sixth sibling					
How often does their sixth sibling speak English with your child?	0	1	2	3	4	How often does their sixth sibling speak your home language with your child?	0	1	2	3	4
How often does your child speak English with their sixth sibling?	0	1	2	3	4	How often does your child speak your home language with their sixth sibling?	0	1	2	3	4
How often do you speak English with their sixth sibling?	0	1	2	3	4	How often do you speak your home language with their sixth sibling?	0	1	2	3	4

Appendix F. Stimuli of Experiment 1

Verbal stories:

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
hours fly by	birds fly by	Sarah is a bit bored so she goes outside to feed the pigeons in the back garden. There are always a lot of pigeons in the garden. They have always really liked the seeds she feeds them. Sarah is happy to see all the pigeons! However, the hours / birds fly by so she decides to go inside for tea.	Why does Sarah decide to go inside?	Because it is soon tea.	Because the pigeons leave.	Where does Sarah go to feed the pigeons?	In the back garden.	To the playground.
song keeps playing in his mind	song keeps playing in his room	Tom doesn't like maths and he doesn't know how to do	Why can't Tom concentrate?	Because he keeps thinking	Because there is music playing.	What kind of homework is Tom trying to do?	His maths homework.	His biology homework.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		today's homework. His dad explains it to him. Tom's dad talks for a long time. Tom wants to listen but a song keeps playing in his mind / room and he can't concentrate so when his dad asks him to finish the homework, Tom does not know how to do it.		of a melody.				
minutes go by	friends go by	Amy wants to go outside to play. She asks her mother 'Mum, can I go out?' Her mum says 'Yes, you can go out, but come back for dinner when my mates arrive'. Amy happily goes out	Why is Amy late for dinner?	Because she did not notice how long she was playing for.	Because she did not notice that her mum's mates had arrived.	Where does Amy want to go to play?	Outside.	In the living room.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		to play. Amy does not realise how many minutes / friends go by and when Amy comes home, she is late for dinner.						
Billy passes his illness to Julie	Billy passes his sweets to Julie	Billy is sick and Julie goes to visit him. Billy has many delicious lollies. However, he is not allowed to eat any today because he is unwell and his mum thinks eating lollies can make you sick. While Julie is at Billy's, Billy passes his illness / sweets to Julie and she feels too poorly to go to school the next day.	Why does Julie feel to poorly to go to school?	Because she got Billy's illness.	Because she had Billy's lollies.	Where does Julie meet Billy?	At Billy's.	At school.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
He falls asleep	He falls down	Oliver has been playing outside all day long. When his mum calls him inside for dinner, he can barely keep his eyes open and he almost trips over his feet several times. When he is told to sit down at the dinner table and wait for everyone, he falls asleep/down and his mum has to carry him to bed.	Why does Oliver's mum carry him to bed?	Because Oliver was napping.	Because Oliver tripped over.	How long has Oliver played outside for?	All day long	During the afternoon,
he passes on his idea to Susie	he passes on his gift to Susie	It is Andy's birthday today. His parents have bought him a lot of presents. Susie is Andy's best friend. It is also Susie's birthday. But Susie's parents	What does Andy do to make his mother send a present to Susie?	He tells his mother that he wants to share his present.	He gives his mother the present he wants to give to Susie.	How many presents did Andy get?	He got a lot.	He got very few.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		can't buy her any presents. Andy thinks he could give some of his presents to Susie. He passes on his idea / gift to his mother and Andy's mother sends a present to Susie.						
don't let your spirit drop	don't let your gaze drop	Sarah plays in a football match. Her team is close to winning. However, she becomes tired, loses her motivation and fails to keep looking at the goal. Her coach sees this and when she is in front of the goal, he shouts: 'Don't let your spirit / gaze drop and score the goal!'	What did the coach remind Sarah to do?	To keep up her attention	To keep looking at the goal.	What sport does Sarah play?	She plays football.	She plays rugby.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		And then, Sarah did it.						
Linda’s argument collapses	Linda’s building collapses	Linda and her friend love technology classes. For weeks now, they have been programming a robot. They want it to build a Lego tower but it doesn't work. When the tower falls down again, Linda tells her friend that it would be good idea to try making a house instead. However, Linda's argument / building collapses and she rebuilds the tower.	Why does Linda rebuild the tower?	Because her idea is nonsense.	Because her house doesn't stand.	Where do Linda and her friend programme the robot?	In a technology class.	At her friend's place.
time drips	water drips	It is a really hot summer day and Tom hates the	Why does Tom get out of the pool?	Because after a while,	Because the pool	What does Tom think about the heat?	He hates it.	He loves it.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		heat. His mum says: ‘I will put water in the paddling pool so you can sit in there and cool down’. Tom goes to sit in his paddling pool. He really likes being cool. However, time / water drips and he gets out of the pool.		Tom gets bored.	is leaking.			
the things her mum wanted have disappeared from her mind	the things her mum wanted have disappeared from the store	Lucy’s mum asks Lucy to buy milk and bread from the nearby grocery shop. On the way to the shop, Lucy sees her friends. Lucy and her friends stop for a while to talk. When Lucy arrives at the shop, the things that her mum wanted	Why does Lucy return without buying anything?	Because Lucy forgot what her mum wanted her to buy.	Because the shop did not have what her mum wanted her to buy.	What does Lucy's mum ask her to buy?	Milk and bread.	Sugar and salt.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		have disappeared from her mind / the store and Lucy returns home without buying anything.						
the time hurries	his mum hurries	John got a new book for his birthday. Yesterday, he started reading a great story about dragons. He really wants to know how the story finishes. He sits down to finish it before lunch. While John is reading the time / his mum hurries and when his mother calls him for lunch, he has not finished his story.	Why does John not finish his story before lunch?	Because it is soon lunch.	Because his mother is cooking quickly.	What is the story about that John is reading?	About dragons.	About whales.
the lesson limps along	his friend limps along	Sam wants to go see his grandma after school. Sam	Why can't Sam wait for school to finish?	Because he is	Because his friend cannot	When will Sam go to see his grandma?	After school.	On the weekend.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		does not get to see her very often and is keen on seeing her soon. Normally, he really likes PE class because he gets to run around a lot with his mate. Today however, the lesson / friend limps along and Sam can't wait for school to finish.		getting bored.	run with him.			
to put more ideas into fewer lines	to put more words into fewer lines	Danny has to write a story on a sheet of paper to put on the classroom wall. However, he has problems finishing the story on only one sheet because he comes up with so many things and his handwriting is too big. His	Why can Danny finally write the story on only one sheet?	Because he says more things in fewer words.	Because he writes more letters in less space.	Where does Danny's teacher want to put his story?	On the classroom wall.	In the newspaper.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		teacher tells him to put more ideas / words into fewer lines and then Danny can do it.						
he never diverged from his goals	he never diverged from his track	Jim loves running. He trains three times a week to learn to run faster and faster. This Sunday, there is an important race that he and his team run in. Jim runs the final for his team and he finishes first. His trainer says he never diverged from his goals / track and that this is why he came in first.	Why does Jim's trainer think he has won?	Because Jim always kept his aims in mind.	Because Jim always kept in his lane.	What day is the race on?	On Sunday.	On Wednesday.
they run out of time	they run out of roses	Lisa and Andrew picked many beautiful flowers. Everyone wants	Why do Lisa and Andrew have to close their shop?	Because it is getting late.	Because they had no flowers left.	Where do Lisa and Andrew set up their shop?	On the street.	In their room.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		their pretty flowers. Lisa and Andrew have the idea to set up a little shop and sell their flowers on the street. Many people come to buy flowers from Lisa and Andrew. But they run out of time / roses and then they have to close their shop.						
she shoulders too many tasks	she shoulders too many toys	Samantha's class is organising a game day for the whole school. Samantha is very excited because she really likes games. She helps wherever she can, making sure everyone is doing their tasks and carrying playthings to	Why is Samantha too tired to play herself?	Because she has taken on too many duties.	Because she has carried too many things.	What is Samantha's class organising?	A game day.	A Halloween party.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		where they are needed. But she shoulders too many tasks / toys and in the end she is too tired to play herself.						
shake off this fear	shake off this spider	Mary is afraid of insects. When she's playing with other children in the little forest near her house, she suddenly notices a black little something crawling over her arm. Mary gets frightened and starts crying. An older girl who has seen her comes over to Mary and says: 'You should shake off this fear / spider and you will feel better.'	What does the older girl suggest Mary should do to feel better?	Free herself from her fright.	Get rid of the insect.	Where are Mary and the other children playing?	In a forest.	At a lake.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
grandpa’s pains go away	grandpa’s animals go away	Nick's grandpa takes care of homeless cats and dogs. Normally, he also plays a lot with Nick. Recently, however, he couldn't play with Nick. Nick's grandpa says it's because of his knee problems, but Nick's mum thinks his grandpa doesn't have time because of the cats and dogs. Soon, grandpa's pains / animals go away and he can play with Nick again.	Why can Nick's grandpa eventually play again with Nick?	Because his knee problems stopped.	Because his cats and dogs aren't with him anymore.	Why does Nick's grandpa take care of cats and dogs?	Because they are homeless.	Because they are sick.
follow the advice	follow the signs	Amy's class are doing a scavenger hunt in the forest. Amy's group is	How did Amy's group manage to finish second?	Because they did what the other	Because they used the clues on the trees.	Where is Amy's class doing the scavenger hunt?	In the forest.	In the mountains.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		last so they ask other groups for help. One group tells them to use clues on the tree trunks but another group tells them that it's quicker to ignore the clues and just go north. Amy's group decides to follow the advice / signs and they finish second.		group suggested.				
struck by his honesty	struck by the balls	Nathan wants to join the Cricket team. Whoever wants to join has to hit well. Nathan thinks he can do it easily and doesn't pay attention. The captain decides to have Nathan play as batsman and then tell him if he's good or	Why does Nathan promise to train more?	Because the team captain's open words impress him.	Because he is hit by the team captain's throws.	What sport does Nathan want to do?	Cricket.	Hockey.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		not. Nathan is struck by his honesty / the balls and he promises to train more.						

Nominal stories:

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
they are in the same boat	they are in the same group	Tom and his friends have handed in stories copied from a book to a writing competition. All children in the competition read their stories to judges. After Tom has read his, the judge gets angry because he recognises the book from other stories he has heard. Tom and his friends realise that they are in	What makes Tom sigh?	That he and his friends have similar problems now.	That he and his friends have read their story to the same judge.	What competition do Tom and friends take part in?	A writing competition.	A music competition.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		the same boat / group and Tom sighs.						
Jane is the queen-bee	Jane is the music-lover	Jane is playing with her friends. Her friends like to draw but Jane thinks that playing the guitar is much more fun. Because of Jane's comment, everyone starts playing the guitar and singing along. Jane's mum sees what is happening and thinks how Jane is the queen-	Why does Jane's mum smile?	Because Jane is the leader of her friend circle	Because Jane likes playing the guitar so much.	What do Jane's friends like to do?	They like to draw.	They like to dance.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		bee / music-lover of the group and she smiles to herself.						
Andrew is over the moon	Andrew is over the line	Andrew plays in a rugby match and the score is tied. Andrew is running towards the goal and he throws. He would be happy if he scored the	Why does Andrew scream?	Because he is really happy that he has scored the goal.	Because he is sad that he has run out of the field of play.	What match is Andrew playing in?	A rugby match.	A football match.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		goal, but if he is out of the field of play the goal will not count. After he has thrown Andrew is over the moon / line and he screams.						
Jack's face is beetroot	Jack's face is bruised	Jack is in his PE class. Today, they are learning how to do high jump. When it is Jack's turn, he jumps and lands off the mat. The teacher is worried that Jack might have hurt himself but the other	Why does Jack run out of the class?	Because he is embarrassed.	Because he hurt himself.	What is Jack learning in class today?	How to do high jump.	How to do a handstand.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		children laugh. When he gets up, Jack's face is a beetroot / bruised and he runs out of the class.						
the new boy is a bad apple	the new boy is a bad person	Jamie and his mates are playing cards. A new boy joins them. The new one does not like losing and starts cheating. Because the other boys don’t want to lose either, they start cheating, too. Jamie thinks that the guy is a bad apple /	Why does Jamie decide to stop playing with the new boy again?	Because the boy makes the whole group cheat.	Because the boy wouldn't be a good mate.	What are the boys playing?	Card games.	Board games.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		person and he decides that he will never play with him again.						
her thoughts are rockets	her thoughts are creative	Tina has to write a story as homework. She sits down and starts thinking about what could happen in her story. At first she really does not know what to write about. Suddenly, however, she has a lot of ideas and she starts writing really	Why is Tina pleased?	Because her ideas come very quickly.	Because her ideas are so great.	What does Tina have to do for her homework?	She has to write a story.	She has to draw a picture.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		quickly. She thinks that her thoughts are rockets / creative and she is very pleased.						
grades are lotteries	grades are misleading	Vicki is really nervous about last week's maths test. When she gets the test back, Vicki has a bad grade so she thinks she is really bad at maths. At home, she tells her mum that she thinks	What is it that makes Vicki feel better about herself?	She understands that marks can be random.	She sees that she has probably misunderstood the meaning of the mark.	What was the test in?	Maths.	Physics.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		she is useless at maths. Her mum says that sometimes, grades are lotteries / misleading and Vicki feels better about herself.						
the stores are jungles	the stores are bustling	David goes shopping with his family. They go to a shopping mall and his mum says 'You can go look at video games while I will look at books'. David loves video games so he	Why does David take long to find his mum?	Because the shops are confusing.	Because the shops are very busy.	What does David want to look at?	Video Games.	Books.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		wanders off. Later, he wants to join his mum again. However, the stores are jungles / bustling and he takes a long time to find his mum.						
her friends' smiles are razors	her friends' smiles are faked	Laura likes painting her cat and she wants to show her paintings to her friends. However, Laura is afraid that they will make fun of her or not tell her what they really think. She	Why does Laura feel sad?	Because they are unkind to her.	Because they are not honest with her.	What does Laura like to paint?	Her cat.	Her dog.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		finally decides to show the paintings to her friends. She asks them if they like them. Her friends' smiles are razors / faked and Laura is sad.						
books are lights	books are hard	Daniel wants to know how planes fly. He asks his dad to tell him. His dad says: 'Read this book. I want to see if you can understand by yourself. Then come tell me if you do'. Daniel reads	What will Daniel tell his dad?	That you can learn a lot through reading.	That reading is difficult.	What does Daniel want to know?	How planes fly.	How big planes are.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		all there is in the book about planes. He finds that books are lights / hard and goes to tell his dad.						
her words are daggers	her words are friendly	Sam likes Anna. One day, he tells her during the lunch break. If she is happy about it, he would invite her to go to the cinema together. Sam tells her and waits for her answer. When she responds, her words are daggers / friendly and	Will Sam invite Anna to go to the cinema tomorrow?	No, he won't because what she said hurt him a lot.	Yes, he will because what she said was nice.	When does Sam talk to Anna?	During the lunch break-	In the morning.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		then the bell rings and the two have to go back to class.						
their homework is a headache	their homework is a pleasure	Hannah and her friend got a difficult physics paper to do for next week. They keep trying all afternoon and they read books and websites, but they can't figure it out. Hannah decides to ask her mother if she can help because she has studied physics.	Why do Hannah and her friend sigh?	Because Hannah's mum agrees that their physics paper is hard.	Because Hannah's mum thinks that the paper is fun.	What do Hanna and her friend read to try and solve the physics paper?	Books and websites.	Journal articles.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		Hannah's mother says that their homework is a headache / pleasure and Hannah sighs.						
their stomachs are barrels	their stomachs are stuffed	Jamie and her sister Emma are at their grandma's. They love their grandma's cake. They have both eaten one piece already but they would like another one. Their grandma says: 'You're only allowed another piece if you	Have Jamie and Emma eat more cake?	Yes, they will because their stomachs are very large.	No, they won't because their stomachs are full.	Where are they eating cake?	At their grandma's.	At their uncle's.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		have enough space for it in your bellies'. Jamie looks at Emma and says that their stomachs are barrels / stuffed and Emma nods sheepishly.						
her mother is a rainmaker	her mother is considerate	Cindy and her parents had just sat down for dinner when the doorbell rings. It is their neighbours who have had a power cut and are in the dark! Her mum goes to help them. Using a few tools	Why is Cindy's dad proud of her mother?	Because her repairs have produced excellent results.	Because she was kind with their neighbours.	When do the neighbours knock?	When Cindy and parents sit down for dinner.	When Cindy gets ready to go to school.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		and an old manual, she fixes the problem. Cindy's dad says that her mum is a rainmaker / considerate and he looks proud.						
the next week will have to be a spring	the next week will have to be perfect	Claudia and her friends have been rehearsing for months to perform at the school festival. One day, the head teacher tells them that for organisational reasons, the festival will take place one week earlier than	Will Claudia and her friends be able to perform at the school festival? What does the teacher think?	That they will have to work quickly.	That they will have to be very good.	How long have Claudia and her friends been rehearsing for?	For months.	For weeks.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		planned. He asks if they still want to perform. The teacher says that the next week will have to be a sprint / perfect and everyone agrees.						
the discussion is a wrestle	the discussion is a success	Holly shares a bike with her sister. Unfortunately, they often want to ride their bike at the same time and then they get cross with each other. Their mother suggests they should find a way of sharing	Why do Holly and her sister stop discussing?	Because their conversation is so difficult.	Because they have managed to find a solution.	What does Holly share with her sister?	A bike.	A book.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		the bike so that each of them can use it when they want. The two sisters sit down to decide what they will do. Their discussion is a wrestle / success and they stop.						
the story is a babble	the story is a thriller	Adam has a new book. His sister wants him to read out a good chapter to her. He says he will first read the next chapter to himself to see if it is worth reading out loud. When	Why does Adam's sister frown?	Because the chapter is meaningless.	Because the chapter is scary.	Is the girl Adam's older or younger sister?	His younger sister.	His older sister.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		Adam is done, his sister asks him what he thinks. Adam says that the story is a babble / thriller and his sister frowns.						
the car is a giggle	the car is a Kombi	Sam and his friends are playing outside. Suddenly, they see a car coming down the street. The car stops and a man gets out of the car. Sam and his friends are still staring at the car. The man	Why does the driver grin so broadly?	Because Sebastian finds the car weird.	Because Sebastian has recognised the car type.	How is the car coming down the street?	It is coming at high speed.	It is coming very slowly.

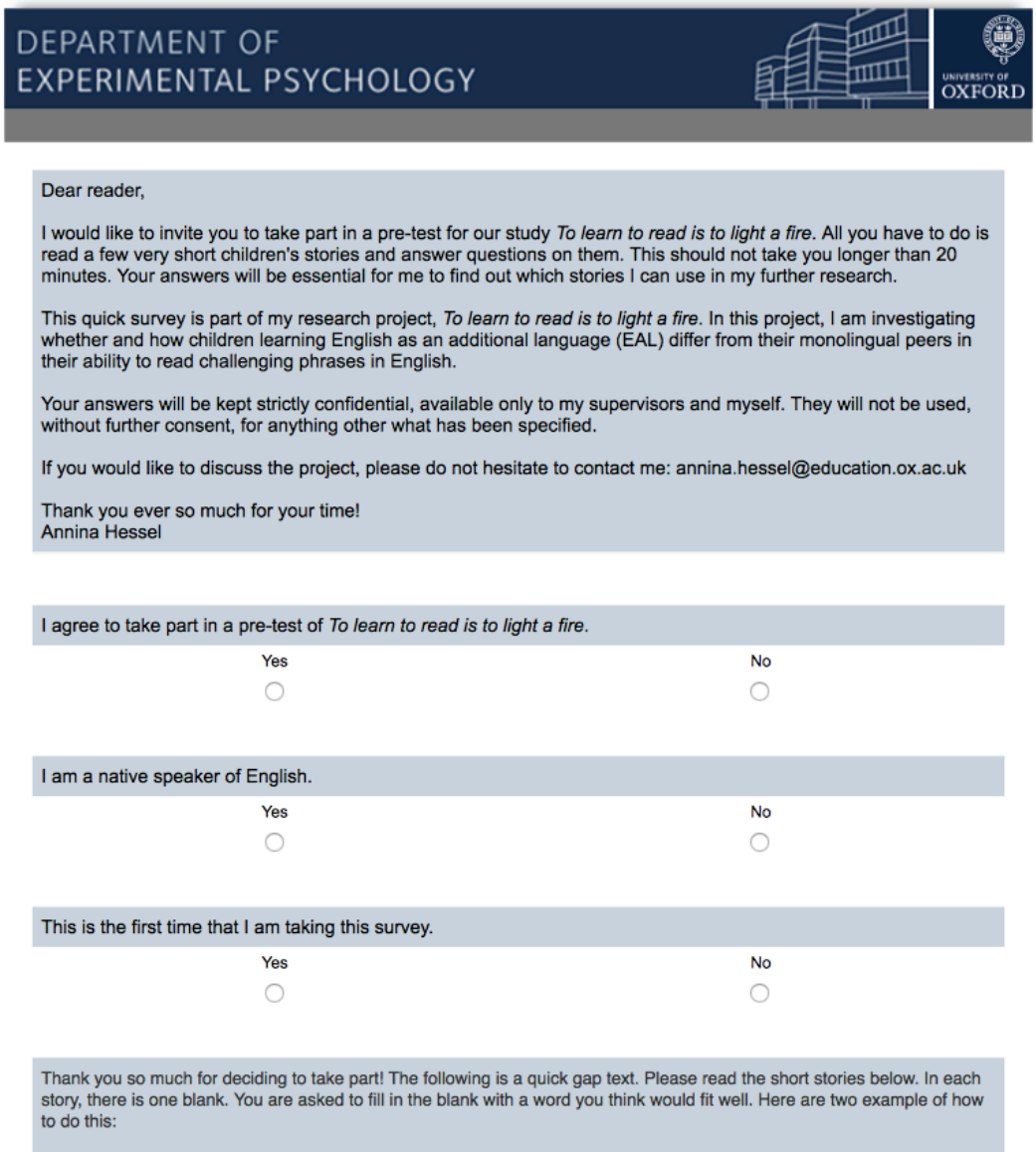
Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		asks why they are looking at his car. Sam says that the vehicle is a giggle / Kombi and the man smiles.						
the movie was a flop	the movie was a drama	Tonight, Fred's friends are going to a surprise showing at the cinema where you don't know what movie will be shown. Fred doesn't have enough money to go with them. He would be sad if it was a new action movie and	Why does Fred feel better?	Because the movie he missed was bad.	Because the movie he missed was not an action movie.	What showing are Fred's friends going to?	To the surprise showing.	To the late night showing.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		his friends saw it without him. The next day, his friends say that the film was a flop / drama and Fred feels better.						
the performance was a smash	the performance was a failure	Martin is organising an evening to earn money for a hospital. He has invited a theatre group to play. Martin knows that it is important that the people like the theatre group so that they give money for the	Why does Martin sigh?	Because people liked the group.	Because people didn't like the group.	What is Martin trying to earn money for?	The hospital.	The school.

Metaphor	Literal control	Text	Comprehension question 1	Correct answer for metaphors	Correct answer for literal controls	Comprehension question 2	Correct answer	Incorrect answer
		hospital. After the group has played, Martin overhears someone saying that the performance was a smash / failure and he sighs.						

Appendix G. Stimuli Norming Questionnaire

Screenshot of Online Survey



Example items from Norming Study.

Predictability through cloze-test

QID5 Thank you so much for deciding to take part! The following is a quick gap text. Please read the short stories below. In each story, there is one blank. You are

asked to fill in the blank with a word you think would fit well. Here are two example of how to do this: Example 1: Tom is playing with his dog. After a little while, Tom is _____ and he decides to go inside. Possible answer: 'tired' or 'bored' Example 2: Sarah and her friend are chatting about school. At some point, _____ says she has had problems with the last homework and the two friends decides to ask their teacher. Possible answer: 'her friend' or 'Sarah'*Note: If you need to insert an article or possessive (such as the, his or a) with your one word you insert in order to make the sentence work, please do so. There are no right or wrong answers for this gap filling task so please follow your intuitions on what word(s) you would expect to find in the gap. Please fill in the gaps in the stories below.

V1i Sarah is a bit bored so she goes outside to feed the pigeons in the back garden. There are always a lot of pigeons in the garden. They have always really liked the seeds she feeds them. Sarah is happy to see all the pigeons! However, the _____ fly by so she decides to go inside for lunch.

V2ii Tom doesn’t like maths and he doesn't know how to do today’s homework. His dad explains it to him. Tom’s dad talks for a long time. Tom wants to listen but a song keeps _____ in his mind and he can't concentrate so when his dad asks him to finish the homework, Tom does not know how to do it.

[….]

Naturalness through rating

Thank you so much for deciding to take part! The following is a quick naturalness rating. Naturalness refers to how normal an utterance sounds in the context it appears in, or how likely it is that a speaker might spontaneously express an idea in this particular manner. Please read the stories below and rate: How natural do the underlined phrases sound in the story context? Please indicate your decision by choosing a number between 1 and 6 (from 1 = " Very unnatural to 6 = " Very natural).

	1 - Very unnatural	2	3	4	5	6 - Very natural
Sarah is a bit bored so she goes outside to feed the pigeons in the back garden. There are always a lot of pigeons in the garden. They have always really liked the seeds she feeds them. Sarah is happy to see all the pigeons! However, the hours fly by so she decides to go inside for lunch.						
Tom doesn't like maths and he really doesn't know how to do today's homework. His dad explains it to him. Tom's dad talks for a long time. Tom wants to listen but a song keeps playing in his room and he can't concentrate so when his dad asks him to finish the homework, Tom does not know how to do it. (V2B) (2)	○	○	○	○	○	○

Transparency Rating

Almost done! The following is another brief judgement task that is unrelated but similar to the previous one. Do not be influenced by your naturalness ratings in the ratings you make for this task. In the examples below, we have provided you with the meaning of the underlined phrases. We would like you to assess how the word or phrase relate to the meaning, in other word you will be asked to rate how much the individual words and the overall meaning of the following phrases are related from 1- 'Totally unrelated' to 6- 'Very related'. Here are a few examples of how to do this: There are phrases which do not bear any relation to their meaning. For example, consider the phrase by and large Meaning: 'when everything about a situation is considered together'. One cannot find any meaningful relation between the meanings of the individual words by and large and the phrase meaning 'when everything about a situation is considered together'. You might want to score this phrase as 1 = 'totally unrelated'. On the other hand, there are phrases where the meaning relation between its constituent words and its overall meaning is clearer. For example, consider the phrased away with your fright Meaning: 'get rid of your fear'. One can find a clear link between do away and 'get rid of' as well as between fright and 'fear'. You might want to score this phrase as 6 = 'very related'. Also, consider the phrase reach the end of the road Meaning: 'for a situation to reach its end'. There seems to be a metaphorical relation between the literal meaning of the phrase reaching the end of the road and for a situation to reach its end. You might want to score this phrase as a 4 or 5. Of course, there are many other kinds of meaning relations between an idiom and its elements apart from those of the examples mentioned above. Please rate for the following stories: How much do you

think the meanings of words and the overall meaning of the underlined phrases below are related? Choose your response from 1- 'Totally unrelated' to 6- 'Very related'."

	1- Totally unrelated	2	4	4	5	6 - Very related
Sarah plays in a rugby match. Her team is close to winning. However, she becomes tired, loses her motivation and fails to keep looking at the goal. Her coach sees this and when she is in front of the goal, he shouts: 'Don't let your spirit drop and score the goal!' And then, Sarah did it. Meaning: To keep up her attention.	❍	❍	❍	❍	❍	❍
Oliver has been playing outside all day long. When his mum calls him inside for dinner, he can barely keep his eyes open and he almost trips over his feet several times. When he is told to sit down at the dinner table and wait for everyone, he falls down and his mum has to carry him to bed. Meaning: Oliver tripped over.	❍	❍	❍	❍	❍	❍

Appendix H. Adapted Materials for YARC in Group Testing

(YARC = York assessment of reading for comprehension) (Snowling et al., 2009)

Administration

The following oral instructions were adapted from the YARC manual given to the whole class in beginning of the testing:

"I will give you three short passages to read. Each of them has the questions at the end both on the side of the story and on its backside. The answers to the questions are in the passage, and you can look back at the passage when you are answering your questions. The passages are not long but please read them carefully. When you are not sure how to answer a question take a guess, but please give only one answer per question. First, please write your name on top of the page. You have 20 minutes to read and answer as much as you can."

Whenever children asked for help in answering the questions, they were told to "[L]ook in the text, all the answers are there".

Adapted materials (see following pages)

1 **Name:** _____

In Australia, the reptiles that are known worldwide as monitor lizards, are called goannas. Goannas range in length from 20 centimetres to over 2 metres, but all have the same distinctive shape: a flattened body, long neck with loose skin under the throat, strong legs with long toes and sharp claws, and a long tail.

Mostly they are ground dwellers, hunting close to the burrows in which they live, but they are also good tree climbers and strong swimmers. Being carnivores, they eat lizards, snakes, small mammals, bird, and eggs; often swallowing the animal whole. They hunt their prey by tracking and attacking it, or by using their sharp claws to excavate animals and eggs hidden in the ground.

Like most lizards, goannas lay eggs, usually in a nest or burrow and between seven to thirty-five eggs at a time. They hatch in eight to ten weeks.

Although bulky in appearance, goannas can run swiftly on two legs. They also rear up on two legs when threatened, inflating flaps of skin around their throats, hissing, and lashing out with their powerful tails.

Questions:

What is a goanna commonly known as?

What sort of homes do goannas live in?

Why would it be wrong to say that goannas stay only on or underground?

How do you know that goannas are carnivores?

Name one thing for which the goanna would use its long toes and claws?

How long do goanna eggs take to hatch?

When would a goanna hiss and inflate the skin around its throat?

How do you know that goannas are not slow, clumsy creatures?

2 **Name:** _____

The house was dark and no one was home. A burglar was in the bushes. He crept around to the back of the house and peered through the window. Carefully he broke the glass.
Once inside, he gazed at the contents of the room. He slipped some silver from the mantelpiece under his overcoat and found some jewellery in a drawer. Suddenly, he heard a key in the door. The burglar quickly squeezed under the large desk. A man walked in to the house. He put his briefcase down and went upstairs.
The burglar decided to get out quickly. As he crept towards the door he tripped over something. The dog was startled and began barking loudly. A light came on upstairs. Before the burglar could escape, he felt the man's steely grip on his shoulder. The man had a truncheon and handcuffs. Unfortunately for the burglar, he had broken into a policeman's house.

Questions:

Outside, where did the burglar hide?

How did the burglar get into the house?

What did he try to steal?

When did the burglar hide under the desk?

Who went upstairs?

Why did the owner come downstairs?

What does steely grip mean?

When do you think this story takes place?

3 **Name:** _____

Grandparents, aunts, uncles and cousins, spilled out of their cars to lay claim to the family's portion of the campsite. Uncle Hasan issued instructions to the pole bearers, but George ignored him. He hadn't wanted to come on this holiday in the first place. He threw down his pole, then, with head down and shoulders hunched, he slouched over to a low stone wall and slid down out of sight. He could hear mum giggling and saying "Grump-a-lot is off again!" He hated it when they teased him and used pet names like this. It was going to be a terrible holiday.
Looking up, George was startled to see a dog sprinting past him, trailing a line of sausages; and, in hot pursuit, a little, grey haired, old woman, brandishing a knife. "Gran?"
Instinctively he lunged forward and caught the dog's collar; the Gran grabbed the sausages and skilfully flicked the knife. George released his grip, allowing the dog to flee with a single sausage.
Everyone cheered as Grandpa hoisted George onto his shoulders and proclaimed him the Breakfast Hero. George grinned. He realised his judgement had been premature.

Questions:

What do you think Uncle Hassan was telling the other members of the family to do?

Where did George hide?

In this story, what does 'pet name' mean?

What did mum think about George throwing down the pole and slouching off?

Who put George onto his shoulder?

Why was George called the Breakfast Hero?

At the end of the story, how did George feel about the holiday?

In the context of the story, what does 'premature' mean?

Appendix I. Pre-registration document for Experiments 2 and 3

(uploaded on Open Science Framework on 10/01/2018)

Pre-registration EAL reading and vocabulary

Reading between the lines: How do English as an additional language learners process metaphors and semantic inconsistencies when reading?

Table of Contents

Study Information 1

Title **1**

Authorship **1**

Research questions **1**

Hypotheses by research questions **3**

Description of Sampling and State of Analyses 6

Explanation of existing data **6**

Data collection procedures **6**

Sample size **6**

Sample size rationale **7**

Variables 7

Manipulated variables **7**

Measured variables **7**

Indices **9**

Design Plan 11

Study Type **11**

Blinding **11**

Study design **11**

Randomization **12**

Analysis Plans 12

Statistical models **12**

Transformations **15**

Follow-up analyses **15**

Inference criteria **16**

Data exclusion **16**

Missing data **16**

Exploratory analyses **16**

References 19

Appendix 21

Experiment 2: Changes from Experiment 1 ..... **21**

Experiment 1 design and results **22**

Design Experiment 1 22

Research questions Experiment 1 22

Study Information

Title

Reading between the lines: How do English as an additional language learners process metaphors and semantic inconsistencies when reading?

Authorship

Annina K. Hessel
Victoria A. Murphy
Kate Nation

Research questions

The present document presents the research questions and predictions for two experiments in a series of three experiments. The first experiment has already been conducted and analysed (see Appendix on p.22 for summary of study and results), and the data of the other two experiments have been collected, and will be analyses as outlined here. The two experiments discussed in this document comprised standardized measures of reading and language, and two experiments in which eye movements were monitored as children read sentences. One experiment (Experiment 2) investigated the reading and comprehension of metaphors, while the other experiment (Experiment 3) investigated the reading of semantic inconsistencies. The following research questions are listed according to which of those measures served as outcome variables.

Differences in language skills between EAL and monolingual children

Children's language profiles will be established using vocabulary size and reading fluency, as outlined the section Measured Variables (see p. 7). Vocabulary and reading fluency will be further used as predictors for reading behaviour and comprehension in later research questions (see Research Question 3, 4 & 6).

RQ 1. Do EAL and monolingual children differ significantly in their vocabulary size, but not in their word reading fluency?

Experiment 2 (Metaphors)

The metaphor task in the current experiment was adapted from our previously conducted experiment, referred to as Experiment 1 in this report (see Appendix for adaptations made, p. 21). Related to the metaphor task, we asked the following research questions:

RQ 2. Do children read and comprehend nominal metaphors (e.g. *grades are lotteries*) in story context differently from matched control phrases (e.g. *grades are misleading*)? In other words, are there online and offline effects of metaphorness?

RQ 3. Do the offline effects of metaphorness (accuracy in answering questions) vary by individual differences (vocabulary size, word reading fluency) and group differences (EAL vs. monolingual)?

RQ 4. Do the online effects of metaphorness vary by individual differences (vocabulary size, word reading fluency) and group differences (EAL vs. monolingual)?

Experiment 3 (Semantic inconsistency)

Alongside Experiment 2, we decided to add Experiment 3 where we manipulated semantic consistency, comparing words that are consistent (e.g., a barking puppy) or inconsistent (e.g., a barking kitten) with the story context and readers' world knowledge. Stimuli were adapted[1] from Connor et al.'s experiment which showed different reading patterns (as revealed by eye movements) in 10-year-old readers of differing language abilities (Connor et al., 2015)[2]. The children's reading of metaphors in Experiment 2 are compared to their reading of semantic inconsistencies in Experiment 3. This comparison allows for the examination of whether the individual and group differences observed in Experiment 1 are specific to metaphor processing (for example, because EAL leaners lack the vocabulary knowledge or comprehension monitoring skills necessary to process metaphors), or if the effects are the similar on comparable discourse-level manipulations (here, on semantic inconsistency, which also requires noting an inconsistency, but relies more on world knowledge).

All children completed the Experiment 3 after they had finished Experiment 2. Related to Experiment 2, we asked the following research questions:

---

[1] Stimuli were kept as much as possible true to the stimuli used by Connor and colleagues, but adapted to fit the following criteria:

1) The stimuli from Connor and colleagues were adapted for use with scanpath by controlling the position of target words on the screen , and controlling the number of words and characters before (13 words, between 69-77 characters) and after the target word (6 words, between 27-42 characters).
2) Stimuli were judged by 3 British English native speakers for appropriateness to use in a British context, and words specific to American English were replaced with closely match British English alternatives (e.g. *nail* for *bolt*).

[2] Connor, C. M., Radach, R., Vorstius, C., Day, S. L., Mclean, L., Morrison, F. J., & Mcdonald, C. (2015). Individual Differences in Fifth Graders ' Literacy and Academic Language Predict Comprehension Monitoring Development : An Eye-Movement Study. Scientific Studies of Reading, 19(2), 114–134.

RQ 5. Do children read implausible words (e.g. a barking kitten) in story context differently from plausible words (e.g. a barking puppy)? In other words, is there an online effect of semantic (in)consistency?

RQ 6. Does the effect of inconsistency vary by individual differences (vocabulary size, word reading fluency) and group differences (EAL vs. monolingual)?

Hypotheses by research questions

The following hypotheses about the research questions asked in Experiment 2 and 3 are proposed based on results of a preceding experiment 1 (see above), as well as on relevant studies, which will be cited where applicable.

Verbal skills of EAL and monolingual children

RQ 1. Do EAL and monolingual children differ significantly in their vocabulary size, but not in their reading fluency?

Hypothesis 2 We predict that EAL children will have smaller vocabularies than their monolingual peers, but will not differ significantly from them on their reading fluency (cf. Oller, Pearson, & Cobo-Lewis, 2007; Paradis & Kirova, 2014, as well as results Experiment 1, see p. 24).

Experiment 2

RQ 2. Do children read and comprehend nominal metaphors (e.g. *grades are lotteries*) in story context differently from matched control phrases (e.g. *grades are misleading*)? In other words, are there online and offline effects of metaphorness?

Hypothesis 2 Based on the findings from Experiment 1, we predict that children will read nominal metaphors more slowly overall (total time) than literal controls, and that those effects will show particularly for the second word of metaphors, and the post-metaphor region (see Indices on p.9 for definition of target areas). We expect that the overall slow-down on metaphors will be due to increased re-reading in the metaphor condition, that is, a higher probability to regress back into the first word of a metaphor, and to reread the second word of a metaphor, as well as the region after the metaphor. We do not predict differences between metaphor and control phrases on early reading measure (gaze duration).

Based on the findings from Experiment 1, we further predict that children aged 9-10 answer questions about metaphors less accurately than questions about literal phrases.

RQ 3. Do offline effects of metaphorness (accuracy in answering questions) vary by language group (EAL vs. monolingual) and individual differences (vocabulary size, reading fluency)?

Hypothesis 2 Based on the results of Experiment 1, EAL and monolingual children are not expected to differ in their accuracy on comprehension questions about nominal metaphors compared to literal phrases (but see predictions made for individual differences in Hypothesis 3b. However, the null effect from the previous experiment may not replicate in Experiment 2 since, given the older and larger sample and more controlled stimuli, the statistical power is likely to be higher in Experiment 2 than in Experiment 1. If, against the current predictions, the null effect from Experiment 1 fails to replicate in a higher-powered study, the following effects are plausible based on related research:

If the null effect of group differences found in Experiment 1 does not replicate, it is plausible to assume that the EAL readers would be particularly disadvantaged in comprehending metaphors. Based on related studies, EAL children can be expected to be significantly less accurate on metaphor comprehension than their monolingual peers (Burgoyne, Whiteley, & Hutchinson, 2013; Hessel, 2014; Martinez & Murphy, 2011).

Hypothesis 3 Based on the results of Experiment 1, it can further be expected that there will be an interaction between reading fluency and metaphorness indicating that with increased reading fluency, questions on metaphors will be answered as accurately as questions on literal phrases (thereby decreasing the predicted difference between metaphors and literals, see predictions made for RQ 2). It is furthermore predicted that larger vocabularies will be associated with better comprehension, particularly in the older sample of Experiment 2, where reading fluency is likely to be higher overall and children's vocabulary more important for their reading comprehension than in Experiment 1 (Babayigit & Stainthorp, 2014).

RQ 4. Do online effects of metaphorness (reading patterns) vary by language group (EAL vs. monolingual) individual differences (vocabulary size, word reading fluency)?

Hypothesis 4a. Based on the results from Experiment 1 (see Exp 1, RQ4 p. 21), we do not expect there to be an overall group difference on the online reading of metaphors compared to literal phrases. Instead, and following the exploratory findings from Experiment 1, we expect differences to show based on continuous measures of verbal skills (see Exp 1, RQ5, p.25). However, given the considerations about developmental effects and power outlined in RQ 3.Hypothesis 2, it might be that the results of Experiment 2 do not replicate those seen in Experiment 1.

To see an interaction effect between language groups and metaphorness on online reading would not be out of step with some previous findings. Previous offline studies have found EAL learners to struggle with integrating word meanings during text reading, and with repairing an interpretation failure (Burgoyne et al., 2013; Kan, 2014; Martinez & Murphy, 2011). In eye movements, later measures (particularly longer rereading durations and higher rereading and regression probabilities) indicate that readers engage in semantic integration, and monitoring and repairing interpretations (Pickering, Frisson, McElree, & Traxler, 2004). Hence, an eye movement record that would replicate the offline findings of less re-analysis in EAL readers would be if monolingual readers show larger differences between metaphor and literal phrases in later eye-tracking measures. Such eye movements, possibly together with stronger comprehension, would be a sign of stronger re-analysis in monolingual than in EAL readers.

Hypothesis 4b. Based on results from Experiment 1, children's individual differences are predicted to influence their metaphor reading. It is predicted that with increasing reading fluency, children will read more quickly and regress less overall. On metaphors, more fluent readers are expected to reread metaphors more quickly as a sign of higher processing ease. This effect is expected to be smaller in EAL children. Further differences are expected between language groups in the role that vocabulary plays for their metaphor processing for comprehension, particularly for the rereading of metaphors. Specifically, EAL readers are expected to show fewer signs of re-analysing metaphors when rereading than monolingual. Signs of metaphor re-analysis found in Experiment 1 were for example that higher vocabulary monolinguals reread metaphors increasingly quickly, indicating that a re-analysis was undertaken and got more efficient, while EAL learners reread metaphor at equal speed regardless of their vocabulary size, which indicates EAL reader may not have undertaken a semantic re-analysis at all (cf. findings in Ashby, Rayner, & Clifton, 2005).

Experiment 3

RQ 5. Do children read implausible words (e.g. a barking kitten) in story context differently from plausible words (e.g. a barking puppy)? In other words, is there an online effect of semantic (in)consistency?

Hypothesis 5. Based on previous findings (Connor et al., 2015), we predict that children will read inconsistent target words more slowly than literal counterparts (gaze duration as a sign of noticing inconsistencies, following Connor et al.).

RQ 6. Does the effect of inconsistency vary by and language group (EAL vs. monolingual) and individual differences (vocabulary size, word reading fluency)?

Hypothesis 6. Based on Connor and colleagues' results (2015), and consistent with findings from Experiment 1 on a comparable paradigm, higher verbal ability readers (larger reading fluency and vocabulary, and possibly also for monolinguals compared to EAL readers) are predicted to adapt their rereading behaviour more to the inconsistency difference than lower verbal ability readers. In line with Connor et al, the differences in rereading on the inconsistent targets are thought to evidence that higher but not lower ability readers try to resolve the inconsistency (longer rereading durations and higher rereading and regression probabilities on target word).

Description of Sampling and State of Analyses

Explanation of existing data

The data for the described analyses has already been collected. No summary statistics have yet been computed that are relevant to the research questions and hypotheses described above.

Data collection procedures

The data were collected in two Oxford primary schools. Children were tested individually. The data consists of standardized test results (vocabulary size, word reading fluency, and non-verbal reasoning), and experimental measures (two eye-tracking experiments, one with stories containing metaphors or control phrases, the other one with stories containing semantic inconsistencies). Stimuli in both Experiment 2 and 3 were presented following a Latin Square design, with 10 stimuli/ condition being read by each child.

Sample size

Sixty-four individual children aged 9;01-10;10 and enrolled in year 5 of UK primary schools took part in the study. 42.19% of the children were classified as EAL (n=27) and the others were monolingual learners (n=37). 46.88% of the children were female, with 13 girls in the EAL and 21 girls in the monolingual group.

Sample size rationale

We aimed to test as many children as was possible within the recruitment (opt-in procedure was followed in one school) and time constraints (three weeks of testing were possible in second school).

Variables

Manipulated variables

Experiment 2: Metaphor vs. literal phrases

Experiment 3: Semantically consistent versus consistent words

Measured variables

- Language group
  - Language group is established as a binary variable following the classification of the Department for Education guidelines (Department for Education, 2016). Every child was asked in a questionnaire-based interview if they spoke any other language at home or in their community. If they reported to be exposed to any other language but English, they were classified as EAL, if they reported to only be exposed to English, they were classified as monolingual.
- Vocabulary size

- Vocabulary size is measured using the British Picture Vocabulary Scale, Version 2 (Dunn, Dunn, Whetton, & Burley, 1997). For the individual differences analyses, vocabulary size is taken as a proxy of oral comprehension abilities (see Research Questions 4 and 6, p. 3 ff.).

The fact that we use the BPVS as the only standardized measure of comprehension contrasts with the use of three measures in Experiment 1 (vocabulary size, vocabulary depth, and reading comprehension). While there are theoretical reasons that motivate distinguishing different aspects of vocabulary and comprehension (Nation, 2001), Experiment 1 showed that all three measures we used in Experiment 1 converged on the same story about our sample (i.e., that there is a gap between monolingual and EAL learners in their comprehension, but not in their decoding skills). Furthermore, we were not able to use more than one comprehension measure as predictors for performance on the metaphor task because of high collinearity between them. In the interest of reducing testing time (see collinearity in results Experiment 1, Exp 1, RQ5, p.25), we thus decided to reduce the number of tests in experiment 1 to one test of vocabulary, which would also serve as a proxy for oral comprehension. We chose the BPVS since it is a widely-used test of receptive vocabulary size which is often as a proxy for comprehension skills in the field of EAL research. This makes it possible to compare our study to previous studies.

- Reading fluency
  - Reading fluency is measured using the Test of Word Reading Efficiency, Second Edition (Torgesen, Rashotte, & Wagner, 2012).
- English Language Input Quality and Quantity
  - The quality and quantity of English language input outside school were assessed through the EAL students' self-report on the amount of English they speak at home (quantity), and the different ways in which they encounter English, for example by speaking it with friends, or through reading and other media (quality) (Paradis, 2016).
- Online reading (both Experiment 2 and 3)
  - Online reading was measured in eye movements on the screen. Measures were calculated as outlined in the next section (see Indices).
- Offline comprehension in Experiment 2[3]

[3] Children answered comprehension questions for one third of all trials in Experiment 3 as well. However, questions in Experiment 3 did not actually check the comprehension of the semantic inconsistencies but were only included to encourage attentive reading, and children's answer accuracy on them is thus not included as an outcome measure.

- Offline comprehension is operationalised as 1 (correct answer) and 0 (incorrect answer), based on children's button press in answer to the comprehension questions in Experiment 2.

- Confidence about comprehension levels (both Experiment 2 and 3)
  - Confidence about comprehension levels is operationalised through the self-assessment that each reader completed after having finished Experiment 2 and 3. Children were asked to rate their comprehension on a four-point scale, which will be compared to their actual level of comprehension (operationalised as their accuracy in answering comprehension questions that followed each trial).

- Self-reported reading behaviour
  - Self-reported reading behaviour is again operationalised through the children's answers in the structured interview they completed after having finished both experiments (adapted for use with children from Hyönä & Nurminen, 2006). In this interview, children were asked to rate four statements about their own reading behaviour (e.g. "I reread words when I don't understand them") on a 3-point scale to answer if the statement sounded "Not like me", "A little like me", or "A lot like me". They were also given the option to answer "Don't know".

Indices

Online reading measures will be computed using the SR Research Dataviewer software, following conventions in the field. The following measures will be computed (Dataviewer names in brackets behind measure):

- Gaze duration (IA_FIRST_RUN_DWELL_TIME)
- Rereading duration (IA_DWELL_TIME – IA_FIRST_RUN_DWELL_TIME)
- Total time (IA_DWELL_TIME)
- Rereading probability (IA_DWELL_TIME – IA_FIRST_RUN_DWELL_TIME > 0 = 1, 0 =0)
- Regression in probability (IA_REGRESSION_IN)[4]

[4] Note that only regression in and rereading probability will be used for confirmatory analyses since no effects of individual differences or metaphorness were found for regression out probability in experiment 1. It is likely regression out probability is not an appropriate measure to see the effects in the current paradigm because of the large inter-personal variability in when the decision to regress back to previous context is made (von der Malsburg, Kliegl, & Vasishth, 2015).

Pre-registration EAL reading and vocabulary

Reading measures will be computed based on following interest areas:

Experiment 2- Metaphor experiment

	Word 1 of metaphor		**Word 2 of metaphor**	**Post-metaphor region**
Metaphor	/ grades/	/ are/	/ lotteries /	/ and she feels[5]/
Literal	/ grades/	/are/	/ misleading/	/ and she feels/

In order to see if results from Experiment 1 replicate (see *Results Experiment 1*, p. 23), the following reading measures were chosen to look at for each interest area:

Word 1 of metaphor - Regression in probability

Word 2 of metaphor - Gaze duration, rereading duration, total time, rereading probability, and regression in probability

Post-metaphor region - Gaze duration, rereading duration, total time, rereading probability, and regression in probability

Experiment 3 - Inconsistency task

		Pre-target word		**Target word**	**Post-target region**	
Inconsistent	/ Rover/	/ barks/	/ at all the passing animals on the	/kitten/	/in the whole [6]//	/neighbourhood in fact./

[5] Post-metaphor region is defined as three words out of the six words that separate every metaphor region from the end of the sentence, which is also the end of the story. The remaining three words are the final words of the story/ sentence, and are not included in the computation of this measure to avoid including wrap-up effects.

[6] As in the metaphor task, the post target region is defined as three words out of the six words that separate every target region from the end of the sentence, which is also the end of the story. The remaining three words are the final words of the story/ sentence, and are not included in the computation of this measure to avoid including wrap-up effects.

			street.He's the most alert /			
Consistent	/ Rover/	/ barks/	/ at all the passing animals on the street.He's the most alert /	/puppy/	/in the whole [7]//	/neighbourhood in fact./

In order to see if the results from Connor et al. replicate, and to see if similar effects show for semantic inconstancy as for metaphors in Experiment (see *Results Experiment 1*, p. 23, Connor et al., 2015), the following reading measures were chosen to look at for each interest area:

Pre-target word - Gaze duration, rereading duration, total time, rereading probability, and regression in probability

(In)consistent target word - Gaze duration, rereading duration, total time, rereading probability, and regression in probability

Post-target region - Gaze duration, rereading duration, total time, rereading probability, and regression in probability

Design Plan

Study Type

The study is experimental in manipulating text features such as metaphorness (Experiment 2) and inconsistency (Experiment 3), partly observational (in analysing the effect of existing individual and group differences).

Blinding

Children in the study were blind to the manipulation in the eye-tracking experiment. The first author however did all the testing and will run all analyses, who at any point has been aware of the study design and purpose.

Study design

[7] As in the metaphor task, the post target region is defined as three words out of the six words that separate every target region from the end of the sentence, which is also the end of the story. The remaining three words are the final words of the story/ sentence, and are not included in the computation of this measure to avoid including wrap-up effects.

Experiment 2 and 3 both have a mixed design with between-subject factors (individual and group differences) and within-subject factors (metaphor vs. literal in Experiment 2, and consistent vs. inconsistent in Experiment 3), as well as repeated measures (20 trials within each experiment). The observational part of the study is a cohort study in that it compares two existing groups, EAL and monolingual learners, and the effect of individual differences (vocabulary size and reading fluency) on performance.

Randomization

All stimuli were presented in randomized order within both experiments. Children always completed Experiment 2 before Experiment 3 because we feared that otherwise the reading of semantic inconsistencies in Experiment 3 might influence children's reading of metaphors in Experiment 2.

Analysis Plans

Statistical models

We will run t-tests to answer research question 1, and mixed effect models for all other analyses. For mixed models investigating confirmatory research questions, fixed effects will contain main effects that are central manipulations in the experiment (that is, metaphorness and language groups for the metaphor task; and inconsistency and language groups for the inconsistency task), and those variables and interactions that are predicted to be significant and meaningful. Two sets of models will be run for each outcome variable, one which will contain only group effects and the experimental manipulation and the interaction of the two, and a second model where we will also add individual differences and interactions to the fixed effects of the first model. For each research question, that means:

For research question 2, 3 and 4, which are about the metaphor task, we will run two models for each outcome variables that we made predictions for. The outcome variables are offline comprehension accuracy and several online reading measures taken from the eye movement record (gaze duration, rereading time and probability, regression in and out probability, total time. These will be calculated across different regions of interest: the metaphor region 1 and 2, the post-metaphor region (see p.9 for definition of these variables, and which variables will be computed for which regions). For each outcome variable, the fixed effects will be defined as follows:

Model 1- Group effect (EAL vs. monolingual) and manipulation (metaphor vs. literal)

Outcome measure ~ LanguageGroup + Metaphorness + Metaphorness:LanguageGroup
Model 2- Group effects, manipulation, and individual differences:
Outcome measure ~ LanguageGroup + Metaphorness + Metaphorness:LanguageGroup
+Vocabulary*LanguageGroup*Metaphorness +
WordReadingFluency*LanguageGroup*Metaphorness

There are two important decisions that we made when deciding to build the models this way. Firstly, we decided to run two models, first with group, then also with individual effects included. Running the model in two steps means that we can test the prediction that differences will not show for language groups overall (null effect in model 1), but that they will show when individual differences are being taken into account (significant interactions of group with manipulations and individual differences in model 2). The first model thus allows to check how much variance is captured by the binary distinction EAL vs. monolingual; the second model will reveal whether any can be explained additionally by adding in estimates off individual differences. The second decision we made was to enter individual differences as interaction terms with both language group and experimental manipulation. We chose complex interaction terms to see if we can replicate the results of Experiment 1, where we found main effects, as well as two and three-way interactions with individual differences.

For research question 5 and 8, which are about the inconsistency task, we will again run two models for each outcome variables that we made predictions for. The outcome variables are several reading measures, which are gaze duration, rereading time and probability, regression in and out probability, total time (for the target region, the post target region, and the pre-target region, see again p.9 for definition of those variables and regions). For each of those outcome variables, the fixed effects will be defined as follows (following the same logic as for the models on the metaphor task):
Model 1- Group effects (EAL vs. monolingual) and manipulation (consistent vs. inconsistent)
Outcome measure ~ LanguageGroup + Inconsistency + Inconsistency:LanguageGroup
Model 2- Group effects, manipulation, and individual differences:
Outcome measure ~ LanguageGroup + Inconsistency + Inconsistency:LanguageGroup
+Vocabulary*LanguageGroup*Inconsistency +
WordReadingFluency*LanguageGroup*Inconsistency

Model checks

Residuals will be checked for normality (inspecting QQ-plots and histograms of the residuals gained from initially fitted models) and heteroscedasticity. For linear models, this will be done using base R (Baayen, 2008), and for binomial models, using the DHARMa package (Hartig, 2016). The package influenceME checked whether any outliers had a large influence on the model estimates and significance values (Nieuwenhuis, Pelzer, & Grotenhuis, 2016). The influenceME package takes an existing MEM and reruns it several times, removing either one subject or one item from the model calculations each time. The beta estimates and significance values are then compared across all of these leave-one-out simulations. Large differences of beta and/or significance values between the original model and a leave-one-out simulation suggest that an item or child had a large influence on the model outcome. When a model has a large number of influential cases (e.g. between 25-50% of all cases), influential cases will be inspected to see what defined those influential cases (e.g., to see if the estimate changes strongly if subjects with large vocabularies are not included in the model). If the model estimates does not change much across leave-one-out simulations, the model will be considered reliable across observations. The outcomes of the leave-one-out simulations will be taken into account when evaluating the model reliability, and discussing the results.

When a model fails to converge, several checks will be run to identify and solve the reason of non-convergence. Singularity will be checked to see if the random structure should be simplified. If singularity will be a problem, the random structure will be reduced, starting with the least explanatory[8] and most highly correlated higher-order term in the random structure[9]. If singularity is not a problem, the package numderiv will be used to improve the precision of the gradient estimate, and to restart the model from the more precise start value (Gilbert & Varadhan, 2016). If a more precise gradient estimation does not solve the problem either, the package optimx will be used to run the model using different optimizers than the default bobyqa (Nash, 2015). If despite all these efforts, the model will not converge, it will be considered not converged, and its results not considered further. On convergence, the final model will be evaluated for model fit against the criteria described above.

[8] For example, the in a list of random effects that explain a variance of either "0.21" (random 1), "0.39" (random 2), or "0.003" (random 3), the random effect 3 would be considered to have the smallest explanatory power, and would be a first candidate to be taken out.
[9] For example, if there was a correlation of "1" between the random main effect for metaphorness and the metaphorness-metaphor type interaction nested within subject, the

While we are aiming to describe our decisions for data modelling a priori, we want to acknowledge that in many cases, it will be impossible to predict all eventual outcomes of data modelling, and all future decisions that will need to made to adapt to challenges in the data. We hope, however, that the guidelines outlined above will provide a clear insight into our a priori plans. In case unexpected adaptations to the plan will be necessary, we will acknowledge that in our reporting of results when and where appropriate.

Transformations

Eye-tracking data was manually pre-processed before extraction to correct for gaze drift. Gaze drift was defined as systematic a deviation of the recorded fixations from the line of letters, for example due to head tilting of the recorded child, or poor calibration of the eye-tracking. The existence of gaze drift in a trial was judged based on notes from the time of time collection together with visual judgement of the recorded reading patterns. If we decided that there was a gaze drift in the trial, fixations were corrected by up to 10 pixels on the vertical axis to move fixations back onto the line of letters. In a final step, long and short fixations were deleted as is convention in eye-tracking studies (see Data exclusion p. 16) before extracting the data from Dataviewer.

All predictor variables will be centered (by subtracting the mean) and rescaled (dividing by 2 standard deviations), and all reading duration measures will be converted from milliseconds to seconds in order to support model convergence. Contrasts for binary predictors will be coded using ANOVA contrast coding so that the intercept represents the overall mean, and the beta estimates the distance of each level of measurement (e.g. being EAL) to it. Data will be transformed where necessary based on model fit (e.g. log-transformations or taking the square root) using the BoxCox test (cf. Keshtiari & Vasishth, 2013).

Follow-up analyses

To follow up from interaction effects, we will inspect data visually, look at the effects descriptively, or run models on a subset of the data, where appropriate.

---

interaction effect would be taken out first, since it is the higher-order term of the two highly inter-correlated random effects.

Inference criteria

We will use a two-tailed criterion (t or z ≤ 1.96 SE) for all inference making, corresponding to a 5% Type 1 error cut-off (Vorstius, Radach, Mayer, & Lonigan, 2013). Given that eye-tracking data analysis requires running multiple models for different outcome measures (different eye-tracking measures to account for early and late processes of reading, and multiple areas of interest to account for phrase-level effects), the significance criterion will be adapted for multiple comparisons using a Bonferroni correction to reduce the risk of type 1 error (von der Malsburg & Angele, 2015).

Data exclusion

Following conventions in the literature, fixations shorter than 80ms or longer than 1200ms will be discarded from analyses. Outliers will not be tested for or removed prior to running the models, but influential cases on model predictions will be checked using the influence.ME package (Nieuwenhuis et al., 2016).

Missing data

Partially missing data (e.g. in case of temporary tracker loss within a trial, or missing trials within a child) were left as missing data, and the trial or participant remained in the dataset.

Exploratory analyses

Linking online reading to offline comprehension in Experiment 2

The link between online reading and offline comprehension will be explored in Experiment 2 through the exploratory research questions 1 and 2 below. We will also investigate the possibility that relationships are not linear (see comments on non-linear modelling in the section on Statistical Models, p. 12).

Expl.RQ1. Is children's rereading of target phrases linked to their ability to subsequently answer questions about those target phrases correctly? In other words, is there an effect of online reading on offline comprehension?

Expl.RQ2. Does the effect of online reading on offline comprehension vary by individual and group differences, and for metaphors compared to control phrases?

Linking effects in Experiment 2 to effects in Experiment 3

Expl.RQ3. Are the online effects of metaphorness and inconsistency associated (cf. effects predicted in RQ 2 and RQ 5)? In other words, is there a link between children's sensitivity to semantic manipulations of plausibility and their sensitivity to figurativeness?

Expl.RQ4. Is the link between the online reading of metaphors and implausibilities (see **Error! Reference source not found.**) less strong in EAL children? More specifically, can we see a particular difficulty of EAL learners with metaphor processing, as opposed to processing semantic inconsistencies?

Expl.RQ5. Are children confident that they understand figurative vocabulary, and the semantic inconsistency in the stories? Are EAL learners more confident than monolingual children of their comprehension, despite lower accuracy scores, particularly on metaphors? (Martinez & Murphy, 2011)

Expl.RQ6. Are quality and quantity of English and home language input outside of school related to EAL learners' performance on the standardised and experimental tasks in the experiment? (Paradis, 2016). In other words, does extending the binary distinction between EAL and monolingual children capture more variance in our measures of reading comprehension and processing?

Expl.RQ7. Does a non-linear model describe the links between individual differences, online reading, and offline comprehension better than a linear mixed effect model?

Reading monitoring

Expl.RQ8. Are children confident that they understand figurative vocabulary, and the semantic inconsistency in the stories?

Expl.RQ9. For Experiment 2: Is there a mismatch between children's actual and self-perceived comprehension?

Expl.RQ10. Are children able to describe their reading patterns accurately, indicating awareness of reading styles?

Expl.RQ11. Does children's ability to monitor their reading (accuracy in judging comprehension cf. Expl.RQ9, and awareness of reading styles, cf. Expl.RQ10) vary by individual differences (vocabulary size, word reading fluency) and language groups (EAL vs. monolingual)?

Scanpaths in Experiment 2 and 3

Expl.RQ12. Does analysing children's scanpaths using the package Scacism (von der Malsburg & Vasishth, 2011) rather than using traditional eye movement analyses reveal a clearer picture of the rereading on metaphors and semantic inconsistencies, and how they link to group and individual differences, and to comprehension?

More specifically, the following research questions would be asked:

Expl.RQ10a. Do scanpaths triggered by semantic manipulations (metaphors and semantic inconsistency) differ between EAL and monolingual readers, and higher and lower vocabulary readers? In other words, it is possible to identify qualitatively different scanpaths on semantic manipulations for different reader groups?

Expl.RQ10b. Are the scanpaths from RQ 1 differently associated with comprehension success? In other words, is it possible to identify scanpaths that are linked to more accurate comprehension?

Expl.RQ10c. Does the link between scanpaths and comprehension success (see RQ 3) interact with readers' individual and group differences? In other words, are there scanpaths that are successful in some readers, but not in others? (Kaakinen & Hyönä, 2007)

References

Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2013). The role of background knowledge in text comprehension for children learning English as an additional language. *Journal of Research in Reading*, *36*(2), 132–148. http://doi.org/10.1111/j.1467-9817.2011.01493.x

Connor, C. M., Radach, R., Vorstius, C., Day, S. L., Mclean, L., Morrison, F. J., & Mcdonald, C. (2015). Individual Differences in Fifth Graders ' Literacy and Academic Language Predict Comprehension Monitoring Development : An Eye-Movement Study. *Scientific Studies of Reading*, *19*(2), 114–134. http://doi.org/10.1080/10888438.2014.943905

Department for Education. (2016). *School census 2016 to 2017*.

Dunn, L. M., Dunn, D. M., Whetton, C., & Burley, J. (1997). *The British picture vocabulary scale II*. *BPVS II* (Second Edi). Windsor: NFER-Nelson.

Hessel, A. K. (2014, July). *Exploring Metaphor Comprehension in UK Primary School Children with English as Their First or Additional Language*. University of Oxford.

Hyönä, J., & Nurminen, A.-M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology (London, England : 1953)*, *97*, 31–50. http://doi.org/10.1348/000712605X53678

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory*, *15*(6), 634–646. http://doi.org/10.1080/09658210701457096

Kan, R. (2014). *Effects of frequency and idiomaticity on second language reading comprehension in children*. University of Oxford.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990. http://doi.org/10.3758/s13428-012-0210-4

Martinez, R., & Murphy, V. a. (2011). Effect of Frequency and Idiomaticity on Second Language Reading Comprehension. *TESOL Quarterly*, *45*(2), 267–290. http://doi.org/10.5054/tq.2011.247708

Murphy, V. (2014). *Second language learning in the early school years: Trends and contexts*. Oxford: Oxford University Press.

Nieuwenhuis, R., Pelzer, B., & Grotenhuis, M. (2016). *Package " influence . ME ": Tools for Detecting Influential Data in Mixed Effects Models*.

Oller, K. D., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, *28*(2), 191–230. http://doi.org/10.1017/S0142716407070117.Profile

Paradis, J. (2016). Parent report data on input and experience reliably predict bilingual development and this is not trivial. *Bilingualism: Language and Cognition*, (March), No-Specified. http://doi.org/10.1017/S136672891600033X

Paradis, J., & Kirova, A. (2014). English second-language learners in preschool: Profile effects in their English abilities and the role of home language environment. *International Journal of Behavioral Development*, *38*(4), 342–349. http://doi.org/10.1177/0165025414530630

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*(2), 110–149. http://doi.org/10.1016/j.dr.2013.03.001

The British National Corpus, version 3 (BNC XML Edition). (2007). Accessed through Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium.

Torgesen, J. K. (2012). *TOWRE 2: Test of word reading efficiency*. (C. A. Rashotte, R. K. Wagner, & P. Pro-Ed (Firm), Eds.)*Test of word reading efficiency* (Second edition). Austin, Texas : Pro-Ed.

Von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*(2), 109–127. http://doi.org/10.1016/j.jml.2011.02.004

Wechsler, D. (2011). *WASI -II: Wechsler abbreviated scale of intelligence - second edition*. (D. Wechsler 1896-1981, Ed.)*WASI*. San Antonio, Tex.: Psychological Corporation.

Wiig, E. H., & Secord, W. (1992). *Test of word knowledge (TOWK): Complete kit*. Pearson.

Appendix

Experiment 2: Changes from Experiment 1

Experiment 2 was adapted from our previously conducted experiment that also tested metaphor reading and comprehension, referred to as Experiment 1 in this report (see Appendix, p. 21). The experiment was adapted in the following ways:

a) Stimuli were reduced to only nominal metaphors and their literal controls. This was decided for several reasons: The first goal was to reduce testing time so that other measures could be included alongside the metaphor task (such as the inconsistency task, see below). Another goal was to reduce the variance in children's reading times as far as it was caused by item heterogeneity (such as different lengths of interest areas between verbal and nominal metaphors in Experiment 1). When choosing which metaphor type to keep for Experiment 2, we decided to focus on nominal metaphors compared to literals since it was for those stimuli that we saw clearest differences in children's rereading patterns, also in interaction with individual differences. To further improve the nominal metaphor stimuli used in Experiment 1 for use in Experiment 2, stimuli were matched more closely within the metaphor-control pairs on their individual word frequency (The British National Corpus, version 3, 2007), and on the numbers of syllables and morphemes, and we made sure that all individual words contained in a metaphor had a maximum age of acquisition of 10 years (the average age of our participants) (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012).

b) We further decided to run Experiment 2 with an older and more age-constrained sample than the one tested in Experiment 1 (year 5 as opposed to year 3-4 students). Model checks in Experiment 1 suggested that individual subject's differences masked differences between language groups. It was thus decided to test older children to reduce the subject variance, a variance which in Experiment 1 is likely to have been caused by the fact that a proportion of readers in year 3-4 were still learning to decode.

c) To reduce testing time further, we decided to reduce the lengths of the context that preceded metaphors. Given that in Experiment 1, only one comprehension question was actually measuring a variable of interest (metaphor comprehension), while the second question was to ensure children were reading the whole story attentively, the reduction of the preceding context in the stimuli also meant that we could reduce the number of comprehension questions to one. Thus, every trial is followed by one comprehension question that measures the comprehension of the metaphor / control region.

d) Finally, we adapted all stimuli to make them amenable to scanpath analyses (See Exploratory Analyses, p. 16). An important prerequisite for scanpath analyses that stimuli are matched in their layout on the page. That is, we matched stories for the number of words and characters preceding a target region (40 words, between 220-228 characters), and for words and characters following the target word (6 words, between 30-35 characters). As target region, we here define the second word of the metaphor (see Indices on p.9)

Experiment 1 design and results

Hypotheses about the research questions asked in Experiment 2 are informed by results of Experiment 1 of the first author's doctorate, alongside cited literature. Experiment 1 preceded the current Experiment 2 and both follow a similar methodology, as described below.

Design Experiment 1

The children in Experiment 1 were 44 children aged 7;00-9;08 enrolled in year 3-4 of local UK primary school, approximately half of them EAL (n=21) and monolingual learners (n=23). 54.5% of the children were female, with 10 girls in the EAL and 14 girls in the monolingual group. We compared EAL and monolingual readers on two within-subject factors, metaphorness (metaphor vs. literal) and metaphor type (verbal vs. nominal). Stimuli were presented in two versions (metaphors vs. literal) and were counterbalanced on two lists using a Latin square design, with every child reading only one version of a story. There were 42 stories in total, 2 of which were practice trials and source for global reading measures, and 40 were experimental trials. Half of the stories contained verbal metaphors, which took the form (subject) noun phrase + verb or verb + (object) noun phrase, e.g. *hours fly by* or *shake off this fear*, and the other half contained nominal metaphors, which took the form A is B, e.g. *Jane is the queen-bee*, or their literal control phrase. Every child read 10 stories/ condition (e.g., 10 metaphorical verbal stories). Beyond the grouping into EAL and monolingual learners, we further used standardised tests to test children's individual differences in word reading fluency (Test Word Reading Efficiency Second Edition; Torgesen, 2012), vocabulary size and depth (Test of Word Knowledge, Expressive Vocabulary (size) and Multiple Contexts (depth) subtasks; Wiig & Secord, 1992), and non-verbal reasoning (Wechsler Abbreviated Scale of Intelligence matrices subtask; Wechsler, 2011). Results from exploratory analyses of the differential effect of word reading fluency and vocabulary on EAL and monolingual children's reading and comprehension of metaphors motivate central a number of predictions for Experiment 2.

Research questions Experiment 1

Exp 1, RQ1. Do EAL and monolingual children differ significantly in their comprehension-related skills (vocabulary size and depth, reading comprehension), but not in their reading fluency?

Exp 1, RQ2. Do children read and comprehend metaphors (e.g. *grades are lotteries*, or *hours fly by*) in story context differently from matched control phrases (e.g. *grades are misleading*, or *birds fly by*)? In other words, are there online and offline effects of metaphorness?

Exp 1, RQ3. Do children read and comprehend nominal metaphors vs. literals (e.g. *grades are lotteries*, or *misleading*) differently from verbal metaphors vs. literals (*hours* or *birds fly by*, and *follow the advice*, or *signs*)? In other words, does the metaphorness effect vary by the metaphor type?

Exp 1, RQ4. Do EAL and monolingual children differ in their reactions to metaphorness and metaphor type in their online reading and offline comprehension? In other words, does do the effects of metaphorness and metaphor type vary by language groups?

Exp 1, RQ5. Do reading fluency and vocabulary predict online and offline metaphorness and metaphor type effects, and are there differences between EAL and monolingual children?

Results Experiment 1

The following are the results from Experiment 1 that informed hypotheses for Experiment 2[10].

Exp 1, RQ1. Do EAL and monolingual children differ significantly in their comprehension-related skills (vocabulary size and depth, reading comprehension), but not in their reading fluency?

- EAL children had smaller and less deep vocabularies and weaker reading comprehension than their monolingual peers, but did not differ significantly from them on their reading fluency. This is in line with previous findings that the gap between EAL and monolingual learners is specific to vocabulary and comprehension related abilities, but not reading aloud and decoding (Oller et al., 2007).

[10] Note that part of the significant results reported here were only significant at a traditional two-tailed criterion (t or z ≤ 1.96 SE), and not when models were corrected for multiple comparisons. For the purpose of hypothesis generation, those results are however listed here alongside more reliable effects, and will be subjected to a replication test in the experiment 2 data.

Exp 1, RQ2. Do children read and comprehend metaphors (e.g. *grades are lotteries*, or *hours fly by*) in story context differently from matched control phrases (e.g. *grades are misleading*, or *birds fly by*)? In other words, are there online and offline effects of metaphorness?

- Children spent longer (total time on word 2 and whole metaphor region) reading metaphors than literal phrases, an effect that also showed in rereading time (regressions into word 1, rereading probability of word 2), and they also comprehended metaphors marginally less well than literal phrases.

Exp 1, RQ3. Do children read and comprehend nominal metaphors vs. literals (e.g. *grades are lotteries*, or *misleading*) differently from verbal metaphors vs. literals (*hours* or *birds fly by*, and *follow the advice*, or *signs*)? In other words, does the metaphorness effect vary by the metaphor type?

- The effect that children took longer to read metaphors compared to literal phrases was carried by nominal metaphors, both on the phrase itself (more regressions back into word 1, more rereading and marginally longer rereading as well as longer total time word 2) and in the immediate post-metaphor region (more rereadings). This was not true, however, for children's comprehension. Additionally, there was a non-predicted speed-up in the region following verbal metaphors compared to literal controls (shorter total time, less rereading).
- For Experiment 2, it was decided to drop verbal metaphors since children did not react to them differently than to their literal controls, beyond a speed-up following the phrase. Nominal metaphors, however, would be investigated further in Experiment 2 to see if the effects of nominal metaphors on increased re-analysis of the phrase and its immediately following area can be replicated.

Exp 1, RQ4. Do EAL and monolingual children differ in their reactions to metaphorness and metaphor type in their online reading and offline comprehension? In other words, does do the effects of metaphorness and metaphor type vary by language groups?

- EAL and monolingual children did not differ in their reading or comprehension of metaphors. Checks of influential data points suggested, however, that this null effect may have been due to large variability in the data, and to individual differences within language groups particularly.

For Experiment 2, it remains to be seen if the overall null effect of language group would replicate in a sample of older and more advanced readers, and if the same link to individual differences will show, as discussed in research question 5.

Exp 1, RQ5. Do reading fluency and vocabulary predict online and offline metaphorness and metaphor type effects, and are there differences between EAL and monolingual children?

- For research question 5, reading fluency and vocabulary size were investigated as predictors, additional to the effects investigated in research question 2-4. It was decided to focus on reading fluency and vocabulary since only two predictors could be entered for reasons of collinearity. Reading fluency and vocabulary were chosen both have been found to make distinct contributions of either to reading comprehension, while only vocabulary differed between language groups. It is thus plausible that comparing the effect of the two on the reading and comprehension of EAL and monolingual readers would show different, and meaningful results.

Before describing the results, it is important to note, however, that the following results stem from exploratory research, and warrant a confirmatory replication in a higher-powered study.

- With increasing reading fluency, all children read faster and were less likely to regress into earlier text parts (across measures and regions). Reading fluency furthermore interacted with metaphorness in that children with higher reading fluency were quicker to reread metaphors but not literals (whole metaphor region), and answered comprehension questions on metaphors as well as questions on literals (while children with lower reading fluency answered questions on metaphors less well than questions on literals). It appears that higher reading fluency facilitated readers' ease of text reading, which particularly supported the reading and comprehension of those more demanding phrases. Stronger reading fluency is thus associated with greater ease in text processing. Interactions with language groups suggest, however, that EAL readers benefit less from higher reading fluency for their phrase reading fluency (less steep decrease of rereading time on whole metaphor region), and that the link between higher reading fluency and longer rereading and better comprehension of metaphors is carried by monolinguals, but not EAL readers. Hence, the facilitation effect of reading fluency on phrase-level reading was stronger in the monolingual group, possibly because EAL learners were limited by smaller vocabularies limited beyond general decoding skills.

- As for reading fluency, larger vocabularies were associated with better comprehension, supporting the idea that vocabulary is important for successful text comprehension. The idea that differences in vocabulary knowledge explain particularly EAL reading patterns further was supported by three-way interactions between vocabulary, metaphorness, and language groups. The effects indicate that with increasing vocabulary, monolingual and EAL readers adjusted differently to metaphors and literal phrases. For example, EAL learners take less time to reread literal phrases with increasing vocabularies, while their rereading of metaphors is not affected by their vocabulary size at all. This is reversed for monolingual readers who take less time to reread metaphors with increasing vocabularies, while their rereading of literals is quick, and not affected by vocabulary. Previous research suggests that when faced with challenging phrases that surpass a readers' current comprehension skills, they may not engage in selective re-analysis of those challenging phrases, but adopt an incomplete interpretation of the phrase, and possibly rely more on context to improve their comprehension (Rayner et al., 2006). Different from monolingual readers, EAL readers may thus engage only in selective re-analysis of literal phrases, but fail to do so for metaphors.

Given that EAL and learners have considerably smaller vocabularies than their monolingual peers, they may still be at an earlier stage in their reading processing abilities where their processing on metaphors differs. Differences may arise because the children adapt dynamically to their available resources (Kaakinen & Hyönä, 2007; Rayner et al., 2006), or because of differences in their top-down regulation of reading standards of coherence (Kaakinen, Lehtola, & Paattilammi, 2015) or comprehension monitoring (Connor et al., 2015).

- Taken together, the effects of word reading fluency on metaphor processing suggest that reading fluency supports ease of text processing, which seemed particularly important for reading and understanding metaphors. Furthermore, the ability to notice and efficiently adapt seems to vary between EAL and monolingual learners, possibly because EAL learners are at an earlier stage of their vocabulary development. Since all these effects are still exploratory, they will form the basis of confirmatory analyses in Experiment 2 and 3.

Appendix J. Ethics application for Experiments 2 and 3

University of Oxford

CENTRAL UNIVERSITY RESEARCH ETHICS COMMITTEE (CUREC)

CUREC/1A Checklist for the Social Sciences and Humanities

The University of Oxford places a high value on the knowledge, expertise, and integrity of its members and their ability to conduct research to high standards of scholarship and ethics. The research ethics clearance procedures have been established to ensure that the University is meeting its obligations as a responsible institution. They start from the presumption that all members of the University will take their responsibilities and obligations seriously and will ensure that their research on human subjects is conducted according to the established principles and good practice in their fields and in accordance, where appropriate, with legal requirements. Since the requirements of research ethics review will vary from field to field and from project to project, the University accepts that different guidelines and procedures will be appropriate. Please check the CUREC website to ensure that you have the correct form for your project.

This form does not cover research governance, satisfactory methodology, compliance with the requirements of publishers when administering their tests or questionnaires, or the health and safety of employees and students. As principal investigator, it is your responsibility to ensure that requirements in these areas are met. Please carry out a risk assessment of the project, in consultation with all researchers involved, using the checklist and CUREC's other documentation.

The use of an asterisk in this form indicates a phrase defined in the glossary. The glossary and further information on the University's research ethics procedures are available from the CUREC website:

www.admin.ox.ac.uk/curec

This form is designed largely for research that falls within the Divisions of Social Sciences and Humanities and which does not involve a high level of risk to the subjects. Elite interviews, field work and oral history are included in the CUREC process. Please take a moment to read through it and if you have any questions or doubts as to whether it is the appropriate form, please review Section A or consult the CUREC website.

Note on anonymised data and audit: you do not need to obtain ethical approval for your study if:

- you are using previously collected **anonymised data** about people which neither you nor anyone else involved in your study can trace back to the individuals who provided them (e.g. census data, administrative data, secondary analysis). Please refer to the definition of *personal data in the glossary and FAQ A4 for further guidance; or
- you are conducting research on behalf of or at the request of a service provider that matches the definition of *audit in the glossary.

If your research is audit or uses prior-anonymised data, please check this box: ☐

You do not need to seek ethical approval from CUREC, and you do not need to complete any more of this form. However, please check with your department's requirements, as some departments require you to lodge this form with them.

Office use only: IDREC Ref. No. _____

Date of confirmation that checklist accepted on behalf of IDREC: // //

SECTION A	Yes	No
1) Are you using research methodologies commonly used in biomedical or behavioural laboratory sciences?		x
2) Is there a significant risk that the research will induce anxiety, stress or other harmful psychological states in participants that might persist beyond the duration of any test or interview in which they are participating?		X
3) Will the research involve human participants recruited by means of their status as present or past NHS patients or their relatives or carers?		X
4) Does the research involve *human participants aged 16 and over who do not have *capacity to consent for themselves? See the Mental Capacity Act 2005		X
5) Is the study to be funded by the US National Institutes of Health or another US federal funding agency?		x

If you answered 'yes', please **stop** work on this checklist and

- **for questions 1 and 2**, complete CUREC/1 instead (available from www.admin.ox.ac.uk/curec);
- **for questions 3 and 4**, submit your proposal to the appropriate NHS ethics committee (see www.nres.nhs.uk and www.admin.ox.ac.uk/researchsupport/=ctrg for further information);
- **for question 5**, or **if you answered 'yes' to questions 1, 2 or 4 and your research will take place outside the EU and is a biomedical study**, submit your proposal to OXTREC, which uses separate documentation. **Applications to OXTREC using this form will not be accepted.** If your research is not a biomedical study and does not have US funding, but will take place outside the EU, you may use this form to submit your application for approval to the Social Sciences and Humanities IDREC.

If you have answered 'no' to all questions in Section A, please complete Sections B-E. This form and any supporting materials should be typewritten.

SECTION B

Principal investigator/ supervisor/student researcher (title and name):	***Supervisor: Professor Victoria Murphy, Department of Education & Professor Kate Nation, Department of Experimental Psychology Student researcher: Annina Hessel, MSc& MEd***
Name of supervisor (STUDENT RESEARCH PROJECTS ONLY):	***Professor Victoria Murphy & Professor Kate Nation***
Degree programme, e.g. DPhil, MPhil, MSc (STUDENT RESEARCH PROJECTS ONLY):	***DPhil in Education***
Department or institute:	***Department of Education***
Address for correspondence (if different):	
Email and phone contact:	***Annina.Hessel@education.ox.ac.uk***
Title of research project:	***Metaphor comprehension in English as an additional language (EAL) learners***

My research focuses on the language learning of students learning English as an additional language (EAL). I am researching the comprehension processes that EAL children engage in when reading metaphor. Some EAL children experience comprehension difficulties that impede their literacy development, which in turn can affect their overall academic attainment (Hutchinson, Whiteley, Smith, & Connors, 2003; OECD, 2006). Furthermore, EAL learners' comprehension problems can be linked to difficulties understanding metaphors (Bountrogianni, 1984; Johnson, 1991). Importantly, different metaphors seem to cause different problems (Hessel, 2014; Kan, 2014). Measuring reading comprehension in EAL learners helps to understand these difficulties and develop targeted teaching. I thus will measure the processing of different metaphors online to see whether and how EAL children differ from monolinguals in their processing and comprehension of metaphor.

Participants will include typically-developing primary school children in year 3 (8 years of age), including both monolingual and EAL children (≈30 per group). They will be tested for vocabulary knowledge through the Test of Word Knowledge (Wiig & Secord, 1992). In order to make sure they are able to do the reading metaphor test, their literacy skills will be tested through the Test of Word Reading Efficiency 2nd Edition (Torgesen, 2012). These tests are quick to complete and have been developed and validated for this age range. For inclusion into the EAL group, language background questionnaires adapted from previous studies (McKendry & Murphy, 2011) and (Paradis, Emmerzeal & Duncan, 2010) will be administered to both the parents and children to establish the degree of exposure to home language (see Appendix A). If necessary, the questionnaire will be translated into the parents' first languages. To test metaphor comprehension, the children will read stories on a computer screen (see Appendix B). The stories will be 55-63 words to avoid tiring the children. Children will read 30 literal stories and 30 stories that contain metaphors. After every story, they will answer a comprehension question (see Appendix B). During the story reading, children's eye movements will be filmed through a table mounted Eyelink eye tracker.

All testing will be done in the children's schools. Children will be tested individually in a calm area of the school across two sessions of no more than 30-40 minutes each. Piloting with monolingual and EAL children (n ≈10 per group) will help to make sure that the tasks are age-appropriate. In case the tasks appear too difficult for the children or if any other problem is detected during the testing, the tasks will be considered age-inappropriate and will be adapted to meet the children's needs and abilities. During the first testing session, the standardised tests and the language background questionnaire will be administered on one day together with approximately 20 stories from the metaphor test, with the remaining 40 stories of the metaphor comprehension test being administered on another day. All data will be saved under a code assigned to the participants from the first data collection session onwards so that no personal information will be retained in the data sheets. I will keep the list of children's names and numbers in a password-protected computer file, and I will erase this list at the end of the project.

Metaphors will be pretested for ecological validity and frequency in the Oxford Children's Corpus (Banerji, Gupta, Kilgarriff, & Tugwell, 2013). Stories will be piloted with adult native speakers (N≈150) in order to make sure they sound natural and that literal and metaphorical stories don't vary in predictability (see Appendix C). There will be five pre-tests that investigate the literal and the metaphorical version of the story respectively or different aspects of the phrases (for cloze tests). Each survey will take approximately 15-20 minutes and will be filled out by approximately 20-30 individuals. The first participants taking these questionnaires will be asked if they are happy with being contacted for a one-one-one conversation to comment on potential difficulties connected to stimuli and tasks (n=2-3 per pre-test). In recruitment, I will rely on a convenience sample of students and members of societies and groups I am a member of, such as the Oxford University Ballet Society. The data will be treated as described above for children's data.

For recruitment of monolingual and EAL children, I am cooperating with a local charity called Jacari providing tutoring for EAL children. As tutors, Jacari recruits University of Oxford and Oxford Brookes students who volunteer to teach an EAL child for one year. My language testing will be done as part of a project that Jacari has asked me to implement to help them examine their work with EAL students. To that end, I will send out online questionnaires via email to Jacari tutors (see Appendix F) to explore the tutors' development, their evaluation of Jacari support, and their relationship with parents and students. The tutors who Jacari will approach for this online questionnaire will be both the tutors of the children I will be testing as well as other a few other tutors working with children from other year groups. The same questionnaire will be sent out at two time points, once in the beginning of the year and then at the end of the tutoring. The questionnaire will be distributed through the online platform Qualtrics. The tutors' answers will be analysed for the three areas named above. Furthermore, the questionnaire responses of the tutors whose Jacari students I will have tested in the above described study will be linked to their students' language abilities. Linking the tutors' experiences to the students' current state of knowledge will provide an overview of the current state of tutoring to Jacari. The data will be presented in a clear and professional way that helps Jacari to evaluate and improve their own services without compromising any participant's anonymity (either tutor or pupil). Jacari tutors will be given the option to opt in for discussing their tutoring experience within a focus group of several tutors organised by Jacari. For those tutors opting in to the focus group, a feedback will be provided on their development between the two testing points, for example on how their self-efficacy in teaching has developed from the beginning of the year to the end of their tutoring. Jacari will help establish contacts to the schools and families they work with in Oxford which will help me to reach the intended population of my study in the ways outlined below. Additionally, I will recruit other local schools by first emailing the head teachers and, upon obtaining their consent to do my study in their school, I asking the parents for consent.

I have received a CRB check in December 2013 that discloses no convictions, cautions, final warnings or reprimands. The described research project may best be conducted by following the protocol P25 provided by the central university research ethics committee. In order to comply with this protocol, the following procedures in recruiting participants and securing consent will be adopted.
All participants will receive information on the project together with an opt-in form to take part in the online questionnaire for tutors and NS in the pilots (integrated in the questionnaires, see Appendix C and F) and to take part in the language testing for teachers, children and their parents, respectively (see Appendix D and E). The child participants will be recruited and tested in local primary schools. The schools' head teachers will be contacted by myself together with Jacari and will be asked for permission to conduct the study in their school (see Appendix D). With the head teacher's consent, the aforementioned information letters will be given out to the parents. In case the EAL children's guardians are not in the position to understand the information provided in English, translations will be provided. The need for such additional material will be established in consultation with head and class teachers. If a child shows signs of being unwell or distressed during testing, the session will be stopped, and the event reported to the child's teacher.

List actual or probable location(s) where project will be conducted, if known:	Testing in local primary schools in and around Oxford in combination with online questionnaires
If your research involves overseas travel or fieldwork, have you completed and returned a travel risk assessment form? (Bear in mind that this may be necessary to ensure that the travel or fieldwork is covered by the University's travel insurance – see http://www.admin.ox.ac.uk/finance/insurance/travel/.)	**No overseas travel planned.**
Anticipated duration of project:	__32__ months
Anticipated start date:	20 / 01 / 16
Anticipated end date:	30 / 9 / 18
Name and status (e.g. 3rd year undergraduate; post-doctoral research assistant) of others taking part in the project:	No other researchers involved.
Please indicate what training on research ethics the researchers involved with this study have received, e.g. the title of the online or in-person course, and date completed (online training available at www.admin.ox.ac.uk/researchsupport/integrity/human/):	Philosophy of Education course as part of PRS training in Department of Education

SECTION C

Methods to be used in the study (**tick** as many as apply: this information will help the committee understand the nature of your research and may be used for audit).

	Please tick
Interview	
Questionnaire	X
Analysis of existing records	
Participant performs verbal/paper and pencil/computer based task	X
Measurement/recording of motor behaviour	X
Audio recording of participant	
Video recording or photography of participant	
Physiological recording from participant	
Participant observation	
Covert observation	
Systematic observation	
Observation of specific organisational practices	

Other (please specify)	

SECTION D

Have you read one or more of the following professional guidelines and do you undertake to use the principles listed there as a guide for your own work? Please note that this is not intended to be an exhaustive list. Links to the guidelines listed below are included on the CUREC website.

	Please tick
British Society of Criminology: Code of Ethics for Researchers in the Field of Criminology [www.britsoccrim.org/codeofethics.htm]	
British Educational Research Association Ethical Guidelines for Educational Research [www.bera.ac.uk/guidelines]	x
Academy of Management's Code of Ethics [www.aomonline.org/aom.asp?ID=&page_ID=242]	
Association of American Geographers Statement on Professional Ethics [www.aag.org/cs/resolutions/ethics]	
Oral History Society of the UK Ethical Guidelines [http://www.oralhistory.org.uk/index.php]	
American Political Science Association (APSA) Guide to Professional Ethics in Political Science (Section H) [www.apsanet.org/content_9350.cfm]	
Political Studies Association Guide to Good Professional Conduct [www.psa.ac.uk/psa]	
British Psychological Society Code of Ethics and Conduct [www.bps.org.uk/what-we-do/ethics-standards/ethics-standards]	
Ethics Guidelines of the Association of Social Anthropologists of the UK and Commonwealth [www.theasa.org/ethics/guidelines.shtml]	
Social Research Association: Ethical Guidelines [http://the-sra.org.uk/sra_resources/research-ethics/ethics-guidelines/]	
Statement of Principles of Ethical Research Practice from the Socio-Legal Studies Association [www.slsa.ac.uk/content/view/247/270/]	
Statement of Ethical Practice for the British Sociological Association [www.britsoc.co.uk/about/equality/statement-of-ethical-practice.aspx]	
Other professional guidelines (please specify):	

SECTION E

Please put a tick in the yes/no column as appropriate to indicate your response.

1) Will you obtain informed consent according to good practice in your discipline before participation?	**Yes**	**No**
	X	
2) Will you ensure that *personal data collected directly from participants or via a *third party is held and processed in accordance with the provisions of the Data Protection Act?	**Yes**	**No**
	X	
3) Does the research involve as participants *people whose ability to give free and	**Yes**	**No**

informed consent is in question? ***Note: participants aged under 16 are generally considered to require consent of a parent or guardian (only answer 'no' to this question if you can cite one of the protocols listed under 'children'). For participants aged 16-17, consult FAQ C13***	**x**	
4) As a consequence of taking part in the research, will participants be at serious risk of rendering themselves liable to criminal prosecution (e.g. by providing information on drug abuse or child abuse)?	**Yes**	**No**
		x
5) Does the research involve the *deception of participants, as part of the investigation/experiment?	**Yes**	**No**
		x
If any of your answers above are in a shaded box, please indicate whether those aspects of your project are fully covered by the following.		
6) Research protocol(s) which has/ve received IDREC/CUREC approval?	**Yes**	**No**
If yes, please give protocol number(s): **DREC_2014_P25**	**x**	
7) Professional guidelines that you will be following, as noted under Section D?	**Yes**	**No**
I will be following the *British Educational Research Association Ethical Guidelines for Educational Research.*	**x**	

If any of your answers in Section E are in a shaded box and are not covered by a protocol or by professional guidelines, please complete CUREC/2, available to download from the CUREC website. Then submit both this form (you need not complete section F) and the CUREC/2 to the Social Sciences and Humanities IDREC.

If all your answers in Section E are in the unshaded boxes or your answers in shaded boxes are covered by a protocol or professional guidelines, complete Section F and submit this form and any accompanying documents to the Social Sciences and Humanities IDREC or to the relevant officer/committee at departmental level (see notes and address below).

FINAL CHECK

Please check each of the following before submitting the checklist. **If the appropriate supporting documentation is not included with your application, you will then be asked to provide this separately. This may well delay the ethical review process, and thus the start of your research.**

Have you completed Sections A-E? ☐

Have you defined all technical terms and abbreviations used? ☐

If you have produced any documentation in support of your application (which might include questionnaires, participant information, consent forms/form or note of procedure for recording oral consent, advertisements and surveys), have you attached a copy of these? ☐

Are all pages (including appendices and attachments) numbered? ☐

SECTION F

You can submit this checklist by email and/or as a signed hard copy; if it is being sent by email only, the checklist, and any email from the head of department or nominee separately endorsing its submission, must be sent from a University of Oxford email address (i.e. as a minimum, the checklist and supporting documents must be submitted by the head of department or nominee indicating his/her approval from a University of Oxford email account).

Complete this section only if you do not need to submit form CUREC/2.

I understand my responsibilities as principal researcher/supervisor/student researcher as outlined in the CUREC glossary and guidance on the CUREC website.

I declare that the answers above accurately describe my research as presently designed and that I will submit a new checklist should the design of my research change in a way which would alter any of the above responses so as to require completion of CUREC/2 (involving full scrutiny by an IDREC). I will inform the relevant IDREC if I cease to be the principal researcher on this project and supply the name and contact details of my successor if appropriate.

Signed by principal researcher/supervisor/student researcher:  ……

Date: 14/01/2016

Print name (block capitals) Annina Hessel

Signed by supervisor:……(for student projects)

Date:……

Print name (block capitals)……Victoria Murphy……

I understand the questions and answers that have been entered above describing the research, and I will ensure that my practice in this research complies with these answers, subject to any modifications made by the principal researcher properly authorised by the CUREC system.

Signed by associate/other researcher: ……

Print name (block capitals)……

Date ……

I have read the research project application named above. On the basis of the information available to me, I:

(i) consider the principal researcher/supervisor/student researcher to be aware of her/his ethical responsibilities in regard to this research;

(ii) consider that any ethical issues raised have been satisfactorily resolved or are covered by relevant professional guidelines and/or CUREC approved protocols, and that it is appropriate for the research to proceed without further formal ethical scrutiny at this stage (noting the principal researcher's obligation to report should the design of the research change in a way which would alter any of the above responses so as to require completion of a CUREC/2 full application);

(iii) am satisfied that the proposed project has been/will be subject to appropriate *peer review and is likely to contribute something useful to existing knowledge and/or to the education and training of the researcher(s) and that it is in the *public interest.

(iv) [FOR DEPARTMENTS/FACULTIES WITH A DEPARTMENTAL RESEARCH ETHICS COMMITTEE (DREC) OR EQUIVALENT BODY - PLEASE DELETE IF NOT APPLICABLE] confirm that this checklist (and associated research outline) has been reviewed by the Department's Research Ethics Committee (DREC)/equivalent body, and attach the associated report from that body.

Signed:……
(Head of department or nominee e.g Chair of DREC, Director of Graduate Studies for postgraduate student projects)

Print name (block capitals)……

Date:……

If your research involves participants recruited by means of their status as current or former NHS staff, or the research will, in whole or in part, be carried out on NHS premises, use NHS facilities or assess NHS facilities or services, please see FAQ B3 (www.admin.ox.ac.uk/curec/faqs/).

Please check with your department about its procedures for the approval of CUREC forms. If your department indicates that the checklist should be submitted directly to the IDREC, please send it, together with any supporting documentation, to the following address(es), keeping a copy for yourself:

Secretary of the Social Sciences and Humanities IDREC Email: ethics@socsci.ox.ac.uk
University of Oxford Social Sciences Division
Hayes House, George Street
Oxford, OX1 2BQ

IDRECs and/or CUREC will review a sample of completed checklists and may ask for further details of any project.

Revised July 2013

FINAL CHECK

Please check each of the following before submitting the checklist. **If the appropriate supporting documentation is not included with your application, you will then be asked to provide this separately. This may well delay the ethical review process, and thus the start of your research.**

Have you completed Sections A-E? ☐

Have you defined all technical terms and abbreviations used? ☐

If you have produced any documentation in support of your application (which might include questionnaires, participant information, consent forms/form or note of procedure for recording oral consent, advertisements and surveys), have you attached a copy of these? ☐

Are all pages (including appendices and attachments) numbered? ☐

SECTION F

You can submit this checklist by email and/or as a signed hard copy; if it is being sent by email only, the checklist, and any email from the head of department or nominee separately endorsing its submission, must be sent from a University of Oxford email address (i.e. as a minimum, the checklist and supporting documents must be submitted by the head of department or nominee indicating his/her approval from a University of Oxford email account).

Complete this section only if you do not need to submit form CUREC/2.

I understand my responsibilities as principal researcher/supervisor/student researcher as outlined in the CUREC glossary and guidance on the CUREC website.

I declare that the answers above accurately describe my research as presently designed and that I will submit a new checklist should the design of my research change in a way which would alter any of the above responses so as to require completion of CUREC/2 (involving full scrutiny by an IDREC). I will inform the relevant IDREC if I cease to be the principal researcher on this project and supply the name and contact details of my successor if appropriate.

Signed by principal researcher/supervisor/student researcher:  .....

Date: 14/01/2016

Print name (block capitals) Annina Hessel

Signed by supervisor:.....(for student projects)

Date:.....

Print name (block capitals).....Victoria Murphy.....

I understand the questions and answers that have been entered above describing the research, and I will ensure that my practice in this research complies with these answers, subject to any modifications made by the principal researcher properly authorised by the CUREC system.

Signed by associate/other researcher: .....

Print name (block capitals).....

Date .....

I have read the research project application named above. On the basis of the information available to me, I:

(i) consider the principal researcher/supervisor/student researcher to be aware of her/his ethical responsibilities in regard to this research;

(ii) consider that any ethical issues raised have been satisfactorily resolved or are covered by relevant professional guidelines and/or CUREC approved protocols, and that it is appropriate for the research to proceed without further formal ethical scrutiny at this stage (noting the principal researcher's obligation to report should the design of the research change in a way which would alter any of the above responses so as to require completion of a CUREC/2 full application);

(iii) am satisfied that the proposed project has been/will be subject to appropriate *peer review and is likely to contribute something useful to existing knowledge and/or to the education and training of the researcher(s) and that it is in the *public interest.

(iv) [FOR DEPARTMENTS/FACULTIES WITH A DEPARTMENTAL RESEARCH ETHICS COMMITTEE (DREC) OR EQUIVALENT BODY - PLEASE DELETE IF NOT APPLICABLE] confirm that this checklist (and associated research outline) has been reviewed by the Department's Research Ethics Committee (DREC)/equivalent body, and attach the associated report from that body.

Signed:……..
(Head of department or nominee e.g Chair of DREC, Director of Graduate Studies for postgraduate student projects)

Print name (block capitals)……

Date:……

If your research involves participants recruited by means of their status as current or former NHS staff, or the research will, in whole or in part, be carried out on NHS premises, use NHS facilities or assess NHS facilities or services, please see FAQ B3 (www.admin.ox.ac.uk/curec/faqs/).

Please check with your department about its procedures for the approval of CUREC forms. If your department indicates that the checklist should be submitted directly to the IDREC, please send it, together with any supporting documentation, to the following address(es), keeping a copy for yourself:

Secretary of the Social Sciences and Humanities IDREC Email: ethics@socsci.ox.ac.uk
University of Oxford Social Sciences Division
Hayes House, George Street
Oxford, OX1 2BQ

IDRECs and/or CUREC will review a sample of completed checklists and may ask for further details of any project.

Revised July 2013

Appendix K. Ethical Clearance from Central University Research Ethics Committee for Experiments 2 and 3

Research Office 1 March 2017 at 13:31
FW: Amended CUREC application Details
To: Victoria Murphy, Cc: Annina Hessel

Hi Vicki

Please note that Liam has approved this.

BW
Jackie

From: Liam Gearon [mailto:liam.gearon@education.ox.ac.uk]
Sent: 01 March 2017 13:27
To: Education Research Office
Subject: RE: Amended CUREC application

Hi Jackie,

I am signing this off – have they the right forms though. They didn't actually need to fill out whole new forms.

Kind regards

Liam

From: Research Office [mailto:research.office@education.ox.ac.uk]
Sent: 01 March 2017 13:08
To: Liam Gearon
Subject: FW: Amended CUREC application

Appendix L. Recruitment letters used in Experiments 2 and 3

Parent Information Leaflet

DOES MY CHILD HAVE TO TAKE PART?

I am using an 'opt –out' consent policy, such that all children will have the opportunity to take part unless parents prefer that they do not. However, there is no obligation for your child to take part. **If you would prefer for your child NOT TO take part, please fill in the form at the bottom of the attached letter and return it to your child's class teacher as soon as possible.** I will not begin our project until at least two weeks after this letter is sent to allow you sufficient time to opt out. If you would like to discuss the research with someone beforehand (or if you have any questions afterwards) please contact me, Annina Hessel (contact details at the bottom of this page). If now or at any time you have any concerns about this research and you are not satisfied with the answers I have given you, you can contact my supervisor, Professor Victoria Murphy: victoria.murphy@education.ox.ac.uk

Once the study has begun, your child is still able to withdraw at any time, without giving a reason, if they do not wish to continue.

WHAT HAPPENS NEXT?

If you would prefer for your child **not to** take part then please fill in the 'opt-out' consent form on the attached letter and return to your class teacher. If you are happy for your child to participate then you do not need to do anything at all and your child may be seen by me at a time arranged with their class teacher, fitting in with classroom activities.

If you would like to discuss the project, please do not hesitate to contact me:

Annina Hessel
Department of Education
University of Oxford
15 Norham Gardens
OX2 6PY
annina.hessel @education.ox.ac.uk

Phone: 07802883835



LEARNING TO READ PROJECT

My name is Annina Hessel, and I am a doctoral research student at the University of Oxford. In partnership with the University of Oxford, your child's school has agreed to take part in my project on child language learning. I would like to invite your child to participate in the project. Before you decide whether you are happy with this, please read this letter to know why the research is being done and what it involves, and ask me any questions you would like.

WHO IS DOING THE RESEARCH?

The research is led by me, Annina Hessel, a doctoral research student at the University of Oxford. I possess a full, up-to-date DBS certificate and have extensive experience in working with young children, both as a teacher and a researcher. The research is supervised by Victoria Murphy in the Department of Education and Kate Nation in Experimental Psychology at Oxford University. The study has been approved by the University of Oxford's Central University Research Ethics Committee.

WHAT IS THIS PROJECT TRYING TO FIND OUT?

The research project explores how children make sense of texts of varying difficulty, and how that is linked to their other skills, such as the way they read, the number of words they know, and if they learn English as an additional language.

A better understanding of the processes involved in learning to read, will help us to develop ways of optimising reading practice to promote successful learning. I also hope to identify why some children find reading difficult and to develop methods to help these children.

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024
Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk
www.education.ox.ac.uk

Director: Professor Jo Anne Baird

WHY AM I INVITING YOUR CHILD TO TAKE PART?

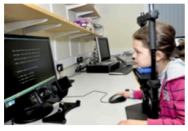
You have been contacted because your child's school is participating in our research. We are writing to all parents with children aged between 8 and 10 years to invite their child to take part. We are inviting children of all reading abilities to take part in the study, so receiving this letter does not indicate that your child has reading difficulties.

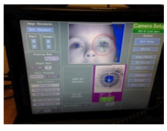
DOES YOUR CHILD HAVE TO TAKE PART?

Your child does not have to take part in this project. If you or your child decides against taking part, they will not be at any educational disadvantage. You can even change your minds later: your child can stop participating at any time, without having to give me a reason – you just need to let me know.

ARE THERE ANY POTENTIAL RISKS IN TAKING PART?

No, there are no potential risks in your child taking part in this experiment. The eye tracker is a completely unobtrusive device. Your child will gain extra one-on-one reading time with an adult outside the classroom, which children enjoy.





How do children learn to understand texts?



A PhD research project by

Department of Education and Experimental Psychology, University of Oxford

Funded by:

Economic and Social Research

WHAT WILL HAPPEN IF YOUR CHILD DOES TAKE PART?

If you agree for your child to take part, I will see them at school for approximately 40-50 minutes individually in a quiet area near their classroom.

Your child will read some stories, do a few word-reading and logic-related exercises and answer some questions about their reading or whether they speak languages other than English outside of school.

They will also read stories and while they read. I will measure their eye movements whilst reading using an eye tracker, a camera device connected to a computer. Children rest there chin on a comfortable pad and the eye tracker sits in front of them (see photos to the left). The eye tracker only captures their eye region and no recordings or photos will be taken of children's faces. Afterwards, I play back a short video so that children can see what their eyes were doing whilst reading. Children in the past have really enjoyed taking part in my studies. If you want to see what using the eyetracker looks like, please have a look at this short video by my supervisor Kate Nation:

https://www.youtube.com/watch?v=Ozrmj28p6SA

WHAT WILL HAPPEN TO THE RESULTS OF THE PROJECT?

All the information I collect will be kept strictly confidential. I will give each child a code number, and in the data, they will only be identified through this number When I have finished the project, I will give a summary of findings about all children to the school, and you are welcome to have this summary too. I also aim to publish our findings in scientific journals, but this may be a year or two from the end of the study.

Letters to parents offering opt-out

University of Oxford
Department of Education

15 Norham Gardens, Oxford OX2 6PY
Mobile: +44(0)7802883835
annina.hessel@education.ox.ac.uk

14/03/2017
Dear Parent or Guardian,

In partnership with researchers at the University of Oxford, your child's school has agreed to take part in the study *How do Children Understand Texts*? I am writing to you because I would like to invite your child to be part of this research.

I very much hope that you would like your child to take part, but before you decide, it is important that you understand why the study is being done and what it will involve. Therefore, please take a moment to read the enclosed 'Parent Information Sheet'. I hope that after reading this you are keen for your child to be involved in the research.

Since we are using an 'opt-out' consent policy, please only fill out the form below if you **DO NOT** want your child to participate and return it to the school by **the 28th March.** If I do not receive an opt-out form from you, your child may be invited to take part in this study, as described in the accompanying information sheet.

Please get in touch (contact details at the top of this letter) if you have any questions at all before, during or after the experiment. I will also be in school on year 5 parents' consultation days on the 20th, 21st and 29th March from 3:00-4:30pm, and I will be happy to answer all your questions and discuss the project in person there.

Yours sincerely,

Annina Hessel

(Doctoral Research Student)

I, the undersigned, hereby DO NOT give permission for my child to take part in the above study.
Name of child:

_____

Name of parent/guardian:

_____

Signature: _____ Date: _____
If you would like to receive a newsletter summarising the research, please fill in your email address:

University of Oxford
Department of Education

15 Norham Gardens, Oxford OX2 6PY
Mobile: +44(0)7802883835
annina.hessel@education.ox.ac.uk

21/03/2017
Dear Parent or Guardian,

I recently sent you a letter about a study that I am conducting investigating how children learn to read and the development of these skills. I will be starting the *How do children learn to understand texts?* study soon. If you would prefer for your child **not to** take part, it is important that you complete the form below and return it to the school by **the 28th March.** If you are happy for your child to take part in our study, then you do not need to do anything else.

I very much hope that you would like your child to take part, but before you decide, it is important that you understand why the study is being done and what it will involve. Therefore, please take a moment to read the enclosed 'Parent Information Sheet'. I hope that after reading this you are keen for your child to be involved in our research.

Please get in touch (contact details at the top of this letter) if you have any questions at all before, during or after the experiment. I will also be in school on year 5 parents' consultation days on the 20th, 21st and 29th March from 3:00-4:30pm, and I will be happy to answer all your questions and discuss the project in person there.

Yours sincerely,

Annina Hessel

(Doctoral Research Student)

I, the undersigned, hereby DO NOT give permission for my child to take part in the above study.
Name of child:

_____

Name of parent/guardian:

_____

Signature: _____ Date: _____

If you would like to receive a newsletter summarising the research, please fill in your email address:

Leaflet for parents offering opt-in

WHAT SHOULD YOU DO NEXT?

If you would like to discuss the project, please do not hesitate to contact me:

If now or at any time you have any concerns about this research and you are not satisfied with the answers I have given you, you can contact my supervisor, Professor Victoria Murphy: victoria.murphy@education.ox.ac.uk

If you are happy for your child to take part, please tick the box on the slip at the bottom of this page, fill in your child's name and return the leaflet to your child's teacher

Yours sincerely,
Annina Hessel

☐ Yes, I am happy for my child to participate in this project.

My child's name is

Signed by parent or guardian:

_____

Name of parent/ guardian:

Date:



LEARNING TO READ PROJECT

My name is Annina Hessel, and I am a doctoral research student at the University of Oxford. In partnership with the University of Oxford, your child's school has agreed to take part in my project on child language learning. I would like to invite your child to participate in the project. Before you decide whether you are happy with this, please read this letter to know why the research is being done and what it involves, and ask me any questions you would like.

WHO IS DOING THE RESEARCH?

The research is led by me, Annina Hessel, a doctoral research student at the University of Oxford. I possess a full, up-to-date DBS certificate and have extensive experience in working with young children, both as a teacher and a researcher. The research is supervised by Victoria Murphy in the Department of Education and Kate Nation in Experimental Psychology at Oxford University. The study has been approved by the University of Oxford's Central University Research Ethics Committee.

WHAT IS THIS PROJECT TRYING TO FIND OUT?

The research project explores how children make sense of texts of varying difficulty, and how that is linked to their other skills, such as the way they read, the number of words they know, and if they learn English as an additional language.

A better understanding of the processes involved in learning to read, will help us to develop ways of optimising reading practice to promote successful learning. I also hope to identify why some children find reading difficult and to develop methods to help these children.

Annina Hessel
Department of Education
University of Oxford
15 Norham Gardens
OX2 6PY
annina.hessel @education.ox.ac.uk

Phone: 07802883835

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024
Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk
www.education.ox.ac.uk

Director: Professor Jo Anne Baird

WHY AM I INVITING YOUR CHILD TO TAKE PART?

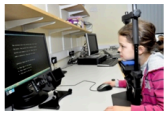
You have been contacted because your child's school is participating in our research. We are writing to all parents with children aged between 8 and 10 years to invite their child to take part. We are inviting children of all reading abilities to take part in the study, so receiving this letter does not indicate that your child has reading difficulties.

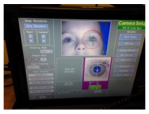
DOES YOUR CHILD HAVE TO TAKE PART?

Your child does not have to take part in this project. If you or your child decides against taking part, they will not be at any educational disadvantage. You can even change your minds later: your child can stop participating at any time, without having to give me a reason – you just need to let me know.

ARE THERE ANY POTENTIAL RISKS IN TAKING PART?

No, there are no potential risks in your child taking part in this experiment. The eye tracker is a completely unobtrusive device. Your child will gain extra one-on-one reading time with an adult outside the classroom, which children enjoy.





How do children learn to understand texts?

PLEASE READ & SIGN ON LAST PAGE IF ALLOW CHILD TO TAKE PART



A PhD research project by

Department of Education and Experimental Psychology, University of Oxford

Funded by:

Economic and Social Research Council

WHAT WILL HAPPEN IF YOUR CHILD DOES TAKE PART?

If you agree for your child to take part, I will see them at school for approximately 40-50 minutes individually in a quiet area near their classroom.

Your child will read some stories, do a few word-reading and logic-related exercises and answer some questions about their reading or whether they speak languages other than English outside of school.

They will also read stories and while they read. I will measure their eye movements whilst reading using an eye tracker, a camera device connected to a computer. Children rest there chin on a comfortable pad and the eye tracker sits in front of them (see photos to the left). The eye tracker only captures their eye region and no recordings or photos will be taken of children's faces. Afterwards, I play back a short video so that children can see what their eyes were doing whilst reading. Children in the past have really enjoyed taking part in my studies. If you want to see what using the eyetracker looks like, please have a look at this short video by my supervisor Kate Nation:

https://www.youtube.com/watch?v=Ozrmj28p65A

WHAT WILL HAPPEN TO THE RESULTS OF THE PROJECT?

All the information I collect will be kept strictly confidential. I will give each child a code number, and in the data, they will only be identified through this number When I have finished the project, I will give a summary of findings about all children to the school, and you are welcome to have this summary too. I also aim to publish our findings in scientific journals, but this may be a year or two from the end of the study.

Appendix M. Language Background Questionnaire Experiments 2 and 3

Questions asked:

Language use/ EAL or not

"For my experiment, I am also interested in the languages you learn. I know you can speak English because we are speaking it right now."

A Language Learning History

1. ... but do you or your parents speak or hear another language but English at home?

[Not show this just ask.]

Yes	No
1	0

[If Yes: continue. If no: stop here.]

1. Which language do you or your parents speak? _____

[From now on: insert language given by child instead of *your home language*]

2. How old were you when you first lived in the UK? _____
3.
4. How old were you when you started learning English? _____

B Languages Spoken with Family Members

"Please think back over the last month and when you used and heard *your home language* and when you used English. I will tell you times when you might use English or your home language, and I would like you to tell me how often you did this in *your home language*."

"I would like you to tell me if this happened

Never- Less than half the time- More than half the time- All the time"

[Show them response scale and point to it when explaining]

Trial: Tell me how often you had sandwiches for lunch this week/ last week.

Never	Less than half the time	More than half the time	All the time
1	2	3	4

	I don't know.
How often did your parents speak in HL to you, for example to ask you a question or tell you about something?	
How often did you talk to your parents in HL, for example to tell them something or respond to a question?	
Do you have brothers or sisters?	
If yes: *How often did you talk in HL to your brothers or sisters?*	
How often did your brothers or sisters talk in HL to you?	
How often did you use HL with your friends?	

C Media Exposure

"Thank you, that is really helpful to know. Now I would like to ask you about three things that you might also do when you are at home- like watching TV or reading books. I would like you to think back over the last month again and tell me: How often you do this from

Never - On one or three days a week- On Four to six days a week- Every day

And then, how often you do this in English or compared to how often you do things in your HL from

Always in home language- More often in home language than in English -About the same- More often in English than in home language- Always in English"

[SHOW VISUAL SCALES WHILE TALKING]

Trial: How often did you

Never	Sometimes	Often
1	2	3

Always in home language	More often in home language than English	About the same	More often in English than in home language	Always in English
1	2	3	4	5

I don't know.

	How often do you do this?	How often do you do this in English compared to your Home language?
1. Watch TV or movies?		
2. Read Books yourself?		
3. Your parents read books to you?		

Scales shown
(one DinA4 page each scale)

Never	Less than half the time	More than half the time	All the time
1	2	3	4

I don't know.

Never	Sometimes	Often
1	2	3

Always in home language	More often in home language than English	About the same	More often in English than in home language	Always in English
1	2	3	4	5

I don't know.

Appendix N. Stimuli Experiment 2

Metaphor	Literal control	Text	Comprehension question	Correct answer for metaphors	Correct answer for literal controls
they’re in the same boat	they’re in the same group	Tom and his friends all bring the same story to a writing competition. They hope no one will notice it because different judges will hear each story. When they arrive, Tom and his friends realise that they're in the same boat/ group and Tom worries others will notice.	Why does Tom worry?	Because he and his friends have similar problems now.	Because he and his friends will read their story to the same judge.
Imogen is the queen-bee	Imogen is the music-lover	Imogen and her friends are playing together. First, they don't know what to do. Imogen's friends want to draw, but Imogen convinces them to play the guitar. Imogen's mom sees what is going on. She thinks that Imogen is the queen-bee/ music-lover of this group and she smiles.	Why does Imogen's mom smile?	Because Imogen is the leader of her circle of friends.	Because Imogen likes playing the guitar so much.
Florence is over the moon	Florence is over the line	Florence is playing in an important football match. She wants to score the next goal. Florence knows that if she accidentally runs out of the field of play, her goal won't count. After she has kicked, Florence is over the moon / line and she starts screaming out loud.	Why does Florence start screaming out loud?	Because she's happy that she has scored the goal.	Because she has run out of the field of play.
Michael’s face is beetroot	Michael’s face is swollen	In PE class, Michael tries his first high jump. He jumps off the floor and falls on his front. The other children laugh at Michael. His teacher however worries whether Michael is hurt. When he gets up, Michael's face is beetroot / swollen and he runs out of class.	Why does Michael run out of the class?	Because he is embarrassed.	Because he hurt himself.

the new boy is a bad apple	the new boy is a bad mate	Lauren and her friends are playing cards together nicely until a new boy joins. The new boy starts cheating and since Lauren's friends don't want to lose, they start cheating, too. Lauren thinks that the new boy is a bad apple / mate and she stops playing with him.	Why does Lauren decide to stop playing with the new boy?	Because he makes the whole group cheat.	Because he wouldn't be a good friend.
her thoughts are rockets	her thoughts are super	Tina struggles with a story she writes for school. For very long, she thinks and thinks but doesn't have any ideas. Suddenly, she has a lot of ideas and she starts writing really quickly. She thinks that her thoughts are rockets / super and she is pleased with herself.	Why is Tina pleased?	Because her ideas come very quickly.	Because her ideas are so great.
grades are lotteries	grades are misleading	Libby always thought she was good with numbers. But today, she got an odd grade in maths. When she comes home, Libby tells her mother that she now thinks she's useless at maths. Her mother says that sometimes, grades are lotteries / misleading and then Libby feels much better.	Why does Libby feel better?	She sees that grades can be random.	She sees that she has probably misunderstood the meaning of her grade.

the stores are jungles	the stores are crowded	George goes shopping with his mother. Soon, he gets bored. His mother says that he may look around by himself. After some time alone in the video games section, George wants to join his mother again. However, the stores are jungles / crowded and he can't find his mother.	Why does George take long to find his mum?	Because the shop layout is confusing.	Because the shops are very busy.
his smile is a razor	his smile is a fake	Sophia enjoys painting and wants to show her paintings to her friend. However, she's afraid he might mock her or even hide from Sophia what he really thinks. When Sophia eventually shows the paintings to him, his smile is a razor / fake and Sophia walks away, feeling sad.	Why does Sophia feel sad?	Because her friend is unkind to her.	Because her friend is not honest with her.
books are gateways	books are unhelpful	Muhammad asks his father how planes fly. His dad encourages him to read a book and find out for himself. Muhammad reads all there is in the book about planes. While reading the book, Muhammad slowly realises that books are gateways / unhelpful for learning a lot about planes.	What does Muhammad realise?	That reading is useless for learning about planes.	That you can learn a lot through reading.
her words are daggers	her words are friendly	Joshua really likes Jasmine. He decides to tell her about his feelings after school today. If Jasmine says she likes him, Joshua will invite her to go and watch a movie with him tomorrow. When Jasmine answers, her words are daggers / friendly and Joshua knows what to do.	Will Joshua invite Jasmine to go to the cinema with him?	No, because what she said hurt him a lot.	Yes, because what she said was nice.

their homework is a headache	their homework is a pleasure	Hannah and her friend have a difficult physics paper to finish for school. They are trying hard, but they can't figure it out themselves. Hannah decides to ask her mom for help. Hannah's mother says that their homework is a headache / pleasure and Hannah and her friend sigh.	Why do Hannah and her friend sigh?	Because Hannah's mum agrees that their physics paper is hard.	Because Hannah's mum thinks that the paper is fun.
their stomachs are pits	their stomachs are stuffed	Freya and her sister Emma have both already had lots of cake. However, their grandmother offers them one more slice if they think their stomachs are big enough for it. Freya looks at Emma and says that their stomachs are stuffed and Emma agrees with her sister.	Have Freya and Emma decided to eat more cake?	Yes, because their stomachs are very large.	No, because their stomachs are full.
her mother is a star	her mother is a helper	Matilda's neighbours ask for help- they have had a power cut and are completely in the dark! Matilda's mother manages to fix their wires using some tools and an old manual. Matilda's dad says that her mother is definitely a star / helper and he smiles proudly at Matilda.	Why is Matilda's dad proud of her mother?	Because she is really good at repairing things.	Because she was so helpful with the neighbours.
the next week will to be a chase	the next week will to be an effort	Claudia's class have been rehearsing for months to perform at their school festival, but now the festival will be earlier than planned. The teacher says that if they still want to perform, the next week will have to be a(n) chase / effort and the children agree with him.	What will Claudia's class have to do to perform at the school festival?	They will have to rehearse quickly.	They will have to work very hard.
the discussion is a fight	the discussion is a success	Phoebe and her sister always get cross with each other because they want the same toys. Their mother suggests they should sit down and find a better way of sharing them. The discussion between	Why do Phoebe and her sister stop discussing?	Because their conversation is so difficult.	Because they have managed to find a solution.

		Phoebe and her sister is a fight / success and the two girls stop discussing.			
the story is a washout	the story is a thriller	Matthew is reading a book. His sister cannot read yet but she's curious. She therefore asks whether Matthew thinks she might like the book, and whether he would read it to her. However, Matthew answers that the story is a washout / thriller and his sister isn't keen anymore.	Why is Matthew's sister not keen anymore?	Because the story doesn't make sense.	Because the story is scary.
the car is a joke	the car is a Jaguar	Sebastian and his friends are playing outside. Suddenly, they notice a car coming down the street. Sebastian and his friends stop and stare at it. The driver asks why they're looking at his car. Sebastian says the car is a joke / Jaguar and then the driver grins broadly.	Why does the driver grin so broadly?	Because Sebastian finds the car funny.	Because Sebastian has recognised the car type.
the movie was a drag	the movie was a drama	Alexander cannot go to the surprise showing at the cinema. He would be really unhappy if they showed the new action movie, and his friends saw it without him. The next day, Alexander's friends mention that the movie was a drag / drama and Alexander feels a lot better.	Why does Alexander feel better?	Because the movie he missed was bad.	Because the movie he missed was not an action movie.
the show was a smash	the show was a failure	Henry is organizing a charity evening to raise money for a hospital. He knows that it's important that people like the show so that they give more money. After the performance, Henry overhears someone saying that the show was a smash and Henry takes a deep breath.	Why does Henry take such a deep breath?	Because the people liked the show.	Because the people didn't like the show.

Appendix O. Confidence about comprehension levels

Questions asked

I would like you to remember how it was when you read these stories and how much you understood. Did you understand [POINT TO SCALE]
1= Nothing
2= Something
3= Most of it
4= All of it
How much do you think you understood from these texts?

Metaphor	
Implausibility	

Scale shown

(one DinA4 page)

Nothing	Something	Most of it	All of it
1	2	3	4

Appendix P. Stimuli Experiment 3

Inconsistent	Consistent control	Text	Comprehension question	Correct answer	Incorrect answer
plant	dress	For her aunt's wedding in Spring, Linda wore her best outfit. The colourful plant / dress was one of her absolute favourites.			
nail	blade	Jeremy carefully sawed through a large oak tree in his garden. The steel nail /blade became dull after only two minutes.			
truck	plane	Last week William flew to visit his family in another country. The large truck / plane was spacious and quickly transported them.	How long did the journey take?	It was quick.	It was slow.
book	ball	Charles tossed his new toy outside the garden. He became upset when the book / ball went over the blue garden fence.			
broom	sponge	Marcus was washing dishes every night to earn his pocket money. The new broom was great for removing the dirt.	Why does Marcus wash dishes?	To be kind.	To earn his pocket money.
pencil	paper	Before school Jack printed his report for class. He became angry when the pencil / paper got stuck and he couldn't print.			
kitten	puppy	Rover barks at all passing animals on the street. He's the most alert kitten / puppy in the whole neighbourhood in fact.	Does Rover make a lot of noise?	Yes, he does.	No, he is very quiet.
handsaw	hammer	In the evening Nicholas nailed two boards together in his garage. His new handsaw / hammer was proving a really useful tool.			
plate	glass	Tyler cautiously sipped his hot and freshly brewed green tea. He dropped the plate / glass because it was really too hot.			

boots	sheets	Yesterday Joshua slept well for many hours, never once waking up. His new boots / sheets were super cosy and very comfortable.			
saddle	paddle	Thomas and Charlie rowed along the river in the park. Thomas lost his saddle / paddle in the middle of the river.	Where did Thomas and Charlie row?	On the seaside.	In the park.
jogging	football	Freddie threw the ball to his friend during the game. He really liked jogging / football much better than any other sport.			
spanner	needle	Last night Jenny sewed patches onto her old jeans. Sadly, she lost the spanner / needle and could not finish her work.			
hand	foot	When she got annoyed, Charlotte stomped on the floor. It hurt when her hand / foot hit the ground with such force.			
minister	officer	Last night Bob arrested a mean criminal as everyone watched. As an able minister / officer he handled the situation very well.	Was Bob alone when he arrested the criminal?	No, everyone watched.	Yes, he was alone.
movie	novel	Annabelle sat outside and read about a man named Arthur. She loved the movie / novel about ancient times very very much.			
blower	shovel	When planting her tree, Elisabeth dug a hole in the ground. Her sturdy blower / shovel helped make the job much easier.			
shoes	kettle	Michelle poured a fresh cup of coffee for her friend. She dropped the shoes / kettle and got very annoyed with herself.			
kayak	taxi	Today Janet rode with her best friends to the shop nearby. The yellow kayak / taxi quickly got them to the store.	Where did Janet go with her friends?	To school.	To the shop.
bike	boat	They were all really happy as they sailed along the coast. The swift bike / boat raced near the very beautiful beach.			

Appendix Q. Reading behaviour questionnaire

Questions asked

At the end of all reading: Asked children about (conscious) reading behaviour
I will tell you some things that people do when they are reading. I would like you to think if you do this when you are reading, and then tell me if this sounds [PRESENT SCALE]
1 = Not like you
2 = A little like you
3 = A lot like you
If you don't think you can answer this, you can also say "I don't know" [point to box]

Practice item " I like ice-cream". Does this sound… (practice with answer scale until child has understood and can rate the practice item.

	Not like me	**A little like me**	**A lot like me**	**Don't know**
I reread words when I don't understand them.				
I reread each sentence before moving to the next.				
I look back at earlier sentences when I need to.				
I go back to words and sentences that I have found difficult to understand when I first read them.				

Scale shown

Not like me	A little like me	A lot like me
1	2	3

I don't know.