Measuring pragmatic language in children from diverse linguistic backgrounds

Lydia Eleanor Snowdon Gunning

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Psychology

November 2020

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First and foremost, I would like to thank my supervisors. I thank Mark Mon-Williams and Amanda Waterman for their invaluable advice, direction and enthusiasm. This thesis would not have been possible without them or my primary supervisor, Ekaterini Klepousniotou, who made Qatar seem like it was just down the corridor and who should really think about charging for her counselling skills. You, and the safe space your office created, made these last few years just that little bit easier.

I am so grateful to all the children who participated in this research and the schools who made me feel so welcome. I would also like to thank all the undergraduate students, Masters students and Born-in-Bradford interns who assisted in the data collection including Charlotte Holding, Imogen Crook, Jessica Booth, Chloe Bullock, Eloise Boden, Emily Earp, Sarah Nottingham, Charlotte Naylor, Jamie Barron, Laura Bricklebank, Rebecca Ritchie, Xinyi Dan, and Anqi Dai. I'd especially like to thank Annabel Kersley who made testing feel like a team effort.

To all the friends I have made in my time at Leeds who have been a constant support in both the ups and the downs of PhD life - thank you for never/rarely judging me for my level of tea and chocolate consumption when it all got a bit much, or for complaining about the sauna-like conditions of the office.

Thank you to my incredible family for understanding that my stress levels directly correlated with my Whatsapps, calls and visits home, and to Kate, for allowing me to whine about every minor inconvenience but not letting me lose sight of the small victories – thank you for being the family on my doorstep.

Thank you to Megan, who always knew the difference between when tea or wine was needed and who painstakingly taught me the correct placement of apostrophes (numerous times) - I truly wouldn't have got to this point without you.

Finally, this thesis is dedicated to my sister, Michelle, who will forevermore be my inspiration to see the good and persist. You are special.

Abstract

The extent of one's bilingualism affects various aspects of language development, however, little research has investigated how a child's language exposure impacts their pragmatic awareness. Chapter 2 of this thesis involved truncating the CELF-4 UK and developing two questionnaires. One enables researchers to quantify a child's language experience whilst mitigating the need for parental report, a factor which cannot always be relied upon in school settings. The second questionnaire gathers detailed information regarding a child's classroom-based language competencies from the child's teacher. All three measures enable quicker measurement of both language skills and language experience in the classroom. Chapter 3 collected decomposition scores for 121 idiomatic expressions as well as meaning-specific AoA norms for homonyms, metaphors, metonyms and idioms from an adult British-English sample. This allowed for stimuli selection in subsequent chapters. Chapters 4, 5, 6, and 7 concentrate on homonyms, metaphors, metonyms and idioms, respectively. These chapters explore the relationship between these specific aspects of pragmatic competence and language exposure and the degree in which this relationship may be moderated by factors associated with bilingualism, such as academic attainment, general language development, age and working memory. This involved the administration of a novel battery to 127 children in Years 3 (7-8 years) and 6 (10-11 years) of primary school. Results show age to be the most consistent predictor of pragmatic understanding but this effect to be highly moderated by socio-economic status. Finally, Chapter 8 compared children's knowledge of the four studied pragmatic tropes (homonyms, metaphors, metonyms and idioms), finding metonyms to be the most easily understood of the four tropes in both age groups. Taken together, this thesis adds to the limited information of how children from diverse language backgrounds understand different tropes of pragmatic language and the factors that mediate the relationship between language experience and pragmatic understanding.

Contents

List of Figures

List of Tables

Abbreviations

AoA – Age of Acquisition

BiB – Born in Bradford

LeBLEQ – Leeds-Bradford Language Exposure Questionnaire

LI – Language Impairment

L1 – First language

L2 – Second language

SES – Socio-Economic Status

SLI – Specific Language Impairment

SLT – Speech and Language Therapist

TD – Typically developing

1 General Introduction

The ever evolving, rich diversity of the UK in terms of culture and language is a factor that is largely reflected in schools. Approximately 20.6% of children speak a language in addition to, or other than, English in the home, and over 33.1% identify as being from an ethnic minority. Bilingualism and multilingualism thus make up a substantial part of our society with numbers of those speaking English as an additional language predicted to increase by a ratio of .5 percentage points per annum in the next decade (Department for Education, 2017). Consequently, in today's current climate, speaking more than one language is increasingly becoming the norm, rather than the exception (Harris & McGhee Nelson, 1992). Research with bilinguals, however, remains to generate inconsistent results (Adesope et al., 2010; Antoniou, 2019; Cummins, 1979; Grosjean, 2010; Romaine, 1995). While some literature reports bilingual advantages, other research demonstrates bilinguals to lag behind their monolingual peers and some research finds no differences. Such inconsistencies relate to domains such as phonetic learning (Antoniou et al., 2015), metalinguistic awareness (Ben-Zeev, 1977; Palmer, 1972), cognitive control (Bialystok et al., 2010; Morton & Harper, 2007) and academic attainment (Agirdag & Vanlaar, 2018; Whiteside et al., 2017) to name but a few instances.

This is likely due to bilinguals constituting an extremely heterogeneous population with language skills spanning the full continuum of proficiency (Strand et al., 2015). Bilingualism is highly complex with many impacting factors. It is well known that the quality, quantity and context of language directed to bilingual children is highly varied (De Houwer, 2007) and it is undisputed that it is these variations that affect the language development of bilingual children (Grüter & Paradis, 2014; Hoff et al., 2012; Paradis, 2011; Thordardottir, 2011; Thordardottir et al., 2006; Unsworth, 2013, 2015; Valdes & Figueroa, 1996). These factors include, but are not limited to cumulative exposure to the language in question (De Houwer, 2013; Thordardottir, 2011), order of acquisition (Hermanto et al., 2012), context of acquisition (Ben-Zeev, 1984) and quality of language input (Cornips & Hulk, 2008).

Due to bilingual children's linguistic knowledge being spread across two languages, a bilingual child's knowledge in either one of their languages cannot be assumed to be equal to a monolingual child's knowledge of their only language and differs on several dimensions (discussed separately in the sections below; e.g., Chondrogianni & Marinis, 2011; Gollan et al., 2002; Oller et al., 2007). Consequently, it remains to be advantageous to evaluate the language skills of bilingual children for a multitude of purposes, including educational placement, clinical diagnoses of language disorders and research purposes (often for ensuring inclusion/exclusion criteria are/aren't met).

The factors known to influence the extent of a child's bilingualism and their resultant effects on an individual's language development are described separately in the sections below.

1.1 Factors Affecting the Extent of Bilingualism

1.1.1 Age- and Order-of-Acquisition

Age and order of a child's language onset is a factor that has generated much controversy in child acquisition research (Herschensohn, 2007). While some children are exposed to more than one language at birth, others acquire their second language later in childhood. These children are referred to as simultaneous and sequential bilinguals respectively with literature often demonstrating behaviours of these bilingualism types to differ across tasks (De Houwer, 2013; Meisel, 2009; Unsworth, 2013).

When both languages are considered, simultaneous bilinguals have been shown to follow the same developmental course as monolinguals (Patterson, 1998; Pearson & Fernández, 1994) with research suggesting that, when learnt concurrently, languages are acquired relatively independently for the most part (Döpke, 2000). However, language acquisition in sequential bilinguals is increasingly complicated due to the need to consider the age in which the second language is acquired. A meta-analysis by Birdsong (2005) demonstrated that up to 65% of the variance in long term attainment in vocabulary and grammar can be attributed to the age in which a language was learnt. Likewise, Jia et al. (2002) reported a correlation between the age of English onset and long-term attainment on English grammaticality tasks when studying Mandarin-English bilingual children, with Bedore et al. (2012) positing that the age of onset of a language can account for about 35% of the variance in language dominance scores.

Furthermore, sequential bilinguals are often divided into early learners and late learners. The cut-off age for this divide is often between 4 and 5 years-old, with grammatical differences (Herschensohn, 2007), memory (Meisel, 2009) and knowledge of gender marking (Unsworth et al., 2014) becoming more advanced around this age. Those who have been exposed to language earlier are more likely to perform better on various language-based tasks than those who have acquired language later (e.g., Davison & Hammer, 2012).

1.1.2 Current Exposure vs Cumulative Exposure

Length of exposure refers to the time an individual has been learning a language; it has been shown to be a significant predictor of language performance (e.g., Davison & Hammer, 2012). Exposure is perhaps the most widely considered variable when researching bilingual populations. While earlier research simply focussed on a snapshot of language experience (De Houwer, 2009; Jia & Aaronson, 2003; Paradis, 2011), more recent research focuses on cumulative exposure (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020).

Cumulative exposure provides a more holistic view of a child's language history (Gutiérrez-Clellen & Kreiter, 2003; Unsworth, 2013), taking into account that an individual's relative exposure to each of their languages varies over the course of time. As such, this results in differing degrees of exposure to languages at different timepoints and developmental stages (Gutiérrez-Clellen & Kreiter, 2003; Thordardottir, 2011; Unsworth, 2013). This is particularly relevant in the case of sequential bilinguals whose onset of English may only occur when they are enrolled in childcare or begin schooling. In turn, this variable is highly related to the age of onset of a language, whether language acquisition was simultaneous or sequential, and whether they are early or late learners (as discussed above).

While current exposure to language, and years of exposure to a language are fair tools to gain an indication of language experience, cumulative exposure is considered a much better indicator (Unsworth et al., 2011), and is calculated by multiplying time spent in a certain environment (e.g., at home, at school, at child-minders) with the percentage of exposure to a language in such an environment (Unsworth et al., 2011; Unsworth, 2013). As a result, cumulative exposure to language has recently been widely used across the literature (e.g., Brandeker & Thordardottir, 2015; Sorenson Duncan & Paradis, 2020; Tao et al., 2019).

There are many standardised tools that allow for the measurement of cumulative language exposure. Whilst adults are normally tested using self-report measures (e.g., Li et al., 2006), children's cumulative exposure is usually assessed using parental report (e.g., Cattani et al., 2014; Hardin et al., 2013; Tuller, 2015). However, the latter must overcome many obstacles in the context of research including the low response rates that are often observed when testing in schools (e.g., Carnell et al., 2005) as well as the biases often observed in parents reporting the skill levels of their own children. This has been seen especially in lower socio-economic populations (Roberts et al., 1999) or for children that are known to have atypical language development (Paradis et al., 2010; Restrepo, 1998). Overall, then, language exposure determines bilingual language performance and needs to be captured when designing research on bilingual language ability.

1.1.3 Quality of Input

Bilingual children's language exposure varies not only in amount, as described above, but also in type. As such, both quantity and quality of language exposure matter (Paradis, 2011), with language presentation patterns in the home being the prominent factor in the development of bilingualism (De Houwer, 2007). Various factors are known to contribute to qualitative input differences in bilingual language acquisition, including richness of children's language input (i.e., the variety of sources from which a child hears the language; Jia & Aaronson, 2003; Jia & Fuse, 2007; Scheele et al., 2010). Such measures often include behaviours known to enrich language, for example, reading books, engaging in storytelling or role-play, watching educational TV or partaking in conversations with peers using the language in question.

Further influential factors include whether the child is spoken to in a standard or nonstandard variety (Cornips & Hulk, 2008; Larrañaga & Guijarro-Fuentes, 2013), the diversity of speakers providing the input (Place & Hoff, 2011), and whether this is by native or non-native speakers (Place & Hoff, 2011). For example, in a study on language practices, Paradis (2009) found over 90% of the parents in their sample were raising their children in one-parent one-language households. Such environments result in child receiving a large amount of input from native-speakers of each language with the idea that they will associate each language with a single parent, enabling the child to acquire two languages without confusion (Bain & Yu, 1980). However, while the one-parent one-language is a common approach in households with different language backgrounds (Barron-Hauwaert, 2004), it is not the only way to raise children whose parents are each native speakers of different languages. There are many other patterns of language use, including both parents speaking to the child in both language; their native language and the native language of their partner (i.e., their non-native language). This can become detrimental, however, when a parent chooses to speak their non-native language, in which they may have limited proficiency. This is especially true in the case of sequential bilinguals (Hammer et al., 2003), but also in the case of children from newcomer families (especially those of immigrant or refugee status) whose parents may also be in the process of learning the language themselves (Bohman et al., 2010). Consequently, the quality of the language received by the child may be of low quality, with less than perfect grammar or heavily accented speech, impacting their own language development.

1.1.4 Language Use

Finally, although in recent years the quantity and quality of a bilingual's linguistic input has gained more attention, research into a child's comparative output in their languages has been sparse (Bohman et al., 2010; Unsworth, 2015). This is despite research with bilingual adults demonstrating the use of language to be a central factor in determining the extent of bilingualism (Jia et al., 2002).

The consideration of a bilingual's output, however, is important as, without it, an individual becomes passive in that language, often having sufficient receptive knowledge, but losing expressive ability. For example, research has shown that even when a child is asked a question in their home language in the home environment, the child's response often tends to be in the language spoken in school compared to the home language (De Cat, 2020). In fact, few children use their home language to the same extent as they are exposed to it, being more passive in their home language. This trend is observed regardless of whether a child is a sequential or simultaneous bilingual (De Cat, 2020).

However, a bilingual's passivity in one language tends to lead them to become more proficient in their other language. Actively engaging in language allows focus to be placed on the rules and grammar of the language in question, aiding competence (Bonnell & Eison, 1991). Researchers have often advocated children's English output to be strongly correlated with vocabulary and morphology (Bohman et al., 2010; Paradis, 2011), with suggestions that a child's use of language is even more important than their input as it indicates active language use (Bohman et al., 2010). It has even been postulated that if a child does not learn to speak the language that is spoken in the home, this can detrimentally impact feelings of closeness and intimacy between parents and children (Fillmore, 2000; Portes & Hao, 1998). Subsequently, there have been more recent efforts to include a measure of output when assessing a bilingual child language exposure (Bedore et al., 2012; Bohman et al., 2010; Goldstein et al., 2010; Gutiérrez-Clellen & Kreiter, 2003) and more encouragement in the literature for parents to engage their children in the home language (e.g., Ribot et al., 2018).

1.2 The Impact of Bilingualism

1.2.1 Core Language

Exposure to language is known to strongly affect language growth in both monolingual and bilingual children (Hart & Risley, 1995; Hoff, 2003; Oller et al., 2007; Pearson et al., 1997) with the rate of vocabulary learning shown to be proportional to exposure (Chondrogianni & Marinis, 2011; Houwer, 2007, 2009; Duursma et al., 2007; Gathercole & Thomas, 2009; Hammer et al., 2008; Hoff et al., 2012; Pearson et al., 1997; Thordardottir, 2011; Thordardottir et al., 2006). Children exposed to more than one language receive less exposure to each of their languages compared to monolingual children of the same age, being unable to devote as much time to becoming proficient in each of their languages as they would if they were only learning one – in other words, they have to divide their time between two language. As a result, these bi-/multilingual children acquire each of their languages slower (at least initially) than their monolingual peers (e.g., Hoff et al., 2012) and this is considered to be the main explanation behind the vocabulary gap observed between monolinguals and bilinguals (Oller et al., 2007; Thordardottir et al., 2006).

Though it must be noted that good vocabulary skills are not completely uncommon in bilingual children (e.g., Uccelli & Páez, 2007), bilinguals often attain lower vocabulary scores in comparison to their monolingual, age-matched peers (Bialystok, 2001; Gollan et al., 2002; Hammer et al., 2008; Kimbrough Oller et al., 2007), with this gap being in danger of widening if the child does not receive sufficient support (e.g., Snow & Kang, 2007). This is a finding that has been demonstrated consistently amongst both pre-school children and children of school age but, interestingly, has been shown to differ across bilingualism types. For instance, when comparing the grammatical development of monolinguals to sequential bilinguals, the former outperformed the latter but not gap was observed when comparing the grammatical competence of monolingual and simultaneous bilinguals (Oller et al., 2007; Paradis & Genesee, 1996; Thordardottir et al., 2006). Despite these outcomes, recent literature by Babayiğit and Shapiro (2020) has suggested that a considerable proportion of bilingual children, regardless of their order of acquisition, still may not achieve native-like proficiency in English vocabulary or grammatical skills even by the time they leave primary school due to insufficient exposure.

In addition to the differences in the abilities of monolingual and bilingual children in terms of vocabulary and grammar, literature has also reported bilinguals to perform worse on tasks of semantic fluency (Gollan et al., 2002) and be slower at naming and identifying pictures (Gollan et al., 2005). Furthermore, the development of sufficient oral language has been identified as an important precursor of reading ability, especially during the initial phase of decoding letters into the sounds that form words (Whitehurst & Lonigan, 1998). As such, oral language ability in Reception class, when children start compulsory education, can predict later reading ability up until the fifth year of formal education (Duncan et al., 2007; NICHD Early Child Care Research Network, 2005). Thus, without this vital foundation, reading comprehension, and, ultimately, academic attainment is likely to suffer.

Oral language encompasses both receptive language (language c*omprehension*) and expressive language (language *production*). Monolingual and bilingual children have been shown to understand language better than they are able to produce (Benedict, 1979; Fenson et al., 1994; K. Nelson et al., 1978), a finding that seems relatively intuitive considering infants first demonstrate an understanding of words (receptive language) before they are able to produce them (expressive language; Benedict, 1979). Following this, a child's receptive vocabulary has consistently been shown to be larger than their expressive vocabulary (Fenson et al., 1994), a phenomenon referred to in the literature as the receptive-expressive gap.

However, although receptive and expressive language skills have been shown to be highly correlated in monolingual populations (Mashburn et al., 2009), the effect does not necessarily hold true for bilingual children who exhibit a more pronounced receptive-expressive gap compared to their monolingual peers (Gibson et al., 2012; Yan & Nicoladis, 2009). This phenomenon can be identified in pre-schoolers (Miccio et al., 2003), school-aged children (Gibson et al., 2012; Kan & Kohnert, 2005; Oller & Eilers, 2002; Yan & Nicoladis, 2009), and adults (Muñoz & Marquardt, 2003) alike, and gap is attributed to language exposure.

Gibson et al. (2014) studied Spanish-English 5-year-old bilingual children finding that the amount of language exposure a child received was correlated with the size of the gap between their receptive and expressive language abilities. Namely, less exposure to a language increased a child's receptive-expressive gap. This was further supported by research by (Hammer et al., 2008a), who demonstrated that bilingual children who have increased exposure to English prior to school entry show improved receptive language development in comparison to bilingual children who are not exposed to English until they enter school.

However, although the language of bilinguals has been shown to initially be behind that of their monolingual peers, their rate of development is reported to be faster than their monolingual peers. For instance, over a two-year period, (Hammer et al., 2008a) studied English-Spanish bilingual pre-schoolers who attended a Head Start program in the United States. They found that children who entered school with only Spanish exposure had a faster rate of English receptive vocabulary learning than children who entered with prior exposure to both English and Spanish. Nevertheless, the language gap between monolinguals and bilinguals tends to remain throughout childhood as, even though the rate of bilingual development is quick, bilingual children typically have a lower starting point. As monolinguals gradually also improve, bilingual children are essentially chasing a continuously moving target (Thordardottir & Juliusdottir, 2013). This is a factor that is reflected in the academic attainment of bilingual children, discussed below.

1.2.2 Academic attainment

The relationship between bilingualism and academic attainment is understood to be complicated with mixed results throughout the literature. There is an abundance of literature that has shown the differences between monolingual and bilinguals to be quite notable, with some research demonstrating monolinguals to outperform their bilingual peers (e.g., Agirdag & Vanlaar, 2018), while others report a bilingual advantage (e.g., Hutchinson, 2018; Whiteside et al., 2017).

The observation of monolinguals outperforming bilinguals is often referred to in the literature as an attainment gap. Research has suggested bilingual children to, on average, possess poorer literacy achievements when compared to their monolingual peers (Agirdag & Vanlaar, 2018; Demie, 2003) a trend that has been found not only in the UK (Runnymede Trust, 1998) but in the US (Lloyd et al., 2001) the Netherlands (Tesser et al., 1999) and Japan (Shimahara, 1991) also. This is notable as poor academic attainment, subsequently affects later employment outcomes (Currie & Thomas, 2001). Reports from the Department for Education (Department for Education, 2014a, 2014b, 2014c) revealed this attainment gap to be present across state-funded primary schools, in Reception (i.e., entry in compulsory education; ages 4-5 years-old) Year 2 (i.e., at the end of Key Stage 1 (KS1); age 6-7 years-old) and Year 6 (i.e., in Key Stage 2 (KS2) – the end of primary school; ages 10-11 years-old).

In contrast, research examining monolingual and bilingual children in Year 2, Whiteside et al. (2017) found bilinguals to be increasingly likely to meet academic targets in comparison to their peers who spoke only one language. Similar conclusions were drawn by Dowdy et al. (2011) and Halle et al. (2012) who found bilingual advantages in meeting curriculum targets at Year 2. Hutchinson (2018) demonstrated that children who speak more than one language are often some of the higher achievers within the classroom. However, they did mention that this effect was relatively heterogeneous. For primary school pupils specifically, this is significantly impacted by the language spoken within the home. For example, while some language groups (such as speakers of Pashto, Punjabi, Turkish, and

Slovak) perform below what is expected of them in primary assessments, other language groups (such as speakers of Chinese, Hindi and Tamil) perform *above* expectations.

An additional reason behind the opposing findings may be attributed to the ages of the sample, the controls in place and the designs of the research itself. For example, Strand et al. (2015) observed that only 44% of bilingual pupils obtained a good level of development (GLD) at the end of Reception compared to 54% of monolingual pupils. However, they found the bilingual children to make notably more progress in comparison to their monolingual peers, both between the ages of 7-11 and 11-16. They concluded that although bilingualism may be associated with lower achievement when starting school, this effect reduces with age and is largely eliminated by the time the child leaves compulsory education. The Bell Foundation (2018) supports this view of bilingual academic attainment, specifying that bilinguals indeed have higher rates of improvement across the years, potentially to the extent that bilinguals are later able to outperform their monolingual peers in some cases. After an extensive analysis of national assessment data, Strand et al. (2015) posited that the attainment gap between monolingual and bilingual children narrows as children progress through primary education, and is eliminated by the time children reach the age to leave compulsory education (15-16 years-old) where bilingual children demonstrate better attainment in some areas of the curriculum than their monolingual peers.

Thus, not only does the developmental trajectory of a bilingual child need to be considered, but so do various other contributing factors. For instance, bilingual pupils who are considered fully fluent in English have been shown to achieve significantly higher scores in KS2 SATs (Year 6) and GCSEs (Year 11) than their monolingual peers (Demie & Strand, 2006; Strand & Demie, 2005), a trend that has also been demonstrated in tests across primary school years (Halle et al., 2012; Winsler et al., 2014). Such research indicates language to be of crucial importance to school success (Pace et al., 2019). Without a high level of oral language proficiency in the language of teaching, students are unable to follow their teacher's instructions and participate in lessons (Baker, 2011). Such children will also have difficulties accessing what is referred to as the hidden curriculum, where lessons are learned despite not being openly taught in the classroom (Baker, 2011). Such gaps in the school experience leaves these children unlikely to make similar academic and language achievements to monolinguals (Babayiğit, 2015) or meet school demands (O'Connor et al., 2018). However, as children's oral language skills increase during primary and secondary education, with the gap between monolinguals and bilinguals receptive and expressive language skills tapering, so does the achievement gap between the academic scores of the two groups. This indicates a strong relationship between general language ability and educational attainment, with the scholastic test scores of bilingual pupils increasing with their degree of fluency.

In conclusion, it appears that speaking more than one language can indeed have a considerable impact on academic attainment, but a child's stage of development and general language skill must be taken into account in order to understand the full picture of monolingual and bilingual attainment. While bilinguals have been shown to underperform on some tasks of academic attainment, literature has shown that it is possible for these children to catch up, and in some cases, outperform those who only speak one language and such variances are thought to be attributed to the influence of fluency of the language used in education.

1.2.3 Executive Functioning

The practice of speaking two or more languages daily is a factor that has been shown to produce changes in cognitive performance (Bialystok, 2009). The process by which bilingualism causes this cognitive change is likely due to the need to focus attention to one language, even though the other language is activated to some extent during both comprehension and production (Blumenfeld & Marian, 2007; Rodriguez-Fornells et al., 2002; Thierry & Yan, 2007). Literature has found that there is a bilingual advantage in terms of various cognitive tasks, with bilinguals frequently exceling in tasks of executive functioning (Bialystok et al., 2009; Bialystok & Viswanathan, 2009; Blom et al., 2014; Martin-Rhee & Bialystok, 2008; Morales et al., 2013; White & Greenfield, 2017). Although there are numerous models of Executive Function that considerably differ in their definition, generally speaking, executive functioning includes processes that are related to goal-directed behaviour, or the control of complex cognition (Banich, 2009; Fuster, 1997; Lezak, 1995). Executive function also encompasses other tasks where the bilingual advantage applies, including tasks of enhanced inhibition (Bialystok et al., 2014; Calvo & Bialystok, 2014; Engel de Abreu et al., 2012; Poarch & van Hell, 2012; Prior & Macwhinney, 2010), phonological awareness (Bialystok & Herman, 1999), and task switching (Barac & Bialystok, 2012).

This bilingual advantage has also been found in terms of a child's working memory (e.g., Calvo & Bialystok, 2014). Literature considers working memory to be one of several disparate executive functions that control cognitive performance (Blair et al., 2005; Pennington et al., 1996; Zillmer & Spiers, 2001), with working memory typically being described as the system responsible for active maintenance and manipulation of information over brief time periods (Miyake & Shah, 1999). Working memory is often viewed as being part of larger memory architecture where information is perceived, attended to, and retrieved (Baddeley, 1986; Unsworth & Engle, 2007). However, the type of working memory task should be considered when studying child populations. For instance, Bialystok and Feng (2009) asked children to recall lists of words; they found no differences in the scores of monolingual and bilingual children. However, bilinguals are likely to be at a verbal disadvantage in comparison to their monolingual peers and can, consequently. struggle on tasks that encompass verbal skills, which could be confounding the results and potentially masking a bilingual advantage. For example, when working memory was measured using backward digit recall, Blom et al. (2014) reported bilingual participants to significantly outperform their monolingual peers on the working memory task.

1.3 Pragmatic Language Development

Spoken and written language and their associated components (i.e., receptive and expressive) are systems comprised of individual language domains (i.e., phonology, morphology, syntax, semantics, pragmatics) working together to form a dynamic integrative whole (Berko Gleason, 2005). It is the last of these domains, pragmatics, that is of particular interest of this thesis. Pragmatics is, by definition, an essential component of social communication (Baird & Norbury, 2016; Matthews et al., 2018; Norbury & Sparks, 2014) and refers to the ability of language to be both used and understood to enable effective communication in social interactions (Eigsti et al., 2011).

Though pragmatics encompasses non-verbal skills such as appropriate eye-contact, body language, and tone of voice (e.g., Gorman, 2018), the current thesis pertains to pragmatic skills that relate to language. Pragmatic language is deemed to be an important contributor to the development of communicative competence (Farmer, 1997; Leinonen et al., 2003; McKown, 2007) with enhanced pragmatic skills being positively associated with how well-liked a child is, both within preschool (Gertner et al., 1994) and in school-aged children (Place & Becker, 1991). However, Thompson (1997) argued that there are key aspects of pragmatic competence that distinguish it from all other areas of language. Unlike core language understanding, pragmatic competence requires an understanding of how language is construed and how it can be used, as well as knowledge of the ways in which we share the world with others. Pragmatics also requires picking up cues from others and understanding the rules which govern behaviour in a particular context. Essentially, pragmatic competence necessitates an individual to correctly understand speech and behaviours, to elicit appropriate speech and behaviours in response, to appreciate cultural norms and to inhibit inappropriate speech or behaviours. As such, children develop pragmatic language skills in much the same way that they acquire milestones in other areas of development with pragmatic abilities coinciding with oral language development.

The corresponding development of oral language skills and pragmatic language skills is likely due to the latter largely pertaining to words and phrases that are ambiguous (e.g., Powell et al., 2019). This ambiguity is often caused by words/phrases possessing multiple, often figurative meanings/senses. Fundamentally, knowledge of multiple meanings/senses of a word/phrase is vocabulary knowledge. Vocabulary has been defined as "*the stock of words used by or known to a particular people or group of persons*," where *word* is defined as "*a unit of language, consisting of one or more spoken sounds or their written representation, that functions as a principal carrier of meaning*" (Random House Webster's Unabridged Dictionary, Flexner, 2003). It is well documented that vocabulary is pivotal to children's development (e.g., Anderson & Freebody, 1981; Bleses et al., 2016; Duff et al., 2015; Nagy, 1988; Ouellette, 2006; Schut et al., 2017). Vocabulary forms the cornerstone for language and communication, predicting both concurrent and longitudinal reading and writing skills (e.g., Duff et al., 2015; Ouellette, 2006) as well as academic attainment (Bleses et al., 2016; Schut et al., 2017). Furthermore, the impact of children’s language knowledge spans far beyond their experiences within the educational system, and can predict later life chances (e.g., Spencer et al., 2017). This includes employment prospects, health outcomes and factors associated with criminal offending (Howlin et al., 2000; Rucklidge et al., 2013; Snowling et al., 2006). It is, therefore, vital to sufficiently identify pupils who may encounter difficulties with vocabulary.

However, much of the research that has explored vocabulary knowledge (also referred to as language *depth*) in children has focussed on children’s ability to map single lexical units onto single meanings (e.g., studies involving the British Picture Vocabulary Scale (BPVS; Dunn et al., 1997) or the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007)). Words or phrases possessing multiple meanings are likely to complicate vocabulary acquisition. This is due to the increased challenges that ambiguous words/phrases pose in terms of both the depth and *breadth* of vocabulary; not only whether an individual understands a word, but *how much* they know about it. Research assessing primary-school-aged children’s understanding of the multiple interpretations of words/phrases is, however, relatively limited. Consequently, of particular interest to the current thesis are two lexical-semantic aspects of pragmatics, namely ambiguity and figurative language.

1.3.1 Pragmatic Ambiguity

Linguistic ambiguity arises in instances where an utterance has more than one semantic interpretation. Early research by Bréal (1924 [1897]) stated that, although ambiguous words are highly pervasive in the English language, individuals rarely become confused by the multiplicity of meanings that a word can have, with ambiguity being a task handled reasonably effortlessly and unconsciously.

One theory for why this dispute is rarely encountered is that context informs the correct interpretation of the word in question (Simpson, 1994; Tabossi & Zardon, 1993), suggesting selective access to ambiguous meaning is determined by contextual constraints (MacDonald et al., 1994). Literature has consistently shown that a word is recognised more quickly if it is presented in a context which it is related to, rather than in a context when it appears out of place. This has been found for adults as well as children in lexical decision tasks (Schvaneveldt et al., 1977; Schwantes, 1981) and word naming tasks (Simpson & Lorsbach, 1983; Stanovich et al., 1981; West & Stanovich, 1978).

Other theories propose that when ambiguity is encountered, frequency is more influential than context, with the more frequent meaning/sense of a word being more likely to be activated (Swaab et al., 2003; Syssau et al., 2000). However, research demonstrates that children are less sensitive than adults to frequency effects (Levorato & Cacciari, 1992) possibly due to frequency being related to exposure to language. In particular, as younger language learners had less exposure to language than adults, they have less knowledge of which meaning of an ambiguous word might be the most frequent (Booth et al., 2006; Marmurek & Rossi, 1993; Simpson & Foster, 1986). This limited knowledge of language also means that children are less skilled at processing stimuli and require more assistance from prior information in lexical access than what is required by adults (Stanovich, 1980). Consequently, children are more dependent on context than they are on frequency of ambiguous words. In fact, they rely so heavily on the context of words in language that Friedrich and Friederici (2005) found that infants even as young as 19 months old produce N400 patterns akin to adults when they hear an unexpected word in an otherwise normal sentence.

There are three main models to explain ambiguity processing: the *selective access model* (Simpson, 1981), the *multiple exhaustive access model* (Swinney, 1979) and the *re-ordered exhaustive access model* (Duffy et al., 1988). The *multiple exhaustive access model* states all meanings of an ambiguous word are activated to an equal degree (Simpson, 1984). These parallel activations occur, initially irrespective of contextual bias with context only becoming important in subsequently stages of meaning selection and integration (Onifer & Swinney, 1981; Swinney, 1979).

Smith-Cairns et al. (2004) proposed that, as sentence comprehension is dependent both on the meanings of the individual words it comprises and how these words are organised, a listener must always refer back to their lexicon and apply their experience of grammatical knowledge on the utterance. As such, to detect that a sentence contains a word with two meanings, one must first generate both meanings of an ambiguous word. This idea is supported by research with both adults (Gaskell & Marslen-Wilson, 1997; Joordens & Besner, 1994; Plaut, 1997; Swinney, 1979) and children (Swinney & Prather, 1989). Numerous studies have supported this view of dual activation by demonstrating longer reading times for ambiguous words, whether it be in word-reading tasks (Duffy et al., 1988; Miyake et al., 1994; Rayner & Duffy, 1986) or priming studies (Onifer & Swinney, 1981; Swinney, 1979). The activation of both interpretations of the word means that two sentences are momentarily generated and held in working memory. The sentence that is most logical, whether this is due to experience or context, is selected while the other is discarded, though it still remains somewhat activated.

Alternatively, the *selective access model* holds that contextual clues guide access the relevant meaning of an ambiguous word and only the meaning that fits within the current context will be activated (Glucksberg et al., 1986; Perfetti & Goodman, 1970; Simpson, 1981). For example, in the context of "*my favourite animal is a bat*," only the meaning of *the nocturnal mammal* will be activated, not *the sport equipment that is used in cricket to hit the ball*. The selective access or context-dependent model is supported both by ambiguity detection techniques (e.g., Swinney & Hakes, 1976) and priming methods (e.g., (Simpson & Krueger, 1991; Tabossi & Zardon, 1993; Tabossi, 1988; Tabossi et al., 1987) which suggest lexical processing to be a highly interactive, top-down process. For example, Schvaneveldt et al. (1976) presented adult participants with a sequence of three words, the middle of which was ambiguous. Participants demonstrated faster reaction times in identifying the third word when the first and third words were associated with the second (e.g., SAVE-BANK-MONEY) in comparison to if the first and third words were related to opposing meanings of the second (e.g., RIVER-BANK-MONEY). However, one of the main principles of the selective access model is that there is no activation of the context-inappropriate meaning which research has shown not to be the case (e.g., Conrad, 1974; Marcel, 1980; Oden & Spira, 1983; Warren & Warren, 1976).

Consequently, the *re-ordered exhaustive access model* (Duffy et al., 1988) was developed taking both frequency and context into account. This view states that the multiple meanings of ambiguous words are exhaustively accessed but in the order of their meaning dominance (i.e., their frequency of usage in language). Accordingly, a biased context can increase or decrease the likelihood of the relevant meaning being initially activated. However, if this meaning is seen as incongruent with the given context, the second most frequent meaning will be accessed very quickly after that. This model also finds support from different methodologies, for example with ambiguity detection tasks (Hogaboam & Perfetti, 1975) and eye-movements (Duffy et al., 1988).

Due to all three of these processing theories requiring knowledge to be retrieved from vocabulary stores, it is unsurprising that the memory capacity of individuals can affect understanding of ambiguity (Miyake et al., 1994). Links have also been made between ambiguity detection and early reading skill. Hirsh-Pasek et al. (1978) were amongst the first to observe that poor readers struggle with detecting humour in jokes that play on ambiguity. Research has supported this ambiguity-reading link, with effects of such being witnessed even in the early years of schooling (Tunmer & Hoover, 2018; Wankoff, 1983) with the ability to report the dual meaning of ambiguous words being predictive of future reading skill even when assessing children who have not yet learned to read (Smith Cairns et al., 2004). The association between reading skill and pragmatic ambiguity resolution has been attributed to the metalinguistic nature of the two skills (Wankoff, 1983; Zipke, 2007).

Metalinguistic awareness has foundations in semantic, syntactic, and pragmatic knowledge and has been defined as “the ability to reflect on and manipulate the structural features of language” (Nagy & Anderson, 1995, p. 2). Phonological awareness is perhaps the most well documented form of metalinguistic awareness. Describing the ability to reflect on and manipulate the sound structure of spoken words, phonological awareness tasks are among the most stable predictors of both reading and spelling skills (Ehri et al., 2001; Goswami & Bryant, 1990). One of the lesser researched areas of metalinguistic awareness, however, is semantic awareness; how words are related to meaning. It is semantic awareness that is believed to contribute to the ability to disambiguate the meaning of words with multiple meanings (Zipke et al., 2009). Consequently, it is likely that reading and the ability to resolve ambiguity are linked due to both requiring sufficient metalinguistic skill.

Finally, some of the research comparing the processing of ambiguous and unambiguous words has reported an ambiguity advantage whereby individuals showed faster reaction times for ambiguous words than for unambiguous words in visual lexical decision tasks (e.g., Borowsky & Masson, 1996; Hino et al., 2006). Others, however, have found no such effect (e.g., Forster & Bednall, 1976; Gernsbacher, 1984).

These inconsistencies are likely due to the fact that ambiguity is not a homogeneous phenomenon. Linguistic and psycholinguistic literature makes a clear distinction between polysemic ambiguity and homonymic ambiguity (Frazier & Rayner, 1990; Klepousniotou, 2002; Klepousniotou & Baum, 2007; Williams, 1992). Homonymous words are those which have different meanings, and it is considered to be purely by chance that they share the same orthographic and phonological form (e.g., *bank* to mean both *the financial institute* and *the mound of land alongside a river*). Conversely, polysemous words are words that have more than one senses, but these senses are intrinsically connected (e.g., *I ate a juicy lemon* vs *I painted the nursery lemon*).

Both homonyms and polysemes are reasonably prevalent in everyday language (Parks et al., 1998). Approximately, 7.4% of the most frequent words correspond to more than one entry in the dictionary (i.e., are homonyms), while 84% of the entries have multiple senses (i.e., are polysemes) (Baayen et al., 1993). This demonstrates how most words in English are ambiguous in some form. However, although the distinctions between polysemy and homonymy are relatively easy to formulate, they can still be difficult to apply with consistency and reliability across the literature with disagreement often taking place in regards to just how related/unrelated words are (Kilgarriff, 1992; Lyons, 1981). This results in the two terms often being used interchangeably, and potentially incorrectly in the literature (Klein & Murphy, 2001) despite an abundance of emerging literature outlining the differences between the two tropes (e.g., Beretta et al., 2005; Klepousniotou et al., 2012). Both homonymy and polysemy are discussed separately below.

1.3.1.1 Homonymy

Homonyms are processed much in the same way as described above, with both context and frequency assisting in the selection of the appropriate meaning (such as described in the *re-ordered exhaustive access model* (Duffy et al., 1988). However, homonyms can be categorised as either balanced or unbalanced relating to the frequency and, consequently, the dominance of each meaning. While a balanced homonym has equally frequent meanings (e.g., “fan” to mean *a supporting person* or to mean *the object that moves air for cooling purposes*), an unbalanced homonym has one meaning that is more frequent than the other (e.g., “bank” where *the financial institute* is a more frequent meaning than *the land next to a river*).

Previous research has shown that it is not until the age of approximately 4 years old that children begin to consistently accept non-dominant interpretations of homonymous words (Backscheider & Gelman, 1995; Doherty, 2000; Garnham et al., 2000), suggesting metacognitive flexibility is required to judge multiple form-meaning mappings. Such research typically involves children being presented with a homonymous word and having to identify different interpretations from an array of pictures. However, literature has shown that if children are presented with an ambiguous word in a sentential context (e.g., “They could see the bank in the distance”), and asked to provide the two possible meanings, only children over the age of approximately 6 years-old could accurately detect the lexical ambiguity (Smith Cairns et al., 2004). The difference between these two tasks is that the latter is explicitly metalinguistic; it requires a child to focus on the dual meaning of the homonym and to reanalyse the sentence, a skill which pre-school children are not yet proficient at (Trueswell et al., 1999). Differences in findings are, thus, likely due to differences in task demands. This highlights the importance of consistent testing methodologies when comparing results across studies.

However, despite an abundance of research demonstrating bilinguals to have improved inhibitory control (Bialystok et al., 2012) and the associations between inhibition and ambiguity resolution (January et al., 2009; Novick et al., 2009; Ye & Zhou, 2009), there has been surprisingly little research directly comparing monolinguals and bilinguals on tasks of ambiguity. Furthermore, the little research available is contradictory. While some research has shown bilinguals to outperform their monolingual peers in homonym resolution even as young as children 3-5 years old (Diaz & Farrar, 2018), other research has found the converse to be true with monolinguals outperforming their bilingual peers

(Teubner-Rhodes et al., 2016). Kousaie et al. (2015), however, found no behavioural differences between the two groups. They employed a relatedness judgement task whereby participants read a sentence containing a homonym that was biased towards either the dominant or subordinate meaning; both behavioural and electrophysiological measures were employed. Kousaie et al. (2015) demonstrated electrophysiological, but not behavioural, differences between monolingual and bilingual young adults. In particular, they demonstrated that monolinguals relied on context more than their bilingual peers who simultaneously activated both meanings of the homonym. Kousaie et al. (2015) concluded that there were subtle differences in the processing of monolingual and bilingual individuals, but such differences were so slight that they were not reflected in the outcomes. As such, though monolinguals and bilinguals appear to perform similarly on tasks of homonym resolution, there are likely subtle differences at how they come to their conclusion. Chapter 3 of this thesis focuses on homonymy processing by monolingual and bilingual primary school children.

1.3.1.2 Polysemy

Polysemy, when a single word has two or more related senses, can be further divided into two types (Apresjan, 1974), metaphor and metonymy which shall be described separately below.

Metaphors are words which have the same spelling and pronunciation, but, in opposition to homonyms, the meanings of the senses are related. The first basic, sense of the word is often the more literal interpretation, while the second sense is figurative. As such, metaphor forms part of figurative language and has been used in figurative language research. Metaphors are a widely used tool to explain and understand complex topics (for instance, in textbooks to explaining scientific discoveries) or abstract ideas (for instance, in literature). They are also considered to be great persuasive tools, evident by the fact they are often used in contexts such as politics and advertising (Katz, 2017). In metaphors, two entities from distinct conceptual domains are linked due to common semantic properties (Van Herwegen & Rundblad, 2018). The current research focused on metaphorical words rather than metaphorical units (e.g., ‘time flies’). Children’s comprehension of metaphorical words is relatively under-reported in the literature in comparison to metaphorical units, adding further novelty to the current research.

Vosniadou et al. (1984) reported metaphor comprehension to be acquired by children as young as 3-4 years old. However, other research demonstrated the understanding of such words to be acquired considerably later, not before the age of 9 years old (e.g., Cometa & Eson, 1978). Such inconsistencies are attributed to the methods employed. In particular, while Vosniadou et al. (1984) used multiple-choice tasks, Cometa and Eson (1978) employed the use of open-ended questions. Consequently, it is evident that metaphor comprehension ability may depend on the method in which it is assessed with tasks that use multiple-choice being easier than tasks requiring some form of verbal explanation (Perlini et al., 2018; Pouscoulous, 2011, 2014). This pattern is understandable considering tasks involving verbal comprehension require a higher degree of metalinguistic competence (i.e., the ability to reflect upon and analyse language), expressive language skill and social abilities (Kwok et al., 2015; Lewis et al., 2007; McGregor et al., 2012; Nippold, 2016), making them more difficult.

In line with research on homonyms, however, there has been a small body of work investigating how monolinguals and bilinguals compare on tasks of metaphoric competence. Trosbrog (1985) compared the metaphoric interpretation abilities of monolingual and bilingual adults, finding no differences in comprehension of metaphors. However, monolingual learners were found to produce a higher number of conventional metaphors when compared to their bilingual peers which was attributed to monolinguals being more familiar with conventional metaphors due to increased experience of the language in question. Furthermore, while Johnson and Rosano (1993) also found no differences between monolingual and bilingual adult populations, they did find a significant positive relationship between metaphorical competence in bilinguals and their communication skills, as rated by their lecturers. Such null results, however, are in opposition to what might be expected when considering the aforementioned links between ambiguity resolution and inhibitory control (e.g., January et al., 2009) and the fact that bilinguals have been shown to have improved inhibitory control in comparison to monolinguals (Bialystok et al., 2012). Further research, thus, is necessary to understand better the influence of language experience on metaphoric understanding; this is the focus of Chapter 4 of this thesis.

The other type of polysemy is metonymy which is defined as "the use of a word or phrase to stand for a related concept that is not explicitly mentioned" (Shutova et al., 2013, p11). For example, the meaning of the word *cup* can denote either the container ("she was holding a yellow cup") or the contents ("the recipe called for a cup of milk"). Unlike metaphors, both the basic and secondary senses of a metonymic word are literal. Research reports metonyms to be the easier form of polysemy (Falkum et al., 2017) and recent experimental work has further demonstrated the differential nature of metonyms and metaphors (e.g., Klepousniotou, 2002; Klepousniotou et al., 2012; Klepousniotou & Baum, 2007; MacGregor et al., 2015). In a study by Rundblad and Annaz (2010) individuals from five years old to 37 years old were presented with a novel polysemic task where they were required to listen to 20 short stories containing either a metaphor or a metonym. Participants were then required to answer questions regarding their interpretation of the polysemic word. Unsurprisingly, overall adults performed better on the task than children. However, the findings also supported the idea that metonyms are the less complex form of polysemy with participants across all ages performing around 21% better on the metonymic stories compared to the metaphoric stories. Furthermore, metaphor comprehension was shown to develop at a much slower rate than metonymic knowledge.

Generally, research has shown that the development of metonyms begins relatively early in life, with some stating that children as young as three years old demonstrate some form of metonymic knowledge (Nerlich et al., 1999; Van Herwegen et al., 2013). However, there has been comparatively little research on the metonymic understanding in children. Thus, little is still known about the factors that contribute to poor metonymic comprehension and the populations that are likely to perform worse in tasks requiring metonymic comprehension. Chapter 5 investigates metonymy processing in monolingual and bilingual primary school children.

1.3.2 Figurative Language

A figure of speech is defined as "an expression that uses words in an unusual or nonliteral sense; the meaning is something other than what appears on the surface" (Yeomans et al., 1992, p.272). Figurative language is often subdivided into the five distinct components, namely metaphor, simile, sarcasm, indirect requests and idioms (Carrow-Woolfolk, 1999) which vary in their communicative function and comprehension demands (Colston & Gibbs, 2002; Nippold, 2016). For example, metaphors can be powerful tools used to achieve educational or political goals (Katz, 2017), while idioms add colour to language (Gillett, 2004) allowing language to be less formal and stilted (Whitford & Dixson, 1953). However, both types of figurative language share the fact that the words in the utterance do not overlap with the intended meaning of the phrase (Glucksberg, 2001) confirming the idea that when individuals are speaking figuratively, the intended meaning of their phrase is more than the literal use of their words (Gibbs & Colston, 2012).

Figurative language is a ubiquitous part of both spoken and written language, understood to be vital for successful social participation (Kerbel & Grunwell, 1997; Laval, 2003; Swineford et al., 2014) and establishing intimacy (Gerrig & Gibbs, 1988). Understanding of figurative language is also key for educational achievement (Cain et al., 2005; Kerbel & Grunwell, 1997; Nippold & Martin, 1989). To access lesson and national curriculum resources, children must understand figurative phrases, a requirement that increases in demand as they progress through education (Colston & Kuiper, 2002; Department for Education, 2013b). Though statistics are not available for English schooling, 6% of the sentences of reading programmes for primary schools in the United States are known to contain figurative expressions (Nippold, 1991), with clear positive relationships being found between non-literal comprehension and reading comprehension, as well as academic attainment (Fuste-Herrmann, 2009). Consequently, it is likely that children with poor figurative skills will find school increasingly challenging. The ability to identify such groups, thus, enables the access of resources, such as interventions, which can improve difficulties with figurative language and ameliorate general social functioning.

With figurative utterances comprising approximately 36% of a child’s and 25% of an adult’s exposure to language (Lazar et al., 1989; Van Lancker-Sidtis & Rallon, 2004), Carter (2004) posited that such expressions are one of the most common expressions of creativity in everyday communication, occurring across almost all contexts such as social conversation, teaching, media communication, music and lyrics, politics, newspapers, magazines, blogs and literature (Nippold, 2016).

Figurative language comprehension begins to develop in early childhood, alongside other aspects of linguistic development, progressively improving throughout childhood (Ozcaliskan, 2014; Pouscoulous, 2011, 2014; Vosniadou et al., 1984). Additionally, although forms of figurative competence have been demonstrated across all stages of development, general comprehension of figurative language is thought to rapidly improve between the ages of 7 to 11 years (Levorato & Cacciari, 1992; Levorato & Cacciari, 1995, 1999), becoming adult-akin around the age of 10 years old in typically developing populations (Winner, 1997). Like oral language, however, this development is known to continue even into adolescence and adulthood (Nippold & Duthie, 2003).

Regarding the processing of figurative language, a progression through stages has been suggested (Glucksberg et al., 1982; Norbury, 2005). Firstly, knowledge of individual words in the utterance must be accessed (Evans & Gamble, 1988; Jung-Beeman, 2005; Vosniadou, 1987). Both literal and figurative meanings of the speech can then be integrated (Glucksberg et al., 1982; Keysar, 1989), resulting in the perceived intended meaning being selected (Jung-Beeman, 2005) through a process of inhibiting all other unintended meanings (Glucksberg et al., 1982). However, there has been much debate over whether the literal interpretation of a figurative utterance has a role to play in processing.

Bobrow and Bell (1973) were amongst the first to propose the activation of the literal meaning of a figure of speech to be obligatory. They stated that, when learners encounter a figurative utterance, they will first interpret it literally, only considering the figurative interpretation if the literal meaning does not make sense within the context of the utterance. This view was supported by Grice’s Theory of Conversational Implicature (1989) which posits learners first consider the literal interpretation of an utterance as it is easier. Under this view, processing the figurative interpretation first is comparatively less informative, less clear and harder to access, thus, placing additional cognitive demands on the learner (Liontas, 2002). Consequently, the learner first analyses the literal interpretation, only inferring an alternative meaning if this interpretation is deemed to be contextually inappropriate.

In opposition to the above theories, the Direct Access View (Gibbs & Gerrig, 1989; Gibbs, 1982, 1989, 1994; Hoffman & Kemper, 1987) states that the more literal interpretation does not need to be always processed. Furthermore, though processing the figurative interpretation may take longer in some instances, this is not due to a preliminary stage of literal analysis, but due to increased complexity, or novelty, of the utterances.

Giora (1997), however, noted that this theory did not take into consideration the disparities in processing between familiar and unfamiliar figurative language, and developed the Graded Salience Hypothesis to resolve this oversight. The Graded Salience Hypothesis states it is the salient meanings that are first processed regardless of literality or contextual fit. As such, when alternative meanings are equally salient, both meanings are processed initially (Blasko & Connine, 1993). This corroborates research finding figurative language to be first processed figuratively in figuratively biased contexts (e.g., Katz & Ferretti, 2001), and literal language to be first processed literally in literally biasing contexts (e.g., Giora & Fein, 1999), which could not be explained by the earlier Conversational Implicature or Direct Access View theories.

1.3.2.1 Idioms

Alongside metaphor discussed above, idioms are one of the most extensively studied forms of figurative expression. English is considered a language rich in idioms (Brenner, 2011; Zyzik, 2011) with many advocating that without its rich use of idioms English would lose much of its diversity (De Caro, 2009). The ability to interpret idioms is considered to be a skill essential in the use of language (Conklin & Schmitt, 2008; De Caro, 2009; Sridhar & Karunakaran, 2013) with many advocating that the awareness, interpretation and correct usage of idioms is an indication of full proficiency in a language (Beloussova, 2015; Celce-Murcia, 2007; Saleh & Zakaria, 2013). The ability to produce such expressions is also linked with the increased ability to express thoughts and feelings (Ellis, 1997; Yorio, 1989).

Not only has there been much research on idiom usage in recent years, but there has also been an emergence of several idiom dictionaries enabling language learners to research the literal underpinnings and the origins of idioms (e.g., The Oxford Dictionary for Learners of English, Ashby, 2009; The Collins Cobuild Idioms Dictionary, Hands et al., 2012). Such dictionaries have been created with the assumption that idioms are learnt holistically (Wray, 2008) with increased knowledge of an idiom's origin allowing learners to better appreciate the connection between figurative language and the culture within which they exist (Liontas, 2017). The awareness of an idiom's literal underpinnings has consequently been proposed to aid in learners' retention of such expressions (Boers et al., 2004). There are two main theories that rest upon the latter assumption.

The first theory focuses on an idiom's transparency, stating that figurative expressions which are more clearly associated with their literal interpretations are likely to be acquired more easily (e.g., Steinel et al., 2007). Mantyla (2004) suggested that there are three categories of idiom transparency: transparent idioms, semi-transparent idioms, and opaque idioms. While transparent idioms have associated literal and figurative meanings (e.g., *to give the green light*), the literal and figurative meanings of opaque idioms are completely unrelated (e.g., *to be home and dry*). Finally, semi-transparent idioms have figurative meanings that are related to the literal underpinnings of the phrase, but not so linked that they can be described as transparent (e.g., *to quake in one's shoes*).

A different classification system of idioms was proposed by (Gibbs et al., 1989). According to the idiom decomposition hypothesis (Gibbs, 1992; Gibbs et al., 1997; Gibbs, Nayak, & Cutting, 1989; Gibbs, Nayak, Bolton, et al., 1989; Gibbs & Gonzales, 1985; Gibbs & Nayak, 1989), idioms can be characterised by how easily their meaning is related to the words of which the idioms are comprised. While decomposable idioms can be easily broken down so that the individual words relate to the overall figurative meaning of the phrase (e.g., *to save one's skin*), non-decomposable idioms cannot be analysed in such a way (e.g., *to kick the bucket*). It must be noted however that classification into decompositional and non-decompositional idioms is not simply binary, but a matter of degree with only a small number of idioms lying at the extreme ends of the decomposition scale (Gibbs, Nayak, & Cutting, 1989).

The second theory, known as the Dual Coding Theory, was proposed by Paivio (1986). According to this view, concepts which are considered to be more concrete are more easily retained than more abstract ones due to their increased level of imageability (i.e., how easy it is conjure a mental image of the referent). Although, in principle, the nature of idioms renders them all abstract, understanding the context of the literal sense in which they were once originally used is likely to induce concrete, mental images, leading to the idiom being more memorable (Cacciari & Glucksberg, 1995). Mental images can be generated for both decompositional and non-decompositional idioms using similar processes. Furthermore, after an idiom has been learned and stored as a lexical unit, individuals can access the figurative mental imagery of the idiom without the need to access its literal meaning (Hung, 2010). As such, concreteness is an additional factor that needs to be considered when researching idiomatic expressions.

Consequently, idioms that are less concrete and less decomposable are more difficult to link to their literal underpinnings and, thus, more easily misunderstood (e.g., Boers et al., 2004; Gibbs et al., 1989). Moreover, even if an individual is able to link an idiom's figurative speech and literal underpinning, challenges such as individual word comprehension still persist. This can be attributed to the idiom containing a key word that is not in an individual's vocabulary (e.g., *buck* in *to pass the buck*) or the phrase being complicated further by containing a form of ambiguity, making the phrase increasingly abstruse. For example, the homonym in the idiom *'to follow suit'* may be easily misinterpreted if the learner assumes it refers to an item of clothing.

Additionally, Mantyla (2004) observed speakers to often struggle when encountering idioms due to unfamiliarity. With idiomatic expressions having their roots in various areas of life, from work, to rural life, science and technology, literature, history and beyond, idioms can differ in their diversity and predictability of meaning (Stathi, 2006). This refers back to the frequency of specific idioms and how often they are encountered.

Notwithstanding the differences that are observed amongst individual idioms, there is a strong body of research in recent years regarding how children decipher idiomatic expressions and the age in which these skills develop. There has been a large discrepancy in the ages at which children acquire such expressions, however. Though the ability to understand and use short, literal sentences has been shown to be achieved by approximately 5 years of age in typically developing children (Nippold, 1991), non-literal language skills tend to be developed later, gradually increasing from the age of 5 years old and well into adolescence (Nippold, 2007; Nippold & Rudzinski, 1993). Research suggests that there is a four-stage process through which idioms are fully integrated into an individual’s repertoire (Caillies & Le Sourn-Bissaoui, 2013; Grunwell & Kerbel, 1996; Nippold & Taylor, 2002).

The first of these four phases is the identification stage whereby an individual is able to recognise a phrase as figurative because it does not make sense in the given context (Nikolaenko, 2004). This skill is said to begin developing around the age of five years-old and improves until the child is approximately twelve years old (Nikolaenko, 2004; Spector, 1996). The second phase is idiom interpretation whereby an individual can use contextual cues to interpret the meaning of the idiom, demonstrating comprehension across multiple scenarios (Cain et al., 2009; Norbury, 2004). This ability improves throughout adolescence (Nippold & Taylor, 2002). While research has shown children between the ages of 6-10 years to give plausible explanations of idiomatic phrases, these explanations not only are drawn purely from the contextual cues of the sentence in which the idioms are placed, but are most often inaccurate (Caillies & Le Sourn-Bissaoui, 2013; Cain et al., 2009).

Idiomatic explanation is the third phase, developing between the ages of 6-11 years old (Caillies & Le Sourn-Bissaoui, 2013; Grunwell & Kerbel, 1996; Chiara Levorato & Cacciari, 1995). In this stage, individuals acquire the ability to describe the meaning of an idiom across multiple contexts where the phrase may have different interpretations (Caillies & Le Sourn-Bissaoui, 2013; Le Sourn-Bissaoui et al., 2012; Whyte et al., 2013).

The fourth and final phase of idiomatic understanding is idiomatic use where individuals are able to correctly use an idiom in the appropriate context (Nesi et al., 2006). This is the stage with the least associated literature but is believed to develop last. Overall, the last two stages, namely explanation and use, are deemed to be expressive skills, while the first two stages, identification and interpretation, are more receptive (Benjamin et al., 2020).

However, much of the literature into the development of idiomatic understanding does not consider the stages in which idioms are acquired. Due to the lack of a standardised assessment that can comprehensively measure idiomatic understanding, researchers are required to develop their own materials when investigating idioms. These materials tend to assess the idiom acquisition of different children, employ different stimuli selection procedures, use different methodologies, and assess different stages of children's idiomatic development (i.e., identification, interpretation, explanation or use); yet, results are compared. Consequently, it is unsurprising that research reports inconsistent results regarding the idiomatic understanding of children. For example, Levorato and Cacciari (1999) found that 9-year-olds, but not 7-year-olds, could take advantage of an idiom's transparency in order to understand its meaning. Cain et al. (2009), however, found 7-year-olds capable of understanding transparent idioms in similar contexts. Consistency in the measurement of idiomatic language, thus, remains to be vital in order to ensure the reliability of the results.

Furthermore, despite the idea that idioms are learnt holistically (Wray, 2008), they are rarely included in frequency or age of acquisition lists and limited studies have attempted to measure how idiomatic vocabulary develops across childhood. This is despite the fact that many idioms appear to be constituted of high frequency, early learnt components. Such considerations are rarely standardised across experiments, yet again adding to discrepancy in results found across the literature. With even typically developing, native English speakers demonstrating difficulty regarding the use and interpretation of idioms (Mantyla, 2004; Sornig, 1988), it becomes increasingly detrimental when such inconsistent results are compared to other populations, for example children with SLI or ASD (who the literature has shown to be disadvantaged in such task (e.g., Kerbel & Grunwell, 1997; Norbury, 2004) or children who speak more than one language.

With regards to bilinguals, their ability to understand idioms is further complicated due to idioms being culture specific (Glucksberg, 2001). Not only do differing cultures boast differing idioms, but individuals learning idioms outside of their native language are prone to misinterpreting idioms in their second language due to knowledge and associations that are not shared across cultures. For example, research by Hu and Fong (2010) demonstrated individuals whose first language was Chinese to habitually misinterpret idioms containing the words heart and mind. The authors attributed this to the native-Chinese speakers having limited knowledge about western views of emotion where the heart is often used to describe emotion whereas the use of mind signifies some form of reasoning. As this duality is absent form Chinese culture, it is consequently unsurprising that individuals whose first language was Chinese had difficulty with such phrases.

Idioms have consequently been shown to pose comprehension problems for adult bilinguals in terms of learning, comprehension and production (Abel, 2003; Charteris-Black, 2002; Cieślicka, 2006; Conklin & Schmitt, 2008; Durrant & Schmitt, 2010; Nekrasova, 2009; Tabossi et al., 2008; Wray, 2008). For instance, Cieślicka (2006) conducted a cross-modal priming study whereby Polish-English bilinguals were instructed to listen to sentences containing English idioms and look at visual targets. They were instructed to make a lexical decision response. Researchers found participants to demonstrate more priming for literal targets in comparison to idiomatic targets, suggesting participants to have difficulty understanding idiomatic expressions in their second language. In fact, bilinguals have been shown to be prone to interpreting phrases literally despite contextual mismatch (Bishop, 2004), with many not even recognising the gaps in their own idiomatic knowledge (Martinez & Murphy, 2011). This becomes increasingly problematic as idiomatic phrases are rarely explicitly taught within the educational context. However, research regarding the degree of this association between idiomatic understanding and language experience is relatively sparse especially for bilingual child populations most likely due to the limited tools available to assess understanding. Chapter 6 addresses idiom processing.

1.3.3 Pragmatic Assessment

There are standardised pragmatic assessments in distribution that solely measure pragmatic competence (e.g., the Test of Pragmatic Language; TOPL-2; Phelps-Terasaki & Phelps-Gunn, 2007). Furthermore, there are various comprehensive batteries that contain subsets intended to capture pragmatic development such as the Pragmatic Judgement subset of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999) and the Pragmatic Profile subset of the Clinical Evaluation of Language Fundamentals-4 (CELF-4; Semel et al., 2003). However, while such subsets do capture some aspect of pragmatic competence, they do not do so in a comprehensive manner, failing, thus, to provide a full picture of pragmatic competence.

Available standardized measures can reliably identify children's pragmatic language impairment (Bishop & Baird, 2001; Geurts et al., 2004; Laws & Bishop, 2004; Norbury et al., 2004; O'Neill, 2007; Young et al., 2005) with much of the literature focussing almost exclusively on special populations. For example, previous literature has largely focussed on children with autism (e.g., Martin & McDonald, 2003; Young et al., 2005), ADHD (McKown, 2007), conduct disorder (Gilmour et al., 2004) or Williams Syndrome (Gillberg & Rasmussen, 1994; Laws & Bishop, 2004). Presently, however, there are no social pragmatic assessments standardized for use with bilingual and multicultural children. Furthermore, all existing pragmatic assessments offer a fine-grained profile of a person's communicative skills meaning they are usually quite lengthy (90 minutes on average), often presenting difficulties in administration and scoring. Consequently, their use is not always feasible within the classroom environment when taking children away from lessons, or for research purposes where children need to be tested on a large scale.

A short and concise pragmatic assessment tool that could be used to screen efficiently and effectively for difficulties in homonyms, polysemes (metaphors and metonyms) and idiomatic expressions would, thus, be a highly valuable tool within both academic and research environments. Though the creation of pragmatic language assessment tool is not a primary focus of this thesis, the development of such a tool has begun in this thesis as a by-product of exploring language abilities and its use shall be discussed in the following chapters.

1.4 The Born in Bradford Cohort

Most of the studies constituting this thesis are nested studies conducted within the Born in Bradford (BiB) cohort Raynor et al., 2008). Based in the city of Bradford, West Yorkshire, the cohort is situated in the sixth largest metropolitan district in the UK (Wright et al., 2013) and was initially established in response to the alarmingly high rates of childhood mortality and morbidity observed within the area. BiB has since become one of the world's largest research studies, combining detailed information from various healthcare and education records, and subsequently capturing each child's socio-economic characteristics, ethnicity, lifestyle, environmental risk factors, and physical and mental health (Raynor et al., 2008). However, the reach of the cohort goes far beyond its longitudinal research and the 13,776 mothers recruited from the maternity wards of Bradford Royal Infirmary between March 2007 and December 2010. In the last decade, numerous links have developed between the cohort and schools within the city leading to more expansive research networks. Much of the work described in this thesis benefits from these pre-established links, working with children within Bradford schools, rather than solely the children within the cohort.

Interestingly, in the last 60 years, Bradford has experienced both large socio-economic decline and unprecedented migration patterns. These migration patterns that once were highlighted as prodigious, however, are akin to those observed in various areas throughout the world today. Subsequently, in numerous ways, Bradford can be seen as a superb microcosm of the pressures placed on cities through notably high rates of bilingualism and low Socioeconomic Status (SES), both of which are factors that impact language development, making the cohort a particularly interesting point of study for the current research.

1.4.1 Ethnicity and Bilingualism within the Cohort

In addition to being among the most deprived areas in the UK, Bradford has some of the highest numbers of bi or multi-lingual inhabitants making it a superb area of study for language research. The national average of primary school aged children from an ethnic minority is estimated to be at approximately 32.1%, with 33.3% of these children being of Asian heritages (Department for Education, 2017). In Bradford, these numbers are substantially higher, with 60.0% of children in the city being from ethnic minorities and 74.1% of these being form Asian heritages. Moreover, within the city, 42.3% of primary school aged pupils are exposed to a language known, or believed to be, other than English in their home, substantially higher than the national average of 20.6% (Department for Education, 2017).

With approximately 153 languages spoken throughout the city (Department for Education, 2016), Bradford is known to have the highest number of bilingual dominant schools in the North of England (NALDIC, 2015). Moreover, the linguistic make-up the cohort is also quite different to the make-up of most studies exploring the effect of language experience. While the typical participants in studies exploring bilingualism tend to be either Spanish-English speakers from the USA (e.g,. Baron et al., 2018; Oller & Eilers, 2002), Basque-Spanish speakers from Basque County (e.g., Larrañaga & Guijarro-Fuentes, 2013; Pérez-Tattam et al., 2019) or Welsh-English speakers from Wales (e.g., Chondrogianni & Kwon, 2019; Gathercole et al., 2016), most of the bilinguals in Bradford speak south

Asian language. This includes Punjabi, Urdu, Bengali, Pashto, Gujarati, and Arabic (Bradford Census, 2011).

As such, the bilingual population within the city tends to be relatively heterogeneous with vastly different language experiences. Consequently, studying the differences between these bilinguals, and specifically how they vary from their monolingual peers who are exposed to the same level of education and have similar socio-economic backgrounds, can enable better understanding of how language experience affects language development. Additionally, research with more diverse populations allows for a broader perspective of how language experience impacts children.

1.4.2 SES within the Cohort

A review by Johnson and Kossykh (2008) identified the most notable influencers of early child development to be parental education, the home learning environment and quality of pre-school education. These all reflect one's Socio-economic status (SES). SES is considered a multifaceted compound, comprising of education, occupation, and income (Kohn, 1963). In child development research, maternal education is the most commonly used proxy of SES (Ensminger & Fothergill, 2003) due to the idea that mothers' language use and knowledge of child development varies according to their educational level (Hoff, 2003; Pan et al., 2005; Rowe, 2008). More recently, research has determined that children's SES can be identified on the basis of the school they attended (Balladares et al., 2016) and eligibility for state/government support (Suskind et al., 2016), such as entitlement to free school meals (Locke et al., 2002). In the UK, entitlement to free school meals coincides with those who are entitled to Pupil Premium. The Pupil Premium initiative was introduced in the UK in April 2011 to help close the SES attainment gap (Abbott et al., 2015) and improve identification of children from low SES backgrounds, in order to provide extra resources to aid their academic development.

Bradford's urban areas are among the most deprived in the country, with the city, as a whole, being ranked as the 18th most deprived in the UK on the Index of Multiple Deprivation 2015 (Department for Communities and Local Government, 2015). The Indices of Deprivation are based on 37 indicators, organised across seven distinct domains of deprivation[1]. These domains are then combined, using appropriate weights, to calculate a city's ranking. Additionally, every *neighbourhood* in England is ranked according to its level of deprivation relative to that of other areas. In this instance, Bradford ranks 11th in the ordering of cities in the UK which contain the highest number of deprived neighbourhoods, with 32.6% of the city's neighbourhoods residing in the 10% most deprived neighbourhoods in the UK, further emphasizing the notably low SES of the city (Department for Communities and Local Government, 2015).

Another indicator of SES used in the literature (e.g., Babayiğit, 2015) relates to Free School Meals entitlement, also known as Pupil Premium. This initiative, which is a government scheme, allows any child with a parent or carer in receipt of means-tested benefits to be entitled to free school meals[2]. Across the UK, approximately 14.7% of primary-school pupils are known to be claiming free school meals (Department for Education, 2017). In Bradford, however, this number is considerably increased with 17.2% of children qualifying for Pupil Premium, further demonstrating the high level of deprivation across the city.

Consequently, due to the overarching deprivation observed within the city, the Government has recently highlighted Bradford as an Opportunity Area (Department for Education, 2016). The primary goal of the Opportunity Areas is to focus local and national resources on increasing social mobility (predominantly through education) in areas most in need of additional support. Bradford is one of 12 Opportunity Areas within England with the key aim of increasing young people's knowledge, skills and opportunities to ultimately help them reach their full potential.

The socioeconomic make-up of the cohort remains vital to the study of language development due to its impact on such. Conger and Donnellan (2007) state that SES is linked to child development through a series of parenting behaviours and child-rearing practices. Mothers from higher socio-economic strata speak

---

[1] These seven domains include: Income Deprivation; Employment Deprivation; Health Deprivation and Disability; Education, Skills and Training Deprivation; Crime; Barriers to Housing and Services; and Living Environment Deprivation

[2] These benefits are: Income Support; Income-based Jobseekers Allowance; Income-related Employment and Support Allowance; Support under Part VI of the Immigration and Asylum Act 1999; the guaranteed element of State Pension Credit; Child Tax Credit; Working Tax Credit; and Universal Credit.

markedly more to their children than mothers from lower socio-economic groups with the quality of their speech has also been found to be conceptually different. Higher SES mothers have been observed to engage in more conversational speech that would elicit a response from the child and less speech focussed on directing their child's attention to objects (Hoff-Ginsberg, 1998; Hoff-Ginsberg, 1991; Hoff, 2003, 2006; Hoff et al., 2002; Huttenlocher et al., 2010). Furthermore, the size of these SES effects are substantial with significant differences across families from low-, mid- and high-SES strata (Hart & Risley, 1995), highlighting the influence of SES on development.

Moreover, children from higher SES environments are read to more frequently than children in lower SES families (Fletcher & Reese, 2005), with more syntactically complex and lexically-rich interactions occurring in such instances (Hoff-Ginsberg, 1991; Snow et al., 1976). High SES parents have been associated with a parenting style characterised by more warmth and responsiveness, and less parental stress, assisting in making the home language environment more conducive to language learning (Evans et al., 1999; Perkins et al., 2013). High SES parents also have increased time and financial resources available for their children (Sohr-Preston et al., 2013) resulting in more opportunities to receive linguistically rich input and be exposed to more literacy related materials, supporting both language acquisition and, as a consequence, educational attainment (Bradley & Corwyn, 2002; Hoff, 2006).

Subsequently, parental SES impacts a child's environment, affecting their language input and experiences which, in turn, impact their own development. One of the most notable factors affected are lexical skills. There is a substantial body of research that has evidenced a strong association between higher SES and more advanced language skills, with the most widely reported observation being a child's vocabulary (Fernald et al., 2013; Hart & Risley, 1995; Hoff-Ginsberg, 1998; Hoff, 2003; Huttenlocher et al., 2010; Oller & Eilers, 2002; Pan et al., 2005; Rescorla & Alley, 2001; Rowe, 2008). Hoff and Tian (2005) found SES to account for as much as 5% of the variance in the size of the vocabularies the children used. SES has also been associated with grammatical development with children living in higher SES households producing longer, more complex responses to speech (Arriaga et al., 1998; Snow, 1999), performing significantly better on tasks of receptive and expressive syntax (Huttenlocher et al., 2002) and standardised measures of general grammatical competence (Dollaghan et al., 1999). High SES has even been linked with reading ability, accounting for approximately 5% of the variance in school achievement (White, 1982).

Finally, effects of SES have been shown to be both robust and substantial regardless of the measure of SES used. Parental education, employment status, household income, school attended and entitlement to government subsidiaries have been employed across the literature either individually, in combination (e.g., Bradley & Corwyn, 2002) or as composite scores (e.g., Gatt et al., 2020) with the relationship between SES and a child's language environment, and resultant linguistic skill, consistently being demonstrated (Suskind et al., 2016).

However, this association between SES and linguistic ability is more complicated in the case of bilingual children, like those contained within BiB and, consequently, the current thesis. Not only does the external factor of SES have a significant impact on the development of a child's first language (L1) in terms of vocabulary, syntax and literacy skills (Duncan & Seymour, 2000; Hart & Risley, 1995; Hoff, 2003; Huttenlocher et al., 2002; Netten et al., 2016), but similar findings have been demonstrated for a child's second language (L2) also (Golberg et al., 2008; Hammer et al., 2012; Hoff, 2006; Scheele, 2010). Consequently, while living in less affluent households remains to be detrimental for both monolingual and bilingual children alike, bilinguals from lower SES families risk facing a double disadvantage (Hoff, 2013; Neuman et al., 2018).

Literature reported differences in the developmental trajectories of high SES children and low SES children depending on their language experience. For instance, Strand et al. (2015) reported that low SES bilingual children scored lower on KS1 SATs than those of higher SES. However, the difference in scores was smaller in bilingual children than between low SES and high SES monolingual children. Furthermore, there have also been reports that the effects of bilingualism may be more pronounced at some SES levels than at others (Woodard & Rodman, 2007). However, findings must be interpreted with caution as differences in linguistic abilities have been observed between ethnicities too (Mount-Weitz, 1996; Roberts et al., 1999) and much of the bilingual literature uses multi-ethnic samples. Furthermore, research has shown bilingual children to be more likely to come from households of lower SES and to be less likely to attend early childhood education programmes (Karoly & Gonzalez, 2011), further disadvantaging their development.

1.5 Summary Thesis and Outline

The extent of an individual's bilingualism is known to affect various aspects of language development (e.g., Hoff et al., 2012; Paradis, 2011; Thordardottir, 2011; Unsworth, 2013, 2017), yet little research has been conducted on how the degree of exposure to a language impacts pragmatic awareness. The research contained within this thesis explores pragmatic awareness in the form of a child's understanding of meanings/senses of a word/phrase. Fundamentally, this relates to vocabulary knowledge. It is well documented that vocabulary is vital to children's development, impacting reading and writing skills (e.g., Duff et al., 2015; Ouellette, 2006), and, resultantly, academic attainment (Bleses et al., 2016; Schut et al., 2017). However, ambiguous words/phrases complicate vocabulary acquisition, posing a challenge for both the breadth and depth of vocabulary knowledge. Furthermore, bi-/multi-lingual children already have less vocabulary knowledge within each of their individual languages compared to monolinguals, a factor attributed to their exposure being distributed across languages (e.g., Bialystok et al., 2010; Farnia & Geva, 2011). Therefore, the challenge imposed by words/phrases with multiple meanings is likely more pertinent for children who speak more than one language. Consequently, the chapters contained in this thesis attempt to explore the relationship between specific aspects of pragmatic competence and language exposure. In particular, homonyms, metaphors, metonyms and idiomatic expressions were investigated. In addition, the studies contained within the thesis investigated the degree to which this relationship may be moderated by factors which are also known to be linked with bilingualism, such as academic attainment, general language development, age and working memory.

Chapter 2 describes the development of pre-requisite materials required for the more experimental work described in later chapters of the thesis. The first of these measures was a general language battery developed to be used in place of the gold-standard The Clinical Evaluation of Language Fundamentals: Fourth Edition (CELF-4 UK; Semel et al., 2003). The novel battery encompassed many of the same items contained in the CELF-4 but was reduced in such a way that it could be administered in half the time. This was deemed necessary to avoid fatigue of the participants and was possible as the battery was not intended to be used for diagnostic purposes, but rather to gain an indication of a child's core language skills, their receptive and expressive language abilities, in addition to an indication of their working memory.

The second set of materials were two questionnaires developed out of necessity. The first of these questionnaires allowed for the quantification of the bilingual language experience and consequently determined the degree to which pragmatic competence is reliant on language exposure. This questionnaire was designed to be used with child populations in situations where parental report cannot be relied upon. The second questionnaire works in collaboration with the children's language exposure questionnaire, gathering detailed information from teachers with regards to a child's classroom-based competencies.

Chapter 3 details two norming studies. The first of the norming studies gathered decomposition scores for 121 idiomatic expressions. The second of the surveys generated of Age of Acquisition (AoA) norms for homonyms, metaphors, metonyms and idioms. Obtaining such information was necessary in order to select appropriate stimuli for the experimental studies.

Chapters 4, 5, 6 and 7 describe experimental research into pragmatic abilities, namely homonymy, metaphors, metonyms and idiomatic expressions respectively. The homonyms chapter (Chapter 4) contains both balanced and unbalanced homonyms. Chapters 5 and 6 discuss research on polysemy, with the former discussing metaphors and the latter discussing metonyms. The chapter of idiomatic expressions (Chapter 7) contains both decompositional and non-decompositional items in the same study. The chapters aimed to determine the degree of the relationship between performance on these pragmatic tasks and language experience, as well as investigate the factors that mediate this relationship, such as general language ability, academic skills, working memory and SES. Due to the lack of available assessments, novel assessments were created that could assess both monolingual and bilingual children on multiple tropes of pragmatic knowledge. Stimuli were carefully selected in each of the pragmatic tasks for comparison across subsets.

All four experimental studies were developed to contain both receptive and expressive components, thus enabling for pupils' pragmatic scores to be compared to the Receptive and Expressive scores generated by the language battery developed in Chapter 2. Assessing accuracy rates across different methods of response also allowed for a wealth of data to be obtained. While the receptive component comprised of multiple-choice questions, the expressive component was open-ended, providing additional insight into the processing of lexical-pragmatics. Including both an open-ended and multiple-choice component followed the same protocol as similar language measures, such the CELF5-Metalinguistic (Clinical Evaluation of Language Fundamentals, Fifth Edition, Metalinguistics; Wiig & Secord, 2014) and the TLCE (Test of Language Competence - Elementary; Wiig & Secord, 1989). This methodology enabled within-group comparisons in order to determine at which method of assessment children performed better, and whether this was consistent regardless of a child's age or language experience. While the Open-ended component is likely more challenging, literature has shown children over the age of approximately 6 years-old are competent when assessed in this way (Smith Cairns et al., 2004). Possessing sufficient understanding of the advantages of each method of assessment is crucial in research contexts when selecting the most appropriate method to use for a particular population.

Finally, in Chapter 8, children's performance across all four tasks is compared. The choice of tasks in the present thesis allowed for lexical-pragmatics to be considered from the word level (homonyms) to the sentence level (idioms), with polysemes (metaphors and metonyms) acting as a mediator between the two.

2 Development of Materials

The current chapter describes the development of pre-requisite materials required for the more experimental work described in later chapters of the thesis. Details of how the gold-standard language assessment battery the Clinical Evaluation of Language Fundamentals: Fourth Edition (CELF-4 UK; (Semel et al., 2003) was truncated will be discussed (Section 2.1). The development of two questionnaires used to quantify language exposure (Section 2.2.1.1) and gain an indication of language proficiency (Section 2.2.1.2) will then be presented.

2.1 Adapting and Streamlining the CELF-4 UK for More Time-Effective Testing in Schools

The communication abilities of children have long been an area of interest in educational settings. Deficits have increasingly been recognised as a public health issue, with school staff stating that approximately half of the children who start school possess insufficient communication skills that would enable them to learn in the classroom environment (Bercow, 2008). This is particularly notable as language difficulties are known to result in long-term disadvantages (Law et al., 2009). Children whose language does not develop at a typical rate may be at risk of a Specific Language Impairment (SLI) diagnosis and are likely to struggle with tasks of phonological short-term memory (Gathercole & Baddeley, 1990), verbal memory (Weismer et al., 1999), and grammar (Leonard & Bortolini, 1998). A broad term, SLI has high comorbidity with various other disorders/delays such as Autism (e.g., Bishop, 2010), Attention Deficit/Hyperactivity Disorder (ADHD; e.g., Hutchinson et al., 2012) general speech delays (e.g., Shriberg et al., 1999) and dyslexia (e.g., Vandewalle, et al., 2010).

However, a factor that can complicate the diagnosis of SLI is bilingualism with confusion of what a ‘typical’ bilingual should look like often leading to misdiagnosis (Bedore & Peña, 2008; Grimm & Schulz, 2014; Salameh et al., 2002). Generally, when only one of a bilingual child's languages is tested, performance on various language assessments tends to be substantially lower than that of their monolingual peers (Thordardottir et al., 2006). This trend led many researchers to conclude that for a more comprehensive view of bilingual children’s language ability, individuals should be tested in each of their spoken languages (American Speech-Language-Hearing Association, 2004). This is indeed feasible in many contexts where the bilingual sample is homogenous and the languages are well-documented, such as Spanish-English samples in the USA (e.g., Baron et al., 2018; Oller & Eilers, 2002), Basque-Spanish bilinguals in the Basque County (e.g., Larrañaga & Guijarro-Fuentes, 2013; Pérez-Tattam et al., 2019) and Welsh-English speakers in Wales (e.g., (Chondrogianni & Kwon, 2019; Gathercole et al., 2016). However, due to the large diversity of languages spoken in many countries, and even cities, assessing both of the languages of a child is not always practical. Although this is predominantly due to the availability of bilingual speech pathologists or skilled interpreters, limited financial resources and instruments are also clear obstacles.

Furthermore, when assessing the language ability of bilingual children, the importance of appropriate bilingual norm categories has been emphasised (Gathercole et al., 2013). This is vital to distinguish not only between monolingual and bilingual individuals but, additionally, the variation within the bilingual population. Similarly, Thordardottir (2015) argued that the amount of exposure a child has to the tested language should also be taken into consideration during assessments with lower cut-off criteria implemented if their weaker language is being assessed. However, in her guidelines for the assessment of bilingual norms, Thordardottir (2015) also stressed that merely adapting the norms of standardised assessments may not always be sufficient and that knowledge of how a particular child performs on other measures related to language is important in assessing the capabilities of bilingual children. In this view, non-word repetition tasks are widely considered to be highly informative in identifying if a child might be atypically developing. The same is true for detailed language and development histories.

Consequently, the inclusion of such measures has become the norm in language testing, with language researchers appreciating the value of gaining such detailed information, even if just for matching groups of participants. A prime example of gaining such detailed information on bilingual children is research conducted by Paradis et al. (2013) who tested both typically developing (TD) bilingual English-language learning children and age-matched children with a SLI in the US. TD and atypically developing (ATD; i.e., those with a SLI) children were tested on a range of measures, including English standardised tests of morphology and vocabulary, and a non-word repetition task. Additionally, parents were given the Alberta Language Development Questionnaire (ALDeQ; Paradis et al., 2010) to gain information on the child's early language milestones, current first-language abilities, activity preferences and behaviour, and family history. Predictably, a large number of the TD bilingual children scored below the age expectations when using monolingual methods, indicating the misfit of the measurement for these children who would likely be misidentified as having a language impairment (LI). To this end, more appropriate bilingual norms were generated by applying linear discriminant function analyses. This was to enable further understanding of which measures (or combinations of measures) could discriminate between the TD children and the children with LI. When employing these bilingual norms a more reliable identification of the children with LI could be made. This simply emphasises how the combination of language assessments, non-word repetition tasks and detailed histories can help in differentiating bilingual children with a SLI from their more typically developing peers. Consequently, this combined methodology is thought to offer an appropriate solution to the obstacles of assessing children in both languages.

The Clinical Evaluation of Language Fundamentals: Fourth Edition (CELF-4 UK; Semel et al., 2003) is deemed to be a valid tool that uses this combined methodology (Paslawski, 2005). Designed to assess the nature and extent of language difficulties in children and adolescents, the battery is one of the most widely used tools of Speech and Language Therapists diagnosing LI's. The CELF-4 UK has long been used to study language proficiencies in under 16 year-olds, both within research with a main focus on language (e.g., research on the language performance of children who were born pre-term; Barre et al., 2011) and outside of it (e.g., research on children with Epilepsy; Currie et al., 2018). Furthermore, although not designed for bilingual populations, the CELF is frequently used to evaluate the language competence of bilingual children (e.g., Babayiğit, 2014; Barac & Bialystok, 2012; Bowyer-Crane et al., 2017). A recent study by Oxley et al. (2019) has indeed found the CELF-4 UK to be the most popular language assessments standardised on monolinguals but used with bilingual children too (e.g., Barac & Bialystok, 2012; Iluz-Cohen & Armon-Lotem, 2013; Paradis et al., 2003). However, although the CELF-4 UK is one of the most consistently used assessments, its uses outside of a clinical sample are limited. Despite being a reliable tool for the assessment of language disorders or delays in children, the battery can only be administered by trained clinicians. This is a problem as many speech and language therapy (SLT) services have limited capacity for providing school-based input due to limited funding (Lindsay et al., 2002; Pring et al., 2012). Furthermore, its administration time is considerably lengthy; depending on subsets administered, the battery has the potential to take over an hour, which, in the classroom environment, is not always feasible.

2.1.1 The Current Study

The aim of the present study was to streamline the gold-standard CELF-4 UK into a reduced 30-40 minute assessment that would be more feasible to administer in the classroom environment. This was conducted by administering several of the tasks in the battery in full, and subsequently running post-hoc analyses to determine which, if any, of the subsets could be administered in its alternate item form (i.e., only administering every other item in the subset). This method of reduction was deemed plausible as the items increased incrementally in difficulty in each subset. Thus, administering alternate items respected the design of the original battery in incrementally increasing the difficulty of each subtest. Furthermore, as the battery was not intended for diagnostic purposes, or to determine difficulties in a specific child, but rather as an indication of language proficiency or to investigate patterns across groups of children, such a reduction in the items administered would not lead to erroneous assumptions in clinical diagnosis. Administering only alternate items has been a method often employed in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. The Weschler Intelligence Scale for Children (3rd UK edition; WISC–III; Wechsler et al., 1992), a language and cognitive assessment, was consistently streamlined in this way (Chang et al., 2014; Golding et al., 2021; Horwood et al., 2008; Osimo et al., 2020; Stautz et al., 2016; Taylor et al., 2017). Researchers within ALSPAC have deemed this method of streamlining to be robust for both the WISC and for the Short Mood and Feelings Questionnaire (SMFQ; Edwards et al., 2014).

Resultantly, a battery that would take only 30-40 minutes to administer, that could be easily administered by teachers, or teacher assistants, to monolingual or bi-/multi-lingual children, would be especially useful in the classroom environment.

This would be particularly useful to those who want to determine if a child's language proficiency may be behind their peers and if they might benefit from additional help in the classroom or referral to a clinician.

2.1.2 Method

2.1.2.1 Participants

The study tested 70 participants (34 males, 36 females) from two primary schools in West Yorkshire. All children were in Year 3 and between the ages of 7 years and 7 months to 8 years and 6 months (M = 8 years, SD = 3.85 months). No children had any known hearing problems or developmental delays. Table 1 summarises the language status (monolingual/bilingual) of the individuals as well as the distributions of age and gender.

Table 1

Distribution of age and gender of the sample (years;months)

	Gender	Min	Max	Mean	St.Dev.
	Female (n = 24)	7;7	8;6	7;11	3.79
Bilingual	Male (n = 22)	7;7	8;6	8;0	3.85
	Total (n = 46)	7;7	8;6	8;0	3.81
	Female (n = 10)	7;7	8;6	7;11	4.33
Monolingual	Male (n = 12)	7;7	8;6	8;1	4.02
	Total (n = 22)	7;7	8;6	8;0	4.16

School consent was obtained from the headteachers of each school. Once this written consent was obtained, information sheets and consent forms were distributed to all parents of children in Year 3 via classroom teachers. Parents were given the option to withdraw from the study by returning an opt-out consent form. If the opt-out form was not returned, consent was implied. Ethical Approval was granted by the University of Leeds Ethics Committee (PSC-586, approved 23/01/19).

Socio-economic Status (SES) was determined using the Income Deprivation Affecting Children Index (IDACI; Communities & Local Government, English Indices of Deprivation, 2015), again based on the school's postcode, and can be seen in Table 2.

Table 2

IDACI ranks, deciles and scores of the three schools sampled

	IDACI Rank	IDACI Decile	IDACI Score
School 1	3933	2	0.368
School 2	21138	7	0.101

IDACI Rank = 32,844 neighbourhoods in England are ranked from the most deprived (a score of 1) to the least deprived (a score of 32,844). IDACI Decile = Ranks are divided into ten equal groups where 1 indicates the most deprived and 10 indicates the least deprived. IDACI Score = Percentage of children in that area classed as deprived

2.1.2.2 Materials

Receptive language abilities (language *comprehension*) and Expressive language abilities (language *production*) were assessed using specifically selected subsets of the CELF-4 UK. The subsets chosen allowed for a Receptive Language Score (RLS), an Expressive Language Score (ELS) and, subsequently, a Core Language Score (CLS) to be generated as stated in the original CELF-4 UK. Furthermore, two further tasks, Phonological Awareness and Digit Recall, were administered to evaluate a child's knowledge of the language's sound structure as well as their working memory. All subsets were administered in their entirety, but the battery was later streamlined during analysis (as detailed in the Results section).

2.1.2.2.1 Receptive Language

To capture receptive language abilities, three subtests were chosen from CELF-4 UK, namely Concepts and Following Directions, Word Structure and Recalling Sentences.

The Concepts and Following Directions subtest was used to evaluate participants' ability to follow directions of increasing length and complexity. Ascertaining this ability in children is particularly important as issues with comprehending, recalling and acting upon spoken directions in the classroom increases the likelihood of task failure, which can, ultimately, result in missed learning opportunities (Gathercole et al., 2008). During this task, children were presented with an array of pictures and were asked to point to items named by the researcher in the order that the researcher stated them (e.g., "Before you point to the car, point to the ball").

The Word Structure subset was administered to evaluate participants' ability to employ morphology to aid the marking of inflections, derivations, and comparisons as well as to select the appropriate pronouns to refer to people, objects, and possessions. Researchers presented the children with two pictures, named the first and asked the child to name the second picture which differed in number, time or possessive relationship (e.g., "Here is one mouse. Here are two… [*mice*]").

The Recalling Sentences subset was designed to evaluate a child's ability to listen to sentences of increasing length and complexity and to repeat them back without altering them. This ability taps into various levels of sentence and language processing and can give insight into phonological, morphological and semantic representations with successful comprehension being necessary for correct repetition (Marinis & Armon-Lotem, 2015). The researcher read a sentence to the child and the child was asked to repeat it verbatim (e.g., "The student who won the award at the art show was very excited").

2.1.2.2.2 Expressive Language

To capture expressive language abilities, three different subsets were chosen from the CELF-4 UK, namely Formulated Sentences, Word Classes and Sentence Structure.

The Formulated Sentences subset was administered as part of the language proficiency battery to evaluate a child's ability to formulate complete, semantically and grammatically correct sentences of increasing length and complexity using given words and contextual constraints imposed by pictures. Children were presented with a picture and a target word and had to formulate a sentence that would describe the picture using that word (e.g., “Make a sentence about this picture using the word ‘gave’ ”).

The Word Classes subset was used to evaluate the ability to understand and explain the links between related words. Children were presented with three pictures and were asked to state which two were related and how. Thus, this subset is unique as it comprises both receptive and expressive components. The Word Classes subset has two forms; Word Classes 1 (WC1) which is to be administered to children 5 to 7 years old, and Word Classes 2 (WC2) to be administered to children 8 to 16 years. As the participants in the current study were on the cusp of the two age groups, it was decided that all participants, regardless of their age, would complete both versions. This was to enforce uniformity amongst all children tested. The format of WC2 was much the same as that of WC1, except there was no visual referent for the child to refer to and all items consisted of four objects. Examples of both are below:

(1) Word Classes 1: *“Look and listen carefully to the words that I say and tell me the words that go together best - slide, swing, flag.”*
(2) Word Classes 2: *“Listen carefully to the words that I say and tell me the words that go together best - fence, window, glass, rug.”*

In both cases, once the child responded, they were asked: “*Why* do those two words go together?”

Finally, the Sentence Structure subset was administered to evaluate a child’s ability to interpret spoken sentences of increasing length and complexity and select the pictures that relate to that sentence. The children were presented with an array of pictures and were asked by the researcher to point to an object that was either minimally described or described in detail (e.g., “Point *to ‘The spotted puppy is in the box’*”). In most cases, there was a similar foil (e.g., a picture depicting a spotted puppy outside of the box or an all-white puppy in the box). In the classroom, these abilities are used in creating stories and creating meaning in relation to illustrations.

2.1.2.2.3 Additional Measures

The first of these subsets was Phonological Awareness which was used to evaluate the child's knowledge of the sound structure of English words and how these sounds can be manipulated. In the classroom, these skills relate to the ability to produce rhymes and understand how letters form sounds, thus having a strong impact on literacy skills. This subset comprises 17 subsections, all of which are outlined in Table 3.

Table 3

Subsection titles and examples of items

	Subsection	*Example*
A	Syllable blending	*If you put the two words "Rain" and "Coat" together, what new word does that make?*
B	Initial Syllable Identification	*What sound does "sit" begin with?*
C	Rhyme Detection	*Do "cake" and "lake" rhyme?*
D	Final Phoneme Identification	*What sound does "him" end with?*
E	Two-Syllable Deletion	*Say "starfish" without saying "fish"*
F	Rhyme Production	*What word rhymes with "book"?*
G	Syllable Segmentation	*How many syllables are in the word "rainbow"?*
H	Phoneme Blending	*What word do the sounds "n-e-s-t" make?*
I	Initial Phonemes Substitution	*Say "tap" but swap the "t" for "n"*
J	Sentence Segmentation	*How many words are in the sentence "see me play"?*
K	Three-Syllable Deletion	*Say "sunflower" without saying "sun"*
L	Final Syllable Deletion	*Say "exercise" without saying "cise"*
M	Medial Phoneme Identification	*What is the sound in the middle of the word "make"?*
N	Initial Phoneme Deletion	*Say "seat" without saying "s"*

O	Medial Phoneme Substitution	*Say “hut” but swap the “u” for “o”*
P	Final Phoneme Substitution	*Say “make” but swap the “k” for “l”*
Q	Phoneme Segmentation	*What are the sounds that make up the word “spot”?*

The second of these additional subsets from the CELF4-UK was Number Repetition; a cognitive task. This was used to evaluate the children’s memory and their ability to repeat random number sequences of increasing length, both forwards and backwards. While forwards digit recall taps into working memory, backwards digit recall is typically used to asses working memory. Children were asked to listen carefully to a sequence of numbers and were then asked to repeat them back to the researcher either exactly as they heard them, or in reverse (e.g., “Repeat 24957163”). The maximum number of digits the children were required to repeat in the forward's repetition task was nine. The maximum number of digits the children were required to repeat in the backwards repetition task was eight.

2.1.2.3 Procedure

Children were taken from their classrooms and tested individually in a quiet space. Each child was tested for 40 minutes on two consecutive days to avoid fatigue and boredom. All questions were presented orally and recorded verbatim by the researcher. Responses to all questions were recorded on both the corresponding worksheets and digitally via a COOSA High-Quality Digital Voice Recorder placed on the table directly in front of the child in case of the need for later transcription. Scoring procedures for each of the tasks can be seen in Table 4.

Table 4

Scoring procedure of each of the subsets in the CELF-4 UK

Task	*Coding*	Requirements of a correct score	Discontinue Rule
Receptive language tasks			
Concepts and Following Directions	*Correct (=1) or Incorrect (=0).*	Identify all items stated by the researcher in the correct order.	Seven consecutive scores of zero but the counting of zero scores only begins with item 24.
Word Structure	*Correct (=1) or Incorrect (=0).*	Correctly manipulate inflections/derivations or use appropriate pronouns when talking about the scene presented. If the child's response contained the appropriate targeted form, but with a non-targeted noun or verb that was meaningful to the context, this was also recorded as correct.	All items administered.
Recalling Sentences	*Correct (=3), one error given (=2), two/three errors given (=1), four or more errors given (=0).*	Repeat verbatim sentences read to the children.	Five consecutive scores of zero.
Expressive language skills			

Formulated Sentences	*Correct (=2), a complete sentence that demonstrates correct structure but has one or two deviations in syntax or semantics (=1), Incorrect (=0).*	Create a semantically and syntactically complete sentence with correct structure. A sentence that was not logical, didn't relate to the picture, didn't contain the required stimulus word, had more than two deviations in syntax or semantics or was not complete was marked as incorrect.	Five consecutive scores of zero.
Word Classes*	*Correct (=1) or Incorrect (=0).*	On the *receptive* element of the task: correctly identify the two items that best paired. On *expressive* element of the task: logically demonstrate *why* these items were best paired.	Five consecutive scores of zero on the receptive element of the task.
Sentence Structure	*Correct (=1) or Incorrect (=0).*	Identify the picture corresponding to the sentence spoken by the researcher.	All items administered.
Additional measures			
Phonological Awareness	*Correct (=1) or Incorrect (=0).*	Correctly manipulate the sound as per the researcher's instructions.	Four consecutive scores of zero on three consecutive item sets.
Number Repetition	*Correct (=1) or Incorrect (=0).*	Repeat the number of sequences verbatim with no errors.	After zero scores on both parts of an item

*Due to the participants in the current study spanning the age ranges of the two subsets, all children, regardless of their age, started with the first item in WC2. If an incorrect score was obtained in the first five items of WC2 however, WC1 was administered in reverse until three consecutive correct scores were achieved. WC2 was continued until three consecutive scores of zero were obtained in the receptive section.

2.1.3 Results

Firstly, scores for each child, on each subset of the full CELF-4 UK were generated. These were created by totalling the number of correct scores, as instructed in the CEL-4 UK manual.

Secondly, an alternate item score was developed for each child, on each task. This score was calculated by totalling the correct number of responses from the alternate items (i.e., only counting correct scores from items 1, 3, 5, 7 etc.) and then, in most cases, doubling that number. The exceptions to this were Word Classes 1, Phonological Awareness and Word Structure. In the case of Word Classes 1, the complete subset was made of 21 items - an odd number. As only the odd alternate items were counted, this gave 11 items to assess. The alternate item score was thus generated by totalling the correct number of responses from the alternate items and then, multiplying this by 1.9.

In the case of Word Structure, there were 17 subsections with varying numbers of items in each. As each subsection assessed slightly different skills (e.g., Subsection A focussed on pluralisation of regular nouns while Subsection B focussed on pluralisation of irregular nouns), all odd numbered items in each subsection (i.e., the first, third, fifth etc. of each subsection of this subtest) were totalled. This number was then multiplied by 1.68 to create the alternate item score for this subset.

Finally, in the case of Phonological Awareness, there were 17 subsets, and each had five items. While in every other subset, the odd-numbered items were included in the analysis, here the even-numbered items were used to create the alternate item score to reduce the item further (i.e., instead of items 1,3 and 5 from each subsection, items 2 and 4 were used). This number was then multiplied by 2.5 to generate the alternate item score.

A series of t-tests were performed to determine if there were any statistically significant differences between the proportion of correct scores gathered from the children on the full CELF-4 UK that was administered and the alternate item scores. Only participants with complete data (i.e., every item) were included in the analyses (which explains the differences in the degrees of freedom). Table 5 shows the results of the t-tests for both monolinguals and bilinguals.

Table 5

Significance values for streamlining subsets

Subset	df	t-value	Sig.	Mean Alternate Item Score	Mean score when all items administered
Concepts and Following Directions	67	0.67	0.505	36.68	36.47
Word Structure	67	0.23	0.816	23.57	23.53
Sentence Structure	67	0.34	0.736	22.00	21.93
Recalling Sentences	66	1.04	0.302	51.97	51.55
Formulated Sentences	59	3.46	0.001*	32.53	30.83
Word Classes	67	-1.85	0.068	22.02	22.82
Phonological Awareness	68	1.63	0.107	73.04	72.42
Number Repetition	67	0	1	12.26	12.26

* p< .005

As can be seen in Table 5, the t-tests revealed no significant differences between the full CELF-4 UK subsets and the alternate item scores for seven of the eight chosen CELF-4 UK subsets. This signified that these eight subsets (six assessing language proficiency and two additional measures) could be administered in their alternate item forms, and valid results would still be gathered. However, for the remaining subset, Formulated Sentences, there was a significant difference between the scores obtained in the full, unabridged CELF-4 UK and the alternate item scores. It was therefore deemed inappropriate to streamline this subset in this way and was kept in its full version in subsequent testing. Finally, it must be noted that while the Word Classes subset approached significance (p =.068), it was still above the pre-determined threshold of significance (p> .05). As the p-value of the Word Classes subset was above this pre-determined threshold of significance, the difference between the full, unabridged subset and the truncated version of the subset was determined to be statistically similar. It was, therefore, deemed appropriate to truncate this subset in the same manner as all other subsets which were also over the .05 threshold. These seven alternate-items subtests plus the full Formulated Sentences subtest created what shall henceforth be referred to as the Streamlined-CELF.

2.1.3.1 The Streamlined-CELF

Although coding and the requirements for a correct score in the streamlined-CELF remained the same as in the full, unabridged version of the CELF-4 UK (see Table 6), the discontinue rules and number of items in each subset differed.

Changes in the discontinue rules reflect the changes made in the subsets and, in most cases, the number of zero scores needed to be obtained to discontinue were halved (and rounded up where necessary). Furthermore, in the Concepts and Following Directions subset, the zero scores in the CELF-4 UK start to be counted at item 24. This was amended so that, in the Streamlined-CELF, zero scores start being counted at item 12 – which is essentially the same item. The amended discontinue rules and the new number of items in the Streamlined-CELF can be seen in Table 6.

Consequently, in the Streamlined-CELF, the number of items in Concepts and Following Directions, Recalling Sentences, Word Classes 2, Sentence Structure and Number Repetition, were reduced by 50%. The number of items in three of the subsets, Word Classes 1, Word Structure and Phonological Awareness were reduced by 47.62%, 40.62% and 60%, respectively. This was due to the number of items in the task or the layout of the sections within the task. Formulated Sentences was not reduced at all. Overall, this resulted in the Streamlined-CELF being comprised of seven language proficiency measures and two additional assessments, reduced, on average, by 55.36% from the CELF-4 UK (plus one unreduced language measure).

Table 6

Comparison of the CELF-4 UK discontinue rules and number of items in each subset and those in the new, alternate-item version

Task	CELF-4 UK		Streamlined-CELF	
	Number of items	Discontinue Rule	Number of items	Discontinue Rule
Concepts and Following Directions	54	Administer all items up to, and including item 24. After this, discontinue after seven consecutive scores of zero.	27	Administer all items up to, and including item 12. After this, discontinue after three consecutive scores of zero.
Word Structure	32	All items administered.	19	All items administered.
Recalling Sentences	32	Five consecutive scores of zero.	16	Three consecutive scores of zero.
Word Classes 1	21	Five consecutive scores of zero on the receptive element of the task.	11	Three consecutive scores of zero on the receptive element of the task.
Word Classes 2	24	Five consecutive scores of zero on the receptive element of the task.	12	Three consecutive scores of zero on the receptive element of the task.
Sentence Structure	26	All items administered.	13	All items administered.
Phonological Awareness	85	Four consecutive scores of zero on three consecutive item sets.	34	Two consecutive scores of zero on three consecutive item sets.
Number Repetition	30	After zero scores on both parts of an item.	15	A single score of zero.

2.1.4 Discussion

The current study aimed to streamline the gold standard language battery, the CELF-4 UK, to shorten its administration time. Results from a series of t-tests revealed this to be possible by administering only every other item in each subset. Although one of the selected subsets was unable to be streamlined in this way (Formulated Sentences), the remaining five language proficiency measures (Concepts and Following Directions, Word Structure, Recalling Sentences, Word Classes and Sentence Structure), and both of the additional measures (Phonological Awareness and Number Repetition) produced statistically similar results if only every-other item in the subset was administered. This enabled a reduction in the number of items in the battery by over a third. Consequently, the CELF-4 UK battery, usually taking up to 80 minutes to administer, could be administered in 35-40 minutes. This makes it less resource- and time-expensive, and thus more feasible in the classroom environment compared to the CELF-4 UK in its full, unabridged form.

It must be noted, however, that while the CELF-4 UK is among the most commonly used language assessments in the UK when assessing both monolingual and bilingual populations (Oxley et al., 2019), the battery was only standardised on a monolingual sample. Consequently, when assessing bilingual children, the validity of the outcomes must be considered when using either the full, unabridged version of the CELF or the Streamlined-CELF, especially when the aim is of a diagnostic nature. However, while this is a factor that does need to be taken into consideration, the aims of the current study focussed on simply gaining an indication of the children's' linguistic competency, and not on diagnosing children with any form of language disorder/difficulty.

The reduction of items in the CELF-4 UK was deemed appropriate as the current project does not intend the battery to be used for diagnostic purposes. Instead, the battery is intended to be used as an indicator of language proficiency and the areas of language proficiency which a group of children may particularly struggle with. Furthermore, although this Streamlined-CELF is applicable for monolingual children, it still complies with Thordardottirs' (2015) guidelines for the assessment of bilingual norms, including both language and cognitive assessments.

Therefore, in subsequent studies in this dissertation, the streamlined CELF was used, knowing it still accurately captured children's language proficiency.

2.2 Quantifying a bi-/multi-lingual child's language experience when parental response is low: Development of the LeBLEQ

Bilingualism is a complex phenomenon. This idea is rife within language research, with even the criteria differentiating between monolingual and bilingual groups being "fuzzy at best" (Luk & Bialystok, 2013, p. 605). Due to this difficulty in classifying bilinguals and monolinguals into truly distinct groups, and the heterogeneity of bilingual language experience (influenced by the factors discussed in Chapter 1, such as context or age of exposure), the literature shows the linguistic capabilities of bilingual children to be highly varied. Consequently, it is becoming ever more apparent that a more fluid approach to bilingualism should be adopted. In recent years, there has been increased debate that bilingualism and multilingualism are best conceptualised as variables along a continuum, rather than dichotomously (Baum & Titone, 2014; Kašćelan et al., 2020; Kaushanskaya & Prior, 2015; Serratrice & De Cat, 2020). Questions regarding the *extent* of an individual's bilingualism (i.e., where on the bilingual continuum they may lie) are not uncommon and the ability to identify this positioning is informative in various contexts. For example, from a research perspective, the ability to classify an individual as being either monolingual or bilingual, and determining the extent of their bilingualism, is often imperative in participant selection (i.e., ascertaining whether an individual meets certain inclusion criteria, or whether they need to be excluded). Additionally, it might be that parents or teachers are concerned about a particular child's behavioural or developmental patterns and the knowledge of where the child sits along this continuum can help decide whether that child should be referred to a clinician. The clinician can then determine whether they may benefit from additional support.

Previous literature has attempted to address the idea of a bilingual continuum by suggesting thresholds, whereby an individual who has been exposed to a language below a certain level should be classed as a monolingual. (Pearson et al., 1997), for example, suggested a 75:25 ratio whereby children with less than 25% of exposure to a language tend to not have the same fluency that would be expected of a native/native-like speaker and, therefore, should not be classed as bilinguals in research settings. These children are often referred to in the literature as '*functional* monolinguals' (Bedore et al., 2012). Additionally, perhaps the most widely accepted bilingual threshold was proposed by Thordardottir (2011), who suggested that bilinguals who had been exposed to a language more than 60% of the time were able to perform on par with their monolingual peers. In her study with French-monolingual and French-English bilingual 5-year olds, bilinguals receiving exposure to a language in excess of this threshold behaved akin to monolinguals on various tests of expressive language when controlling for age, SES status and non-verbal communication skills. The bilinguals who received less than 60% exposure to one language, however, demonstrated much lower expressive scores than their matched monolinguals. It must be noted, however, that both of these thresholds are primarily focussed on a bilingual child's expressive language – i.e., their production abilities.

Building upon Thordardottir (2011) concept of a 60% threshold, Cattani et al. (2014) conducted research taking into consideration both expressive and receptive language abilities. Cattani and colleagues studied the linguistic experiences of typically developing 2.5 year olds across the UK with the aim of determining the minimum proportion of English exposure a bilingual child must receive in order to perform as well as a their monolingual peers on a test standardized for monolingual English-speaking children. Similarly to Thordardottir's (2011) conclusions, it was determined that a bilingual who receives input to a language at least 60% of the time will perform equivalently to their monolingual peers on a variety of both receptive and expressive language tasks. Consequently, this research gives a more comprehensive view of the level of language exposure bilingual children must receive in order to behave akin to their monolingual peers on both expressive and receptive based tasks. However, as previously discussed, variability within the bilingual language experience is high, and thus thresholds may not always be specific enough to give the full picture of a bilingual person's capabilities.

It must be noted, however, that not all researchers agree with the use of thresholds, as monolingual standards are not always appropriate to assess bilingual competence (e.g., Grosjean, 1985; 1985). Consequently, in contrast to the use of thresholds, an abundance of literature has attempted to quantify the extent of bilingualism using a numerical value. The Language and Social Background Questionnaire (LSBQ: Anderson et al., 2018) allowed for the assessment of the degree of bilingualism in adult populations. Not only did this work highlight the difficulties experienced when attempting to classify individuals into dichotomous groups of language experience (i.e., monolingual versus bilingual) but also the lack of reliable methods available to position them along a continuum. Using Latent-Factor analysis, Anderson et al. (2018) were able to use the LSBQ to construct a composite score which could be used both as a continuous variable and a criterion to define groups categorically. Furthermore, outputs were found to predict performance on a range of executive function tasks, further demonstrating that a continuous measure of language experience can be a valid predictor of language competence. Similarly, De Cat and colleagues (2021) have recently highlighted the need for a common approach in quantifying bilingualism which has led to the development of the Q-Bex project (Quantifying Bilingual Experience). Though the project is still ongoing, it aims to produce a user-friendly, online questionnaire that will return measures of current and cumulative language experience.

Information related to language exposure is most often gathered via self-reported questionnaires which are considered to be convenient, time- and cost-effective tools in adult populations. Furthermore, research supporting the idea that self-reported measures are indicative of linguistic ability (Ross, 2006) has led to self–report measures becoming one of the most widely used methods for gaining information on language proficiency in recent years (Li et al., 2006). In child populations, however, parental-report methods are most commonly used. Parental-report measures are widely considered valid tools for monitoring language development (Jackson-Maldonado et al., 1993; Squires et al., 1997), identifying language impairments (Restrepo, 1998), obtaining language histories (Adler, 1991; Siren, 1995), and informing of early language experiences (Camaioni et al., 1991; Patterson, 2000).

Additionally, Paradis (2011) reported substantial correlations between parental report of a child’s English vocabulary and the amount of L1-L2 seen to be used in the home environment. After noting the challenges faced by researchers when attempting to obtain information on language use in multilingual contexts, Paradis and colleagues (2011) developed a parental-report questionnaire on bilingual language development in English. This questionnaire was designed to not be specific to a particular language but be able to be used in various multilingual contexts. Entitled the Alberta Language Environment Questionnaire (ALEQ) the measures aimed to enable examination of the impact of individual differences on children's acquisition of vocabulary, such as quality and quantity of input. The questionnaire revealed that factors such as length of exposure to a language, and richness of the child's language environment were significant predictors of variation in children's language skill. Consequently, the ALEQ is considered to be a reliable, valid tool for capturing the multifactorial view of bilingualism, being particularly useful to those wanting to obtain the language histories of children when their language environment cannot be examined directly. The same applies to many other parental-report measures, including the Parents of Bilingual Children Questionnaire (Tuller, 2015), the Bilingual Information and Observation Questionnaire (Hardin et al., 2013) and the Babylab Language Exposure Questionnaire (Cattani et al., 2014).

However, although studies show parental-report to be a comprehensive method for obtaining information regarding bilingual language development (Espinosa & López, 2007; Gutiérrez-Clellen & Kreiter, 2003), there is no consistent or agreed process for gathering this information, and responses heavily rely on parents' willingness and ability. Low parental response rates are often a substantial issue during testing with child populations, with perpetually low rates observed from poorer socio-economic strata (Couper & Groves, 1996; Curtin et al., 2000; Groves & Peytcheva, 2008; Warriner et al., 2002), particularly from those who live in more urban areas (Couper & Groves, 1996; Eaker et al., 1998). For example, in a study investigating parental opinions of overweight children, responses were as low as 16% in schools with higher proportions of pupils receiving free school meals and where English was not the parents' first language, in contrast to the 52% average across all schools (Carnell et al., 2005). Moreover, ethnicity has been found to influence response rates, with studies finding that white individuals are more likely to respond to, or attend surveys in comparison to non-white individuals (Curtin et al., 2000; Voigt et al., 2003). It has also long been hypothesised that individuals are less likely to respond to a questionnaire if it is not presented in their first language (e.g., Harzing, 1997).

Consequently, parents cannot always be relied upon for information regarding the language experience of primary-school-aged children. This is a factor that is particularly true for cities such as Bradford, and cohorts such as Born in Bradford (BiB), where there are high numbers of individuals of low socioeconomic strata, with diverse ethnic backgrounds. Gaining this information about the quantity, and quality of language exposure, however, is of vital importance as the knowledge of how much language experience a child has assists clinicians and researchers alike in the interpretation of a child’s scores on several direct linguistic measures (Paradis, 2011a; Tuller, 2015). Accordingly, a method of gathering such information independent of parental response is highly valuable.

One way to address this issue is to rely on information gathered from school teachers. Teacher reports have long been deemed to be valid and reliable sources for gathering information on children’s language abilities as teachers tend to spend extended periods with the child themselves and assessing their communicative abilities (Rimfeld et al., 2019). Allowing teachers to provide their input permits an understanding of the child in the classroom setting and the types of communication that are specific to this environment (Gutiérrez-Clellen et al., 2006; Restrepo, 1998; Restrepo & Kruth, 2000). Consequently, due to the developmental knowledge that teachers hold, and the tendency to be less prone to biases in reporting, some researchers consider teacher reports to be one of the most valid methods of gaining information about an individual (e.g., Stone et al., 2010). Teachers are more likely to have a better notion of how well each particular child performs in comparison to their monolingual peers in such academic settings, a factor that parents who only speak the home language with their child are likely not to be privy. In terms of practicality, teacher questionnaires also reduce the possibility that the questionnaires will not be returned or fully understood and can, thus, be administered in instances where low parental response is expected. This, of course, applies only to the language of schooling as teachers are unlikely to have detailed understanding of children’s language skills in their home language.

Literature has also shown children themselves to be reliable in reporting their abilities and internal states. This has been widely found in terms of health-related outcomes (Riley et al., 2004; Tomlinson et al., 2017; Vernon-Roberts et al., 2019), experiences of pain (Castarlenas et al., 2017; Emmott et al., 2017), sleep patterns

(Richdale & Baglin, 2015), amount of screen time (LeBlanc et al., 2015), activity levels (Janz et al., 2008) and quality of life (Egilson et al., 2017). Nevertheless, to the author's knowledge, no attempt has been made to develop a similar questionnaire focussing on language experience.

2.2.1 The Present Study

The present study proposes a new method for obtaining detailed information of an individual's unique language experience. The method, entitled the Leeds-Bradford Language Exposure Questionnaire, comprises two questionnaires and was developed to reliably assess language experience across various ages. Outcomes of the first questionnaire provide a gradient, continuous score of bilingualism while the second component (i.e., the teacher's questionnaire) captures language ability. The two components of the LeBLEQ are: the Leeds-Bradford Language Exposure Question for Children LeBLEQ-C), and the (complementary) Leeds-Bradford Language Exposure Questionnaire for Teachers (LeBLEQ-T). Each of these questionnaires shall be discussed separately below.

2.2.1.1 Leeds-Bradford Language Exposure Questionnaire for Children (LeBLEQ-C)

The inception of the Leeds-Bradford Language Exposure Questionnaire for Children (LeBLEQ-C) was influenced by Paradis (2011) who reported that the total length of exposure, frequency in different contexts, and language used in different types of activities (reading, listening to music, playing games etc.) to be among the most common considerations when developing tools to assess language proficiency. Consequently, much like many prior parental-report questionnaires (e.g., the PaBiQ, ALEQ etc.), factors explored in the LeBLEQ-C included the length of exposure, cumulative exposure, the richness of exposure, frequency of exposure/use in different contexts, and family dynamics. Although there have been differing reports of the contributions of these factors, their effects on L2 development have had consistent support (Paradis, 2011a). The fundamental difference, however, between Paradis (2011) and the LeBLEQ-C is that the LeBLEQ-C is a self-report measure for child populations, thus negating the need for parental input.

Based on the literature outlined above (e.g., Cattani et al., 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020; Thordardottir, 2011), it would be expected for the scores generated from the LeBLEQ-C to be associated with other measures of linguistic competencies, such as academic attainment, language proficiency and cognitive expectations in addition to pre-existing measures of language exposure. Specifically, in accordance with the literature reviewed in the General Introduction in Chapter 1 (e.g., Bialystok, 2001; Fenson et al., 1994; Gibson et al., 2012; Hammer et al., 2008), it would be predicted that monolingual children would receive higher scores on tasks of objective language skill (core language, receptive language and expressive language scores alike), but when focussing on the bilingual children, these scores would decrease proportionately with the increased exposure to another language (Bialystok et al., 2010; Rescorla & Achenbach, 2002). Conversely, bilingual children would be expected to perform better on tasks of working memory (Morales et al., 2013), with the older children outperforming the younger children (Gathercole et al., 2004; Miles et al., 1996).

Ultimately, the study set out to construct a measure that could identify a child's degree of bi-/multi-lingualism and how this exposure may affect their abilities in classroom-based tasks. The ability to obtain such vital information on a child's language experience, without the need to rely on parental report, would be highly valuable in research contexts, but also in education and clinical settings where such information is sparse or currently unavailable.

2.2.1.1.1 Method

2.2.1.1.1.1 Participants

The study aimed to recruit a sample representative of the population within primary schools in West Yorkshire, with a high degree of linguistic heterogeneity. Four schools were recruited and 244 children (126 females) across Year 3 (133 participants, M= 7 years, 11 months, SD= 4.37 months) and Year 6 (111 participants, M= 10 years 9 months, SD= 4.98 months) took part in the current study. No children had any known neurological or language impairments, or hearing problems. Table 7 summarises the distributions of age and gender of the participants.

Table 7

Distribution of age and gender of the sample included in the development of the LeBLEQ-C (years;months)

	Mean	Min	Max	StdDev (months)
Year 3				
Female (n=71)	8;0	7;3	8;8	4.73
Male (n=62)	7;11	7;2	8;8	3.95
Total (n=133)	7;11	7;2	8;8	4.37
Year 6				
Female (n=55)	10;10	10;2	11;8	4.84
Male (n=56)	10;8	10;2	11;8	5.06
Total (n= 111)	10;9	10;2	11;8	4.89

Socio-economic Status (SES) was determined using the Income Deprivation Affecting Children Index (IDACI; Communities & Local Government, English Indices of Deprivation, 2015), based on the school's postcode, and can be seen in Table 8.

Table 8

IDACI ranks, deciles and scores of the four schools sampled in the development of the LeBLEQ-C

	IDACI Rank	IDACI Decile	IDACI Score
School 1	5607	2	0.274
School 2	8685	3	0.221
School 3	5191	2	0.283
School 4	7173	3	0.245

IDACI Rank = 32,844 neighbourhoods in England are ranked from the most deprived (a score of 1) to the least deprived (a score of 32,844). IDACI Decile = Ranks are divided into ten equal groups where 1 indicates the most deprived and 10 indicates the least deprived. IDACI Score = Percentage of children in that area classed as deprived

Consent was first obtained from the Headteachers of the participating schools as all testing would take place in school premises within school hours. Once written consent from headteachers was obtained, information sheets and consent forms were distributed to all parents of children in Year 3 and Year 6 via classroom teachers. Parents were given the option to withdraw their child from the study by returning an opt-out consent form. If the opt-out form was not returned, consent was assumed. Finally, before taking part in the study, children were informed about the tasks and their assent was obtained. The study received approval from the Ethics Committee of the School of Psychology, University of Leeds (ethics reference number: PSC-313, approved 23/01/2019).

2.2.1.1.1.2 Materials

2.2.1.1.1.2.1 The Leeds-Bradford Language Exposure Questionnaire for Children (LeBLEQ-C)

The development of the LeBLEQ-C was influenced by the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller, 2015) discussed below. The LeBLEQ-C consists of four subsets covering the following topics: Handedness, General Language Information, Languages Used in the Home, and Languages

Spoken in Other Contexts. The questionnaire comprised of 52 questions developed in order to obtain information on areas deemed influential in the acquisition of a second language, such as order of acquisition, modalities of use, quality and quantity of exposure and cumulative exposure (e.g., Thordardottir, 2011; Unsworth, 2013). As some questions depended on previous answers provided, not all questions were necessary for all participants. For example, if a child had stated that their father only spoke English, they would not be asked the proportion of the time their father spoke to them in English (as it would be deemed to be always).

Children were asked questions regarding the exposure they received in English and any other languages spoken in the home environment. Children reported their age of onset of exposure to each language and in what context (e.g., home, nursery, pre-school, school etc.). Children were asked if they were born in the UK, the language they preferred to speak and the language they believed themselves to be better at speaking. Furthermore, they were asked about the interlocutors they had the most interactions with (such as parents, or other family members that they lived with), including the language in which they received their input, but also their language of response (i.e., their language use). Finally, the children were asked questions regarding the types of activities undertaken in each of their languages (e.g., reading, listening to music, playing games), as well as details of any formal language teaching.

Although some of the questions were influenced by those contained in the Parents of Bilingual Children Questionnaire (PaBiQ; Tuller, 2015), a number of items were also developed specifically for the LeBLEQ-C (e.g., "do you go to school for your home language?"). Furthermore, it was fundamental for the purpose of the questionnaire that the language used was age-appropriate, with simple phrasing and clear intentions so that the children could respond to the questions themselves. For this reason, researchers were trained to rephrase questions to aid understanding further if necessary, as well as to ask follow-up questions if partial answers were given or the original question was misunderstood. The questionnaire took approximately 5-10 minutes to complete.

Scoring scales for the LeBLEQ-C were kept constant across subsets in order to increase consistency, facilitate sufficient completion and, consequently, researchers' interpretation. The LeBLEQ-C comprised of 52 questions: eight of them were open-ended, semi-structured questions; 21 were dichotomous questions (yes/no); and 23 were questions on Likert scales. Seventeen of these questions were on a 5-point scale ('Never'=0, 'Rarely'=0.25, 'Half of the time'=0.5, 'Often'=0.75, 'Always'=1) whilst the remaining six questions were on a 3-point scale ('Minimal language skill'=1, 'Moderate language skill'=2, 'Excellent language skill'=3). The questionnaire took approximately five minutes to administer for bilinguals but less for monolinguals who were only required to complete the first subset (Handedness) as all other exposure was English.

Estimates for a child's cumulative language input were calculated on the basis of their answers about the languages spoken within the home environment, taking into consideration any regular outside activities that the child reported participating in (the hours attended, the years participating and the language spoken in these environments). It must be clarified that learning of modern foreign languages did not automatically contribute to the score and to the classification of children as 'bilingual'. For example, while all children in the sample had weekly French lessons as part of the curriculum, this did not contribute to the children's final scores; only lessons in the child's heritage language were taken into consideration and contributed to their final score.

Furthermore, calculations were based on waking-hours only, calculated using the national sleep guidelines for each age (Hirshkowitz et al., 2015). Thus, for the Cumulative Language Input Score (CLI score), the scale ranged from 0-1. A lower score would indicate limited experience to a language other than English (where a monolingual would score 0), whilst a higher score would indicate increased exposure to another language. So, for example, a bilingual child who has spoken their Home Language (HL) all their life, and always speaks the HL outside of school, would score 1.

2.2.1.1.1.2.2 The Questionnaire for the Parents of Bilingual Children (PaBiQ)

The PaBiQ (Tuller, 2015) was initially developed as part of the COST Action IS0804 to collect background information on bilingualism factors such as quality and quantity of a child's language input, but also risk factors for Specific Language Impairment (SLI). Questions focussed on developmental history, languages used with and by the child, education and occupation of the parents, as well as family history of speech and/or language difficulties. Although the questionnaire was initially developed for use with younger children, it has since been used with older populations too (e.g., (Bosma et al., 2019; Hansen et al., 2017).

The PaBiQ was selected for inclusion in this battery of testing because it is relatively short and uncomplicated and can be completed by parents themselves without interviewer intervention, in contrast to most existing language exposure questionnaires (e.g., the ALEQ) which are intended for face-to-face administration with parents. However, the PaBiQ still relies on parental input and was, thus, deemed unfit for sole use in the current set of studies given the low response rates observed in low socio-economic strata and those who speak English as an additional language (Carnell et al., 2005; Eaker et al., 1998; Harzing, 1997; Voigt et al., 2003; Warriner et al., 2002). Nevertheless, in order to assess the newly developed LeBLEQ and compare it to existing questionnaires that rely on parental input, the PaBiQ was distributed to the parents of the children included in the current study alongside a brief information sheet of how to complete the questionnaire. Parents were asked to return the paperwork to the school by the end of the week when it was due to be collected by the researcher.

In the PaBiQ, individual questions are assigned points which can be grouped together for section scores and an overall score. Section scores include (1) Positive Early Development, (2) Positive Family History, (3) No Risk Index, (4) L1 Use and Richness and (5) L2 Use and Richness. The PaBiQ is traditionally scored using various composite scores (such as linguistic richness in the HL, total early exposure to the language and both amount and quality of exposure). However, the aims of its inclusion in the current study were slightly different. Firstly, the parental questionnaire was distributed in order to obtain corroborating information given by the teachers, that no children included in the sample had any history of speech and language issues or any problems with their hearing. Secondly, the questionnaire was used to ensure the reliability of the answers given by the children themselves in regard to the input received in the home environment. The questionnaire took parents 10-15 minutes to complete.

2.2.1.1.1.2.3 Language and Cognitive Measures

Several measures of English language proficiency were collected to explore different aspects of language competence and working memory. Firstly, the streamlined version of the CELF-4 UK, as discussed in Section 2.1.3.1, was administered. This included the tasks: Concepts and Following Directions; Word Structure; Word Classes; Sentence Structure; Phonological Awareness, and Number Repetition subsets in their alternate-item forms. Recalling Sentences and Formulated Sentences remained as full, unabridged versions. The streamlined battery allowed for the assessment of core, expressive and receptive language ability, as well as knowledge of sounds and nonverbal short-term memory.

An adapted version of the Children's Test of Non-word Repetition (CNRep) Task (Gathercole et al., 1994), previously used by the Avon Longitudinal Study of Parents and Children (ALSPAC Study Team 2001; http://www.bristol.ac.uk/alspac/), was used to further assess short term memory. This comprised of 12 non-words, all of which consisted of either three, four or five syllables. Each word conformed to sound combination rules present in the English language. The words were: pranstutiary; pennerriful; shimitet; empliforvent; zubinken; perlisteronk; instradontally; frescovent; tridercory; donderificam; brasterer, and doduloppity. Though this measurement is consistently used in the literature with monolingual English speakers (e.g., Bishop et al., 1996; Cilibrasi et al., 2018; Pigdon et al., 2020) for the purpose of this research, stimuli were also assessed in a short pilot study to ensure none of the non-words included in the task approximated any true words in other languages that were likely to be encountered during testing. To this end, the 12 stimuli were recorded by a female native English speaker using Audacity (version 2.1.2.0) to ensure uniformity of presentation. These were then presented to adult participants who were asked to state whether each word resembled a real word in their native language. Languages checked were: Spanish, French, Urdu, Hindi, Arabic, Punjabi, Polish, Mirpuri, Gujarati, Bengali, Mandarin, and Greek. All participants confirmed the non-word status of all the stimuli.

The CNRep task was presented aurally to participants, via Audacity, through headphones in order to minimise other distractions. The child listened to each of these non-words once and was asked to repeat them back to the researcher as accurately as possible. The accuracy of this repetition was scored in two ways. Firstly, responses were recorded as either correct (=1) or incorrect (=0). Secondly, the repetition was scored on a syllable-by-syllable basis in order to form a percentage of the proportion of the word that was correctly repeated.

Key Stage 1 (KS1) SAT scores (Reading and Writing) for each child were provided by the schools. In the case of the Year 6 pupils, access was also granted to the Key Stage 2 (KS2) SAT scores. KS1 and KS2 scores were coded as one of three categorical variables: 1= Below Expected, 2= Expected, 3= Above Expected.

2.2.1.1.1.3 Procedure

Children were taken from their class and tested individually in a quiet classroom. Each child was tested for 40 minutes each on two consecutive days to avoid fatigue and boredom. If children's attentions appeared to be flagging, a small break was taken. This time frame included the administration of the language proficiency measures of the Streamlined-CELF, the cognitive measures and the LeBLEQ-C. The order in which these tasks were administered were randomised across participants to allow for order effects.

All questions were presented orally and recorded verbatim by the researcher, except for the CNRep task which was presented via headphones. Responses to all questions were recorded on both the corresponding worksheets and digitally via a COOSA High Quality Digital Voice Recorder placed on the table directly in front of the child for later transcription. Upon completion of testing, participants were thanked with a certificate of participation and a reward sticker.

2.2.1.1.2 Results

Overall, 244 children were tested using the LeBLEQ-C and had Cumulative Language Input Scores (CLI scores) calculated. Analyses were then conducted to investigate the relationship between CLI score and the language and cognitive tests used. Results for each of the proficiency tests (language proficiency, academic attainment and working memory) are first reported descriptively to compare the performance of the monolingual and bilingual children. Secondly, to model the relationship between English language, cognitive proficiency and academic attainment, and cumulative exposure to a language other than English, linear regression analyses were performed. All statistical analyses were conducted using SPSS (version 26).

Among the children tested, there were 14 children who reported speaking a language other than English, yet their CLI scores remained at 0. This was due to the children reporting speaking only English in the home and having no outside activities in their HL. Upon further exploration of the data provided by the LeBLEQ-C, it was revealed that these children often stated to prefer speaking English, rated their English to be better, and often struggled to find the necessary word enabling them to adequately communicate in their HL (while this was not the case for English). Essentially, these children had few words in their HL repertoire and only used these on occasion. When these children were compared to the children who initially identified themselves as monolingual, their scores were shown to be statistically similar. This was determined by performing nine independent groups t-tests. T-tests were the most appropriate statistical analysis in this instance as only group (monolingual/bilingual) differences across the tasks were of interest (i.e., not the interaction between group and task type). Performing nine separate analyses, thus, had no effect on the likelihood of observing Type-I errors. T-tests demonstrated no statistically significant differences (i.e., in all cases, $p > .05$) between these 14 children and those who reported to be monolingual. The variables explored focussed on language proficiency, determined by the Streamlined-CELF (Core Language Score, Receptive Language Score and Expressive Language Score), working memory (Number of accurate recalls during the CNRep task and Number Repetition) and academic attainment (KS1 Reading, KS1 Writing and, in the case of the Year 6, KS2 Reading and KS2 Writing SATs). Consequently, these 14 children were deemed to be functionally monolingual and were included in the Monolingual group when the data were later stratified into monolingual and bilingual populations. This resulted in the Year 3 group (M= 7 years, 11 months, SD= 4.37 months) consisting of 98 bilinguals and 35 monolinguals and the Year 6 group (M= 10 years, 10 months, SD= 4.98 months) consisting of 84 bilinguals and 27 monolinguals. Linguistic demographics of the sample, as determined by the LeBLEQ-C, can be seen in Table 9

Table 9

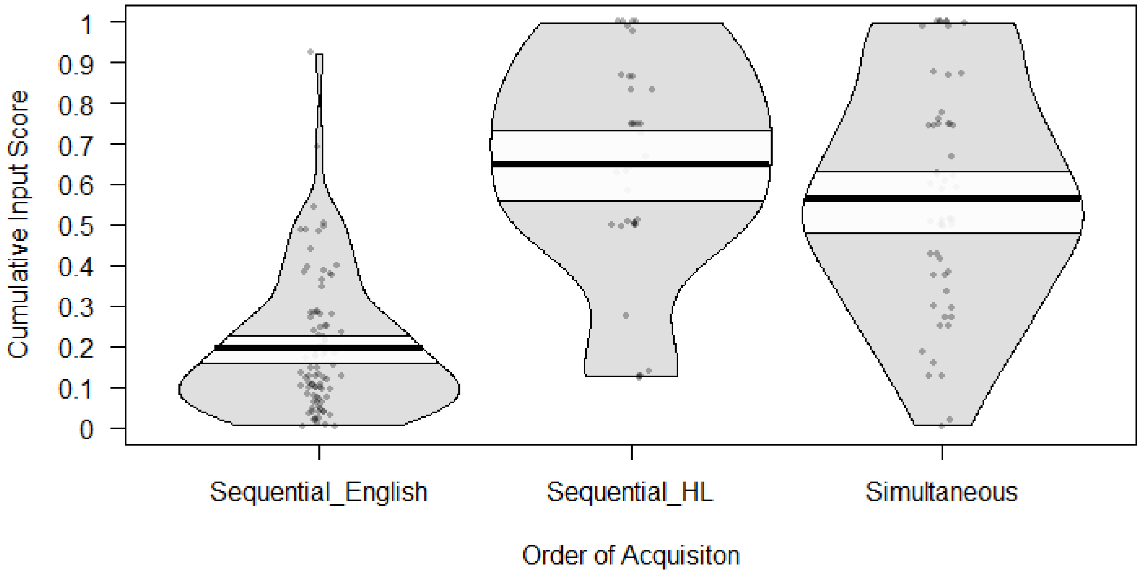
Distribution of age, gender and language experience of the children who were administered the LeBLEQ-C (years;months)

			Mean	Min	Max	StdDev (months)
Year 3						
	Bilinguals	Female (n= 56)	7;11	7;3	8;8	4.95
		Male (n= 42)	8;0	7;2	8;7	3.99
		Total (n= 98)	7;11	7;2	8;8	4.55
	Monolinguals	Female (n= 15)	8;1	7;4	8;6	3.82
		Male (n= 20)	7;11	7;3	8;8	3.86
		Total (n= 35)	7;11	7;3	8;8	3.91
Year 6						
	Bilinguals	Female (n= 40)	10;11	10;2	11;9	5.18
		Male (n= 44)	10;09	10;2	11;8	5.04
		Total (n= 84)	10;10	10;2	11;9	5.17
	Monolinguals	Female (n= 15)	10;10	10;2	11;3	3.83
		Male (n= 12)	10;10	10;3	11;9	5.24
		Total (n= 27)	10;10	10;2	11;9	4.42

The CLI scores for the bilingual children resulted in a sample that ranged across an evenly distributed continuum of language experience (see Figure 1).

Figure 1

Distributions of Cumulative Language Input Score of the bilinguals, stratified by Order of Acquisition



Cumulative Language Input Score ranges from 0-1 where higher scores indicate more exposure to a language other than English. As all monolinguals scores were at "0", they were not included in the current figure.

In particular, children who learnt English later in life (i.e., they were sequential bilinguals who learnt their HL first; Sequential-HL) had more cumulative experience of their HL and grouped slightly higher in the scoring system. Children who learnt their HL later in life (i.e., they were sequential bilinguals who learnt English first; Sequential-English) had more cumulative experience of English and grouped towards the lower end of the scoring system. Finally, as expected, the CLI scores of the simultaneous bilinguals (i.e., children who reported learning English and their HL at the same time; Simultaneous) were found to be more distributed.

2.2.1.1.2.1 Language Proficiency

Of the 244 children tested, 172 children were also tested on the Streamlined-CELF. Descriptive statistics showed children's performance on tasks capturing

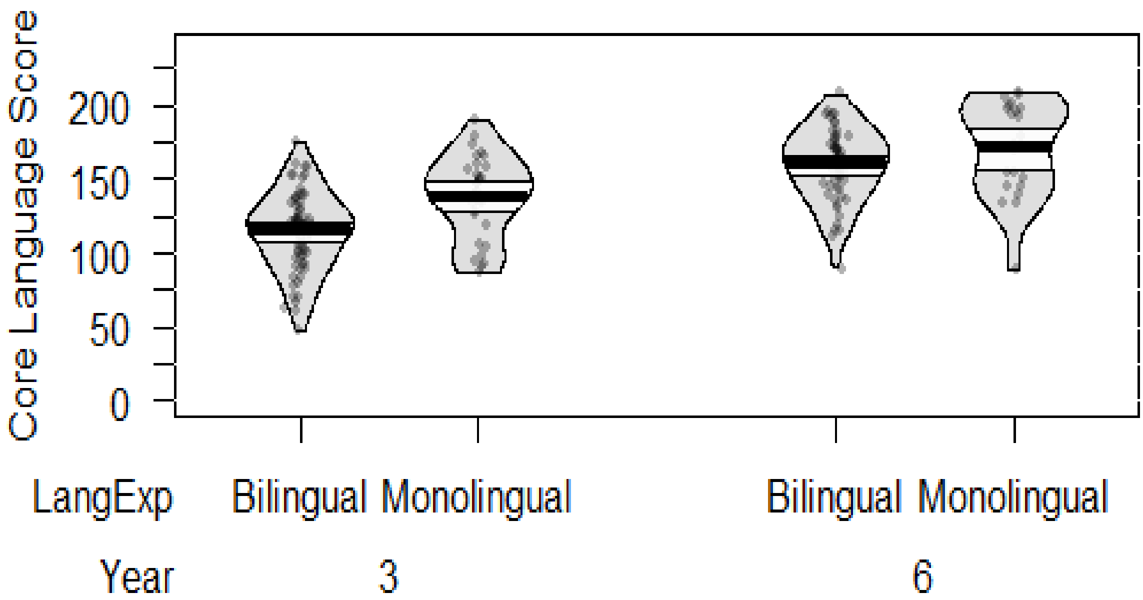
Core Language, Receptive Language, and Expressive Language. Details of how these scores were generated can be found in Section 2.1.3 of this chapter.

Three two-way ANOVAs were conducted to ascertain whether there were any main effects of age (2 levels: Year 3/Year 6) or binary language classification (2 levels: monolingual/bilingual), or an interaction between the two variables. The three IVs were a child's Core Language Score, Receptive Language Score and Expressive Language Score. As can be seen in the discussion below, similar effects were seen across all three measures.

Firstly, when focussing on the Core Language Score (CLS), a two-way ANOVA demonstrated significant main effects of both age and language experience, yet no interaction. Monolinguals (M= 151.63, SD= 33.69) were shown to perform significantly better than bilinguals (M =134.32, SD= 34.25; $F(1, 171)= 15.13$, $p< .001$), and Year 6 pupils (M= 162.61, SD= 27.58) significantly outperformed Year 3 pupils (M= 121.90, SD= 29.23; $F(1, 171)= 73.32$, $p< .001$) (see Figure 2).

Figure 2

Core Language Scores from the Streamlined-CELF stratified by language experience and year group.

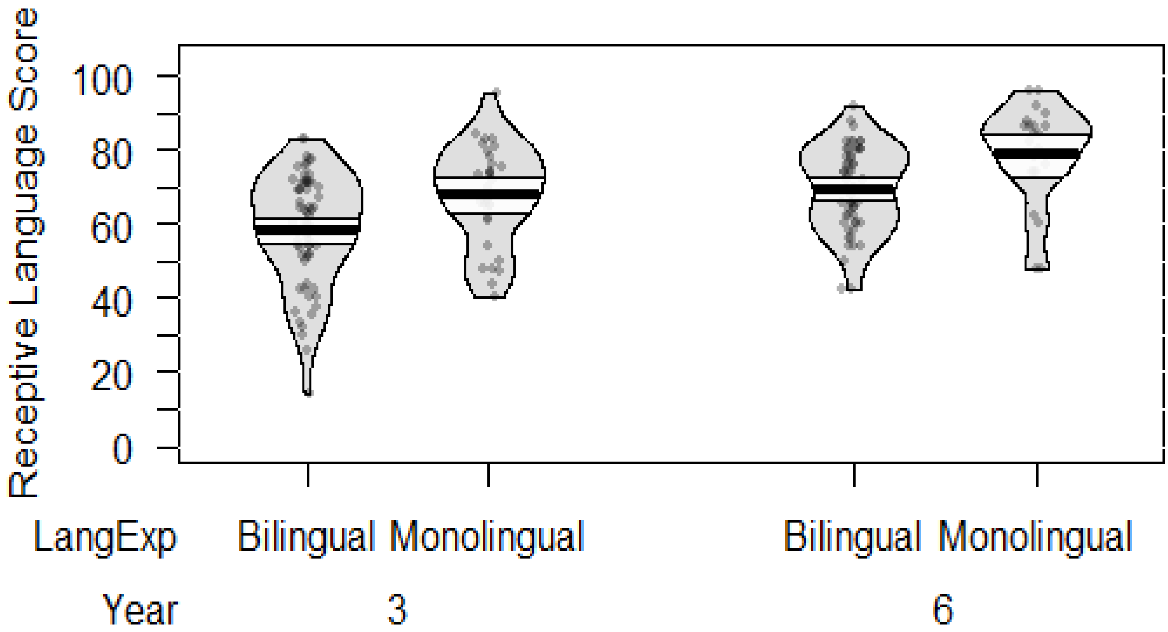


LangExp = Language Experience. Highest possible Core Language Score a child could have received was 238. Higher scores indicate superior core language skill.

This effect was mirrored in the Receptive Language Scores (RLS) where the two-way ANOVA demonstrated significant effects of both age and language experience on RLS, yet no interaction. Monolinguals (M= 72.26, SD= 14.49) performed significantly better than bilinguals (M= 63.31, SD= 14.60; $F(1, 171)= 16.83$, $p< .001$), and Year 6 pupils (M= 72.11, SD= 12.81) significantly outperformed Year 3 pupils (M= 61.38, SD= 15.14; $F(1, 171)= 23.39$, $p< .001$) (see Figure 3).

Figure 3

Receptive Language Scores gathered on the Streamlined-CELF, stratified by language experience and year group

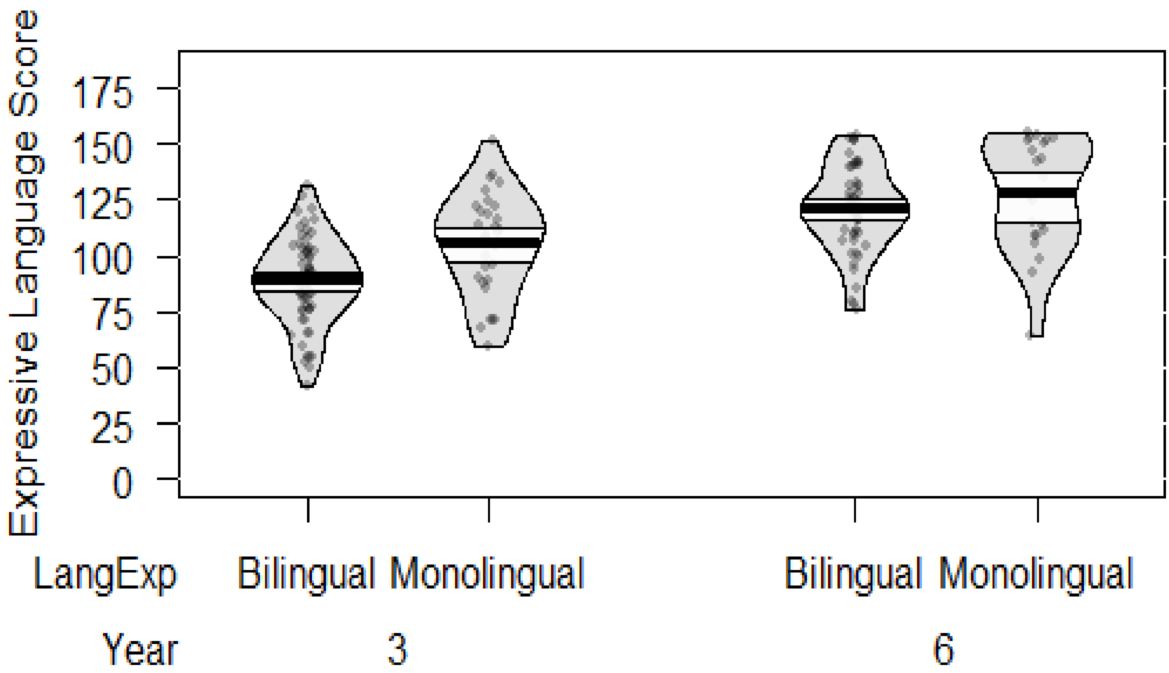


LangExp = Language Experience. Highest possible Receptive Language Score a child could have received was 104. Higher scores indicate superior receptive language skill.

Similarly, the two-way ANOVA for Expressive Language Score (ELS) demonstrated significant effects of both age and language experience, but no interaction. Monolinguals (M= 114.43, SD= 25.93) performed significantly better than bilinguals (M= 102.97, SD= 25.11; $F(1, 171)= 11.13$, $p< .005$), and Year 6 pupils (M= 122.77, SD= 21.32) significantly outperformed Year 3 pupils (M= 93.92, SD= 21.73; $F(1, 171)= 61.48$, $p< .001$) (see Figure 4).

Figure 4

Expressive Language Scores gathered on the Streamlined-CELF, stratified by language experience and year group



LangExp = Language Experience. Highest possible Expressive Language Score a child could have received was 184. Higher scores indicate superior expressive language skill.

Overall, then, across all three measures of language ability (i.e., CLS, RLS and ELS), monolinguals performed better than bilinguals, and Year 6 pupils better than Year 3 pupils but there was no significant interaction between language experience and school year group/age.

When considering only the bilingual participants, a Pearson's product-moment correlation-coefficient showed a significant, negative relationship between a child's CLI score and their RLS ($r(171)= -.195$, $p< .05$). Results of a subsequent linear regression indicated a child's CLI score to significantly predict their RLS [$F(1, 170)= 6.69$, $p< .05$] with an R-square of .038, indicating that children who spend a larger proportion of time at home speaking their home language performed worse on receptive language tasks in English. Finally, although the data met the assumptions of homogeneity of variance and linearity, and the residuals were approximately normally distributed, there was no evidence of a significant linear relationship found between a child's CLI score and their CLS or their CLI score and their ELS.

2.2.1.1.2.2 Academic Attainment

One hundred and sixty-two participants had KS1 Reading and Writing SAT scores. Year 3 and Year 6 children were analysed together for the KS1 scores. As the KS2 SATS are completed in Year 5, only the Year 6 pupils had KS2 scores; these were analyses separately. It must be noted, however, that only 68 children from Year 6 had associated KS2 Reading and Writing SAT scores.

Four separate Chi-squared tests of independence were performed to examine the relationship between language experience and performance on the SATs. Such analyses were both preliminary and necessary for inferences made further below when comparing the SAT scores to a child's language experience score (as determined by their Cumulative Language Input Score). These four analyses examined the differences between monolinguals and bilinguals for (1) KS1 Reading, (2) KS1 Writing, (3) KS2 Reading, and (4) KS2 Writing. There were no significant differences between the scores of the monolinguals and bilinguals in terms of KS1 Reading, KS1 Writing, or KS2 Reading SAT scores. There was, however, a significant difference between monolinguals and bilinguals in terms of their KS2 Writing scores ($X^2(2, N= 68)= 7.210, p< .05$) with bilinguals being more likely to be at the Expected level of attainment than their monolingual peers. These results taken together indicate that although the monolinguals and bilinguals in this sample performed equally well on measures of academic attainment taken at the end of Year 2 (KS1 SATs), the bilinguals in Year 6 obtained better scores in comparison to their monolingual peers in Writing tests taken at the end of Year 5.

Four separate multinomial logistic regressions were performed to model the relationship between bilingual language experience (as measured by their CLI score) and SATs scores. The traditional .05 criterion of statistical significance was employed for all tests and, in all cases, the reference category for the outcome variable was 'Achieving Expected'; both of the other two categories ("Performing below Expected" and "Performing above Expected") were compared to this reference group. Accordingly, CLI scores had two parameters, one for predicting the likelihood of a child performing below expected rather than expected, and one for predicting the likelihood of a child performing above expected.

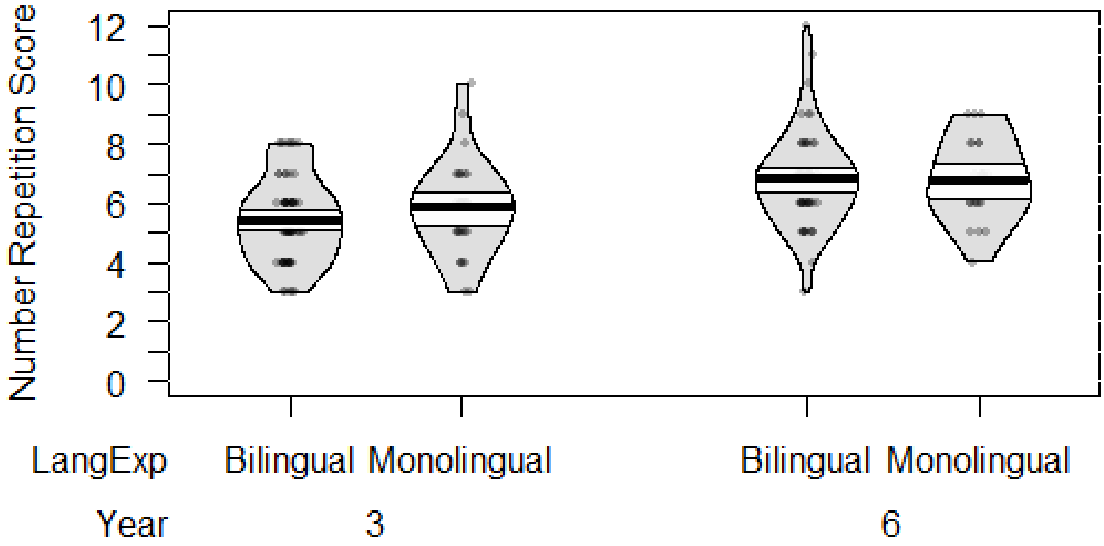
The first regression aimed to predict KS1 Writing scores using the CLI score as a predictor. The addition of this predictor to a model that contained only the intercept was found to improve the fit between model and data with near significance (χ^2 (2, N =119) = 5.736, Nagelkerke R^2 = .064, p= .057). In comparing the coefficients between those achieving an Expected result and those achieving a Below Expected result, the CLI score was found to be a significant predictor (B= 1.689, SE = .719, $p < .05$). Children with higher CLI scores were more likely to achieve a Below Expected score on the KS1 Writing SATs. For every 1 unit increase in their CLI score (which indicates more exposure to the HL) a child's odds of scoring Below Expected on the KS1 Writing SATs increased by 5.416 points. There were no other significant relationships between any of the other SAT scores and the child's CLI score.

2.2.1.1.2.3 Cognitive Expectations

Working memory was determined in two ways. Firstly the children's Number Repetition scores on the Streamlined-CELF were analysed. An initial two-way ANOVA showed no significant differences between the performance of monolinguals and bilinguals on the Number Repetition task. However, there was a significant effect of year of schooling, where the Year 3 pupils (M= 5.47, SD= 1.46) were outperformed by the Year 6 (M= 6.81, SD= 1.56) pupils [$F(1, 171)$= 26.83, $p < .001$]. There was no interaction between the two variables (see Figure 5).

Figure 5

Pirate plot showing number repetition scores, stratified by age and language experience



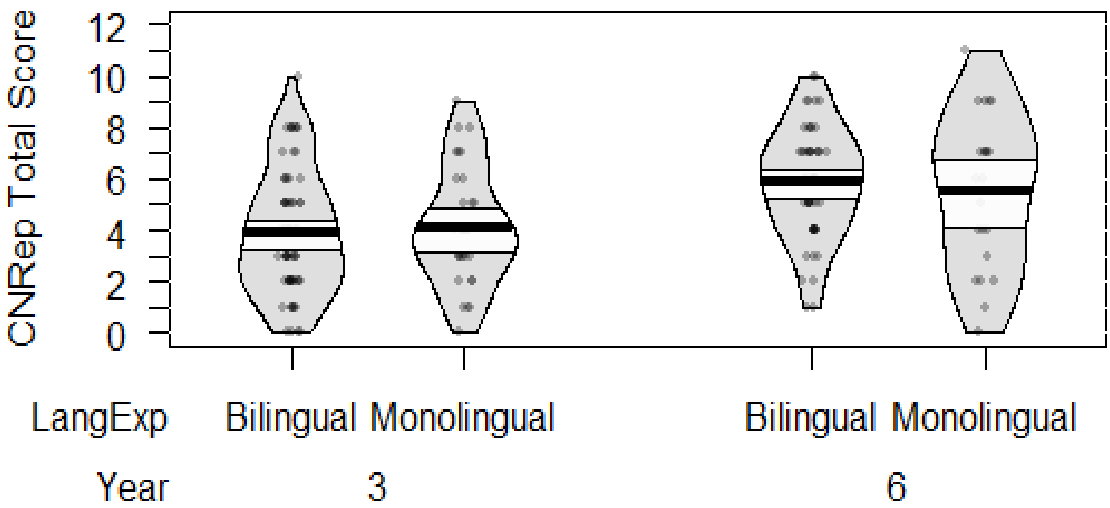
LangExp = Language Experience. Number Repetition Score is out of a possible 12.

Secondly, the children's performance on the CNRep task was analysed. Two subsequent two-way ANOVAs revealed similar effects for both the Total Score (how many of the twelve words the children were able to correctly reproduce) and the Syllable Score (how many of the 48 syllables they were able to correctly reproduce across all words). Specifically, although there were no significant differences for either of these scores between the monolingual and bilingual children, the Year 3 pupils were again outperformed by the Year 6 pupils. In terms of the Total Score, the Year 3 children were able to correctly reproduce 3.98 (SD= 2.36) of the non-words compared to an average of 5.75 by the Year 6 children (SD= 2.43; $F(1,171) = 17.01$, $p < .001$), and a total of 28.55 (SD= 10.35) syllables compared to 32.41 (SD= 11.64) by the Year 6 pupils [$F(1,171) = 3.97$, $p < .05$]. There were no significant interactions present. Results can be seen in Figure 6.

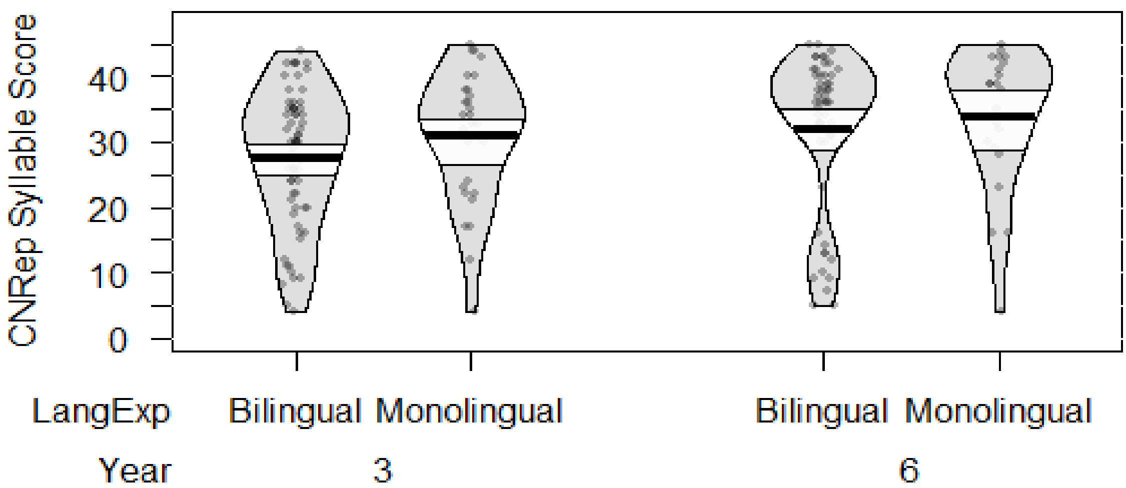
Figure 6

Total words correct (a) and number of syllables correct (b) in the CNRep task, stratified by age and language experience

(a)



(b)



LangExp = Language Experience. The Total CNRep score is scored out of a possible 12 points. The CNRep Syllable Score is out of a possible 48 points.

When considering only the bilingual participants, a Pearson's-moment correlation-coefficient revealed that there was no linear relationship between a child's CLI score and any of their working memory scores. No analyses of regression were thus performed.

2.2.1.1.2.4 Concurrent Validity

Concurrent validity was explored by studying the relationship between children's scores/responses on the LeBLEQ-C and those provided by their primary caregivers using the Parents of Bilingual Children Questionnaire (PaBiQ; Tuller, 2015).

Of the 244 questionnaires distributed to the parents/guardians of the children that participated in the study, only 22 PaBiQs were returned, a 9.02% response rate. Demographic information of the children whose parents returned PaBiQ's is presented in Table 10.

Table 10

Distribution of age, gender and language experience of the PaBiQ sample (years;months)

			Mean	Min	Max	StdDev (months)
Year 3						
	Bilinguals	Female (n= 9)	7.11	7;4	8;7	4.86
		Male (n= 3)	7;11	7;4	8;3	6.08
		Total (n= 12)	7;11	7;4	8;7	4.89
	Monolinguals	Male (n= 1)	8;4	-	-	-
Year 6						
	Bilinguals	Female (n= 6)	10;10	10;6	11;8	5.17
		Male (n= 2)	11;0	10;8	11.3	1.36
		Total (n= 8)	10;11	10;6	11;8	4.84
	Monolinguals	Female (n= 1)	11;0	-	-	-

Further exploration of the parents' completion of the PaBiQ also revealed that, on average, 18.85% of the questions contained in the questionnaires were omitted.

Moreover, despite the questionnaire being sent home with information on how it should be completed, there were still various errors. For example, in a question pertaining to the amount of time the child was exposed to each of their languages before the age four years (question 2.7), a large majority of the parents stated that they spoke English 100% of the time at home, but then also went on to state they spent 100% of the time speaking their HL in the home. Another example of this was a question asking the age at which children started speaking in sentences where one parent answered: “I want food”, (presumably the first sentence said by the child).

Due to the differing natures of the two questionnaires (and the multitude of data gathered in both), only items which were deemed vital to the creation of the CLS score were analysed (where available). Consequently, the agreement between four items on the two questionnaires was assessed; (1) country of birth, (2) language spoken (3) ages in which languages were acquired, and (4) proportion of the time HL is used in the home.

Results revealed that all the children (n= 2) who identified themselves as monolingual were confirmed to be so by their parents. However, based on the PaBIQ responses, there were three children that identified themselves as bilingual, but their parents classified them as monolingual. These three children were excluded from the analyses below. The remaining 17 children who identified themselves as bilinguals had this information confirmed by their parents. Omitting instances where responses were not given by parents in the PaBiQ, 94% of the children were correctly able to state their place of birth and 82% were correctly able to report the name of the language spoken in the home (when compared to parental report).

When analysing ages in which languages were learnt, it was found that only 50% of the children were able to state if they were simultaneous (i.e., they had learnt more than one language from birth) or sequential (i.e., they learnt languages at different ages) bilinguals. However, this finding should be interpreted with caution as the age at which languages were acquired was one of the instances where parents omitted information (9%). Additionally, on more than one occasion, parents (perhaps inadvertently) stated that their children had not been exposed to any language at all, before the age of two to three years (45%). It was assumed that this question had been misunderstood and, thus, incorrect information was provided.

A paired samples t-test demonstrated that, on average, there were no significant differences between the exposure scores generated from the PaBiQ and the LeBLEQ-C. This suggests good concurrent validity between the LeBLEQ-C and the PaBiQ, and, consequently, that the LeBLEQ-C which relies on children's self-reports is a viable method of capturing important information on language experience when parental input is limited or unavailable.

2.2.1.1.2.4.1.1 Discussion

The current study aimed to construct a novel measure that would enable the quantification of a child's degree of bilingualism/language exposure that could be gathered without the need for parental report.

As predicted, significant differences were found between the monolinguals and bilinguals, as determined by the LeBLEQ-C, in terms of their objective language scores, as measured on the Streamlined-CELF. This was true for core language, receptive language and expressive language skills. However, when considering only the bilinguals, a child's self-reported level of exposure to their home language (their Cumulative Language Input Score – CLI score) was able to significantly predict only their Receptive Language Score (not their Core Language or Expressive Language Score). Here, children who spent a higher proportion of time in the home speaking English performed better on receptive tasks. Although the predictive power of the CLI score was hypothesised to be apparent across all three language scores, retrospectively, it is logical that the CLI score would be best associated with receptive language due to its underpinning constructs. In particular, the CLI score was developed as an indication of the distribution of language spoken by *others* in the home environment of the child. Thus, it solely takes into consideration the languages the child is expected to comprehend (receptive language), not respond in (expressive language). Hence, although it is clear that the LeBLEQ-C can assist in stratifying functional monolinguals from bilinguals, thus highlighting the difference between the expressive scores of the two that has long been shown in the literature (e.g., Rescorla & Achenbach, 2002), the CLI score, which is calculated from a subset of the LeBLEQ-C questions, is more closely associated to receptive than expressive language.

The negative association with receptive language and a child's English receptive language skills is a finding in opposition to research by Cummins (1979). Cummins (1979) stated that language skills are transferred from a child's first language to their second language through a process called 'linguistic interdependence' where there is a common underlying proficiency across languages. Under this theory, it would be expected that children's receptive language would increase when exposure to another language increases; however, this pattern of results was not observed in the current study. Instead, the present findings are more in line with research that suggests increased exposure to another language is associated with lower English receptive proficiency (e.g., de Bot, 2014; Pfenninger, 2014). This is likely due to the notion that increased exposure to one language results in less exposure to another; i.e., a trade-off. As children are hearing less English, their English language skills are more limited than their monolingual peers who receive all their input in one language (Gathercole & Thomas, 2009).

Furthermore, contrary to expectations, although there was a significant difference in the working memory scores of the Year 3 children and the Year 6 children (as predicted), monolingual and bilingual children performed similarly on all three tasks (forward digit recall, backward digit recall and CNRep) of working memory. Moreover, there was no relationship between a bilingual child's CLI score and their performance on such tasks. Thus, there were no significant differences between monolingual and bilingual children in the working memory tasks used. However, it must be noted that the tasks used in the study are relatively uncomplicated memory tasks which may not be demanding enough to exert differences in the performance of monolinguals and bilinguals. In particular, Bialystok et al. (2004) presented monolingual and bilingual adults with a Simon task where they were required to indicate the colour of a square by pressing the appropriate response key. This task contained a working memory component whereby there were two-stimulus or four-stimulus conditions. Consequently, in the latter, participants were required to hold more stimulus-response pairings in mind. There were no differences in response time between monolinguals and bilingual in the 2-stimulus condition. However, monolinguals took longer to respond in the 4-stimulus condition than bilinguals with a larger difference in reaction times being apparent in older adults than younger adults. Bialystok et al. (2004) suggested that this difference may have been due to task demands and the task being too simplistic for the differences in monolinguals and bilinguals to be apparent. More recent research has similarly found no difference between monolingual and bilingual performance in tasks comparing *simple* working memory skills (Bialystok & Feng, 2009; Bonifacci et al., 2011; Engel de Abreu, 2011). The present findings add to this body of literature.

As predicted, analyses determined that there were no significant differences between monolinguals and bilinguals in terms of the majority of the SAT scores. This is a useful finding as it suggests that although bilinguals receive less exposure to English, they are not detrimentally affected in terms of their academic attainment. SAT scores are vital in the academic setting for a multitude of reasons. Primarily, SAT results are vital for identifying the children who either begin primary education at a lower level than their peers or develop at slower rate than expected. SAT results, thus, help identify the children who may benefit from additional support within the classroom environment and measures that can be put in place to assist them. Furthermore, the use of SAT scores are not just limited to a child's primary school education. As such, a child's KS2 attainment is often used by secondary schools to place children in the subject sets that are appropriate for the child's level of ability. However, when it then comes to GCSEs, children placed in lower sets often do not have the opportunity to take the higher tier papers, meaning their attainment levels are capped and the top grades are not attainable. Lower GCSE grades can then, subsequently, impact the children's wider opportunities after leaving secondary education in terms of further education and job opportunities.

However, it must be noted that language exposure did impact children's KS1 and KS2 Writing SATs. As such, children with higher CLI scores (i.e., increased exposure to the home language, determined using the continuous score generated using the LeBLEQ-C) were likely to achieve a lower grade in the KS1 Writing SATs, but bilinguals (using a binary classification system) were more likely to a achieve higher grade than their monolingual peers in KS2 Writing SATs. This could be due to a number of reasons. Firstly, it could be that the bilinguals did indeed demonstrate increased improvement between KS1 and KS2 which corroborates previous literature (e.g., Hutchinson, 2018). Alternatively, the results could simply be due to the cross-sectional nature of the data. It is possible that this effect might disappear if the data were collected longitudinally. Further research should address this to provide a more detailed view of the bilingual language experience.

Finally, results from the LeBLEQ-C were compared to those obtained from a standardised parental-report method developed to collect information regarding children’s language experience – the PaBiQ (Tuller, 2015). Importantly, moderate consensus between the two measures was found verifying that the LeBLEQ-C is a useful tool in capturing important information on language experience when parental input is limited or unavailable. It must be noted that identical responses were not expected between the two questionnaires. Parents have long been found to be reliable sources in reporting information of their children (e.g., Espinosa & López, 2007), while children, in comparison, may be more limited in answering such detailed questions. Thus, the aims of the LeBLEQ-C were not to develop a method of obtaining information on language experience identical to that which could be obtained from parents, but rather gaining an *indication* of language experience that could be employed when such responses from parents were not available.

Furthermore, only a small percentage of the PaBiQs were returned. Though this provides a very limited sample size for investigating the relationship between the newly developed LeBLEQ-C and a pre-existing measure of language exposure, this marginal response rate serves to emphasise the limited response rates in parental reports. In addition, even in instances where the PaBiQ questionnaires were indeed returned to the researchers, there was a substantial number of questions left unanswered or incorrectly interpreted. Although this omission of certain questions did not affect the method of scoring used in the current study, if the traditional scoring system of the PaBIQ would have been employed, this would have led to a considerable gap in knowledge and, resultantly, a sizeable amount of children's scores unable to be obtained. Consequently, it is argued that although the LeBLEQ-C may not necessarily be the most detailed method in gathering information of language exposure, it may be researchers’ best, and in some cases, only, method in gathering such information in populations where parental engagement is known to be limited.

Additionally, a nonresponse bias may have caused a skew in the data. Children's responses on the LeBLEQ-C could be compared to the PaBIQ and validated only if a PaBIQ questionnaire was also returned, yet only a very small percentage of parents returned the PaBIQ. Although it is not known why more parents/guardians did not return the PaBIQ, the factors mentioned in the introduction of the current chapter may be possible contributors. Potential reasons include the low SES of the group which is known to affect the rate of parental response (Curtin et al., 2000; Groves et al., 2000; Warriner et al., 2002) or Bradford being a highly urbanised city which is also known to affect parental response (Eaker et al., 1998). Furthermore, the city of Bradford consists of a diverse population with over a third of its population from ethnic minority backgrounds and 153 languages spoken by children attending its schools (Department for Education, 2016). With research stating individuals to be less likely to respond to questionnaires or surveys if they are from non-white populations (Curtin et al., 2000; Groves et al., 2000; Voigt et al., 2003) or if the questionnaire is not presented in their first language (Harzing, 1997), it is understandable why such low response rates were observed. This only provides further reasoning as to why a measure of language experience that does not rely on parental involvement would be highly informative in a multitude of situations.

In conclusion, the LeBLEQ-C has been shown to have sufficient levels of construct and concurrent validity, with a child’s language experience being indicative, to some degree, of components of language ability. Though care must be taken when using a child’s Cumulative Language Input score (CLI score), due to the nature of self-report methods, the output has been shown to provide a reliable indicator of language experience in settings where parental engagement is low. Such a tool is highly valuable in both clinical and educational settings where gathering such vital information would have previously been near impossible. Consequently, the added knowledge of a child’s degree of bilingualism can assist in the identification of children whose linguistic skills are below what would be expected from their self-reported level of exposure. Interventions can subsequently be put in place in order to ensure such children are not at a detriment in classroom environment due to poor oral language skills.

2.2.1.2 Leeds-Bradford Language Exposure Questionnaire for Teachers (LeBLEQ-T)

The Leeds-Bradford Language Exposure Questionnaire for Teachers (LeBLEQ-T) was developed to be complementary to the LeBLEQ-C. Aiming to gain information on the development of verbal and non-verbal skills displayed in the classroom environment, as well as contexts in which reduced language performance is likely to occur, the questionnaire was designed to be given to the classroom teachers of the children being studied, again circumventing the need for parental report. Questions were constructed with the aim of determining the degree to which a child was thought to perform on various tasks involving reading, writing, speaking, listening, as well as some more general aspects of language (e.g., body language, and the ability to empathise). Though the LeBLEQ-T was not designed to obtain a score of language exposure per se, it works in collaboration with the LeBLEQ-C to give a more comprehensive view of a child's language skills.

The scores generated from the LeBLEQ-T were expected to be associated with other measures of linguistic competencies, such as academic attainment and language proficiency, as well as the exposure to English a child may receive. Specifically, children who are rated as being more linguistically proficient by their teachers, are likely to also have increased Core, Receptive and Expressive Language skills (calculated from the Streamlined-CELF), higher SAT scores, and (due to higher exposure to the English language) a lower Cumulative Language Input Score (CLI score); calculated from the LeBLEQ-C). These predictions are based on research by Rimfeld et al. (2019) who found teachers to be reliable predictors of children's achievement (as discussed in the General Introduction). Specifically, in the case of the SAT scores, it was hypothesised that Reading and Writing scores would be associated with the Total LeBLEQ-T Score, while the LeBLEQ-T Reading Score may be an additional predictor in the case of the Reading SATs. It was also predicted that the LeBLEQ-T Writing Score may be an additional predictor of the Writing SATs. Finally, it was predicted that there would be no difference between the LeBLEQ-T scores of the children in Year 3 and Year 6 as the measure was not designed to be indicative of actual skill level across different school year groups, but rather of performance in accordance to within-school year expectations.

Ultimately, the questionnaire aimed to explore the reliability of teacher's opinions of a child's language experience and how this relates to their performance on standardised tests (SATs) within the context of a classroom. The initial LeBLEQ-T was first piloted in one school population before a shorter, more refined version was administered. This was then administered to a different school population to explore the questionnaire's concurrent validity with measures of linguistic ability, academic attainment and language experience.

2.2.1.2.1 Piloting of the LeBLEQ-T

In order to validate the LeBLEQ-T and its viability to be administered in the classroom environment, the first stage in the development was to pilot the first version of the questionnaire. This would allow for the streamlining of the questionnaire.

2.2.1.2.1.1 Methods

2.2.1.2.1.1.1 Participants

Three schools were recruited for the piloting phase of the LeBLEQ-T and data were gathered on 134 children in Year 3, between the ages of 7 years, 7 months to 8 years, 8 months (M= 8 years, 1 month, SD= 3.80 months). No children had any known hearing problems or developmental delays. Participants were stratified into monolingual and bilingual groups on the basis of responses on the LeBLEQ-C. Both monolinguals and bilinguals were recruited from the same schools. Children with any exposure to a second language outside of the school environment were classed as bilingual, while those who had no exposure to an additional language were classified as monolinguals (determined using the LeBLEQ-C). All other children were classes Table 11 summarises the distributions of age, gender and language experience.

Table 11

Distribution of age and gender of the sample (years;months)

	Gender	Mean	Min	Max	St.Dev. (months)
Bilinguals (n= 104)	Female (n= 54)	8;1	7;7	8;8	3.92
	Male (n= 50)	8;1	7;7	8;7	3.60
	Total (n= 104)	8;1	7;7	8;8	3.76
Monolinguals (n= 30)	Female (n= 16)	8;0	7;7	8;6	3.96
	Male (n= 14)	8;1	7;7	8;8	4.13
	Total (n= 30)	8;1	7;7	8;8	4.00

All children received schooling in the English language, and all bilingual children were exposed to another language in their home environment to varying extents. There was a total of 11 home languages in the sample: Punjabi *(n=40),* Bengali *(25),* Urdu *(23),* Arabic *(8),* Czech *(2),* Pashto *(2),* French *(1),* Gujarati *(1),* an unspecified dialect of South African *(1),* and unspecified dialects of Indian (*3*).

Socio-economic Status (SES) was determined using the Income Deprivation Affecting Children Index (IDACI; Communities & Local Government, English Indices of Deprivation, 2015), again based on the school's postcode, and can be seen in Table 12.

Table 12

IDACI ranks, deciles and scores of the three schools sampled

	IDACI Rank	IDACI Decile	IDACI Score
School 1	3933	2	0.368
School 2	21138	7	0.101
School 3	6466	2	0.304

IDACI Rank = 32,844 neighbourhoods in England are ranked from the most deprived (a score of 1) to the least deprived (a score of 32,844). IDACI Decile = Ranks are divided into ten equal groups where 1 indicates the most deprived and 10 indicates the least deprived. IDACI Score = Percentage of children in that area classed as deprived

Classroom teachers completed one LeBLEQ-T questionnaire for each child included in the sample. There were, thus, three teachers who assisted in the completion of the LeBLEQ-Ts, one from each class. No further demographic information was gathered about these teachers.

School consent was first obtained from the headteachers of each school. Once this written consent was obtained, information sheets and consent forms were distributed to all parents of children in Year 3 via classroom teachers. Parents were given the option to withdraw from the study by returning an opt-out consent form. If the opt-out form was not returned, consent was implied. Finally, before taking part in the study, children were informed about the tasks they would be asked to do and their assent was obtained. The study received approval from the Ethics Committee of the School of Psychology, University of Leeds (ethics reference number: PSC-313, approved 23/01/2019).

2.2.1.2.1.1.2 Materials

The initial version of the LeBLEQ-T was designed to obtain information on teachers' perceptions of a child's verbal and nonverbal development, including contexts in which reduced language performance is likely to be seen (e.g., "The child struggles following spoken directions"). Teachers were asked how often they observed a child to be performing well/poorly on a number of tasks. The design of the initial LeBLEQ-T was influenced from various pre-existing questionnaires such as the CELF-4 UK Pragmatics Profile and Observational Rating Scale (Semel, Wiig & Secord, 2003), but also included a number of novel items (e.g., “To the best of your knowledge, what language do you believe to be spoken most often in the home of this child?”). All questions were designed so that they could be answered by teachers or teaching assistants who spend a substantial amount of time with the child in question. Thus, only questions regarding skills observable in the classroom environment were asked.

The original questionnaire comprised of 40 items; one binary question at the beginning of the initial LeBLEQ-T (“To the best of your knowledge, does this child speak another language in the home?”) and 39 multiple-choice questions. These were organised into eight subsets: (1) Current Skills; (2) Conversation Skills; (3) Exchanging Information; (4) Non-verbal Communication; (5) Listening; (6) Speaking; (7) Reading; and (8) Writing. An example of the type of question in each subset is presented in Table 13.

Table 13

Examples of questions in each subset of the initial version of the LeBLEQ-T

	Subset	Example Question
(1)	Current Skills	How often does this child incorrectly pronounce words?
(2)	Conversation Skills	How often does the child ask appropriate questions during a conversation?
(3)	Exchanging Information	How well does the child ask for help appropriately?
(4)	Non-verbal Communication	How often does the child correctly respond to facial cues?
(5)	Listening	How often does the child struggle to follow spoken directions?
(6)	Speaking	How often does the child speak in complex sentences?
(7)	Reading	How often does the child struggle comprehending what they have read?
(8)	Writing	How often does the child use poor grammar when writing?

Due to the LeBLEQ-T being designed to be completed by teachers in their own time, and not administered by researchers, response options had to be carefully developed to avoid ambiguity and/or misinterpretation. Past research has shown that though the number of item responses and the levels to which they correspond (e.g., 'Easy,' 'Moderate,' 'Difficult') can vary throughout a questionnaire (Streiner & Norman, 2003), too many response options in a self-report questionnaire can contribute to error (Bond, 2003), whilst too few can reduce responsiveness (Bovolenta et al., 2009). Moreover, if there are too few response options, this can contribute to an increase of floor and ceiling effects, limiting the value of the information gathered (Merbitz et al., 1989). Thus, the 39 ordinal questions of the LeBLEQ-T were designed so that they could be scored on the same five-point Likert scale used in the LeBLEQ-C ('Never,' 'Rarely,' 'Half of the time,' 'Often,' and 'Always').

2.2.1.2.1.1.2.1 Scoring of the LeBLEQ-T

One of the questions in the first subset was a closed question requiring a binary response (yes/no). The other was an open question ("what is the name of the language the child speaks in the home?"). The remaining questions were on the same Likert scale as in the LeBLEQ-C. Response options were again assigned values ('*Never'*=0, '*Rarely'*=1, '*Sometimes'*=2, '*Often'*=3, '*Always'*=4) in order to generate overall scores from the LeBLEQ-T.

2.2.1.2.1.1.2.2 LeBLEQ-T Appraisal Questionnaire

In addition to the LeBLEQ-T, teachers were required to complete a short appraisal form. This appraisal form contained four questions regarding both the content and the layout of the LeBLEQ-T. Questions included were (1) *Did any of the questions seem unclear?*; (2) *Did all of the questions follow a logical order?*; (3) *Did you feel uncomfortable answering any of the questions?;* and (4) *Do you have any further comments regarding the questionnaire?.* All questions were open-ended and consequently, the Appraisal Questionnaire provided qualitative data that aided the evaluation and further development of the first version of LeBLEQ-T.

2.2.1.2.1.1.3 Procedure

The LeBLEQ-T was presented as a pen and paper questionnaire to teachers on the first day of testing at the school. Teachers were able to take the questionnaires and appraisal forms away and return them within two weeks. Each LeBLEQ-T took approximately six minutes to complete. Though researchers were not present at the time of completion, teachers were instructed to complete the questionnaires in a reasonably quiet space to minimise distractions. The appraisal questionnaire was distributed to the teachers alongside the LeBLEQ-T. All three teachers completed one Appraisal Questionnaire each, taking approximately a minute to complete.

2.2.1.2.1.1.4 Statistical Analyses

Firstly, the Appraisal Questionnaire was considered, and any issues teachers had with the items were addressed, either by omitting or amending items. Secondly, aspects of traditional Rasch Analysis (Rasch, 1960) were incorporated in order to investigate the structure of the questionnaire and the appropriateness of responses gathered. Rasch Analysis is a statistical procedure used to develop psychometrically sound measures (Rasch, 1960). Originally based on Item Response Theory (IRT) and using Guttman scaling (Guttman, 1950), Rasch Analysis determines the probability of a person endorsing a particular item based on both the difficulty of the item and the ability of the person (Tennant & Conaghan, 2007).

However, due to the multifaceted and interrelated nature of language and the overlap anticipated between subsets (e.g., skills that draw upon reading are likely to be associated with skills related to writing), the questionnaire possessed a high rate of multicollinearity, thus violating a primary assumption of the Rasch statistical test. As the aim of the questionnaire was to capture information about all aspects of language (i.e., speaking, listening, reading, writing, pragmatics), including questions that appeared similar yet drew upon different linguistic skills was deemed important. As such, although various Rasch procedures were employed to increase the soundness of the LeBLEQ-T, (which shall be discussed further in the results section), the correlation between questions remained high and the multicollinearity persisted. Consequently, it can be said that only a *partial* Rasch Analysis was conducted as all principles of Rasch were addressed, apart from this assumption multicollinearity. RUMM2030 software was employed to analyse the data (version 5.4 for Windows; Andrich & Sheridan, 2011).

2.2.1.2.1.2 Results and Discussion

The first phase in streamlining the LeBLEQ-T was to address the issues highlighted by the teachers in the Appraisal Questionnaire. All three appraisal questionnaires were returned and two indicated a single item with which the teachers were concerned, namely: "*How often does the child respond to or offer expressions of affection*?" with one teacher stating it was "inappropriate for a teacher to comment on". Consequently, this item was removed from the revised questionnaire. Secondly, the questions deemed ineffective, or repetitive were removed. Following post-hoc inspection of the questionnaire, two items were deemed to be highly similar by the researchers. These questions were "does the child introduce appropriate topics of conversation" and "does the child ask appropriate questions during conversation." These two questions were consequently amalgamated into one question.

Finally, aspects of Rasch analysis were then employed to inform how best to streamline the tool. Assumptions of Rasch Analysis and the number of items removed due to violating such assumptions are outlined in Table 14. All analyses were conducted using the appropriate measures on RUMM software (see Bond et al., 2020, Bond & Fox, 2007, or Tennant & Conaghan, 2007 for a more detailed explanation of the assumptions checked and statistical tests conducted).

Table 14

Explanations of the assumptions of Rasch Analysis and the number of items removed from the LeBLEQ-T after piloting due to violating these assumptions

Assumption	Explanation	Number of items omitted
Ordered thresholds	The degree in which response categories are appropriately sequenced. The ordering of responses should be logical, with the highest probability of the sample endorsing a particular category succeeding the highest probability of endorsing the previous item. *E.g., The highest probability of an individual endorsing 'easy' should proceed the highest probability of them endorsing 'moderate.'*	3
Response Dependency	The likelihood of individual's response to one item being likely to predict their response to another. *E.g., If a person is able to endorse an item which asks if they can walk a mile unaided, they will also endorse the item that asks if they can walk ten feet*.	1
Item Response Bias	When different groups within the sample respond differently to particular items, despite equal measures of the underlying trait being measured. *E.g., If males consistently respond differently to females on a particular item.*	2
Person-Item distribution	Indicates sample includes individuals along a wide range of proficiencies and items along a spectrum of difficulty. *E.g., Not all people are of above average intelligence or not all questions are assessing something extremely difficult.*	2

Piloting of the first version of the LeBLEQ-T revealed gaps in the information gathered. Consequently, the first section of the questionnaire saw the addition of two further items asking of teacher's knowledge regarding a child's language experience in the home. These questions were

(1) Please name the other language this child speaks

(2) To the best of your knowledge, what language do you believe the child to speak most often *in the home?*

As a result of piloting and Rasch analysis, the final version of the LeBLEQ-T was revised to include 32 items; two open-ended questions, one closed question (with a binary response) and 29 ordinal questions. This was substantially reduced from the original 40 item questionnaire. Furthermore, the items were restructured to allow for better cohesiveness, consequently, easing completion and reducing the time taken for teachers to complete the LeBLEQ-T. These restructured subsets were: (1) General Language Information, (2) Listening, (3) Speaking, (4) Reading, (5) Writing, and (6) Interpersonal Skills. An example of the type of questions in each subset is presented in Table 15.

Table 15: Examples of questions in each subset of the revised LeBLEQ-T

	Subset	Example Question
(1)	General Language Information	To the best of your knowledge, what language do you believe to be spoken most often in the home of this child?
(2)	Listening	How often does the child struggle to follow spoken directions?
(3)	Speaking	How often does the child struggle to speak in complex sentences?
(4)	Reading	How often does the child struggle to comprehend what they have read?
(5)	Writing	How often does the child use poor grammar when writing?
(6)	Interpersonal Skills	How often does the child appropriately respond to emotion?

These alterations reduced the administration time of the pilot LeBLEQ-T questionnaire from seven minutes to approximately three minutes for the final, revised version.

2.2.1.2.2 The Final LeBLEQ-T

The second stage in the development of the LeBLEQ-T was to ensure it had sufficient concurrent validity with other measures of linguistic ability.

2.2.1.2.2.1 Methods

2.2.1.2.2.1.1 Participants

Three different schools were recruited for the final phase of the LeBLEQ-T validation. Data were collected from 113 children; 65 from Year 3 (M= 7 years, 9 months, SD= 3.61 months) and 48 from Year 6 (M= 10 years, 8 months, SD= 3.98 months). No children had any known hearing problems or developmental delays. Participants were stratified into monolingual and bilingual groups determined by the LeBLEQ-C. Table 16 summarises the distributions of age, gender and language experience.

Table 16

Distribution of age and gender of the sample for the amended LeBLEQ-T (years;months)

		Mean	Min	Max	StdDev (months)
Year 3					
Bilingual	Female (n= 27)	7;8	7;3	8;2	3.07
	Male (n= 16)	7;9	7;2	8;3	4.11
	Total (n= 43)	7;8	7;2	8;3	3.51
Monolingual	Female (n= 7)	7;11	7;4	8;5	3.99
	Male (n= 15)	7;10	7;3	8;5	3.23
	Total (n= 22)	7;10	7;4	8;5	3.47
Year 6					
Bilingual	Female (n= 12)	10;10	10;2	11;5	4.74
	Male (n= 20)	10;7	10;3	11;1	3.51
	Total (n= 32)	10;8	10;2	11;5	4.13
Monolingual	Female (n= 9)	10;11	10;8	11;4	2.54
	Male (n= 7)	10;8	10;3	11;1	3.68
	Total (n= 16)	10;9	10;3	11;4	3.24

All children received schooling in the English language, and all 75 bilingual children were exposed to another language in their home environment to varying degrees. There was a total of 8 home languages (HL) reported in the sample: Punjabi *(n=30),* Bengali *(1),* Urdu *(28),* Pashto *(3),* Polish (*1*), Latvian (*1*), Slovak *(4),* and an unspecified dialect of South African *(1).* Furthermore, there were six children who, when asked, were unsure of the language spoken at home.

Once again, SES was determined using IDACI scoring (Communities & Local Government, English Indices of Deprivation, 2015) based on the school's postcode and can be seen in Table 17.

Table 17

IDACI ranks, deciles and scores of the three schools sampled

	IDACI Rank	IDACI Decile	IDACI Score
School 1	8685	3	0.221
School 2	5191	2	0.283
School 3	7173	3	0.245

IDACI Rank = 32,844 neighbourhoods in England are ranked from the most deprived (a score of 1) to the least deprived (a score of 32,844). IDACI Decile = Ranks are divided into ten equal groups where 1 indicates the most deprived and 10 indicates the least deprived. IDACI Score = Percentage of children in that area classed as deprived

Classroom teachers completed one LeBLEQ-T per child included in the sample. There were, thus, three teachers who assisted in the completion of the LeBLEQ-Ts, one from each class. No further demographic information was gathered about these teachers.

School consent was first obtained from the headteachers of each school. Once this written consent was obtained, information sheets and consent forms were distributed to all parents of children in Year 3 and Year 6 via classroom teachers. Parents were given the option to withdraw from the study by returning an opt-out consent form. If the opt-out form was not returned, consent was implied. Finally, before taking part in the study, children were informed about the tasks they would be asked to do and their assent was obtained. Ethical approval was granted by the University of Leeds Ethics Committee (PSC-586, approved 23/01/19).

2.2.1.2.2.1.2 Materials and Procedure

The Streamlined-CELF was administered as described in Section 2.1.3 of this chapter. Children's KS1 SATs and, in the case of the Year 6 pupils, KS2 SATs

(Reading and Writing scores in both cases) were provided by each school. Finally, the LeBLEQ-C was administered to all participants.

2.2.1.2.2.1.2.1 Scoring of the LeBLEQ-T

One of the questions in the first subset was a closed question requiring a binary (yes/no) response, while the other two were open questions. These questions were "What other language does the child speak?" and "To the best of your knowledge, what language do you believe to be spoken *most often* in the home of this child?". The remaining questions were on the 5-point Likert scale described previously. Responses to the Likert questions were assigned values ('*Never'*=0, '*Rarely'*=1, '*Sometimes'*=2, '*Often'*=3, '*Always'*=4) in order to generate overall scores from the LeBLEQ-T. For 16 of the 29 multiple-choice questions in the final version of the LeBLEQ-T, a raw score of *0* (i.e., the child *never* does this) would imply a good level of ability in that domain. For the remaining 13 Likert questions, a raw score of 0 would imply reduced ability in that domain. These 13 questions were, thus, reverse scored and summed to produce an overall score generated by the LeBLEQ-T, as well as five, domain-specific scores for Listening, Speaking, Reading, Writing and Interpersonal skills. Total scores ranged from 0-116, with scores closer to 0 indicating superior linguistic ability, according to the classroom teachers.

2.2.1.2.2.2 Results

Six two-Way ANOVAs were conducted to determine the effects of age (Year 3 vs Year 6) and language experience (monolingual vs bilingual) on teachers' perceptions of children's abilities within the classroom environment. The dependent variables included the Total LeBLEQ-T score and scores from each of the subsets of the LeBLEQ-T: (1) Listening, (2) Speaking, (3) Reading, (4) Writing, and (5) Interpersonal Skills. For all ANOVAs, there were no main effects of either age group or language experience. However, there were significant interactions between age and language experience for three of the ANOVAs. In particular, the Total LeBLEQ-T score [$F(1, 119)= 5.16$, $p< .05$], the LeBLEQ-T Listening Score [$F(1, 109)= 4.28$, $p< .05$] and the LeBLEQ-T Interpersonal Skills Score [$F(1, 109)= 3.44$, $p< .05$] ANOVAs revealed that bilinguals were perceived to have poorer abilities in the classroom in Year 3, but then outperformed their monolingual peers in Year 6. The same pattern of a shift in abilities between monolinguals and bilinguals across age groups was also seen in the LeBLEQ-T Speaking Score [$F(1, 109)= 3.53$, $p= .063$] and the LeBLEQ-T Writing Score [$F(1, 109)= 3.82$, $p= .053$] ANOVAs (though these were only trends) as well as the LeBLEQ-T Reading Score ANOVA [$F(1,109)= 2.18$, $p= .143$] (though this interaction was non-significant).

2.2.1.2.2.2.1 Cumulative Input of English

A Pearson-moment correlation-coefficient was conducted to determine if a teacher's perception of a child's abilities within the classroom could be mapped onto their cumulative language exposure, as determined by their CLI score generated by the LeBLEQ-C. Seventy children within the sample had associated CLI score scores. A significant positive correlation was found between children's LeBLEQ-T Speaking Score and their CLI score ($r(69)= .297$, $p< .05$). This indicates that the less exposure a bilingual child has to their HL, the better their verbal skills tend to be, as rated by their teacher. No other scores from the LeBLEQ-T correlated with a child's CLI score. A simple Linear Regression was, thus, performed to predict a child's CLI score based on their LeBLEQ-T Speaking Score. A significant relationship was found [$F(1, 68)= 6.48$, $p< .05$], with an Adjusted R-square of .075 ($\beta = .013$), suggesting that the proportion of English used in a child's home environment can be predicted, to some extent, by a teacher's perception of their speaking abilities. However, this does not apply to other skills observed in the classroom.

2.2.1.2.2.2.2 Language Proficiency

A Pearson-moment correlation-coefficient was conducted to determine if a teacher's perception of a child's abilities within the classroom could be mapped onto a child's language ability as determined using the Streamlined-CELF. All five subsets of the LeBLEQ-T as well as the Total LeBLEQ-T Score showed significant negative correlations with a child's Core Language Score (CLS), Receptive Language Score (RLS), and Expressive Language Score (ELS). Results can be seen in Table 18. These results indicate that children who are rated by their teachers as possessing superior Listening, Speaking, Reading, Writing and Interpersonal Skills are also more likely to perform better on objective measures of language ability. This pertains not only to a child's core language skill but also their distinct abilities to understand the language of others (RLS) and produce language of their own (ELS).

Table 18

Correlation matrix of the relationship between LeBLEQ-T and Streamlined-CELF scores

	Streamlined-CELF		
	Core Language Score	Receptive Language Score	Expressive Language Score
LeBLEQ-T Listening Score	-.314**	-.339**	-.291**
LeBLEQ-T Speaking Score	-.286**	-.312**	-.273**
LeBLEQ-T Reading Score	-.410**	-.420**	-.372**
LeBLEQ-T Writing Score	-.411**	-.433**	-.375**
LeBLEQ-T Interpersonal Skills Score	-.301**	-.287**	-.306**
Total LeBLEQ-T Score	-.360**	-.370**	-.343**

**. Correlation is significant at the 0.01 level (2-tailed), n= 113

Consequently, three separate Stepwise Multiple Regressions were conducted to determine if a child's CLS, RLS, and ELS could be predicted by their LeBLEQ-T scores. In all three regressions, all six scores from the LeBLEQ-T (i.e., Listening, Speaking, Reading, Writing, Interpersonal Skills and Total) were entered as predictors. A child's LeBLEQ-T Writing Score was found to be a significant predictor of Core Language Ability [$F(1, 112)= 22.51$, $p< .001$], with an Adjusted R-square of .161 ($\beta = -3.09$). LeBLEQ-T Writing Score was also shown to be a significant predictor for a child's RLS [$F(1, 112)= 25.54$, $p< .001$, $R^2 = .180$, $\beta = -1.47$] and ELS [$F(1,112)= 18.12$, $p< .001$, $R^2 = .133$, $\beta = -2.07$]. These findings demonstrate that a teacher's perception of a child's writing ability is the biggest predictor of their objective language skill as measured by the Streamlined-CELF, both in terms of their Receptive and Expressive abilities as well as the combination of the two.

2.2.1.2.2.2.3 Academic Attainment

Sixty-three of the sampled Year 3 children had KS1 Reading and Writing scores. Forty-four of the Year 6 children had KS1 Reading and Writing scores while 48 had KS2 Reading and Writing Scores. Four initial Chi-squared tests were conducted to determine the differences between monolinguals and bilinguals in terms of their KS1 Reading, KS1 Writing, KS2 Reading and KS2 Writing scores. Analyses revealed monolingual and bilingual children to perform similarly in all four tests of academic attainment (with $p < .05$). Consequently, both monolinguals and bilinguals were included together in the subsequent analyses.

Four separate multinomial logistic regressions were performed to model the relationship between teacher's perception of pupil's classroom-based skills and SAT scores. Both monolinguals and bilinguals were included in analysis and regressions compared scores obtained on the LeBLEQ-T to (1) KS1 Writing scores, (2) KS1 Reading scores, (3) KS2 Writing Scores, and (4) KS2 Reading scores. The traditional .05 criterion of statistical significance was employed for all tests and, in all cases, the reference category for the outcome variable was 'Achieving Expected'; both of the other two categories ("Performing Below Expected" and " Performing Above Expected") were compared to this reference group. Accordingly, the CLI score had two parameters, one for predicting the likelihood of a child performing below expected rather than expected, and one for predicting the likelihood of a child performing above expected.

The first regression aimed to predict KS1 Writing scores using the Total LeBLEQ-T Score and the LeBLEQ-T Writing Score as predictors. The addition of these two predictors to a model that contained only the intercept was found to significantly improve the fit between model and data (χ^2 (4, N= 65) = 36.469, Nagelkerke R^2= .594, p< .001). As shown in Table 19, significant unique contributions were made by both using the Total LeBLEQ-T Score and the LeBLEQ-T Writing Score.

Table 19

Contributions of the Total LeBLEQ-T Score and the LeBLEQ-T Writing Score to a child's KS1 Writing score as identified in a regression analysis

Predictor	χ^2	Df	*p*
Total LeBLEQ-T Score	17.312	2	< .001**
LeBLEQ-T Writing Score	9.545	2	.008**

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

In comparing the coefficients between children achieving an Expected result and children achieving a Below Expected result, only the Total LeBLEQ-T Score was found to be a significant predictor (B= .081, SE= .033, p< .05, OR= 1.085). Children with higher Total LeBLEQ-T Scores were more likely to achieve a Below Expected score on the KS1 Writing SATs. For every 1 unit increase in the Total LeBLEQ-T Score (which indicates poorer overall Language-based skills) the probability of scoring Below Expected on the KS1 Writing SAT increased by 1.085.

The second regression aimed to predict KS1 Reading scores using the Total LeBLEQ-T Score and the LeBLEQ-T Reading Score as predictors. The addition of the predictors to a model that contained only the intercept was found to significantly improve the fit between model and data (χ^2 (4, N= 65) = 29.606, Nagelkerke R^2= .502, *p*< .001). As shown in Table 20, borderline significant contributions were made from the LeBLEQ-T Reading Score to the model.

Table 20

Contributions of the Total LeBLEQ-T Score and the LeBLEQ-T Reading Score to a child's KS1 Reading score as identified in a regression analysis

Predictor	χ^2	Df	*p*
Total LeBLEQ-T Score	1.447	2	.485
LeBLEQ-T Reading Score	5.707	2	.058

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

When comparing coefficients between children gaining an Expected result and those gaining a Below Expected result, only the LeBLEQ-T Reading Score was found to be a significant predictor (B= .442, SE = 0.210, $p< .05$, OR= 1.555). Children with higher LeBLEQ-T Reading Scores were more likely to achieve Below Expected scores on the KS1 Reading SAT. For every 1 unit increase in the LeBLEQ-T Reading Score (which indicates poorer Reading skills) the possibility of scoring Below Expected on the KS1 Reading SAT increased by 1.555. Consequently, although there was only a a significant trend, if a teacher rated a child as possessing poorer reading skills, they were more likely to receive a worse score on their KS1 Reading SAT.

The next Multinomial Logistic Regression aimed to predict KS2 Writing score using the Total LeBLEQ-T Score and the LeBLEQ-T Writing Score as predictors. The addition of the predictors to a model that contained only the intercept was found to significantly improve the fit between model and data (χ^2 (4, N= 48) = 17.584, Nagelkerke R^2= .461, $p< .005$). As shown in Table 21, significant contributions were made from both the Total LeBLEQ-T Score and the LeBLEQ-T Writing Score.

Table 21

Contributions of the Total LeBLEQ-T Score and the LeBLEQ-T Reading Score to a child's KS2 Reading score as identified in a regression analysis

Predictor	χ^2	Df	*p*
Total LeBLEQ-T Score	8.335	2	.015*
LeBLEQ-T Reading Score	6.404	2	.041*

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

Comparison of coefficients between children achieving an Expected result and those gaining a Below Expected result, demonstrated Total LeBLEQ-T Score to be only near significant (B= .093, SE= 0.51, $p= .06$). Children with higher Total LeBLEQ-T Scores were more likely to achieve a Below Expected score on the

KS2 Writing SAT. For every 1 unit increase in the LeBLEQ-T Total Score (indicative of poorer overall language-related skills) the possibility of them scoring Below Expected on the KS2 Writing SATs increased by 1.098.

Finally, the regression that aimed to predict KS2 Reading Scores using the Total LeBLEQ-T Score and the LeBLEQ-T Reading Score as predictors was shown to be non-significant.

2.2.1.2.3 Discussion

The LeBLEQ-T was developed with the aim to provide researchers with a tool that would allow for quick, teacher-reported assessment of a child's classroom-based skills. After extensive, initial piloting of a longer questionnaire, the final measure was created that allowed gathering information on a child's listening, speaking, reading, writing and interpersonal skills in less than three minutes.

In accordance with predictions, there were no significant differences observed in the LeBLEQ-T scores of the Year 3 children in comparison to the Year 6 children. This reinforces the idea that the LeBLEQ-T is not measuring skill level, but rather skills in comparison to expectation.

Results showed an association between scores on subsets of the LeBLEQ-T and an objective measure of language, namely the Streamlined-CELF. Not only were there high correlations between all subsets of the LeBLEQ-T and the three skill components of the Streamlined-CELF (i.e., Receptive, Expressive and Core Language Scores), a child's LeBLEQ-T Writing Score was shown to be a significant predictor of core language, receptive language and expressive language ability. This could be attributed to the multifaceted, overlapping nature of language. As such, a child's understanding of language (listening and reading) is correlated to their production of language (speaking and writing) (Mashburn et al., 2009) as written language skills have also been shown to be dependent on oral language skills (Berninger et al., 2002).

For example, the writing subset of the LeBLEQ-T contained questions that were similar to questions that were contained in other subsets. For instance, the question "Does the child struggle to express their thoughts?" was contained within the Speaking sub-section, while the question "Does the child struggle to write down their thoughts?" was contained within the Writing sub-section. Though the two questions tap into different skills, there is an overlap as they both assess the ability to tap into one's thoughts and communicate them in some way (orally or through writing). Consequently, if a child has difficulties expressing their thoughts verbally, they may also have issue expressing them orthographically. Orthographic expression of thoughts, although still relying on the grammar and vocabulary, is complicated by both punctuation and spelling. Whilst spoken language does not rely on the knowledge of how sound maps to form, orthographical knowledge does. Furthermore, orthographical expression cannot be supported by factors such as body language or vocal emphasis which can aid verbal expression. In the written word, everything must be in the words themselves, and they must be interpretable.

A child's LeBLEQ-T Speaking Score was shown to significantly predict the proportion of exposure to English a child receives in the home environment. Hearing less English in the home environment may lead to reduced speaking abilities as receptive ability is known to precede expressive ability in children. It may be that the type of bilingual children included in the current study do not have impeded listening skills due to receiving a sufficient amount of input in English from school, but their ability to reply is still hindered due to less experience of hearing that language in their daily lives. This is associated with the finding that bilinguals tend to experience more 'tip-of-the-tongue states,' reporting to know items contained in picture naming trials, but experiencing difficulties in trying to access the word from memory and express it (Gollan et al., 2007; Gollan & Acenas, 2004; Ivanova & Costa, 2008). This difference between the receptive and expressive abilities of children is known in the literature as the Receptive-Expressive Gap (discussed in detail in the General introduction) where increased exposure to another language is known to proportionately increase the gap between listening and speaking abilities (Gibson et al., 2014).

Interestingly, no significant differences were found between the monolingual and bilingual children in terms of their LeBLEQ-T scores. This corroborates the finding that there were no statistical differences between the SAT scores of monolingual and bilingual children. The similarity in results on SAT scores and the LeBLEQ-T could suggest that the two are measuring similar concepts. Both the LeBLEQ-T and SAT measure skills linked with classroom activities and, thus, it is possible that if there were no differences between scores obtained on an objective measure (the SAT), there would also be no differences on a questionnaire-based measure. Thus, the present findings demonstrate that, to some extent, the LeBLEQ-T scores can predict performance on the SATs, supporting the link between higher teacher-rated performance on class-based skills and higher attainment on standardised tests of academic attainment.

In conclusion, these findings support the views of Rimfeld et al. (2019) that teachers can be utilised as reliable sources of the skill levels of primary-school-aged children. Results also validate the LeBLEQ-T, which records teachers' perceptions of a child's language-based skills in the classroom, demonstrating it to be a useful novel tool that can be used as a sufficient predictor of linguistic ability of school-aged children. Overall, the LeBLEQ is a particularly valuable tool for use in classroom contexts where time spent assessing children may be limited and parental involvement low.

3 Norming Experiments for Ambiguous Words and Idiomatic Expressions

The present chapter aims to add to the norms available for both ambiguous words (homonyms, metaphors and metonyms) and idiomatic expressions. First, an experimental survey will be described whereby 33 British-English monolingual adults rated 83 idiomatic expressions in terms of their decomposition (Experiment 1). Participants were asked to indicate along a Likert scale (1-7) whether the individual words contained within an idiom contributed to its overall figurative meaning. Secondly, Age of Acquisition (AoA) norms for 231 ambiguous words and 121 idiomatic expressions were obtained (Experiment 2). These ratings were acquired from 100 British-English monolingual adults and were vital for the selection of stimuli in the four main Experimental studies presented in Chapters 4, 5, 6, and 7 (Experiments 3, 4, 5 and 6).

3.1 Decomposition Norms for Idiomatic Expressions

Literature suggests idioms vary along a wide range of dimensions such as familiarity, predictability, degree of ambiguity, literal plausibility, semantic opacity, grammatical well-formedness, and syntactic flexibility (Libben & Titone, 2008). Most notable to the current research, however, is the topic of idiom decomposition, or transparency. According to the Idiom Decomposition Hypothesis proposed by Gibbs and colleagues (Gibbs & Gonzales, 1985; Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989; Gibbs, Nayak, Bolton & Keppel, 1989; Gibbs, 1992; Gibbs, Bogdanovich, & Sykes; 1997) idioms can be either decomposable or non-decomposable.

In the case of decomposable idioms, the meaning of the components in the expression (i.e., the individual words) clearly contribute to the overall meaning of the idiom itself. A well-known example of a decomposable idiom is '*to pop the question'* where the unit's meaning can be derived from the meaning of '*pop'* to imply '*ask'* and '*question'* to imply a specific type of question; a proposal. Conversely, a non-decomposable idiom is an expression where the meaning of the idiom's components is not clearly related to the overall meaning of the idiomatic phrase; for example '*to kick the bucket'* has no relevance to its meaning of '*to die*.'

However, decomposition is not necessarily a fixed or absolute feature and idioms can, in fact, be regarded along a continuum of decomposability. Over a series of studies, Gibbs and colleagues concluded that the degree of decomposability could predict syntactic and lexical behaviour (Gibbs & Gonzales, 1985) as well as processing (Gibbs et al., 1989a) with idioms classed as more decomposable (e.g., ‘to pop the question’) being both easier to understand and quicker to process than less decomposable idioms (i.e., non-decomposable idioms such as ‘to kick the bucket’) (Caillies & Butcher, 2007; Gibbs, 1987). This processing advantage for more decomposable idioms has, likewise, been found in developmental studies investigating the impact of idiom type on children’s comprehension (Gibbs, 1991; Nippold & Taylor 1995, 2002; Nippold & Rudzinski, 1993) with research demonstrating that children also tend to acquire decomposable idioms earlier. For instance, Caillies and Le Sourn-Bissaoui (2008) found that children as young as 5 years-old can understand decomposable idioms in context but it is not until they are 7 years old when they are able to understand non-decomposable idioms.

However, there has been some disagreement to the decomposition hypothesis. Tabossi et al. (2008), for example, attempted to replicate the results of Gibbs et al. (1991), finding no support for the claims for decompositional idioms being more easily processed and, as a result, the decomposition hypothesis. More recently, there has been additional research to suggest a processing disadvantage for decomposable idioms when items are presented in sentence context (Cieslicka, 2013; Zhang, Yang, Gu, & Ji, 2013). However, although the conditions in which decomposable idioms may show a processing advantage are not very clear, it is evident that the degree of decomposition does indeed play a role in processing. Thus, when selecting idiomatic stimuli in linguistic research, it seems imperative to be able to quantify an idioms’ degree of decomposition to control for differences between stimuli.

Recent research by Koleva, Mon-Williams and Klepousniotou (2019) attempted to do so. Koleva et al. (2019) quantified the decomposition of 60 British-English idioms in a British-English sample of eight participants. Participants were instructed to indicate on a Likert scale (1-7) the degree to which they believed the literal meaning of the individual words in an idiom contributed to the overall figurative meaning of the phrase. However, only a small number of idioms were included in the online ratings survey (30 decomposable, 30 non-decomposable), which was tested on a small population sample. Thus, there is scope to improve these ratings both by including more idioms and by administering them to a bigger population sample.

3.1.1 The Current Study (Experiment 1)

The present study aimed to gather decomposability ratings for idiomatic expressions and assist in promoting consistency in the research of how British-English children process idiomatic expressions. Such research employed methods similar to those of Koleva, Mon-Williams and Klepousniotou (2019). Making data of this type widely available is particularly useful for researchers who must otherwise collect their decomposability data on small scales, prior to experimental testing, to determine appropriate stimuli. In addition to these broader applications, this study also aims to generate information that is vital in the selection of stimuli for the experiments in Chapter 7 (Experiment 6).

3.1.2 Methods

3.1.2.1 Participants

Thirty-three monolingual native British-English speakers between the ages of 18 and 60 years (8 males, 25 females) completed the Decomposability survey. All participants were born and resided in the UK all their lives, had no known hearing problems, language-related difficulties, or history of brain injury. All participants had normal or corrected to normal vision and were recruited as volunteers via word of mouth. All participants were given informed consent and were told participation in the study was anonymous. The study received ethical approval from the School of Psychology, University of Leeds Ethics Committee (PSC-586, approved 23/01/19).

3.1.2.2 Stimuli

In total, 83 idioms were rated. Of these 83, 20 were taken from Koleva et al.'s previous work investigating idiom familiarity in order to allow for later cross-validation analyses. An internet search found a further 63 common English idioms, pooled from various sources of idiomatic research and educational information for non-native speakers. No proverbs were included in this list. All idioms can be found in Appendix A.

3.1.2.3 Procedure

Decomposition ratings were collected using an online survey (Google Forms https://docs.google.com/forms/). Participants were informed they would be presented with a series of idioms for which they were required to make a single judgement; to decide to what degree the individual words contained within the idiom contributed to its overall figurative meaning. This was indicated on a 7-point Likert scale, as previously employed by Koleva et al. (2019). A '*1'* on the Likert scale indicated that the individual meanings of the words did not contribute at all to the overall meaning of the idiom. A '*7'* indicated the meaning of the individual words *were* considered to be related to the idiom's overall meaning. Each idiom was presented to the participants in isolation, without context, and the survey took approximately 15 minutes to complete.

3.1.3 Results

Respondents used the 7-point Likert scale in its entirety. Following Koleva et al. (2019), items with an average decomposability rating of below 4 were classed as non-decomposable, while ratings with an average of 4 and above were classed as decomposable. This resulted in the stimuli list consisting of 57 non-decomposable idioms, and 26 decomposable idioms.

The average decompositional value of the decomposable idioms was 2.80 (SD= .74), while the average decompositional value of the non-decomposable idioms was 4.59 (SD= .49). An independent t-test revealed the difference in decomposability ratings between the two decomposition-classifications to be statistically significant ($t(70.52) = -13.12$, $p < .001$).

To determine cross-sectional validity of the results, the 20 decomposability ratings that were sampled from Koleva et al. (2019) were compared to the ratings gathered in the current study. A repeated measures t-test revealed there were no statistical differences between the decomposition scores obtained in the current ratings and those generated in Koleva et al. (2019) for 17 of the items. Only three items (15%) were shown to have statistically different scores. The ratings obtained in the two studies were, thus, deemed to have good concurrent reliability.

3.1.4 Discussion

The current study aimed to provide British-English decomposability scores of 63 idiomatic expressions and thus increase the pool of available norms for studies investigating idioms. An idiom's decomposability has been shown to affect processing (e.g., Gibbs et al., 1989a; Tabossi et al., 2008) and is, thus, an important factor to be considered when drawing conclusions from research into idiomatic comprehension.

As expected, the current decomposability ratings demonstrated good congruence with the ratings gathered by Koleva et al. (2019) and, additionally, provided 63 novel decomposability ratings. These ratings, totalling 83, will be informative when designing studies aimed at investigating the processes involved in the comprehension or production of idiomatic expressions, including research outlined in Chapter 7 (Experiment 6).

3.2 Age of Acquisition norms for Ambiguous Words and Idioms

Ambiguity is considered rife within the English language. Research has shown a significant 44% of words in the English language are considered to be semantically ambiguous (Britton, 1978), with figurative utterances constituting approximately 36% of a child's and 25% of an adult's exposure to language (Lazar, Warr-Leeper, Nicholson, & Johnson, 1989; Van Lancker-Sidtis & Rallon, 2004). It is, consequently, of no surprise that disambiguation skills begin to develop relatively early in life with knowledge that a single word can have multiple distinct meanings being seen in children of 3-4 years old (e.g., Doherty, 2000), and the understanding of some idiomatic phrases being seen in children as young as 5-years-old (Caillies & Le Sourn-Bissaoui, 2008).

The inability to acquire such skills can result in children being detrimentally impacted in terms of their spoken and written sentence comprehension (e.g., MacGregor et al., 2019; Simon et al., 2010), reading comprehension (Fusté-Herrmann, 2008), educational achievement (Cain et al., 2005), social participation (Laval, 2003; Swineford et al., 2014) and ability to establish intimacy (Gerrig & Gibbs, 1988). As a result, it seems critical to understand the processes behind disambiguating words and phrases with multiple meanings and identifying the instances where problems with disambiguation may occur. A pivotal factor in pursuing research into ambiguity and idiomatic expressions, however, lies in selecting appropriate items to serve as stimuli, especially when testing child populations.

Psycholinguistic research, including studies of ambiguous words and phrases, typically involves careful stimuli selection. Some of the most frequently controlled variables when selecting linguistic stimuli include concreteness (the degree to which something can be experienced by our senses), dominance (the degree of control one feels to have on a word), imageability (the degree of effort involved in generating a mental image of something) and familiarity (a measure of a word's objective experience), to name but a few. Additionally, idioms are often controlled for in terms of dimensions including predictability, degree of ambiguity, literal plausibility, semantic opacity, grammatical well-formedness, and syntactic flexibility (Libben & Titone, 2008). Moreover, numerous studies have attested to the importance of considering a word's Age of Acquisition (AoA) when selecting stimuli for research (Brysbaert & Ghyselinck, 2006; Ghyslinck et al.; Johnston & Barry, 2006; Juhasz, 2005).

Age of Acquisition (AoA) refers to the age in which a word or phrase was first learnt and has been demonstrated to be one of the most significant contributors to language and memory processes (Carroll & White, 1973; Morrison et al., 1992). Some research has even suggested that the age in which words are acquired can explain over 5% of the variance in lexical decision times of English (Brysbaert & Cortese, 2011). Though there have been various norming lists established in the last few decades regarding the AoA of non-ambiguous words (Cortese & Khanna, 2008; Khanna & Cortese, 2011; Kuperman et al., 2012;) and English proverbs (e.g., Benjafield et al., 1993) (which are considered to be similar to idiomatic expressions), there has been comparatively little research on the AoA of the numerous meanings/senses of ambiguous words, or idiomatic expressions where sentences can be interpreted either literally or figuratively.

Gilhooly and Logie (1980) was the most notable study to gather ratings for more than one meaning of ambiguous words on a wide scale. However, due to the study being conducted four decades ago, understandably, some of the words have become outdated (e.g., "*topic*" to mean *the discontinued chocolate bar*), with many homonymous words which are now considered quite common not included on the stimuli list (e.g., "*tweet*" to mean *a post made on Twitter*). In addition, Gilhooly and Logie (1980) did not stipulate the variation of English the participants needed to speak in order to be eligible for the study. This was similarly true for the more recent research conducted by Khanna and Cortese (2011). While Khanna and Cortese (2011) collected norms for an impressive 1,208 ambiguous words, a large proportion of these words were homophones (i.e., words that sound the same but do not have the same orthographic representation; e.g., *bass* to mean either *musical instrument* or *fish*). Furthermore, these ratings were gathered in America. With the various differences between British-English and other dialectal variations of English, such as American-, Canadian-, and Australian-English (Armstrong et al., 2015; Trudgill & Hannah, 2002), it would be reasonable to assume that AoA norms collected from other variations of English may not be wholly appropriate for British-English speakers and that this lenient inclusion criteria could have confounded the results gathered.

The same can be said for idiomatic expressions. Despite such norms being gathered in Bulgarian (Nordmann & Jambazova, 2016), Chinese (Li et al., 2016), French (Caillies, 2009; Bonin et al., 2011; Bonin et al., 2013; Bonin, Meot, Boucheix & Bugaiska, 2017), German (Citron et al., 2016), Italian (Cacciari & Corradini, 2015; Tabossi et al., 2011), and Polish (Imbir, 2016), there have been no attempts to gather such norms for British-English idioms.

For this reason, researchers do not always collect AoA norms when working with linguistic stimuli, rather choosing to employ measures of word frequency (e.g., Schmitterer & Schroeder, 2019). This is due to the links between AoA and word frequency with the tendency for more frequent words to be acquired earlier (Zevin & Seidenberg, 2002, 2004). Consequently, literature demonstrate both highly frequency words and words learnt at a younger age being processed more efficiently than those that are heard less often or learnt later in life (Brysbaert & Cortese, 2010; Brysbaert & Ellis, 2016; Brysbaert et al., 1995). However, these generated frequencies often fail to take into consideration the multiple senses of ambiguous words. An example of this would be the Children Printed Word Database (Masterson et al., 2010), a comprehensive database of the vocabulary in reading materials used by British-English 5-9-year-old children. Although this list contains multiple ambiguous words (e.g., "*monkey*"), the specific meanings of these items remain unclear, and thus the associated frequencies become unreliable for use in research focussing on words with multiple meanings.

Maciejewski and Klepousniotou (2016) recently provided researchers with the first meaning-specific frequency ratings of homonymous words in British-English. Based on the American-English eDom norms of Armstrong and colleagues (2012), 100 native British-English speakers living throughout the UK rated 100 homonymous words, providing frequency ratings for the two most frequent meanings of each word. However, despite the aforementioned link in the literature showing the potential for frequency to be used as a proxy for AoA for non-ambiguous words (Brysbaert & Ellis, 2016), there has yet to be any research demonstrating a relationship between meaning-specific AoA norms and meaning-specific frequency norms for homonymous words (such as those gathered by Maciejewski and Klepousniotou, 2016).

Consequently, researchers aiming to avoid using non-meaning-specific frequency as a proxy for the AoA of ambiguous words and phrases must typically perform their own, small-scale norming studies on candidate variables (e.g., MacGregor et al., 2015; Rodd et al., 2010). This is increasingly important when working with child populations. Before testing a child's knowledge of a word, researchers must first ensure that the word is likely to be in the vocabulary of the age range tested. For instance, there would be little benefit of testing children's understanding of words that are known to be generally acquired much later in childhood. However, this is a relatively time-consuming process, with the number of items being rated limited to the number of items participants are willing to rate in a single session. Conversely, the lack of recognised AoA norms for British-English ambiguities has led researchers to use the generic AoA norms (e.g., Jager & Cleland, 2016). Due to the nature of ambiguous words however, using ratings of this kind can lead to conclusions of questionable reliability, since researchers cannot be sure which meaning of the ambiguous word the participants were referring to. Similar practices can be seen in literature focussing on idioms, where the paucity of research considering the AoA norms for British-English idiomatic phrases has led researchers to either use AoA ratings for the individual words contained in the idioms (e.g., Pesciarelli et al., 2014) or to ignore the impact AoA has on units of speech entirely (e.g., Cain et al., 2015).

3.2.1 The Current Study (Experiment 2)

The current study aims to improve this situation by obtaining AoA norms for the two meanings/senses of homonymous, metaphorical and metonymic words in addition to British-English idiomatic expressions. Such knowledge can assist in promoting consistency in the research of how British-English children process ambiguous words and phrases. The study also aims to generate norming information that is vital in the selection of stimuli for experiments in Chapters 4, 5, 6, and 7.

Though using either frequency or non-meaning specific AoA has become common practice in research with ambiguous words (homonyms, metaphors and metonyms), such measures do not take the multiple meanings/senses of ambiguous words into consideration. Consequently, further aims of the current research were to determine whether the meaning-specific AoA ratings of the ambiguous words correlated with either the non-meaning specific AoA ratings of Kuperman et al. (2012) or the non-meaning-specific frequency ratings of the Children's' Printed Word Database (Masterson et al., 2010). Specifically, there was expected to be no correlation between the meaning specific AoA ratings gathered in the current study and either of these measures. The Children's Printed Word Database was selected as it relates to the frequencies individuals would have been exposed to as children which would have impacted their development and knowledge of the words (the use of this was favoured over using norms of the most frequently heard words in adulthood as this would not have impacted the acquisition of the words in question).

Furthermore, meaning-specific AoA ratings of the homonyms gathered in the current study were also compared to the meaning-specific frequency ratings gathered by Maciejewski and Klepousniotou (2016). It was hypothesised that there would be significant correlations between the ratings that related to the individual meanings of the homonymous words. Furthermore, due to the nature of balanced homonyms, and their relatively equal frequencies, the AoAs of the two meanings of balanced homonyms were expected to be relatively similar, while the meanings of unbalanced homonyms were likely to be slightly more divergent with the second meaning acquired later than the first.

Finally, in relation to the idiomatic items contained within the current research, both decompositional and non-decompositional idioms were selected as part of the sample in order to gather a wide range of stimuli. It was predicted that decompositional idioms would be acquired earlier than non-decompositional idioms, in line with past research from other languages (Caillies & Butcher, 2007; Caillies & Le Sourn-Bissaoui, 2008). Furthermore, idioms containing more words were expected to be acquired later due to the increased cognitive load needed for longer sentences (Fanari et al., 2010).

In line with previous research (e.g., Gilhooly & Logie, 1980; Morrison & Ellis, 1995, 2000), the norms were obtained by asking adult participants to evaluate at which age they learnt a particular meaning/sense of an ambiguous word or idiomatic phrase in a rating survey using the same Likert scale. This has been demonstrated as a reliable (Rubin, 1980) and robust indicator of AoA, correlating with naming speed (Morrison et al., 1992) and naming accuracy (Morrison & Ellis, 1995).

3.2.2 Methods

3.2.2.1 Participants

One hundred monolingual native British-English speakers between the ages of 18 and 60 years (19 males, 81 females) completed the AoA rating survey. This uneven gender split was not considered to be an issue with research showing extremely high correlations ($r = .960$) between males and females regarding self-rated AoA (Gilhooly & Hay, 1977). All participants were born and resided in the UK all their lives, had no known hearing problems, language-related difficulties, or history of brain injury. All participants had normal or corrected to normal vision and were recruited as volunteers via word of mouth. All participants gave informed consent and were told participation in the study was anonymous. The study received ethical approval from the School of Psychology, University of Leeds Ethics Committee (PSC-586, approved 23/01/19).

3.2.2.2 Stimuli

3.2.2.2.1 Ambiguous Words

All ambiguous words were pooled from previous studies investigating homonymic, metaphoric and/or metonymic ambiguity (Klepousniotou, Pike, Steinhauer & Gracco, 2012; Maciejewski & Klepousniotou, 2016). All words had two meanings/senses with the same spelling and pronunciation, as outlined by the Wordsmyth dictionary (Parks et al., 1998). Words were excluded from the list if they only made sense in their plural form or could not be placed in the sentence-final position. The stimuli selection criteria excluded ambiguous words that were considered highly infrequent (e.g., "*upset*" to mean *capsize a boat*), outdated (e.g., "*topic*" to mean *the discontinued chocolate bar*), or inappropriate for research with child populations (due to the main aims of the thesis; e.g., "*gin*" to mean *the alcoholic beverage*). Words were also excluded if either of the meanings/senses were deemed to be highly culturally specific (i.e., the word '*register'* to mean *cash till* is generally exclusively used in American-English). This allowed for a list that could be used with all English-speaking populations. This resulted in a total of 231 ambiguous words. These 231 items included 164 homonymous words (55 balanced, 109 unbalanced and a combined 328 meanings), 37 metaphors (74 senses) and 30 metonyms (60 senses). All stimuli can be found in Appendix B.

Each word was presented both in isolation and in the context of a sentence that highlighted its intended meaning/sense. All sentences that placed the ambiguous word in context contained 4-14 words (M= 9.31, SD= .68). This was 4-14 words (M= 9.06, SD= 1.99) for homonyms, 6-14 words (M= 10.08, SD= 2.10) for metaphors and 6-13 words (M= 8.79, SD= 2.47) for metonyms. Ambiguous words were always presented in the sentence-final position.

3.2.2.2.2 Idiomatic Expressions

In total, 121 idioms were rated. Of these 121, 60 of the figurative expressions were taken from previous work investigating idiom familiarity (Koleva et al., 2019). An internet search found a further 61 common English idioms, pooled from various sources (e.g., www.theidioms.com). No proverbs were included in this list. Of these 121 figurative expressions, 38 were decompositional while the remaining 83 were non-decompositional, as determined by the decomposition rating survey described in the previous study (see Experiment 1, Section 3.1).

Each idiom was presented both in context and in isolation and participants were asked to approximate the age at which they learnt the target idiom. All sentences that placed the idiom in context contained 7-20 words (M= 12.87, SD= 2.99) words. The high variation was due to the variability of the number of words in the idiom itself; the idiom occurred in the sentence-final position in all cases. All stimuli can be found in Appendix B.

3.2.2.2.3 Stimuli Lists

Due to the large number of stimuli, the 583 items (as each ambiguous word has two meanings/senses) were divided amongst four surveys: A, B, C, and D. To do so, firstly, the ambiguous words were pseudo-randomly divided into two lists. The first list contained 115 items (28 balanced homonyms, 53 unbalanced homonyms, 19 metaphors and 15 metonyms) while the second list contained 116 items (28 balanced homonyms, 55 unbalanced homonyms, 18 metaphors and 15 metonyms).

From these two lists, four surveys were created. In terms of the ambiguous words, Surveys A and B contained the same words in opposing meanings/senses. For example, participants completing Survey-A might be asked the age in which they learnt the word "*bank*" in the meaning of *the financial institution*, while those completing Survey-B were asked about *the land alongside a river* meaning. Likewise, Survey-C and Survey-D contained the same words but, again, in opposing meanings/senses. In each of the four surveys, half the words referred to the first meaning/sense and the other half referred to the second meaning/sense. The number of words in first- and second- meaning/sense sentences were statistically the same, as were the number of words across all four surveys (all p values > .05). The 121 idioms were divided into four lists with each list containing nine decompositional idioms and 21 non-decompositional idioms (with an extra decompositional item included in Survey-A). Each participant completed only one of the lists resulting in 25 participants completing each survey and, thus, 25 ratings being obtained for each item.

To test for interrater reliability, 10% of the items (36) occurred across all four lists. This included 12 homonyms, 6 metaphors, 6 metonyms and 12 idioms. This resulted in participants rating a total of 171 items in Survey-A and -B (90 homonyms (32 balanced and 58 unbalanced), 23 metaphors, 19 metonyms, and 39 idioms), and 172 items in Surveys-C and -D (92 homonyms (32 balanced and 60 unbalanced), 22 metaphors, 19 metonyms, and 39 idioms).

3.2.2.3 Procedure

The AoA ratings for the ambiguous words and phrases were collected using an online survey (Google Forms; https://docs.google.com/forms/). Each participant was randomly assigned to complete one of the four surveys, and the stimuli were presented in a pseudo-randomised order. This resulted in 25 participants completing each survey (A, B, C or D).

At the onset of the survey, participants were informed they would be estimating the ages they learnt both ambiguous words and figurative expressions which were presented in separate subsections. They were also informed of the scale they would be using. Following Gilhooly and Logie (1980), participants were asked to estimate the age they learnt the specific meaning/sense of a word or phrase in years, using a 7-point Likert scale that used age ranges (0-2 years, 2-4 years, 4-6 years, 6-8 years, 8-10 years, 10-12 years, 13+ years) where “1” signified the lower ages. The scale was always visible to the participant. Additionally, there was an option to state whether a particular phrase or meaning/sense was still unknown.

Participants were given two further sets of shorter instructions; one that preceded the rating of the ambiguous words, and one that preceded the idiomatic phrases. For the ambiguous words, participants were asked to rate words which had more than one meaning/sense. They were instructed to only consider the meaning/sense of the word in the context given. For instance, if the word was 'ring' and the context was 'He got down on one knee and proposed with a ring', participants were asked to state the age they learnt the word for the small circular band worn on the finger (not the sound a telephone may make when someone is calling you). Similar instructions were given at the beginning of the idiomatic expressions, highlighting to the participants that the next items would ask the age they learnt the figurative phrases and reminding participants of the scale.

Half of the participants were exposed to the ambiguous words first while the other half were exposed to the idiomatic phrases first. The items within both the ambiguous words and the idiomatic phrases subsets were presented in a randomised order. This resulted in an administration time of approximately 70 minutes. Participants were asked to minimise distractions and complete the survey in one single session.

3.2.3 Results

AoA ratings corresponding to each meaning/sense of all homonyms, metaphors, metonyms and idioms shall be discussed separately below. Respondents used the Likert scale in its entirety in the case of all four tropes.

3.2.3.1 Homonyms

When considering only the homonyms for which participants knew both their meanings, the average age for learning the homonymous words was 6 years, 11 months (SD= 2 years, 4 months). Balanced homonyms, on average were reported to have been learnt first (M= 6 years, 8 months, SD= 2 years, 4 months), with unbalanced homonyms being acquired later (M= 7 years, 0 months, SD= 2 years, 5 months). An independent t-test revealed this difference to be non-significant. However, meanings of unbalanced homonyms (M= 2.62, SD= 5.27) compared to balanced homonyms (M= 1.04, SD= 2.66) were statistically more likely to be reported as unknown ($t(326)= 3.63$, $p< .005$).

The data also revealed several cases in which participants were unfamiliar with at least one of the meanings of homonymous words. These null responses (i.e., unknown meanings) were observed across 85 participants (85.00%) and 115 meanings (35.06%).

When considering unbalanced homonyms, the first (i.e., the more frequent) meaning was found to be acquired significantly earlier (M= 5 years, 11 months, SD= 2 years, 4 months) than the second meaning of the word (M= 8 years, 2 months, SD= 2 years, 0 months) ($t(215)= -7.52$, $p< .001$). There was also a significant difference in the number of unknowns reported between the first (M= .42, SD= 1.43) and the second (M= 4.83, SD= 6.63) meanings of unbalanced homonyms ($t(118)= -6.77$, $p< .001$); there were more unknown second meanings than first meanings.

Unsurprisingly, due to the relatively similar frequencies, the two meanings of balanced homonyms were shown to be acquired at statistically similar ages (Meaning 1: M= 6 years, 7 months, SD= 2 years, 4 months; Meaning 2: M= 6 years, 8 months, SD= 2 years, 3 months). Furthermore, there was no difference in the number of unknown meanings for balanced homonyms ($M_{Meaning1}$= .82, $SD_{Meaning1}$= 2.15; $M_{Meaning2}$= 1.25, $SD_{Meaning2}$= 3.09).

When the acquisition of *both* meanings of a homonym was considered, most items were reported to have been acquired before the age of 10 years (80.48%). The breakdown of the average age in which both meanings of a homonymous word were acquired can be seen in Table 22.

Table 22

Count of the average age in which both senses of homonymous words were reported to be acquired

	Homonyms learnt at each age	Cumulative Homonyms learnt
0-2 years	0 (0%)	0 (0%)
2-4 years	2 *(1.22%)*	2 (1.22%)
4-6 years	25 (15.24%)	27 (16.46%)
6-8 years	42 (25.61%)	69 (42.07%)
8-10 years	63 (38.41%)	132 (80.48%)
10-12 years	31 (18.90%)	163 (99.38%)
13 years +	1 (0.61%)	164 (100%)

Unsurprisingly, there was also a statistically significant difference between the ages in which participants reported to have learnt both meanings of unbalanced homonyms (M= 8 years, 4 months, SD= 1 year, 11 months) and both meanings of balanced homonyms (M= 7 years, 8 months, SD= 2 years, 0 months), with the former being acquired later ($t(161)$= 2.04, p< .05).

Results of a Pearson correlation indicated that there was a significant negative association between the ages in which the participants in the current study stated they learnt the items, and the meaning-frequency norms generated by Maciejewski & Klepousniotou (2016), (r(178) = -.55, p< .001). This demonstrates that the more frequent a meaning of a homonym, the earlier it is learnt. In contrast, as predicted, no linear relationship was found between the AoAs reported in the current study and the Children's' Printed Word Database (Masterson, Stuart, Dixon, & Lovejoy, 2010), an AoA normed list that failed to take into consideration the distinct meanings of homonymous words.

A final set of Pearson correlations revealed a significant positive association between the AoA norms obtained in the current study when the two meanings were averaged together and the non-meaning-specific AoA norms reported by Kuperman et al. (2012) (*r*(162)= .62, *p*< .001). This relationship remained significant when the homonyms were stratified into their first (r(162)= .68, *p*< .001) and second (r(162)= .26, *p*< .01) meanings. However, as expected, when comparing Kuperman et al.'s (2012) AoA ratings to those generated in the current study, a series of a Paired Samples t-tests revealed significant differences between the non-meaning-specific AoA norms generated by Kuperman et al. (2012) (M= 7.14, SD= 2.08) and Meaning 1 (M= 6.21, SD= 2.44; *t*(161)= -6.43, *p*< .001) or Meaning 2 (M= 7.71, SD= 2.30; *t*(161)= -2.69, *p*< .01) of the AoA generated by the present study. If the two senses were averaged together (M= 6.95, SD= 1.83), then the AoAs were statistically similar to Kuperman et al.'s (2012) (*t*(161)= -1.36, *p*= .175).These findings demonstrate that previous AoA ratings that did not distinguish between the different meanings/sense of ambiguous words reflected cumulative scores which do not represent accurately either meaning/sense of an ambiguous word.

3.2.3.2 Metaphors

When considering only the metaphorical words for which participants knew both their senses, the average age for learning them was 5 years, 6 months (SD= 2 years, 10 months). Furthermore, the first (i.e., the more literal) sense of the metaphor was found to be acquired significantly earlier (M= 3 years, 9 months, SD= 2 years, 5 months), than the second (i.e., more figurative) sense of the word (M= 7 years, 3 months, SD= 2 years, 0 months; (*t*(72) = -6.65, *p*< .001). When the acquisition of both senses of a metaphor was considered, most items were reported to have been acquired before the age of 10 years (94.59%). The breakdown of the average age in which both senses of a metaphorical word were acquired can be seen in Table 23.

Table 23

Count of the average age in which both senses of metaphoric words were reported to be acquired

	Metaphors learnt at each age	Cumulative Metaphors learnt
0-2 years	0 *(0%)*	0 *(0%)*
2-4 years	2 *(5.41%)*	2 *(5.41%)*
4-6 years	7 *(18.92%)*	9 *(24.32%)*
6-8 years	12 *(32.43%)*	21 *(56.76%)*
8-10 years	14 *(37.84%)*	35 *(94.59%)*
10-12 years	2 *(5.41%)*	37 *(100%)*

The data also revealed several cases in which participants were unfamiliar with one of the senses of the metaphorical words. These 'unknown' responses were apparent across 24 participants (24.00%) and 12 senses (32.43%) (M= .40, SD= 1.14). There was a significant difference in the number of unknowns reported between Sense 1 (M= .00, SD= .00) and Sense 2 (M= .80, SD= 1.52) of the metaphors ($t(36.00) = -3.21$, $p < .005$).

Results of a correlational analysis revealed a non-significant, negative association between the age in which both senses of a metaphorical word were learnt and the frequency of the word (assessed using the Children's' Printed Word Database; Masterson et al., 2010) ($r(57) = -.264$). This implies that increasing the number of encounters of a word does not necessarily mean it will be learnt earlier in life. This was also true when the first sense of the word ($r(31) = -.249$, $p = .169$)

and the second sense of the word (*r*(31)= -.137, *p*= .459) were entered in the correlational analysis separately.

A final set of correlations indicated that there was a significant positive association between the AoA norms obtained in the current study when the two senses were averaged together and the non-sense-specific AoA norms reported by Kuperman et al. (2012) (*r*(36)= .55, *p*< .001). Moreover, this relationship remained to be significant when the metaphoric words were stratified into their first (r(36)= .871, *p*< .001) and second (r(36)= .497, *p*= .01) senses.

Though this correlation existed, there were also significant differences between Kuperman et al., (2012) stated AoA and those generated in the current study. A series of a Paired Samples t-tests revealed significant differences between the (non-meaning-specific) AoA norms generated by Kuperman et al. (2012) (M= 5.34, SD= 2.17) and Sense 1 (M= 3.76, SD= 2.47; *t*(36)= 7.95, *p*< .001) and Sense2 (M= 7.27, SD= 2.06; *t*(36)= -5.52, *p*< .001). However, if the two senses were averaged, the AoAs were statistically similar to those specified by Kuperman et al. (2012) (M= 5.51, SD= 1.94; *t*(36)= -.844, *p*= .404), providing further support to our concerns that previous AoA ratings reflected cumulative scores that might not accurately represent either sense/meaning of an ambiguous word.

3.2.3.3 Metonyms

When considering only the metonyms for which participants knew both their senses, the average AoA was 4 years, 6 months (SD= 2 years, 0 months). There was no statistically significant difference (*t*(29)=-.30, *p*= .769) between the age in which children acquired the first sense (M= 4 years, 5 months, SD= 1 year, 11 months) and second sense (M= 4 years, 6 months, SD= 1 year, 2 months) of the metonymic words, demonstrating that the senses were acquired at a similar age. When the acquisition of both senses of a metonym was considered, all were reported to have been acquired before the age of 10 years (100%). The breakdown of the average age in which both senses of a metonym were acquired can be seen in Table 24.

Table 24

Breakdown of the average age in which both senses of metonymic words were reported to be acquired, (items (%))

	Metonyms learnt at each age	Cumulative Metonyms learnt
0-2 years	1 *(3.33%)*	1 *(3.33%)*
2-4 years	11 *(36.67%)*	12 *(40%)*
4-6 years	11 *(36.67%)*	23 *(76.67%)*
6-8 years	3 *(10.00%)*	26 *(86.67%)*
8-10 years	4 *(13.33%)*	30 *(100%)*
10-12 years	0 *(0%)*	-

The average age in which both senses of the metonymic word were acquired was 4 years, 10 months (SD= 2 years, 1 month). The data also revealed some cases in which participants were unfamiliar with one of the senses of the metonymic words. These 'Unknown' responses were observed across 9 participants (9.00%) and 8 senses (26.67%), (M= .15, SD= .48).

Additionally, results of a correlational analysis revealed a significant negative association between the age in which both senses of the metonymic word were reported to have been acquired and the frequency of a word (assessed using the Children's' Printed Word Database; Masterson, Stuart, Dixon, & Lovejoy, 2010) ($r(25) = -.515$, $p < .01$). This effect held true if the frequency scores were compared to both the first sense ($r(25) = .-.530$, $p = .005$) or the second sense ($r(25) = -.574$, $p < .001$) of the metonymic word, suggesting that metonymic words behave like non-ambiguous words - the more they are heard, the earlier they are acquired, irrespective of the specific sense.

A final set of a correlations indicated a significant positive association between the AoA of the metonyms reported in the current study and the AoA of the word in general (i.e., regardless of its sense-specific interpretation) using the Kuperman et al.'s (2012) AoA norms (r(29)= .678, p< .001). This relationship remained significant when only the first sense (r(29)= .700, p< .001) or the second sense (*r*(29)= .683, *p*= .001) of the metonymic word was entered into the correlational analysis.

Despite this correlation, however, three Paired Samples t-tests revealed significant differences between the non-sense-specific AoA norms generated by Kuperman et al. (2012) and the first sense of the metonymic word (t(29)= 4.18, p< .001) and the second sense of the metonymic word (*t*(29)= 3.47, *p*< .005). A final t-test revealed a significant difference between the non-sense-specific AoA ratings reported by Kuperman et al. (2012) and ratings generated in the current study when the two senses of the metonymic words were averaged together (*t*(29)= 4.14, *p*< .001). Means and standard deviations of the Kuperman et al. (2012) norms compared to the norms generated in the current study can be seen in Table 25. This provided support to concerns that previous AoA ratings reflected cumulative scores that might not accurately represent either sense/meaning of an ambiguous word.

Table 25

Means and standard deviations of the Kuperman et al. (2012) norms compared to the meaning-specific norms of polysemic words, stratified by polysemy type (i.e., metaphor, metonymy) and sense, and compared to homonyms (M (SD) in years).

	Kuperman et al (2012)	Meaning/Sense-specific		
	AoA of general sense	AoA of Sense1	AoA of Sense2	Sense1 and Sense2 averaged
Metaphors	5.34 (2.17)	3.76 (2.47)	7.27 (2.06)	5.51 (1.94)
Metonyms	5.56 (1.68)	4.47 (1.96)	4.53 (2.21)	4.50 (2.00)
Polysemes	5.44 (1.96)	4.07 (2.27)	6.04 (2.51)	5.06 (2.01)
Homonyms	7.14 (2.08)	6.21 (2.44)	7.71 (2.30	6.95 (1.83)

3.2.3.4 Idioms

When considering only the idioms participants knew, 80.17% of idiomatic expressions were learnt before the age of 10 years old. The breakdown of the average age idioms were reported to have been acquired can be seen in Table 26.

Table 26

Breakdown of the average age in which idioms were reported to be acquired, (items (%))

	Idioms learnt at each age	Cumulative Idioms learnt
0-2 years	0 *(0%)*	0 *(0%)*
2-4 years	0 *(0%)*	0 *(0%)*
4-6 years	5 (4.13%)	5 (4.13%)
6-8 years	29 (23.97%)	34 (28.10%)
8-10 years	63 (52.07%)	97 (80.17%)
10-12 years	24 (19.83%)	121 (100%)

An independent t-test revealed that the age of acquisition of decompositional idioms (M= 8 years, 6 months, SD= 1 year, 4 months) and non-decompositional idioms (M= 8 years, 8 months, SD= 1 year, 7 months) was statistically similar (t(90.16)=-.58, p=.560), suggesting that decompositional and non-decompositional items are learnt concurrently.

A simple linear regression was also conducted to predict AoA based on the number of words in the idiom. The length of the idiom was not correlated with the age in which the idiom was acquired [F(1, 119)= 2.03, p= .16, R^2 = .017], suggesting that the length of the idiom does not affect the age at which the idiomatic expression is acquired.

However, the data revealed several cases in which participants were unfamiliar with the idiomatic expressions. These null responses were apparent across 68 participants (68%) and 43 items (35%).

3.2.3.5 Interrater Reliability

Thirty-six items (12 homonyms, 6 metaphors, 6 metonyms and 12 idioms) were included across all four surveys to determine the agreement across raters. One of the six metaphors included across the surveys was used in an incorrect context and was thus removed from all subsequent analyses. Consequently, 35 separate one-way ANOVAs were conducted with results showing participants to agree on the age in which the items were acquired across all four surveys in the case of 32 of the items (i.e., all ANOVAs were non-significant with $p > .05$). All three of the items that that participants disagreed on were idioms. These items on were (1) '*to mince one's words'* [$F(3, 88) = 3.71$, $p < .05$]; (2) '*to go downhill'* [$F(3, 99) = 2.99$, $p < .05$], and (3) '*to shake a leg*'' [$F(3, 89) = .388$, $p < .05$]. However, as these three items constituted a small proportion of the overlapping items, with agreement across the remining 91.43%, the rating of the items were deemed to demonstrate good inter-rater reliability.

3.2.4 Discussion

The present study aimed to provide British-English norms for two meanings/senses of 352 ambiguous words and phrases (164 homonyms, 37 metaphors, 30 metonyms and 121 idioms). The ability to understand ambiguous words and phrases is important when comprehending both spoken word and written texts (e.g., Fusté-Herrmann, 2008; MacGregor et al., 2019; Simon et al., 2010) and, consequently, in educational achievement (Cain et al., 2005). Literature has also shown idiomatic knowledge to be vital for participating in social interactions (Laval, 2003; Swineford et al., 2014) and establishing intimacy (Gerrig & Gibbs, 1988). Thus, increasing the pool of available norms for homonyms, polysemes (both metaphors and metonyms) and idiomatic phrases will allow researchers to select better stimuli when investigating ambiguity, especially when working with child populations.

The ratings were provided by a large and diverse group of participants with a high level of agreement across all four lists. It can, thus, be concluded that the AoA norms generated within the current study are reliable indicators of the age in which British-English speakers learn a specific meaning/sense of an ambiguous word. Such norms can be confidently used in further research to aid in stimuli selection. Ambiguous words and idiomatic phrases shall be discussed separately below.

3.2.4.1 Ambiguous Words

In accordance with predictions, participants used the Likert scale in its entirety, demonstrating early knowledge of at least one of the meanings/senses of homonyms, metaphors and metonyms from as early as 2 years old.

In the case of homonyms, metaphors and metonyms, meaning/sense-specific AoA norms obtained in the current study were positively correlated with the widely used Kuperman et al. (2012) non-meaning-specific AoA norms. However, there were also significant differences between these two sets of AoA norms with the Kuperman et al. (2012) norms being significantly higher than the present AoA of the first meaning/sense and significantly lower than the second meaning/sense of the word. It is possible that this was due to the different methodologies employed in the two studies. In the work by Kuperman and colleagues (2012), participants were instructed to state the specific age they understood a word if somebody had used it in front of them. This method did not involve a Likert scale but rather instructed participants to enter their response. The lack of prompts could, thus, have made estimating the age in which a word was acquired more difficult for the raters.

However, it is increasingly likely that the differences observed between the AoA norms obtained in this study and those reported in Kuperman et al. (2012) are due to the ambiguity of the words that were rated. As Kuperman et al. (2012) did not provide participants with context, there is no indication of whether the participants did indeed consider the ambiguity of the word or not. While some participants may have considered the first meaning/sense that came to mind, others may have considered more than one meaning/sense and averaged these. Finally, some participants may even have not have thought of a specific meaning, but rather stated the age they acquired a vague meaning of the word, leading to inaccurate responses. However, as different meanings/senses of ambiguous words can have different AoAs (as the current study has shown and will discuss further below), this would have impacted the ratings the participants in Kuperman et al. (2012) provided. This emphasises the differences between the non-meaning/sense specific AoA norms generated from Kuperman et al. (2012) and the more accurate meaning/sense-specific AoA norms obtained in the present study. The present findings demonstrate clearly that generic AoA norms should not be used in the place of more precise meaning/sense-specific norms when investigating ambiguous words.

Additionally, the AoA of both the metaphors and the homonyms, on the whole, did not correlate with the frequency ratings of the Children Printed Word Database (CPWD; Masterson et al., 2010), an AoA normed list that fails to take into consideration the distinct meanings/senses of ambiguous words. Such frequency scores did correlate with the meaning-specific AoA ratings of the metonyms, however. Such differences between the tropes are attributed to the relatedness of the different meanings/senses of ambiguous words. Research suggests that the meanings of homonyms are separately stored while metaphors may be in a transition from pure polysemy to homonymy and thus have senses that are loosely related and resemble the mental representation of homonyms. On the other hand, metonyms only have a single semantic representation for their multiple closely related senses (MacGregor et al., 2015; Klepousniotou, 2002; Klepousniotou & Baum, 2007). Consequently, if individuals hold only a singular interpretation of a metonym, it is expected that there would be no differences between the ages in which they state they acquired the two senses of a metonym.

It must also be noted that the meaning-specific AoA norms for homonyms generated in the current study did correlate with the meaning-specific frequency ratings for homonyms by Maciejewski and Klepousniotou (2016). These findings were more in line with literature demonstrating links between a non-ambiguous word’s frequency and AoA (e.g., Brysbaert, 2010; Morrison & Ellis, 1995). Such research demonstrates that, in some cases, frequency can indeed be used as a proxy for AoA (or vice versa) but it is imperative that the specific meaning/sense of a word is first considered. Furthermore, it must be noted that such a relationship can only be investigated in homonyms due to the lack of meaning-specific frequency ratings for polysemes. Consequently, more research needs to be conducted into other tropes (metaphors and metonyms) to determine the degree to which there is a relationship between meaning-specific AoA and meaning-specific frequency ratings in ambiguous words other than homonyms.

Consequently, though using non-meaning-specific frequency ratings (such as CPWD) as a proxy for AoA (or vice versa) might be applicable for words with only one meaning (e.g., Zevin & Seidenberg, 2002), the multiple, different meanings/senses of ambiguous words mean that such substitution should be considered poor practice when selecting stimuli for psycholinguistic research investigating ambiguity in any form. Instead, meaning-specific frequency ratings or meaning-specific AoA norms should be employed.

In accordance with predictions, the different meanings of balanced homonyms were acquired at similar ages, as were both senses of the metonymic words. When considering the unbalanced homonyms and the metaphors, however, the second, less frequent meanings were acquired later than the first meanings with more reports of the second meaning still being unknown in adulthood. In the case of the metonyms, this can, again be attributed to these words having only one, rich semantic representation (MacGregor et al., 2015; Klepousniotou, 2002; Klepousniotou & Baum, 2007), with the different senses not competing in the same manner as in metaphors or homonyms. This leads to little conscious distinction between the two senses of a metonymic word and, thus, the ratings of the age of acquisition being statistically the same. Effects of the other tropes are all due to frequency of exposure to these words. Balanced homonyms, by definition, have similar frequencies and they are acquired at a similar age. However, unbalanced homonyms have a dominant (more frequent) and a subordinate (less frequent) meaning, while metaphors have a literal sense and figurative sense. Research has shown that in order to understand the figurative sense, the literal sense must first be acquired/understood (Klepousniotou, Titone & Romero, 2008). It, thus, follows that the first meaning/sense of unbalanced homonyms and metaphors is learnt before the second meaning/sense can be acquired.

3.2.4.2 Idiomatic Expressions

Contrary to predictions, results demonstrated that the decomposition of an idiom did not affect the age of acquisition of the idiomatic expression. This is in contrast to research by Caillies and Le Sourn-Bissaoui (2008) who found that children as young as 5 years-old can understand decomposable idioms in context but it is not until they are 7 years old that they are able to understand non-decomposable idioms. Although unexpected, the present findings support the Direct Access Model of idiom processing (Gibbs 1980, 1985) which states that the figurative meaning of an idiomatic phrase is accessed directly from the mental lexicon, without the need to process the literal meaning of the expression first. In fact, literal meanings may fail to be analysed at all, especially if the expression is highly familiar and immediately recognized as an idiom; the idioms are processed as 'giant lexical units' (Nippold, 1998, p106). So, an idiom's decomposability may not be the most influential factor affecting its age of acquisition, Instead, other factors such as the frequency and context in which an idiom is encountered as well as the degree of explicit instruction of their meaning may have a greater impact on age of acquisition of idioms.

Such a view might also explain why the number of words in the idiomatic expressions was not shown to impact their reported AoA. As such, each word within an idiom does not have to be literally processed in order to make sense of the phrase, but rather the unit is learnt as a whole. For instance, in the idiom "to pass the buck," one does not need to know the meaning of the word "buck" to understand the meaning of the idiom to be "to pass the blame".

3.2.4.3 Conclusions

The current research is the first to gather meaning/sense-specific AoA norms for British-English homonymous and polysemic (both metaphors and metonyms) words alongside AoA norms for British-English idiomatic phrases. This advances knowledge of the most appropriate method of selecting stimuli for psycholinguistics studies. However, it is important to note that, although the ambiguous words and phrases included in the rating survey were considered to be prominent in the English language, there were still cases that were not recognised by the participants in the sample. This emphasises the importance of gathering AoA norms before testing as presumed ubiquity may not necessarily imply understanding. This research also emphasises that, in the case of ambiguous words, using non-meaning/sense specific AoA and frequency ratings to select stimuli is problematic and can lead to misleading results and interpretations.

In conclusion, AoA meaning/sense-specific norms for 231 ambiguous words and 121 idiomatic phrases were generated in the current study providing detailed information about the impact of AoA on the acquisition of the different meanings/senses of ambiguous words and phrases. Data of this type are particularly useful for researchers who work with ambiguous words and idiomatic phrases in British-English populations, and child populations in particular, who must otherwise collect AoA data on small-scale studies, prior to experimental testing, in order to select appropriate stimuli. These ratings are crucial when designing studies aimed at investigating the processes involved in the comprehension or production of ambiguity and will be used in the research outlined in Chapters 4, 5, 6 and 7.

4 The Effect of Language Experience on the Comprehension of Homonymy in Primary-School aged Children

Homonyms are a class of ambiguous words which have more than one meaning. Homonymy is perhaps the most salient form of ambiguity whilst being both incidental and relatively infrequent (Weinreich, 1966) which makes homonyms distinct from other types of ambiguity. Often confused with heterophonic homographs (words which have identical spelling but disparate pronunciation, e.g., bow/bow) and homophonic heterographs (words which have identical pronunciation but disparate spelling, e.g., alter/altar), homonyms have both identical spelling and pronunciation, but unrelated meanings (Eddington & Tokowicz, 2015). An example of a homonym is the word “bank”; it is considered to be simply by chance that its two meanings share the same orthographic and phonological form (i.e., both *the financial institute* and *the grassy knoll by a river*).

The ability to disambiguate homonyms is a skill that has been reported to develop from the age of 4 years and continue through primary school (Backscheider & Gelman, 1995; Doherty, 2000; Garnham et al., 2000; Smith Cairns et al., 2004; Storkel & Maekawa, 2005; Wankoff & Cairns, 2009). Such research typically involves children being presented with a homonymous word and having to identify different interpretations from an array of pictures. Detection of such ambiguity is considered rare before this age (Kidd & Holler, 2009) with the idea that the high-level linguistic knowledge of children before the age of 4 years old has not yet developed sufficiently in order for them to complete such a task (Doherty, 2004). However, literature has shown that if children are presented with an ambiguous word in a sentential context (e.g., “They could see the bank in the distance”), and asked to provide the two possible meanings, only children over the age of approximately 6 years could accurately detect the lexical ambiguity (Smith Cairns et al., 2004). The difference between these two tasks is that the latter is explicitly metalinguistic; it requires a child to focus on the dual meaning of the homonym and to reanalyse the sentence, a skill pre-school children are not yet proficient at (Trueswell et al., 1999).

While an ambiguity advantage was briefly mentioned in the General Introduction (Chapter 1), it must be noted that there are differential effects for polysemes (discussed in Chapters 5 and 6) and homonyms. Namely, a body of eye-tracking studies investigating word comprehension demonstrated that, in the absence of contextual cues, individuals showed slower reading times of homonymous words in comparison to unambiguous words. Such an effect is known as the ambiguity disadvantage and is seen consistently across the literature for homonyms (Duffy et al., 1988; Yasushi Hino et al., 2002; Jager & Cleland, 2016; Pexman et al., 2004; Piercey & Joordens, 2000; Rayner & Duffy, 1986). However, this is often dependent on the frequency of such words. For example, Duffy et al. (1988) conducted a study where they presented participants with either balanced homonyms (e.g., “fan”) or unbalanced homonyms (e.g., “bank”) in sentential context. When unbalanced homonyms were preceded by a neutral context, fixations were similar to unambiguous words. When these same words were preceded by a sentence that biased the reader towards the subordinate meaning of the word, however, longer gaze durations were observed than to controls. Such findings suggest that the more dominant meaning of a homonym is activated at an early stage of processing, and, when the context supports this meaning, the subordinate meaning of the word does not compete for selection. However, a disambiguating context allows for the subordinate meaning to become activated and compete for selection with the dominant meaning. This effect has been consistently observed across many studies (Duffy et al., 2001; Kambe et al., 2001; Pexman et al., 2004; Piercey & Joordens, 2000; Sereno et al., 2006), and is often referred to as the subordinate bias effect.

Such slowing of reaction times to homonyms in reading tasks has been interpreted to mean that the unrelated meanings of homonyms have separate entries in the lexicon, leading to competition between the meanings on retrieval (e.g., Beretta et al., 2005). This competition is affected by the type of the homonym (balanced/unbalanced). The relative frequency of a word meaning has continuously been shown to determine both the time-course and level of activation (see Twilley & Dixon, 2000 for a review), with the recognition of homonyms varying depending on the relative frequencies of the multiple meanings (Armstrong et al., 2012). As such, a slight slowing in lexical decisions was found for balanced but not unbalanced homonyms. For unbalanced homonyms (e.g., “bank”), the activation process is so strongly biased towards the more salient, higher frequency meaning, that the less frequent meaning rarely gets a chance to compete for activation. In contrast, for balanced homonyms (e.g., “fan”) each meaning is activated in parallel, resulting in competition for further activation (Duffy et al., 1988; Kawamoto, 1993; Maciejewski, 2018; Rayner & Duffy, 1986).

The ability to resolve competition between the multiple meanings of a homonymous word is a prominent characteristic of skilled reading (Gernsbacher, 1991, 1997b, 1997a; Gernsbacher, Robertson, et al., 2001; Gernsbacher, St John, et al., 2001; Jenkins et al., 1984; Nation, 2017; Suggate et al., 2018). Such effects are witnessed even in the early years of schooling (Hoover & Tunmer, 2018; Wankoff, 1983), with the ability to report the dual meaning of ambiguous words being predictive of future reading skill even when assessing children who have yet to learn to read. For instance, research has shown that the ability to detect homonyms in Year 1 not only highly correlates with reading readiness measures (Wankoff & Cairns, 2009) but is also a strong predictor of reading ability in Year 2 and Year 3 (Smith Cairns et al., 2004). Research into the comprehension of homonyms is thus important in the understanding of reading, with the ability to use homonymous competence as an identifier of children who are at risk for difficulties with reading.

The majority of the research conducted with homonymy resolution, however, has focussed on comparative studies between those with some form of language impairment (Bishop, 2000), Alzheimer's (Balota & Faust, 2001; Faust et al., 1997), ASD (Eberhardt & Nadig, 2018; Henderson et al., 2011) or Williams Syndrome (Hsu, 2013; Rossen et al., 1996), with research on bi/multilingual populations, especially with children, being limited. Furthermore, in cases where bi/multilingualism has been considered, much of the literature has either focussed on the sensitivities to cross-language form overlap (e.g., de Bruijn et al., 2001; Dijkstra et al., 2000; Dijkstra & van Hell, 2003; Fontes & Schwartz, 2015; Schwartz et al., 2007) or was tested solely on bilingual populations with no monolingual comparison group to see how bilingualism impacts homonymous understanding (e.g., Fontes & Schwartz, 2011; 2015).

Recently Kousaie et al. (2015) conducted a study with both monolingual and bilingual adults using both behavioural (semantic relatedness) and electrophysiological (N400 focussed) measures. Participants read a sentence that was biased towards either the more frequent (dominant) or less frequent (subordinate) meaning of an unbalanced homonym. For example, participants read “The doctor asked her to step onto the *scale*” and made a relatedness decision on whether the target word (always placed in the sentence-final position) was related to the contextually appropriate meaning (e.g., “balance”), the contextually inappropriate meaning (e.g., “skin”), or a word that was unrelated to either meaning (e.g., “shoe”).

Kousaie et al., (2015) found no behavioural differences between the two groups. There were, however, electrophysiological differences. Monolinguals activated both meanings of homonyms to a greater degree (i.e., smaller N400 in response to target sentences related to both meanings) than unrelated targets across the entire N400 window. In contrast, bilinguals only showed this effect early and late in the time window and demonstrated less activation for the subordinate meanings of homonyms relative to dominant meanings during the peak of the N400. Taken together, these results suggest that monolinguals rely more on context than their bilingual peers. Kousaie and colleagues (2015) concluded that, though there are subtle differences in the processing of homonyms in monolingual and bilingual populations, there is no difference in the outcome of such processing. Importantly, the research by Kousaie et al. (2015) adds to a body of literature (e.g., Elston-Gúttler & Friederici, 2005; Love et al., 2003) that has found potential underlying processing differences between monolinguals and bilinguals in the absence of behavioural differences.

It is possible, however, that behavioural differences were not observed between the two groups due to the varied nature of bilingualism. Links between homonymous understanding and language experience have only ever considered the latter as a binary variable (monolingual/bilingual). However, it is becoming ever more apparent that bilingualism may be better suited measured along a continuum (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020). Participants in Kousaie et al. (2015) reported themselves to be highly proficient (English-French) bilinguals, all learning English before the age of six years old. Consequently, though a bilingual’s electrophysiological processes reflected their bilingual status, their behaviour may have been more akin to that of a monolingual due to similar levels of exposure to English by the time they reached adulthood. A binary measure of language experience may thus not be sensitive enough to capture the subtle behavioural differences between the two groups. The lack of literature investigating the links between homonymous understanding and degree of language exposure (i.e., *how* bilingual an individual is) is surprising considering homonymous processing requires knowledge to be retrieved from vocabulary stores (Duffy et al., 1988; Simpson, 1981; Swinney, 1979), with bilinguals possessing smaller vocabularies than their monolingual peers (e.g., Bialystok, 2001). Furthermore, as these vocabulary sizes are proportional to the level of exposure received (De Houwer, 2009; De Houwer, 2007; Hoff et al., 2012; Thordardottir, 2011), it would be expected that knowledge of homonymic words (a type of vocabulary knowledge) would also be proportional to exposure.

Research into the impact of language experience on homonymous understanding is largely underexplored in child populations. Jang et al. (2014) and Kwak et al. (2015) recently investigated effects of homonymic processing on primary-school aged multicultural children, finding a lack of understanding of homonyms in multicultural populations even if the children showed normal vocabulary development. However, the focus of these studies was on “multicultural children” with the level of exposure to another language not being explicitly recorded or studied. This leaves a gap in the literature with regard to understanding bilingual children’s behaviour towards homonyms in comparison to their monolingual peers.

Such gaps in the literature are likely, in part, due to the lack of standardised assessment tools in distribution that can capture the ability of children from various linguistic backgrounds, across a wide age range, to process both balanced and unbalanced homonymy. The present chapter is a step towards this.

4.1 The Current Study (Experiment 3)

The research presented in this chapter aimed to investigate the comprehension of balanced and unbalanced homonyms in both monolingual and bilingual, primary-school-aged children and the factors that contribute to better understanding. Given the lack of standardised measures, a novel homonymous task was developed whereby children were required to respond to homonyms in two ways. Firstly, they were asked to verbally explain a homonymous word in the Open-Ended component of the task, then they were asked to select the correct meanings of the homonyms from a series of options. Due to previous research which showed that children find explaining figurative language (e.g., metaphors and idioms) more difficult than answering multiple-choice questions on them (e.g., Arcara et al., 2019; Papagno & Caporali, 2007; Perlini et al., 2018; Pouscoulous, 2011, 2014), both response methods were employed in order to allow for rich linguistic data collection and reduce the likelihood of ceiling effects. Using both methods attested in the literature with the same materials and participants would also allow direct comparisons between the methods, leading to better understanding of homonym processing and comprehension by monolingual and bilingual primary-school children. It was hypothesised that children would be more likely to know both meanings of balanced homonyms than both meanings of unbalanced homonyms, where they would be more likely to report the more frequent (dominant) meaning.

As the national curriculum for England suggests children should demonstrate an understanding of figurative language from Year 3, with the development of more sophisticated understanding apparent by Year 6 (Department for Education, 2013a), the current homonymy study was tested in these two age groups with older children expected to perform better. However, on the basis of the aforementioned research suggesting the gap between low- and high-SES groups is eliminated by KS2 (Hutchinson et al., 2016), it was also predicted that there would be differences within age groups. Namely, it was hypothesised that lower SES children would be outperformed by higher SES children in Year 3 but such differences would be attenuated in the Year 6 group. Whether or not the child was in receipt of Pupil Premium was used as a proxy for SES grouping. Additionally, it was hypothesised that a child's Cumulative Language Input (CLI) score, would be a more robust predictor of homonymic ability, due to bilingualism being better suited to measures along a continuum, rather than a binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020).

Finally, associations with scores of academic attainment (Reading and Writing SATs) and teacher's perceptions of linguistic competence (LeBLEQ-T) were also investigated alongside scores of language development (as assessed by the Streamlined-CELF discussed in Chapter 2). Of particular interest was the ability of a child's Receptive Language Score (RLS) to predict homonymous awareness due to the links between receptive language and a number of other tropes of ambiguity (e.g., metaphors and idioms; Hessel & Murphy, 2019; Johnson & Rosano, 1993; Segal & Gollan, 2018; Smith & Murphy, 2015; Trosbrog, 1985).

4.2 Methods

The same children participated in all four main experimental studies (homonyms, metaphors, metonyms and idioms) so that their knowledge/performance *across* the tropes could be compared. The children completed these tasks within the timeframe of a single week with tasks split across three sessions of approximately 40 minutes each. The order in which the tasks were administered was counterbalanced. The four main experimental tasks were never administered consecutively; they were separated by the other language measures (the Streamlined-CELF and LeBLEQ). This was to limit the likelihood of children's responses to each experimental study being impacted by their understanding of a previously assessed trope. Details of participants and materials of all four experimental studies are described further below. Detailed information of each experimental task and procedure is included in each respective chapter.

4.2.1 Participants

Three schools were recruited, and data were gathered on 127 children (63 males, 64 females). Sixty-five of these children were from Year 3 (35 Female, M= 7 years, 10 months, SD= 4 months), while 62 were from Year 6 (29 Female, M= 10 years, 9 months, SD= 4 months). No children had any known hearing problems, developmental delays or special educational needs. Participants were stratified into monolingual (n= 47) and bilingual (n= 80) groups. Children with any exposure to a second language outside of the school environment were classed as bilingual, while those who had no exposure to an additional language were classified as monolinguals (determined using the LeBLEQ-C).

Table 27 summarises the distributions of age, gender and language experience.

Table 27

Distribution of age and gender of the sample (years; months)

		Gender (n)	Mean	Min	Max	StdDev
Year 3	Bilingual	Female (25)	7;8	7;3	8;2	0;3
		Male (14)	7;10	7;2	8;3	0;4
		Total (39)	7;9	7;2	8;3	0;3
	Monolingual	Female (10)	8;0	7;4	8;6	0;4
		Male (16)	7;11	7;3	8;5	0;3
		Total (26)	7;11	7;3	8;6	0;4
Year 6	Bilingual	Female (17)	10;9	10;3	11;5	0;4
		Male (24)	10;7	10;2	11;1	0;4
		Total (41)	10;8	10;2	11;5	0;4
	Monolingual	Female (12)	10;10	10;2	11;4	0;4
		Male (9)	10;10	10;5	11;1	0;3
		Total (21)	10;10	10;2	11;4	0;3

All children received schooling in the English language, and all bilingual children were exposed to another language in their home environment to varying extents. There was a total of eight home languages stated in the sample, as outlined in Table 28. In addition to this, there were seven children who were unsure of the name of the language they spoke.

Table 28

Additional languages spoken in the home within the sample and number of children who reported speaking that language.

Language	Number of children
Punjabi	34
Urdu	28
Polish	3
Slovak	3
Pashto	2
Bengali	1
Chinese	1
Latvian	1

Finally, although the majority of the children were born in the UK (105), 22 were born elsewhere: Pakistan (9), Poland (3), Slovakia (3), Afghanistan (1), Italy (1), Latvia (1), Nigeria (1) and Spain (1). Two children could not remember the name of the country in which they were born. Of the children not born in the UK, the average aged moved to the UK was 3 years, 0 months (SD= 1 year, 9 months) for the Year 3 children, and 4 years, 5 months (SD= 2 years, 4 months) for the Year 6 children.

Schools from similar Socio-economic backgrounds were approached to participate in the research. Socio-economic status (SES) was determined using the Income Deprivation Affecting Children Index (IDACI; Department for Communities and Local Government, 2015), based on the school's postcode. The IDACI score reflects the percentage of 0-15-year-old children living in income-deprived households in a specific postcode area. All three schools had low SES indices; the ranks, deciles and scores of the three schools can be seen in Table 29.

Table 29

IDACI ranks, deciles and scores of the three schools sampled

	IDACI Rank	IDACI Decile	IDACI Score
School_1	8685	3	0.221
School_2	5191	2	0.283
School_3	7173	3	0.245

IDACI Rank = 32,844 neighbourhoods in England are ranked from the most deprived (a score of 1) to the least deprived (a score of 32,844). *IDACI Decile* = Ranks are divided into ten equal groups where 1 indicates the most deprived and 10 indicates the least deprived. *IDACI Score* = Percentage of children in that area classed as deprived

Schools interested in taking part in the research were met with and given specific information regarding the aims of the study. Once headteachers had consented, information sheets and consent forms were distributed to all parents/guardians of children in Year 3 and Year 6 via classroom teachers. Parents were given the option to withdraw from the study by returning an opt-out consent form. If the opt-out form was not returned, consent was assumed. The study received approval from the Ethics Committee of the School of Psychology, University of Leeds (ethics reference number: PSC-313, approved 23/01/2019).

4.2.2 Materials

The experimental task had two components to test participants' knowledge of homonymous words. Firstly, participants were presented with a word and asked to define it. They were instructed to respond verbally to the experimenter; the experimenter then typed their answer on the computer. Secondly, participants saw the word alongside four definitions and had to select the correct definition/s. Sometimes only one of the response options would be the correct definition of the target word; other times two of the response options would be correct definitions. Children were instructed to respond using the mouse pad of the laptop.

Both homonymous (i.e., experimental) and non-homonymous (i.e., control) words were selected from Armstrong et al. (2012) for inclusion into the current study.

The experimental words were selected if they had two or more definitions that had the same orthographic and phonological form but unrelated meanings, based on definitions supported by past literature (e.g., Beretta et al., 2005; Klepousniotou et al., 2012). They had also been used in previous literature investigating homonymic processing and understanding (e.g., Klepousniotou, Pike, Steinhauer & Gracco, 2012; Maciejewski & Klepousniotou, 2016). Control words had only a single definition. Only meanings that were part of British-English vocabulary and could be placed in the sentence-final position were included. Words were excluded if they were compound words, or were not easily definable concepts. Furthermore, words were excluded if they were deemed to be outdated (e.g., "*topic*" to mean *the discontinued chocolate bar*), or age-inappropriate. While these decisions were subjective (made by the researcher), they adhered to certain parameters, namely, words were considered inappropriate for child research if they were related to sex/sexual acts, drugs/alcohol or profanities (e.g., words such as "*gin*" or "*sex*"). The resultant list comprised of 24 homonyms (i.e., words with more than one meaning) and 24 control words (i.e., words with a singular meaning). Of the homonymous words, 21 were noun-noun homonyms, while the remaining three were noun-verb homonyms. Furthermore, 12 of the homonymous words were balanced while the remaining 12 were unbalanced.

Due to the large number of stimuli pooled, the words were split into two lists, List-A and List-B. The use of these two lists resulted in tasks identical in terms of format and instructions, while the target items differed. This allowed for a large number of homonyms to be tested but for administration time to remain relatively short. This resulted in 24 items in each list; 12 homonyms and 12 control words. Finally, to allow for inter-rated reliability, and generalisation across lists, each list contained a 10% cross over. Consequently, three items (one control word, one balanced homonym and one unbalanced homonym) were taken from List-A and added to List-B. Likewise, three List-B words were added to List-A. This resulted in participants being exposed to 27 trials; 13 controls, seven balanced homonyms and seven unbalanced homonyms. Target words were pooled and distributed amongst lists in order to be matched on a number of parameters that are known to influence processing.

Firstly, there were no statistically significant differences between the experimental items and the control items, items of List-A and List-B, or Balanced and Unbalanced homonyms in terms of the number of letters, imageability, concreteness and familiarity. The latter three were determined using the MRC Psycholinguistic Database (Coltheart, 1981). There were also no differences in frequency; this was determined using the Kučera and Francis (1967) frequency scores, in addition to the frequency scores from the Children's Printed Word Database scores (CPWD; Masterson et al., 2010). Whilst the Kučera and Francis (1967) frequency norms were established more than 50 years ago, they remain among the most consistently used *written* frequency lists (Brysbaert & New, 2009) and are still widely used in language research to this day (e.g., Boscolo Nale, 2020; Borghesani et al., 2020; Guan, & Fraundorf, 2020).

Finally, there were no significant differences in AoA (Kuperman et al., 2012) (all p values > 0.05 and can be seen in Table 30).

Table 30

Means and standard deviations (M (SD)) for the characteristics of items across Balanced, Unbalanced and control trials (stratified into control and experimental targets).

	Balanced	Unbalanced	Control
Number of letters	4.25 (1.14)	4.33 (1.30)	4.50 (0.72)
Imaginability	544.70 (45.14)	556.82 (45.86)	573.30 (53.61)
Concreteness	541.00 (58.42)	558.09 (46.04)	585.90 (54.31)
Familiarity	530.91 (36.46)	519.73 (48.30)	456.27 (83.37)
Frequency - Kucera & Francis	30.67 (33.61)	43.42 (35.17)	15.80 (28.58)
Frequency - CPWD	60.64 (43.86)	99.50 (146.63)	38.36 (41.15)
Kuperman's AoA (years)	6.55 (1.55)	6.51 (1.84)	6.56 (1.78)
Meaning-specific AoA (years)	5.99 (1.08)	6.33 (1.92)	-

Homonyms with a range of AoAs (3-11 years, as determined by Kuperman et al. 2012) were included, reducing the likelihood of floor- or ceiling effects occurring. Furthermore, the AoA ratings generated in Chapter 3 of this thesis were considered when creating the lists of words. *Both meanings* of the homonymous word were acquired across a wide range of ages (3-11 years) but additionally, no words were included where more than one participants reported to not know one of the meanings. There was no significant difference between the balanced and unbalanced homonyms in terms of the meaning-specific AoA ratings, or between the items included in List-A and List-B (all p values > 0.05). All homonymic stimuli can be seen in Appendix C.

4.2.2.1 Multiple-Choice Response Options

Not only were there two lists of homonymous words, there were also two different methods of response in the multiple-choice component of the task. In line with the aims of the current study, one response type involved participants being presented with four pictures and being asked to select those that depicted the target word. The other response method used four written definitions instead of pictures. These two versions of the task will henceforth be referred to as the Pictorial form of the task and the Orthographical form respectively and shall be discussed separately below.

4.2.2.1.1 Orthographic Response Options

Written definitions were simultaneously presented auditorily to account for differences in participants' reading ability. Four Orthographic Response Options were created for each target word. For experimental trials, two of the four response options were definitions of the target word, with one response option defining the first meaning of the homonymous word and the other response option defining the second meaning. Conversely, for the control words, only one of the response options was a definition, the remaining response options were foils (i.e., incorrect definitions).

All definitions were selected from the Wordsmyth online dictionary, available online at wordsmyth.net (Parks, Ray, & Bland, 1998). It was ensured that all foils, for both experimental and control trials, defined real concepts and were, thus, feasible. Consequently, the identification of the correct definitions of the target word would imply the participant's precise understanding of a word, rather than vague knowledge.

For the experimental trials, each foil created was designed to be a foil for one of the two correct response options. For example, in the case of the homonymous word "*cricket*", the two correct response options were:

(1) a noisy, green insect.

(2) a game played with wickets.

The two foils were:

(1) a tiny, black insect.

(2) a game played with a hoop.

Similarly, all three foils in the control trials were created to be related to the single correct response option. For example, in the case of the control word "*kilt*," the correct response option was *a tartan skirt worn by men* and the three foils were:

(1) an instrument played by Scottish people.

(2) shoes you wear to dance.

(3) a type of dog.

Finally, response options were matched on a number of the same parameters applied to the target words themselves. Namely, the number of words in each response option, the average number of letters of each word contained in the sentence and the average Kuperman's AoA for each word were shown to be statistically similar between (1) the correct and foil response options, (2) the experimental and control trials, (3) List-A and List-B responses and (4) balanced and unbalanced trials (all p values > .05). Means and standard deviations can be seen in Table 31. All homonymic stimuli, including response options, can be seen in Appendix C.

Table 31

Means and standard deviations (M (SD)) for the characteristics of items across Balanced, Unbalanced and Control trials (correct and foil responses).

	Balanced		Unbalanced		Control	
	Correct	Foil	Correct	Foil	Correct	Foil
Words	5.42 (1.10)	4.79 (1.23)	5.04 (0.75)	5.00 (0.90)	5.75 (1.26)	5.51 (0.80)
Letters	4.35 (2.28)	4.26 (2.16)	4.06 (2.24)	4.15 (2.21)	4.23 (2.26)	3.99 (2.12)
Kuperman's AoA (years)	4.66 (1.47)	4.37 (1.10)	4.66 (1.64)	4.52 (1.31)	4.49 (1.63)	4.47 (1.42)

4.2.2.1.2 Pictorial Response Options

To ensure the pictorial and orthographic responses were comparable, foils of each trial were the same between the two versions. As with the orthographic response options, four Pictorial Response Options were created for each target word. For experimental trials, two of the four response options depicted the target word, with one response option depicting the first meaning of the homonymous word and the other response option depicting the second meaning. Conversely, for the control trials, only one of the response options depicted the target word, the other three were distractors. See Figure 8 for a comparison between the Pictorial- and Orthographic-versions of the trial '*pen'*.

The illustrations for the Pictorial Response Options were created specifically for the experiment by a local artist. It was ensured that all four response options in the pictorial version of the task contained the same level of detail in the illustrations. This was stipulated in order to prevent participants from selecting the most visually appealing option.

4.2.3 Procedure

Children were taken from their class and tested individually in a quiet classroom, onsite at school. Although ethical approval from parents had already been obtained, the testing procedure was first explained to the children before the experiment began, and they gave their verbal assent to continue. Testing included a series of tasks (presented in this thesis) administered in counter-balanced order. Overall, testing lasted approximately two hours and was conducted in two sessions over two consecutive days.

4.2.3.1 Language and Cognitive Measures

The Leeds-Bradford Language Exposure Questionnaire for Children (LeBLEQ-C), the Streamlined-CELF and the CNRep task were administered, and scored, as described in Chapter 2. During the CNRep task, participants wore noise-cancelling headphones to listen to the audio and limit environmental distractions. Testing of these tasks lasted approximately 40 minutes. KS1 SAT scores were provided by the schools for children in both Year 3 and Year 6. KS2 SAT scores were also obtained for Year 6 pupils. Finally, The Leeds-Bradford Language Exposure Questionnaire for Teachers (LeBLEQ-T) was given to the teachers of all children tested. It took teachers approximately 2-3 minutes per child to complete.

4.2.3.2 The Homonymic Task

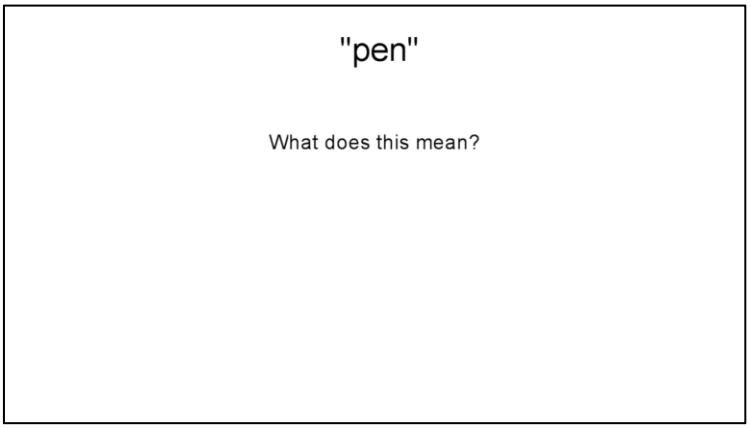
The experiment was created using PsychoPy version 3.0 (Pierce, 2007) and presented to the children on laptops. Trials were presented in random order, as determined by PsychoPy, to prevent order effects. Participants were assigned to one of the four experimental tasks: (1) Orthographical List A, (2) Orthographical List B, (3) Pictorial List A, and (4) Pictorial List B.

Participants were first given instructions to the task. It was explained to the children that that task they were about to participate in regarded "words that looked the same, sounded the same, but sometimes meant more than one thing". Children were given the example of the word "*bat*" which was sometimes used to refer to *a type of animal with wings*, but also used to denote *sports equipment used to hit a ball*; the word "*bat*", consequently, had two meanings. Researchers stated that the study was about the meanings the children knew of the words and, thus, they were required to think carefully about each word. Participants were to be tested on their knowledge of the words in two ways. Firstly, participants were told that they would be presented with a word and asked to define it. They were instructed to respond verbally to the experimenter. Secondly, the participants were informed that they would then see the word alongside four definitions and would be required to select the correct definition/s; sometimes only one of the response options would be the correct definition of the target word, sometimes two of the response options would be correct definitions. Children were instructed to respond using the mouse pad of the laptop. Instructions were presented visually on the screen, one paragraph at a time, lasting 57 seconds in total. After this time, a 'Let's Practice" button was presented allowing participants to proceed to the first of the two practice trials.

Participants were given two practice trials before continuing onto the experimental trials. Each trial comprised of two components: (1) an open-ended question, and (2) a multiple-choice question. In the open-ended component of the trial, the participant was asked to define a word presented on the screen. All experimental trials were presented in the same format as the practice trial of "pen" which is outlined in Figure 7.

Figure 7

Presentation of the Open-ended question in the Homonymy Task

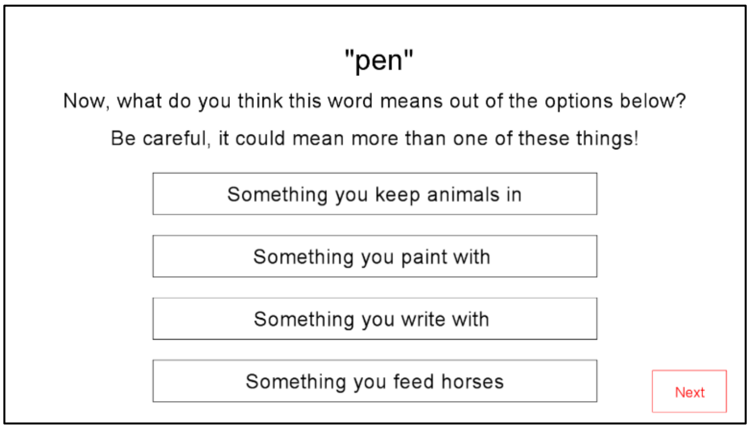


Target words were presented in lower case and participants were able to respond verbally to the experimenter who would key in their answers. In the case of the multiple-choice component of the trials, the format of the Orthographic and Pictorial Tasks can be seen below in Figure 8a and Figure 8b respectively.

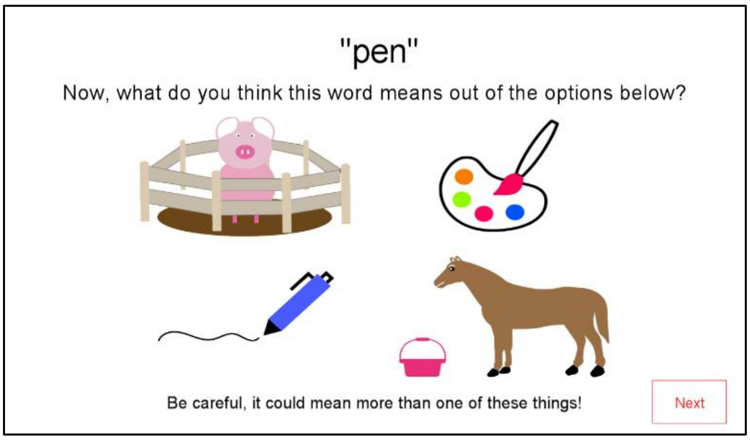
Figure 8

Presentation of the Orthographic (a) and Pictorial (b) versions of the Multiple-Choice question in the Homonymy Task

(a)



(b)



The presentation of the correct and foil response options were randomised by PsychoPy. Participants selected the definition/s of the target word using the touchpad of the laptop. Once selected, the response options would turn green. When the participants had finished their selections, they were instructed to press the 'Next' button in the bottom right corner of the screen. After a short pause (0.2 seconds) during which the screen displayed no text, the next trial began.

After completing the two practice trials (one of these was homonymous, the other was a control), a second set of short instructions were given. Children were told that the practice trials were over and were given the opportunity to ask any questions before the experimental trials began. These instructions lasted 11 seconds. After this, the 'Let's Go!" button was presented allowing participants to proceed to the first experimental trial.

In all cases, instructions and questions were simultaneously presented visually and auditorily to account for differences due to participants' reading abilities. The only period where no audio was presented was in the case of the Multiple-Response phases of the Pictorial task. The audio was recorded by a female, native British-English speaker using Audacity (version 2.1.2). The speaker had a neutral accent.

During the Homonymy task, participants wore noise-cancelling headphones to listen to the audio and limit environmental distractions. The Homonymy Task took approximately 20 minutes to administer. Accuracy of response and response time (in seconds) were recorded automatically by PsychoPy for both the open-ended and multiple-choice components of each trial.

4.2.3.2.1 Scoring of the Homonymy Task

4.2.3.2.1.1 Open-Ended Component

The open-ended component of the Homonymy Task was coded in two ways with responses being assigned both a score and a code. Firstly, the child could either receive a score of 0, 1, or 2. A child received a score of 0 if they failed to give the correct definition of the target word, a score of 1 if they gave one of the correct definitions of the target word, and a score of 2 if they were able to give two correct definitions for the target word.

Secondly, these scores were then coded. Scores of 0 were either coded as an '*Incorrect Definition'* (i.e., the definition they gave was not correct) or as '*Omission'*, where they did not give an answer. For correct scores (i.e., those given a score of either 1 or 2), the meaning given of the definition(s) was coded for the later analysis of the most popular response. Accordingly, each of the meanings of the target word was assigned a number. The first meaning of the word was assigned the code '*Meaning1*', while the second meaning of the word was assigned the code '*Meaning2'*. For example, a definition of '*pen'* as *an object you write with*, would be coded *Meaning1*, while a definition of '*pen'* as *an enclosure for animals*, would be coded as *Meaning2*. If the child was to give a different, yet correct definition of the word, this was scored as '*Other'* (for example, if a child gave the definition of '*pen'* as *the common abbreviation of "peninsula*").

4.2.3.2.1.2 Multiple-Choice Component

Responses to the multiple-choice component of the task were scored using a 0-2 scale. A child received a score of 0 if they had failed to select either of the correct answers; a score of 1 if they identified one of the two correct responses for the target word, or a score of 2 if they were able to identify both correct definitions of the target word.

4.3 Results

Initially, two t-tests (one for the open-ended component and one for the multiple-choice component) were conducted in order to determine if there was an experimental effect. Analysis revealed accuracy rates to be significantly higher in the control condition than the experimental condition in both the Open-Ended ($M_{Experimental}$= .19, $SD_{Experimental}$ = .39; $M_{Control}$= .67, $SD_{Control}$= .47; $t(3246.65) = -32.62$, $p < .001$) and the Multiple-Choice Components ($M_{Experimental}$= .23, $SD_{Experimental}$ = .42; $M_{Control}$= .82, $SD_{Control}$= .39; $t(34.53) = -42.99$, $p < .001$)[1], indicating that, overall, participants found homonymous words significantly more difficult than non-homonymous words. Consequently, all analyses will henceforth only consider experimental trials.

Differences in accuracy rates across the format types (Open-Ended/Multiple-Choice) were subsequently explored, revealing participants to perform significantly better on the Multiple-Choice component (M= .23, SD= .42) in comparison to the Open-Ended component (M= .19, SD= .39; $t(1791) = -4.04$, $p < .001$) of the task.

[1] The same effects were found when analyses were conducted separately for List-A or List-B and Pictorial or Orthographical forms. In all cases, participants had statistically higher scores on the control trials in comparison to the experimental trials.

Subsequently, a series of t-tests were conducted to investigate the differences between the four different versions of the Homonymic Task; Orthographical-A, Orthographical-B, Pictorial-A and Pictorial-B. There was a statistically significant difference between the children who were exposed to the two differing stimuli lists. Despite the lists being well controlled in terms of Age of Acquisition (as determined both by Experiment 2 in Chapter 3 and Kuperman et al., 2012), number of letters in the homonym, imageability, concreteness, familiarity (MRC Psycholinguistic Database; Coltheart, 1981) and frequency (as determined using both the norms generated by Kučera and Francis, 1967, and the Children's Printed Word Database scores by Masterson et al., 2010), participants performed significantly better across all tasks when exposed to List-A ($M_{Open\text{-}Ended}$= .24, $SD_{Open\text{-}Ended}$= .43; $M_{Multiple\text{-}Choice}$= .27, $SD_{Multiple\text{-}Choice}$= .44) compared to List-B ($M_{Open\text{-}Ended}$= .14, $SD_{Open\text{-}Ended}$= .35; $t_{Open\text{-}Ended}(1652.12)$= 5.59, p< .001; $M_{Multiple\text{-}Choice}$= .20, $SD_{Multiple\text{-}Choice}$= .40; $t_{Multiple\text{-}Choice}(1726.93)$= 3.68, p< .001).

However, of the four items that spanned across both stimuli lists, there were no significant differences between the accuracy scores of the participants across Open-Ended and Multiple-Choice components for seven of the eight instances. Means and standard deviations can be seen in Table 32.

Items were consequently considered to demonstrate good inter-rater reliability and all individuals were included in the same analyses, regardless of the specific version of the homonymy task they completed. Interestingly, there was also a significant effect of form of presentation, with participants exposed to the pictorial form (M= .29, SD= .45) outperforming those exposed to the orthographical form (M= .18, SD= .38; t(17.22)= -5.50, p< .001)[2].

Descriptive statistics were first conduced regarding the *type* of response provided by the children. Firstly, this was considered at a group level (including all participants), then the Year 3 children were compared to the Year 6 children. Finally, the monolinguals were compared to their bilingual peers (using the binary classification of language experience). Prior to all t-tests, Levene's tests for

[2] The same effect was shown when considering those exposed to List-A and List-B separately; participants performed significantly better in the pictorial condition.

Table 32

Means and standard deviations (M (SD)) for the characteristics of items appearing across both lists

		Open-Ended			Multiple-Choice		
		List-A	List-B	*p*	List-A	List-B	*p*
Balanced							
	Blow	.02 (.13)	.03 (.17)	.600	.05 (.22)	.09 (.29)	.351
	Match	.65 (.48)	.55 (.50)	.254	.48 (.50)	.33 (.48)	.085
Unbalanced							
	Ball	.10 (.30)	.09 (.29)	.910	.34 (.50)	.36 (.49)	.770
	Park	.42 (.50)	.32 (.47)	.240	.50 (.50)	.30 (.46)	<.041

equality of variances were conducted. In cases where this assumption was violated, degrees of freedom were adjusted and a statistic not assuming the homogeneity of variance was reported.

Subsequently, accuracy and latency data were analysed using a series of Multilevel Linear Models, implemented in IMB SPSS Statistics (Version 26) using a Maximum Likelihood method. The dependent variable was either the accuracy or the latency of the response discussed separately below. The models included the predictors of Gender (male/female), KS1 SAT Reading score, KS1 SAT Writing score, Streamlined-CELF Receptive Language Score (RLS), Expressive Language Score (ELS), Backwards Digit Recall, Total LeBLEQ-T score, Homonymy type (balanced/unbalanced) and Language experience classification (monolingual/bilingual) as fixed factors, as well as an interaction of Year Group and Pupil Premium. Due to the aforementioned differences between stimuli lists, List (A/B) was also included as a variable in all models. Finally, Form of Presentation (orthographical/pictorial) was included as an additional variable in the Multiple-Choice component.

The model was then run for the second time with the same predictors except for language experience classification (monolingual, bilingual) which was replaced with the child's Cumulative Language Input Score generated from the LeBLEQ-C (allowing for a continuous classification of language exposure). The Bayesian information criterion (BIC) of the two models was then compared in order to determine the model with the best fit to the data. In all models, random intercepts were included for each participant. Responses to Open-Ended questions and Multiple-Choice questions are described separately below.

4.3.1 Model Comparison

Bayesian information criterion (BIC) values were compared across the models including the binary (monolingual vs bilingual) and continuous (Cumulative Language Input Score) language experience classifications. BIC values for all models are outlined in Table 33.

Table 33

Bayesian information criterion values for Homonymy models including the binary (monolingual vs bilingual) and continuous (Cumulative Language Input Score) language experience classifications

		BIC	
		Binary	Continuous
Open-Ended Component			
	Accuracy	1046.87	995.47
	Reaction Time	685.54	658.18
Multiple-Choice Component			
	Accuracy	1238.93	1147.69
	Reaction Time	364.85	344.634

Literature considers the model with the lowest BIC preferable (e.g., see Vrieze, 2012). As can be seen in Table 33, the models that include the continuous, Cumulative Language Input Score have lower BIC values, indicating better fitting models. Consequently, only the findings of the continuous models will be discussed below.

4.3.2 Open-Ended Component

4.3.2.1 Response Type Analyses in the Open-Ended Component

First, specific response types were analysed (see Table 34). Responses were coded as ‘0’ (the child failed to respond or repeated the target word without definition, or gave an incomplete or incorrect response), ‘1’ (one correct response), or ‘2’ (two correct responses) depending on their answer.

Table 34

Response types across experimental trials of the Open-Ended Component of the Homonymy Task, stratified by age of child and language experience classification

Response code	Response type	Year 3		Year 6		Total
		Monolingual	Bilingual	Monolingual	Bilingual	
0	Total	29.43%	37.86%	16.33%	22.45%	27.62%
	Omission	7.71%	17.14%	5.10%	6.12%	9.71%
	Incorrect	21.71%	20.71%	11.22%	16.33%	17.91%
1	Total	57.43%	52.68%	53.40%	51.87%	53.46%
	Meaning1	49.71%	45.89%	45.58%	45.58%	46.48%
	Meaning2	4.00%	3.04%	4.08%	3.23%	3.46%
	Other	3.71%	3.75%	3.74%	3.06%	3.52%
2	Total	13.14%	9.46%	30.27%	25.68%	18.92%
	Meaning1 & Meaning2	8.57%	9.46%	30.27%	25.68%	13.00%
	Meaning1 & other	4.00%	3.04%	7.14%	4.42%	4.35%
	Meaning2 & other	0.57%	0.18%	2.38%	3.06%	1.56%

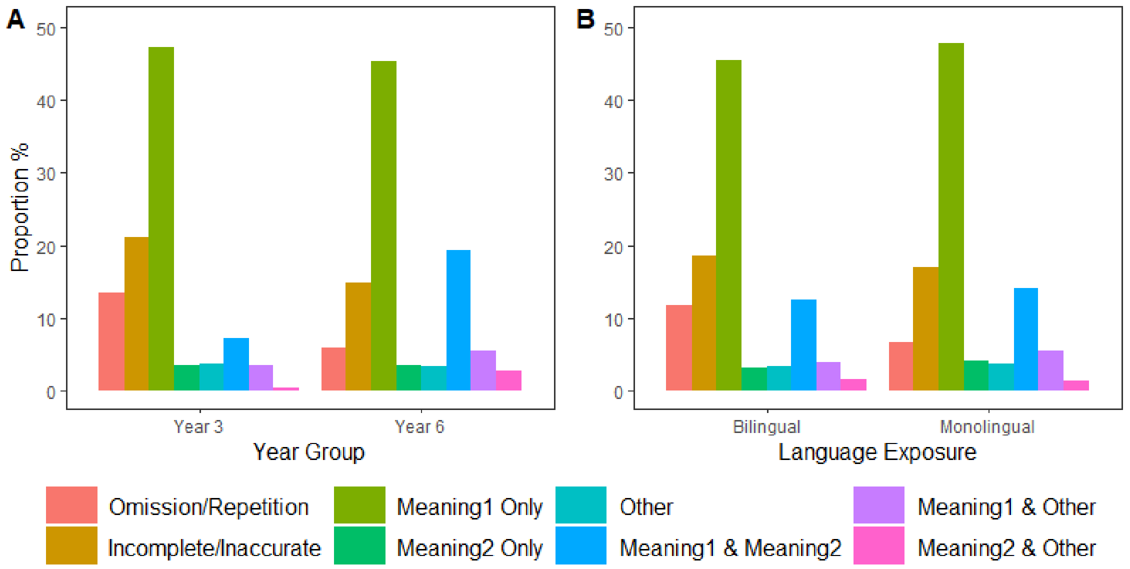
Overall, an ANOVA revealed significant differences in the types of responses given across all participants [$F(2, 383)= 116.05$, $p< .001$]. A post-hoc Bonferroni analysis determined participants to provide more answers with one correct response (M= .53, SD= .14) in comparison to no correct responses (M= .28, SD= .16, $p < .001$) and two correct responses (M= .29, SD= .17, $p < .001$). There were also statistical differences between no correct responses and two correct responses ($p < .001$).

When providing a response coded as ‘0’, participants were statistically more likely to respond with an incorrect or incomplete answer (M= .18, SD= .12) than not respond at all (M= .10, SD= .10; $t(254) = -5.73$, $p < .001$). When giving only one correct response (scored as ‘1’), an ANOVA revealed significant differences in the types of responses given across all participants [$F(2, 383) = 1111.48$, $p < .001$]. Bonferroni analyses revealed this to be attributed to participants providing statistically more Meaning1 only responses (M= .46, SD= .13) than Meaning2 only responses (M= .03, SD= .05, $p < .001$) and ‘other’ responses (M= .04, SD= .05, $p < .001$). No statistical differences were observed between Meaning2 only and ‘other’ responses. Finally, when giving a fully correct response (i.e., two correct responses, scored as ‘2’), an ANOVA revealed significant differences in the types of responses given across all participants [$F(2, 383) = 63.47$, $p < .001$]. Participants were statistically more likely to respond giving the two meanings that were used in the experiment (i.e., the most frequent meanings) (M= .13, SD= .14) in comparison to responding with Meaning1 plus another meaning (M= .04, SD= .05, $p < .001$) or Meaning2 plus another meaning (M= .02, SD= .04, $p < .001$). Participants were also more likely to give a ‘Meaning1 plus another’ response than a ‘Meaning2 plus another’ response ($p < .01$).

Furthermore, a series of t-tests revealed a number of significant differences between children in Year 3 and children in Year 6. While children in Year 6 were more likely to provide two correct responses (M= .26, SD= .19) than those in Year 3 (M= .11, SD= .10; $t(90.83) = -5.96$, $p < .001$), children in Year 3 were more likely to provide no correct responses (M= .34, SD= .15) compared to their older peers (M= .21, SD= .15; $t(126) = 5.22$, $p < .001$). When responding with two correct responses, children in Year 6 were significantly more likely to give every type of correct response, including Meaning1+Meaning2 (M_{Year6}= .19, SD_{Year6}= .15; M_{Year3}= .07, SD_{Year3}= .08; $t(92.05) = -5.51$, $p < .001$), Meaning1+other (M_{Year6}= .05, SD_{Year6}= .05; M_{Year3}= .03, SD_{Year3}= .05; $t(126) = -2.41$, $p < .05$), and Meaning2+other (M_{Year6}= .11, SD_{Year6}= .10; M_{Year3}= .03, SD_{Year3}= .05; $t(88.83) = -3.30$, $p < .01$). When considering the instances where no correct answers were provided, children in Year 3 were significantly more likely to refuse to answer the question, or repeat the target than Year 6 children (M_{Year3}= .13, SD_{Year3}= .12; M_{Year6}= .06, SD_{Year6}= .07; t(110.50)= 4.50, p< .001) as well as give an incorrect or incomplete answer (M_{Year3}= .21, SD_{Year3}= .11; M_{Year6}= .15, SD_{Year6}= .13; t(126)= 2.77, p< .01). There were no differences between Year 3 and Year 6 children in terms of how often they gave only one correct response and the types of such answers (see Figure 9A for a distribution of responses across year groups).

Figure 9

Types of responses given in the Open-Ended component of the Homonymy Task, stratified by year (A) and language experience (B)



Finally, monolingual and bilingual children were equally likely to receive a score of '2' ($M_{Monolingual}$= .21, $SD_{Monolingual}$= .19; $M_{Bilingual}$= .18, $SD_{Bilingual}$= .16), and equally likely to receive a score of '1' ($M_{Monolingual}$= .56, $SD_{Monolingual}$= .15; $M_{Bilingual}$= .52, $SD_{Bilingual}$= .13). However, bilinguals were more likely to receive a score of '0' than monolinguals ($M_{Monolingual}$= .23, $SD_{Monolingual}$= .15; $M_{Bilingual}$= .30, $SD_{Bilingual}$= .16; (t(126)= -2.23, p< .05). Further investigation revealed this to be led by bilinguals refusing to give a response significantly more than monolinguals who at least attempted to provide an answer ($M_{Monolingual}$= .07, $SD_{Monolingual}$= .07; $M_{Bilingual}$= .11, $SD_{Bilingual}$= .11; (t(123.74)= -2.98, p< .01). There were no other significant differences between monolinguals and bilinguals (see Figure 9B for a distribution of responses across monolinguals and bilinguals).

4.3.2.2 Analysis of Accuracy Rates in the Open-Ended Component

When including individuals' Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the Open-Ended trials of the Homonymy task, the relationship between predictors and scores showed significant variance in intercepts across participants, $Var(u_{0j})$= .003, $\chi^2(3)$= 70.73, *p*< .01. There was a significant effect of List [F(1, 92.06)= 12.24, *p*< .01] where more accurate scores were obtained in List-A than in List-B. There was also a significant effect of Homonymy Type [F(1, 1156.94)= 30.08, *p*< .001] with participants more likely to give two correct responses for balanced than unbalanced homonyms. Additionally, both RLS [F(1, 88.98)= 5.26, *p*< .05] and Total LeBLEQ-T Score [F(1, 88.94)= 6.98, *p*< .05] made significant contributions, with higher RLS scores but lower LeBLEQ-T scores (indicating better teacher-rated proficiency) being associated with higher homonym proficiency. Finally, there was an interaction between school year and whether the child was in receipt of Pupil Premium [F(3, 88.95)= 5.23, *p*< .01]. Contributions of all other fixed factors can be seen in Table 35.

The interaction between Year and Pupil Premium was then broken down by conducting separate multilevel models on participants who were in receipt of Pupil Premium and those who were not. The models specified were the same as the main model, but excluding the interaction term of Pupil Premium and, instead, including simply a main effect of Year. These models revealed that, for those who received the subsidiary, there was no significant effect of Year [F(1, 33)= .78, *p*= .382, *b*= -.04, *SE*= .05, *t*= -.87], indicating that although children in Year 6 performed better than children in Year 3 this was not significant. There was a significant effect of Year, however, in children who did not receive the subsidiary [F(1, 55.89)= 11.82, *p*< .01, *b*= -.15, *SE*= .04, *t*= -3.44]. This demonstrated that children in receipt of Pupil Premium do not advance in the comprehension of homonymy between Year 3 and Year 6 to the same extent as their peers that do not receive the subsidiary.

4.3.2.3 Analysis of Reaction Times in the Open-Ended Component

Analyses excluded errors (81.02% of trials) and responses that were 2.5 standard deviations (following recent literature investigating child language, such as Commissaire et al., 2019; Grundy & Keyvani Chahi, 2017; Levy & Hanulíková, 2019; Persici et al., 2019) above/below each participant's mean (a further .12%, resulting in a total 81.14% trials excluded).

The relationship between scores on the Homonymy task and predictors showed a significant variance in intercepts across participants ($Var(u_{0j})$= .000, $\chi^2(3)$= 24.23, $p < .01$) but there was no significant contribution of any of the variables included. All contributions can be seen in Table 35.

Table 35

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Open-Ended component of the Homonymy Task when using a Continuous Measure of Language Experience

	Accuracy				Reaction Time			
	B	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	.02	.10	.22	.827	-.09	.03	-2.68	.009
Cumulative Language Input Score	.03	.05	.67	.506	-.02	.01	-1.65	.110
Gender (Male)	.03	.03	1.09	.280	-.00	.01	-.36	.723
KS1 Reading SATs (below average)	.03	.04	.73	.470	.01	.01	1.34	.189
KS1 Writing SATs (below average)	-.02	.04	-.58	.564	.00	.01	.36	.719
Receptive Language Score	.00	.00	2.29	.024*	.00	.00	1.03	.309
Expressive Language Score	.00	.00	1.25	.216	.00	.00	.05	.962
Working Memory Score	-.02	.01	-1.31	.195	.01	.00	1.39	.174
LeBLEQ-T	-.00	.00	-2.64	.010*	.00	.00	1.03	.308

Year*Pupil Premium (Year 3, without Pupil Premium)	-.11	.04	2.82	.006**	.00	.01	.28	.781
Stimuli List (List-A)	.08	.02	3.50	.001**	.00	.01	.19	.848
Homonym type (Unbalanced)	-.11	.02	-5.48	.000**	-.01	.01	-.94	.348

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

4.3.3 Multiple-Choice Component

4.3.3.1 Response Type Analyses in the Multiple-Choice Component

As with the Open-Ended component, specific response types were first analysed (see Table 36). Responses were again coded as 0' (no correct responses), '1' (only one correct response), or '2' (two correct responses).

Table 36

Response types across experimental trials of the Multiple-Choice Component of the Homonymy Task, stratified by age of child and language experience classification

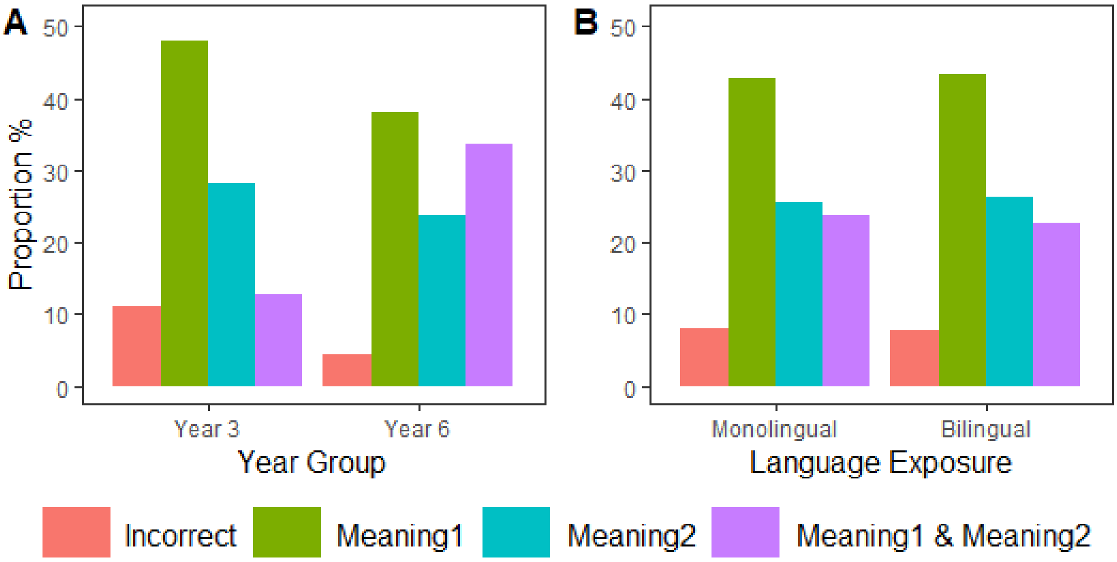
		Year 3		Year 6		Total
		Monolingual	Bilingual	Monolingual	Bilingual	
0	Incorrect	11.43%	10.89%	3.74%	4.59%	7.76%
1	Total	76.00%	76.25%	59.52%	63.27%	69.20%
	Meaning1	47.14%	48.39%	37.76%	38.44%	43.14%
	Meaning2	28.86%	27.86%	21.77%	24.83%	26.06%
2	Meaning1 and Meaning2	12.57%	12.86%	36.73%	32.14%	23.05%

Overall, an ANOVA revealed significant differences in the types of responses given across all participants [$F(2, 383) = 511.46$, $p < .001$]. Post-hoc Bonferroni analyses determined the instances in which participants selected one correct answer ($M = .69$, $SD = .18$) to be significantly more than instances in which they selected no correct answers ($M = .08$, $SD = .09$, $p < .001$) or two correct answers ($M = .23$, $SD = .20$, $p < .001$). Furthermore, they selected no correct answers more often than they selected two correct answers ($p < .001$). When selecting only one correct option (i.e., receiving a score of ‘1’), this was significantly more likely to be the first, more frequent meaning (M= .43, SD= .16) than the second meaning (M= .26, SD= .11; t(225.26)= 10.13, p< .001).

Furthermore, a series of t-tests revealed several significant differences between children in Year 3 and children in Year 6. While children in Year 6 were more likely to provide two correct responses ($M_{Year\text{-}3}$= .13, $SD_{Year\text{-}3}$= .14; $M_{Year\text{-}6}$= .34, $SD_{Year\text{-}6}$= .19; *t*(112.05)= -7.10, *p*< .001), they were less likely to select no correct responses ($M_{Year\text{-}3}$= .11, $SD_{Year\text{-}3}$= .10; $M_{Year\text{-}6}$= .04, $SD_{Year\text{-}6}$= .06; *t*(126)= 4.79, *p*< .001) and less likely to select just a single answer ($M_{Year\text{-}3}$= .76, $SD_{Year\text{-}3}$= .14; $M_{Year\text{-}6}$= .62, $SD_{Year\text{-}6}$= .18; *t*(114.92)= 4.93, *p*< .001). When focussing on instances when only one response was given, this was shown to be driven by higher numbers of Year 3 pupils giving Meaning1-only responses ($M_{Year\text{-}3}$= .48, $SD_{Year\text{-}3}$= .15; $M_{Year\text{-}6}$= .38, $SD_{Year\text{-}6}$= .15; *t*(126)= 3.66, *p*< .001) and Meaning2-only responses ($M_{Year\text{-}3}$= .28, $SD_{Year\text{-}3}$= .10; $M_{Year\text{-}6}$= .24, $SD_{Year\text{-}6}$= .11; *t*(126)= 2.36, *p*< .05) (see Figure 10A for a distribution of responses across years).

Figure 10

Types of responses given in the Multiple-Choice component of the Homonymy Task, stratified by year (A) and language experience (B)



There were no significant differences between monolinguals and bilinguals in terms of types of responses given (see Figure 10B for a distribution of responses across monolinguals and bilinguals).

4.3.3.2 Analysis of Accuracy Rates in the Multiple-Choice Component

When including Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the Multiple Choice component of the Homonymy task, the relationship between predictors and scores showed significant variance in intercepts across participants, Var(u_{0j})= .008, ($\chi^2(3)$= 133.825, *p*< .01). There was a significant, positive contribution of RLS [F(1, 88.89)= 7.32, p< .01] where higher receptive scores indicated higher performance on the Homonymy task. Furthermore, there were negative associations with Working Memory score [F(1, 88.89)= 5.86, p< .05] and Total LeBLEQ-T score [F(1, 88.85)= 4.98, p< .05] where lower scores on the backwards digit recall task and lower scores on the LeBLEQ-T (indicative of a teacher perceiving the child to have better general linguistic competence) were associated with higher scores on the homonymy task. There was also a significant effect of Homonymy type (balanced/unbalanced) [F(1, 1156.84)= 19.73, p< .001] where participants were more likely to select both meanings for balanced in comparison to unbalanced homonyms. Form of presentation was also significant [F(1, 88.85)= 8.43, p< .01] with children being better at selecting the two correct responses when presented with the pictorial form than the orthographical form. The contribution of List also bordered significance [F(1, 93.18)=3.77, p= .055] with participants performing better if they were exposed to List-A than List-B.

Finally, there was a significant interaction between Year of schooling and whether the child was in receipt of Pupil Premium [F(3, 88.86)= 19.73, p< .001]. All contributions can be seen in Table 37. When this interaction was further analysed, there was a significant effect of Year in both the children in receipt of Pupil Premium and those not in receipt of the subsidiary, with those in Year 6 outperforming those in Year 3. However, the improvement in the children with Pupil Premium was shown to be greater [F(1, 33)= 19.03, *p*< .001, *b*= -.23, *SE*= .05, *t*= -4.36], than the improvement in those without [F(1, 55.77)= 14.27, *p*< .001, *b*= -.20, *SE*= .05, *t*= -3.78]. This demonstrates age to affect homonymy processing, but this to be moderated by SES.

4.3.3.3 Analysis of Reaction Times in the Multiple-Choice Component

Due to software malfunction, 25 children were without reaction time data. Analyses then excluded errors (77.60% of trials) and responses that were 2.5 standard deviations above/below each participant's mean (a further .90%, resulting in a total 78.50% trials excluded).

The relationship between scores on the homonymy task and predictors showed significant variance in intercepts across participants, $Var(u_{0j})$= .208, $\chi^2(3)$= 131.50, *p*< .01. There was a significant effect of Form of presentation (orthographical/pictorial) [F(1, 40.49)= 19.53, *p*< .001] with participants performing faster if they were exposed to the pictorial version of the Homonymy task, rather than the orthographical form. There was also a significant effect of Homonymy type (balanced/unbalanced) [F(1, 139.85)= 6.09, *p*< .05] with participants responding faster to balanced than unbalanced homonyms. Working memory also significantly contributed to the model [F(1, 37.49)= 6.98, *p*< .05] with participants with lower scores on the backward number repetition task performing faster on the Homonymy task. Interestingly, gender also made a significant contribution [F(1, 38.98)= 4.25, *p*< .05] with males performing faster than females. Contributions of all other fixed factors can be seen in Table 37.

Table 37

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Multiple-Choice component of the Homonymy Task when using a Continuous Measure of Language Experience

	Accuracy				Reaction Time			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	.31	.12	2.51	.014*	-1.07	1.27	-.84	.403
Cumulative Language Input Score	-.01	.06	-.11	.915	-.14	.38	-.38	.710
Gender (Male)	.00	.03	.10	.924	-.37	.18	-2.06	.046*
KS1 Reading SATs (below average)	.03	.05	.55	.582	.20	.32	.62	.538
KS1 Writing SATs (below average)	-.03	.04	-.77	.443	-.44	.31	-1.42	.163
Receptive Language Score	.00	.00	2.71	.008**	-.01	.01	-.80	.431
Expressive language Score	-00	.00	-.60	.549	.00	.00	.78	.443
Working Memory Score	-.04	.02	-2.42	.018*	.27	.10	2.64	.012*
LeBLEQ-T	.00	.00	-2.23	.028*	.01	.01	1.92	.061

Year*Pupil Premium (Year 3, without Pupil Premium)	-.18	.05	-3.87	.000**	.17	.29	.61	.546
Stimuli List (List-A)	.06	.03	1.94	.055	.18	.16	1.14	.259
Homonym type (Unbalanced)	-.09	.02	-4.44	.000**	.17	.07	2.47	.015*
Form (orthographical)	-.08	.03	-2.90	.005**	.83	.19	4.42	.000**

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

4.4 Discussion

The present research aimed to investigate the lexical ambiguity skills of primary school-aged children in order to understand the impact of bilingualism on the comprehension of homonymy and whether competence could be better predicted using a continuous score of bilingualism rather than a binary classification (monolingual vs bilingual). A novel homonymy task was administered alongside measures of general linguistic ability and working memory (Streamlined-CELF) to Year 3 (7-8-year-old) and Year 6 (10-11-year-old) pupils in schools of relatively low socioeconomic status associated with the Born in Bradford (BiB) cohort. The Leeds-Bradford Language Exposure Questionnaire for Children was administered to record levels of bilingualism of the pupils on a continuous scale (rather than simply using the binary classification monolingual/bilingual). The Leeds-Bradford Language Exposure Questionnaire for Teachers was also administered to obtain an indication of teachers' perceptions of the child's linguistic behaviour in the classroom. Scores on KS1 Reading and Writing SATs were also obtained from the schools, as was information regarding Pupil Premium (which was taken as an indicator of disadvantage).

Primarily, it must be noted that accuracy rates across both response formats were considerably low; 18.98% for the Open-Ended component and 23.05% for the Multiple-Choice component. This is particularly concerning considering the Government's stipulation that by Year 6, children should possess a sophisticated understanding of figurative language (Department for Education, 2013a) which encompasses lexical ambiguity, and, thus, homonymy. The low accuracy rates, however, suggest that this is not the case in the current sample with the children not reaching these goals set out by the Department for Education. Such low accuracy rates are also surprising as the stimuli were specifically designed for use with the current age groups, as determined by the age of acquisition ratings in Chapter 3.

It is possible that the low accuracy rates are due to the population sample in the current research. Whilst socioeconomic status (SES) was accounted for in all models performed (indicated by Pupil Premium receipt), the levels of deprivation seen within the BiB group, and across the city of Bradford as a whole, are still substantial compared to the general UK population. Primarily, all three schools included in the sample were considered to be in low SES neighbourhoods, as indicated by the Income Deprivation Affecting Children Index (IDACI; Department for Communities and Local Government, 2015). Due to the method of allocation of spaces in state schools being determined by catchment areas, it can be assumed that the children in all schools recruited lived in the same neighbourhood as the school, or in those nearby. Thus, it is likely that the children within the sample lived in neighbourhoods with similar deprivation levels to that indicated by the postcode of the school itself.

SES has been demonstrated to be linked to child development through a series of parenting behaviours and child-rearing practices (Conger & Donnellan, 2007). Parents from lower socio-economic strata have been observed to produce less syntactically complex and lexically-rich language when interacting with their children (Hoff-Ginsberg, 1991), engaging in less conversational speech that would elicit a response from the child (Hoff-Ginsberg, 1998; Hoff, 2003; Hoff & Laursen, 2019; Huttenlocher et al., 2010). Such limited interactions impact both language development (Arriaga et al., 1998; Dollaghan et al., 1999; Fernald et al., 2013; Hart & Risley, 1995; Hoff & Tian, 2005; Huttenlocher et al., 2002; Pan et al., 2005; Rowe, 2008) and educational attainment (White, 1982). Consequently, it is likely that these factors could have impacted the language development of the current sample as a whole, due to the disadvantaged conditions in the city of Bradford, not just the children in the sample that were in receipt of Pupil Premium.

In opposition to this, the participants that provided the AoA homonyms ratings obtained in the AoA survey (Experiment 2, Chapter 3) were more likely to closely resemble the general population in terms of SES. Though no specific demographic information for these adults was obtained, a large majority of participants were undergraduate students. Research has shown those from the least represented areas of the country are 32% less likely to continue to higher education (Department for Education, 2016). Such research states that, while 17% of children nationally go on to attend a top third higher education institution (where the current research was conducted), only 10% of children living in Bradford are likely to do so. Consequently, though sampling from the general population, it could be that the SES of the adults that participated in the

Homonymy AoA Rating Study were too disparate from the SES of the children who took part in the main Homonymy Task.

Accuracy rates across the Open-Ended and Multiple-Choice components demonstrated that children found the Open-Ended component more difficult than either of the Multiple-choice components (orthographical or pictorial) of the Homonymy Task. This was in line with predictions as the different difficulty levels across the two tasks was a purposeful manipulation of the study design, with the different components of the task intended to capture different skills. Including both a verbal explanation and a multiple-choice component allowed for the task to be used with a wide age-range, to capture rich, detailed data whilst limiting its susceptibility to ceiling effects. The present study supports previous literature which also demonstrated children to experience greater difficulty in verbal-explanation tasks compared to multiple-choice tasks when assessing the understanding of ambiguous words (Perlini et al., 2018; Pouscoulous, 2014). For example, Perlini et al. (2018) investigated the understanding of both idiomatic phrases and metaphorical words in adult populations, finding increased accuracy rates when participants were presented with multiple-choice questions. Such results highlight the importance of response format when studying ambiguities and suggests an open-ended component is not appropriate for studying pragmatic language within primary-school-aged children. Such an expressive task is too difficult for this age group. The current research builds upon past literature and demonstrates that this is a pattern that emerges regardless of the specific form of ambiguity studied.

Finally, Linear Mixed Effects Models demonstrated various contributors to homonymy understanding. Although older children outperformed younger children on the novel task, different effects were seen between high- and low-SES groups with those from poorer socio-economic strata being adversely impacted. Though there was no effect of language experience when either binary or continuous measures were taken, a model including the continuous measure of the Cumulative Language Input Score was shown to best fit the data. Binary language classification, however, was shown to affect the types of response children gave. Also in accordance with predictions was the finding that children with higher receptive language ability also performed better on the homonymy task, as did children who demonstrated higher teacher-rated linguistic proficiency in the classroom environment. Furthermore, children possessed more knowledge regarding balanced homonyms in comparison to unbalanced homonyms and were better at selecting from pictures rather than sentences in the multiple-choice component. All impacting factors mentioned (e.g., age and language experience), are discussed separately below.

4.4.1 Age of child/School Year and SES

Results demonstrated a clear improvement in homonymy understanding between the two year groups, Year 3 and Year 6, both in terms of accuracy rates and the *type* of responses the children provided. Older children were not only more likely to provide two correct responses but were also less likely to provide no response at all in the Open-Ended component of the task. The latter of these effects is likely due to the increased confidence of the Year 6 pupils (Koriat & Ackerman, 2010); they are less afraid to try to verbalise a response rather than give up. The older children were also less likely to select a single correct answer in the Multiple-Choice component. Consequently, it is clear that the current findings demonstrate a shift in the understanding of homonyms in children between the ages of 7 and 11 years-old where older children are more likely to know two meaning of a homonymous word. Such findings are in line with Smith Cairns et al. (2004) who found children to gradually improve in their understanding of homonymous words throughout primary schooling. Such a shift between these ages is likely due to the introduction of figurative language in schooling (DfE, 2013) with an increase in exposure to literature including homonyms. In addition, the general increase in exposure to language as a whole, in the school environment for older children would also lead to an increase in knowledge of words with multiple meanings (i.e., homonymous words). This is based upon the understanding that exposure to language is known to increase the likelihood of language being learnt (Chondrogianni & Marinis, 2011; De Houwer, 2007, 2009; Duursma et al., 2007; Gathercole & Thomas, 2009; Hammer et al., 2008; Hoff et al., 2012; Pearson et al., 1997; Thordardottir et al., 2006; Thordardottir, 2011).

The current study, however, found the increased competence that occurs with age to be highly moderated by a child's socio-economic position. While children from higher SES homes demonstrate a significant increase of their understanding of homonyms between the two year groups, those from lower SES homes do not advance in their knowledge to the same extent. The findings of the present study oppose original predictions that children in receipt of Pupil Premium (associated with lower SES) would be outperformed by those without in Year 3, but this effect would be eliminated by Year 6 (as per the findings of Hutchinson et al., 2016). Instead, the current study found the gap between children who received Pupil Premium and those who did not to increase in Year 6. Thus, it seems that the acquisition of homonyms follows a different pattern to general vocabulary growth. This is consistent with past research that showed that the acquisition of homonyms is harder than the acquisition of unambiguous words. For instance, Doherty (2004) investigated children's ability to learn new, fictitious meanings of pseudo-homonyms (e.g., a new meaning for the word “fork”). They found children to be significantly more able to learn an intended referent when a nonsense word was attributed it, rather than a pseudo-homonym (such as “fork”). Additionally, they found that when it was highlighted to the child that a word may refer to two different things, children became less accurate at selecting the original meaning of the pseudo-homonyms (i.e., an actual fork). This emphasises how children experience difficulties in associating new meanings to previously existing words (Mazzocco, 1997). As such, the acquisition of both meanings of a homonym could be lagging behind the rate of normal vocabulary development and be increasingly impacted by SES.

This disadvantage found for children of low SES is of particular concern when considering the previously discussed link between homonyms and reading ability, as children who struggle to understand homonyms also underperform on reading tasks (Gernsbacher, 1991, 1997b, 1997a; Gernsbacher, Robertson, et al., 2001; Gernsbacher, St John, et al., 2001). Children from lower socioeconomic strata may, thus, require additional support than their higher-SES peers leaving them at a double disadvantage in terms of both homonymous competence and reading skill. Consequently, it is suggested that interventions which aim to improve homonymous competence (and thus reading performance) should be predominantly targeted at children from lower SES as they are likely to benefit low-SES children to a greater degree. Such interventions may also assist in reducing the educational gap between high and low SES groups.

4.4.2 Language Experience

Perhaps most interesting is that the models shown to fit the data best were those that included the continuous measures of language experience, the Cumulative Language Input Score (from the LeBLEQ-C) indicating that language exposure is indeed likely to contribute to knowledge of homonyms, but a binary classification system may not be sensitive enough to capture such differences. This is in line with literature suggesting bilingualism to be better suited to being measured along a continuum, rather than simply considered to be a binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020). However, it must be noted that although more continuous measures of bilingualism may be preferable, a child's level of exposure to English did not significantly contribute to any of the models predicting homonymy ability. Consequently, findings should be interpreted with caution.

Nevertheless, language was shown to be a significant contributor to the types of responses a child gave. Bilinguals were significantly more likely than monolinguals to not respond to a question rather than attempt to answer it. This could relate to the bilingual children's confidence in their language skills. Macintyre et al. (1998) was among the first to propose a hierarchical system where the willingness of an individual to communicate was built upon an individual's motivational propensities, one of which was their confidence in their L2. It was suggested that bilinguals may choose not to interact due to communication apprehension with the relationship being moderated by level of proficiency in the language (Ozdener & Satar, 2008). Thus, it could be that the bilingual children chose not to answer the questions in the homonymy task due to insufficient knowledge, or simply because they were not as confident as their monolingual peers in their answers.

Consequently, these findings are indicative of differences in processing between these two groups of children and are in line with previous literature that demonstrated monolinguals and bilinguals to perform differently on tasks assessing the understanding of ambiguous words (Hessel & Murphy, 2019; Segal & Gollan, 2018). Segal and Gollan (2018), for instance, found bilinguals to be slower and less accurate on tasks of metaphorical competence in comparison to their monolingual peers. However, until now, there has been limited research on the impact of language experience on homonymic understanding. The differences observed between monolinguals and bilinguals in the current task could be attributed to the increased exposure monolinguals receive in English.

Increased exposure to language has been shown to impact many aspects of language development, including receptive and expressive vocabulary in bilingual children (e.g., Hoff et al., 2012; Paradis, 2011; Thordardottir, 2011; Unsworth, 2013, 2017). Exposure to language is known to be a substantial contributor to the rate of language growth in monolingual and bilingual children (Hart & Risley, 1995; Hoff, 2003;. Pearson et al., 1997) with the rate of vocabulary learning being proportional to exposure (Chondrogianni & Marinis, 2011; De Houwer, 2007, 2009; Gathercole et al., 2013; Hoff et al., 2012; Thordardottir, 2011; Thordardottir et al., 2006). This is attributed to bilingual children experiencing less of each of their languages in comparison to monolingual children of the same age due to being required to divide their time between each of their languages. The present findings, thus, contribute to this body of literature by demonstrating that the same applies to homonymous words; bilingual children's reduced exposure to English negatively impacts knowledge of the distinct meanings of homonymous words.

4.4.3 Receptive Language

The ability of a child's Receptive Language Score to predict their accuracy on both the Open-Ended component and Multiple-Choice component of the task is consistent with the view that increased receptive language ability is related to higher ambiguity competence. Research has identified similar links with metaphoric competence (Deckert et al., 2019; Hessel & Murphy, 2019; Norbury, 2005; Rundblad & Annaz, 2010a) and, to an extent, idiomatic competence (Smith & Murphy, 2015). Comparatively little research is available on the relationship between receptive language and homonymic ability. To the author's knowledge, the only research to identify such links was by Zipke (2011) who found an association between the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4) (Dunn & Dunn, 2007) and the ability to define homonyms in 6-7-year-old children. The present study goes beyond those findings and is the first to use statistical modelling to predict both monolingual and bilingual children's knowledge of homonymous words using their receptive language ability. The present findings demonstrate a clear link between the degree to which a child understands language as a whole (i.e., receptive language ability), and their understating of words with more than one meaning.

4.4.4 Leeds-Bradford Language Experience Questionnaire for Teachers

The Total Score of the LeBLEQ-T was shown to significantly predict children's accuracy rates in both the Open-Ended and Multiple-Choice components. In these cases, lower Total LeBLEQ-T scores, indicative of better general linguistic ability as rated by a child's classroom teacher, were associated with higher accuracy rates in the Homonymy task. The LeBLEQ-T quantifies a teacher's perception of a child's linguistic ability taking into consideration listening, speaking, reading and writing skills of children as well as more pragmatic abilities, thus, providing a comprehensive indication of a child's communicative competence within the classroom environment. Though the LeBLEQ-T is conceptually different from measures normally used within the literature which directly assess oral proficiency of children (such as the British Picture Vocabulary Scale (BPVS; Dunn et al., 1997), or various measures of the CELF), the present findings support previous research demonstrating language skills to be a substantial predictor of homonymous understanding (e.g., Zipke, 2011).

Furthermore, the current research highlights how teacher-report methods, which have been shown to be reliable tools for gathering information regarding child experience (Rimfeld et al., 2019), can be used in place of more direct measurements whilst still providing rich information on the language skills of a particular child. While a direct assessment of oral language proficiency (such as the BPVS or various different measures of the CELF) can take anywhere from 10-40 minutes, the LeBLEQ-T can be completed by classroom teachers in less than three minutes. Consequently, the LeBLEQ-T may be used as a more time-effective measure of language proficiency when testing understanding of homonymy in school environments, whilst still providing a wealth of information regarding a child's linguistic ability.

4.4.5 Homonymy Type

In accordance with predictions, children were more likely to correctly define (in the Open-Ended component) or select (in the Multiple-Choice component) both meanings of a homonymous word if the frequency of its meanings was balanced, rather than unbalanced. Additionally, they were also faster in doing so in the Multiple-Choice component.

In the case of the balanced homonyms, this is likely due to the two distinct meanings of words having similar frequencies and, thus, being equally likely to have been acquired into the individual's vocabulary. Consequently, each distinct meaning is equally likely to be retrieved from memory stores when the child hears the word making them able to provide both definitions. In comparison, unbalanced homonyms are characterised by one meaning (Meaning1) being of significantly higher frequency than the other (Meaning2). As a result, if the more frequent/dominant meaning is in the individual's vocabulary, it is likely to be activated with relative ease. The second meaning is less likely to have been encountered, however, and, thus, the child is less likely to know the definition of the word. Furthermore, even if the second, subordinate meaning has been acquired, as the two meanings in homonymy always compete for activation, it is more likely to lose to the more frequent, dominant meaning. This was supported when analysing the type of responses provided; in cases where participants demonstrated understanding of only one meaning of the homonymous word, this was more likely to be the first (i.e., more frequent) meaning. This effect was observed across both the Open-Ended and Multiple-Choice component of the task. These findings are consistent with those seen in the literature. Fontes and Schwartz (2011) stated that, when presented in a neutral, non-biasing context, distinct meanings of balanced homonyms will be activated at the same time. Conversely, in the case of unbalanced homonyms, the less frequent/more subordinate meaning may be activated at a later timeframe and will not succeed in competition against the dominant meaning.

Furthermore, the current research highlights the importance of including both balanced and unbalanced homonyms when conducting research regarding lexical ambiguity as different effects can be seen between the two types of homonyms. Such research builds upon the research by Kousaie et al. (2015) who investigated the differences in homonymy understanding between monolingual and bilingual individuals, but only used unbalanced homonyms.

4.4.6 Form of Presentation

The form of presentation (orthographical vs pictorial) was shown to be a significant predictor in both the accuracy rate and response time with children performing more accurately and faster when they were exposed to the pictorial version rather than the orthographical version of the task. These findings are in support of past literature that demonstrated the recognition of pictures to be superior to recognition of words. Seminal research by (Paivio, 1971) noted that using pictures or objects as stimuli resulted in superior recall than when words were used. Jones (1973) found this was also the case in children as young as 3 years-old, regardless of form of presentation (pictorial or orthographic). Thus, children exposed to the pictorial form of presentation may be able to associate the word to the response options displayed as images more easily, rather than as definitions in the form of sentences, due to ease of retrieval from memory.

Furthermore, it may be that the use of pictures as response options also reduces the cognitive load placed on the individual. All information was presented visually and auditorily, to control for differences in the reading abilities of participants, for both the orthographic and pictorial versions of the experiment. In the pictorial task, children could easily look at the pictures as they were listening to the definitions in order to make an informed decision. On the other hand, in the orthographical version, quickly glancing at the visual information might have been more demanding. So, children had to either hold each of the four options in their working memory while they retrieved the correct response or re-read the sentences themselves and then make their decision. Either of these options required increased cognitive load which is likely to have impacted both the accuracy and reaction times for children who were exposed to the orthographic task, in comparison to the pictorial task.

It is, thus, clear that the use of different modalities in assessing homonymous awareness can lead to inconsistency in results; this should be avoided if findings are to be compared across the literature. Therefore, standardised methods of assessing the understanding of homonymy are required. This is particularly important due to the aforementioned links between knowledge of homonyms and reading ability (Gernsbacher, 1991, 1997b, 1997a; Gernsbacher, Robertson, et al., 2001; Gernsbacher, St John, et al., 2001) and the potential for the assessment of homonyms providing insight into the reading capabilities of primary-school aged children.

4.4.7 Effects of Stimuli List

Finally, the stimuli list children were exposed to also impacted their accuracy rates in the Open-Ended component of the Homonymy task (and bordered significance in the Multiple-Choice component). This was unexpected considering the lists were matched on a number of parameters known to influence processing including Age of Acquisition (as determined in Experiment 2, Chapter 3), non-meaning specific AoA (Kuperman et al, 2012), number of letters in the homonym as well as imageability, concreteness and familiarity (MRC Psycholinguistic Database; Coltheart, 1981). Furthermore, the words in each list were statistically similar in terms of frequency (determined using both the norms generated by Kučera and Francis, 1967, and the Children's Printed Word Database scores; Masterson et al., 2010).

However, it is possible that the differences observed between the lists could be due to the specific norms applied. For example, the written frequency norms of Kučera and Francis (1967) were obtained over half a century ago. Furthermore, due to the nature of norming studies, all norms were standardised with monolingual populations only; there are currently no norms for the written frequencies of words for bi/multi-lingual populations. These norms, thus, may not be indicative of the type of language the children within the current sample are likely to hear in their everyday interactions.

Alternatively, differences between the lists could be explained by factors other than those controlled for in the current study. For instance, the current research did not take into consideration a homonym’s number of meanings or senses but rather chose to focus on the first and second highest frequency distinct meanings as these indicate the majority of encounters of the word (Armstrong et al., 2012). This is a common practice within homonymy literature due to these two meanings being used to calculate dominance effects (Twilley et al., 1994). However, a recent body of literature has demonstrated that the number of senses of an ambiguous word can affect its processing with a decrease in lexical decision times coinciding with an increase in the number of senses (Beretta et al., 2005; Klepousniotou, 2002; Rodd et al., 2002). Consequently, it is possible that the homonyms contained within the two lists had, on average, dissimilar numbers of associated senses, leading to the different effects seen between the two versions of the experiment. This is a factor to be considered in future research.

4.4.8 Conclusion

Overall, the most notable factors that contributed to a child's understanding of homonymous words included the age of the child and their socio-economic status. These findings were in line with previous literature that suggests an increase in homonymy competence between the ages of 7- and 11-years. The present research advanced such knowledge by demonstrating this relationship to be heavily moderated by socio-economic status. Furthermore, a child's receptive language ability, working memory, and general language ability as reported by their classroom teacher were also shown to be substantial predictors of homonymy understanding. Specific characteristics of the homonyms themselves (such as the frequency of each of their meanings; i.e., balanced/unbalanced) and how they are assessed (i.e., pictorially/orthographically or in tasks requiring verbal-explanation/multiple-choice) were also shown to impact understanding. This emphasises both the importance of controlling for the stimuli in studies using homonymous words, but also the way in which they are presented/assessed.

Interestingly, though language experience did not have a significant impact on accuracy rates, the quality/type of answers a child provided was significantly impacted by their language experience classification (i.e., if they were classed as monolingual or bilingual). Models including a continuous measure of language experience were shown to fit the data better than models only including a binary language classification, suggesting knowledge of homonyms might be associated with language experience to some extent, but not to the degree of other, more influential factors such as age and socioeconomic position.

However, it must be noted that, despite only including stimuli in the task for which both meanings were reported to have been learnt before the age of 10 years-old (as determined by the Norming AoA Study, Experiment 2, Chapter 3), there were still a substantial number of instances where children were unable to identify both meanings of a word. As the ability to correctly define two meanings has been shown to increase with age, this poses the question of whether knowledge of homonymy would significantly increase in older monolingual and bilingual populations and whether the same variables would remain as predictors of understanding.

In conclusion, the current study contributes substantially to the limited literature investigating children's understanding of both balanced and unbalanced homonyms in bi/multi-lingual populations. Knowledge of the factors that impact a child's awareness of homonymy can guide better the identification of children most at risk, allowing targeted interventions to be both developed and implemented. The next chapter explores another trope of ambiguity, metaphor, and the extent to which it is impacted by the same variables as homonymy.

5 The Effect of Language Experience on the Comprehension of Metaphors in Primary-School aged Children

Metaphorical polysemy is unique. Unlike homonymy (discussed in Chapter 4), or metonymy (which shall be discussed in more depth in Chapter 6), the first sense of a metaphoric word is often the more literal interpretation, while the second sense is, at least originally, more figurative. Relatedness between the two senses of a metaphor, however, may not always be obvious (Apresjan, 1974) which can often lead to misinterpretation. Interestingly, however, there has been little research focussing on specific metaphorical words and the age at which the knowledge that a single word can have more than one, related senses is acquired.

Previous research has shown that the understanding of metaphors begins as early as three years old (Van Herwegen et al., 2013). Thus, skills disambiguating polysemy are considered to be acquired relatively early in childhood. Metaphorical understanding has long been considered to play an important role in education, the acquisition of new knowledge and the ability to organise ideas (e.g., Glucksberg, 2008; Ortony, 1975; Vosniadou, 1987). More recently, this has been attributed to metaphorical awareness' association with grammatical competence (Littlemore & Low, 2006) attentional resources (Coney & Lange, 2006), working memory, (Godbee & Porter, 2013) and information processing (Willinger et al., 2019), which are all vital for learning within a classroom environment. Metaphorical comprehension has also been shown to contribute to social intimacy (Bowes & Katz, 2015) and communicative competence (Blasko, 1999; Littlemore & Low, 2006; Winner et al., 1980), assisting children in forming social bonds. Consequently, knowledge about how metaphors are processed, and the factors which may impact this, are key to developing ways in which to ameliorate individuals who experience difficulties.

At the most basic level, metaphors are known to be processed differently to purely literal constructs (Schmidt & Seger, 2009). However, this becomes complicated due to the two disparate senses of the word also being processed differently. Research by Klepousniotou et al. (2012) demonstrated differences in meaning activation patterns between dominant and subordinate metaphorical senses. Klepousniotou and colleagues used event-related potentials (ERPs) to investigate the time-course meaning activation of a range of ambiguity types, focussing on N400 activation. They found the subordinate meanings of metaphors to cause differentially reduced effects moving from left to right hemisphere electrode sites, concluding the right hemisphere to be increasingly involved in processing of figurative meanings of metaphoric words.

It is possible that these subordinate metaphorical meanings are processed more efficiently in the right hemisphere due to this meaning often being less salient (i.e., less familiar, less frequently encountered, or less supported by prior context). In line with this view, the Graded Salience Hypothesis (Giora, 1997, 2002, 2003; Giora & Fein, 1999) states that the meaning salience of a word determines the way in which it is processed, with salient (lexicalised) meanings being processed before non-salient (non-lexicalised) meanings. Several studies have demonstrated this effect in neuroimaging studies investigating metaphorical *units* (e.g., '*how time flies'* or '*I'm on cloud nine'*; e.g., Hessel & Murphy, 2019). In such instances, research has shown both literal and familiar metaphorical units (i.e., speech that has been lexicalised) to be more efficiently processed in the left hemisphere, implying fine semantic coding (Beeman, 1998; Faust & Mashal, 2007; Jung-Beeman, 2005; Segal et al., 2017). Conversely, novel and unfamiliar metaphorical units have been shown to be processed more efficiently in the right hemisphere, implying larger semantic fields of distant associates have been activated by a process of coarse semantic coding (Klepousniotou et al., 2014; Schmidt et al., 2007).

Furthermore, this effect of familiar metaphorical units being processed more efficiently in the left hemisphere, and less familiar processing taking place in the right, has been found to be the case in monolinguals and bilinguals alike. When investigating the N400 event related potentials of monolingual and bilingual adults on a task of semantic judgement, Segal and Gollan (2018) found both groups to show a clear left hemisphere advantage for familiar metaphor processing. Interestingly, they also found left hemisphere processing efficiency to vary as a function of language dominance, with bilinguals who were more dominant in English processing the metaphors more akin to English-monolinguals. Likewise, individual items also exhibited different patterns of left and right processing as would be expected based on theories that propose gradual transfer from right to left hemisphere with increasing levels of familiarity.

However, although the laterisation patterns were the same in both monolinguals and bilinguals, Segal and Gollan (2018) found bilinguals tended to be less accurate, and slower, on this task compared to their monolingual peers. This effect is apparent not only in neuroimaging studies investigating laterisation, but also in research using more behavioural techniques with children. For example, in research including both Year 1 and Year 2 pupils (5-8-year-olds), Hessel and Murphy (2019) found bilinguals to possess weaker metaphorical comprehension than their monolingual peers at both ages, especially when task demands increased.

In addition to these findings, Hessel and Murphy (2019) found metaphorical understanding to be been associated with receptive language skill. In a study of 43 monolingual and 37 bilingual children in Years 1 and 2, they found children demonstrating high metaphorical competence to also achieve high scores on the British Picture Vocabulary Scale II (BPVS; Dunn et al., 1997). Similar conclusions were drawn by Deckert et al. (2019) and Rundblad and Annaz (2010). Norbury (2005) also found this to be true in ASD populations. However, the majority of these studies have employed the use of the BPVS, a rather simplistic tool requiring participants to solely point to an image that best represents a statement heard (excluding the research by Deckert and colleagues (2019) which was conducted in Germany with German measures). Furthermore, only Hessel and Murphy (2019) investigated metaphorical competence in bilingual populations, with results simply demonstrating that those who scored higher on the metaphorical task also scored higher on the BPVS; however, no analyses were conducted to determine whether receptive language ability could reliably predict metaphorical competence. Furthermore, such research also focussed on metaphorical units (such as '*how time flies*') not metaphorical words. Metaphorical units and metaphorical words are conceptually different with the former being more akin to idiomatic phrases (Gibbs, 1992). As metaphors are known to be processed and acquired slightly differently to idioms (e.g., Cacciari et al., 2011) it may be that such units are treated differently to individual metaphorical words, with the potential of opposing conclusions being drawn between the two tropes. Additionally, research into metaphoric comprehension association with expressive language has been widely under-researched. Even as recently as 2019, Deckert and colleagues noted "studies investigating the association between linguistic competence and metaphor processing are scarce," (p. 4) suggesting a need for more research into other aspects of language ability and how this relates to metaphoric knowledge.

While Hessel and Murphy (2019) and Segal and Gollan (2018) demonstrated monolinguals and bilinguals to perform differently on tasks of metaphoric competence, others reported no differences between the accuracy rates of monolinguals and bilinguals (e.g., Johnson & Rosano, 1993; Trosbrog, 1985), creating a discrepancy within the literature. Furthermore, despite the increasing notion that bilingualism is best conceptualised along a continuum, rather that captured as a categorical variable (e.g., Kašćelan et al., 2020; Serratrice & De Cat, 2020), there has been no research to date investigating if the degree of an individual's language exposure can predict their metaphorical ability.

Due to the lack of standardised assessments to test metaphorical awareness, many studies within the literature have developed their own methods for assessing such skills. However, due to insufficient methodological detail (such as specifics of stimuli selection), these studies are not always possible to replicate outside of the lab in which they were developed. The assessments that have been developed across the literature also employ vastly different methodologies. This becomes problematic considering metaphoric understanding has been shown to be dependent on the method in which it is assessed (Perlini et al., 2018; Pouscoulous, 2011, 2014) leading to inconsistencies in the assumptions about the ages in which individuals acquire metaphors. For example, Douglas and Peel (1979) posited that up to 73% of eight-year-olds can demonstrate metaphorical comprehension, a skill which then develops across the primary school years with children being able to comprehend 97% of the metaphors presented to them by the age of 12 years. Ozcaliskan (2007) contradicted this finding however, showing metaphorical understanding to begin to develop even prior to this age, with some 4-5 year olds showing evidence of metaphorical knowledge. Van Herwegen et al. (2013) criticised this even further, demonstrating that children as young as 3 years old can show metaphorical understanding. Thus, further research is needed to address the inconsistencies of the ages in which metaphors are acquired. An assessment that is able to measure the metaphoric skills of children across a wide age range, assessing how the use of different response methods (i.e., open-ended response options compared to choosing from a series of options) impact conclusions drawn about understanding is therefore required.

5.1 The Current Study (Experiment 4)

The research presented in this chapter aimed to investigate metaphoric comprehension in both monolingual and bilingual, primary-school-aged children and the factors that contribute to better understanding of metaphors. Given the lack of standardised measures, a novel metaphor task was developed whereby children were required to respond to a metaphor in two ways. Firstly, they were asked to verbally explain the metaphor in the open-ended component of the task, then they were asked to select the correct meanings of the metaphorical senses form a series of options. Due to previous research which showed that children find explaining metaphors more difficult than answering multiple-choice questions on them (Arcara et al., 2019; Perlini et al., 2018; Pouscoulous, 2014), both response methods were employed in order to reduce the likelihood of ceiling effects occurring and allow for rich linguistic data collection. This methodology was also selected because it could help disentangle the inconsistencies found across the current literature regarding the age in which metaphors first come to be acquired (Douglas & Peel, 1979; Seyda Ozcaliskan, 2007; Van Herwegen et al., 2013). As previously mentioned, these disparities can likely be attributed to the methods in which metaphors are assessed and the stage of metaphor acquisition captured (Perlini et al., 2018; Pouscoulous, 2011, 2014). Thus, using both methods attested in the literature with the same materials and participants would allow direct comparisons between the methods, leading to better understanding of metaphor processing and comprehension by monolingual and bilingual primary-school children.

As the national curriculum for England suggests children should demonstrate an understanding of figurative language from Year 3, with the development of more sophisticated understanding apparent by Year 6 (Department for Education, 2013a), the current study was tested in these two age groups with older children expected to excel. However, on the basis of the aforementioned research suggesting the gap between low- and high-SES groups is eliminated by KS2 (Hutchinson et al., 2016), it was also predicted that there would be differences within age groups. Namely, it was hypothesised that lower SES children would be outperformed by higher SES children in Year 3 but such differences would be attenuated in the Year 6 group. Whether or not the child was in receipt of Pupil Premium was used as a proxy for SES grouping.

Due to inconsistencies in literature when determining a link between language experience and metaphoric understanding (e.g., Hessel & Murphy, 2019; Johnson & Rosano, 1993; Segal & Gollan, 2018; Trosbrog, 1985), the relationship between these two factors was explored in this child population, both in terms of accuracy and reaction time as well as in the type of response a child was required to provide. Additionally, it was hypothesised that a child's cumulative language score (CISs), would be a more robust predictor of metaphoric ability, due to bilingualism being better suited to measures along a continuum, rather than a binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020).

Finally, associations with scores of academic attainment (Reading and Writing SATs) and teacher's perceptions of linguistic competence (LeBLEQ-T) were also investigated alongside scores of language development. Of interest was the ability of a child's Receptive Language Score (RLS) to predict metaphorical awareness. The RLS of the Streamlined-CELF is a composite score assessing receptive ability on more than one task, consequently allowing for a more detailed understanding of the link between metaphorical comprehension and receptive ability.

5.2 Methods

5.2.1 Participants

Participants were the same as those in Chapter 4 (Experiment 3).

5.2.2 Materials

The experimental task had two components to test participants' knowledge of metaphorical words. Firstly, participants were presented with a word and asked to define it. They were instructed to respond verbally to the experimenter; the experimenter then typed their answer on the computer. Secondly, participants then saw the word alongside four definitions and had to select the correct definition/s. Sometimes only one of the response options would be the correct definition of the target word; other times two of the response options would be correct definitions. Children were instructed to respond using the mouse pad of the laptop.

Both metaphoric (i.e., experimental) and non-metaphoric (i.e., control) words were selected from (Armstrong et al., 2012) for inclusion into the current study. The experimental words had one meaning (one orthographic and phonological form) but had more than one senses – a literal and a metaphorical one. Unlike the meanings of the homonymic stimuli included in the previous chapter, the senses of the metaphoric stimuli were related by analogy in a metaphorical way. This was based on the definitions of metaphors attested in past literature (e.g., Beretta et al., 2005; Klepousniotou et al., 2012). The experimental stimuli contained within the current research had also been used in previous literature investigating metaphoric processing and understanding (e.g., Klepousniotou et al., 2012).

Control words had only a single definition. Only meanings that were part of British-English vocabulary and could be placed in the sentence-final position were included. Words were excluded if they were compound words, or were not easily definable concepts. Furthermore, words were excluded if they were deemed to be outdated (e.g., "*topic*" to mean *the discontinued chocolate bar*), or age-inappropriate. While these decisions were subjective (made by the researcher), they adhered to certain parameters, namely, words were considered inappropriate for child research if they were related to sex/sexual acts, drugs/alcohol or profanities (e.g., words such as "*gin*" or "*sex*"). The resultant list comprised of 12 metaphors (i.e., words with more than one meaning) and 12 control words (i.e., words with a singular meaning). All stimuli can be found in Appendix D.

There were no statistically significant differences between the experimental items and the control items in terms of the number of letters ($M_{Experimental}$ = 4.17; $SD_{Experimental}$= .83, $M_{Control}$= 4.58, $SD_{Control}$= .90; $t(22)=1.18$, $p=.252$) or phonemes ($M_{Experimental}$ = 3.50; $SD_{Experimental}$ = 1.80, $M_{Control}$ = 4.11, $SD_{Control}$ = .92; $t(17)=1.32$, $p=.206$).

In addition, the words were controlled for imageability ($M_{Experimental}$ = 520.30, $SD_{Experimental}$= 37.14; $M_{Control}$= 543.00, $SD_{Control}$= 98.68), concreteness ($M_{Experimental}$ = 602.30, $SD_{Experimental}$ = 30.01; $M_{Control}$= 608.33, $SD_{Control}$ = 23.35), and familiarity

($M_{Experimental}$ = 598.60, $SD_{Experimental}$ = 22.51; $M_{Control}$ = 596.33, $SD_{Control}$ = 22.01) using the MRC Psycholinguistic Database (Coltheart, 1981) Again, there were no statistically significant differences between control and experimental words based on these parameters ($t_{Imageability}(3.35)=.45$, $p=.682$; $t_{Concreteness}(11)=.317$, $p=.757$; $t_{Familiarity}(11)=-1.54$, $p=.881$ respectively).

Furthermore, there was no statistical difference between the frequency of the experimental and control words in terms of Kučera and Francis (1967) ($M_{Experimental}$ = 24.36, $SD_{Experimental}$= 24.12, $M_{Control}$= 27.67, $SD_{Control}$= 34.30, $t(15)= -.23$, $p= .819$) and Thorndike and Lorge (1944) written frequency scores ($M_{Experimental}$ = 157.27, $SD_{Experimental}$= 174.41, $M_{Control}$= 188.00, $SD_{Control}$= 255.87, $t(16)= .30$, $p= .765$). This was also determined using the frequency scores from the Children's Printed Word Database scores (CPWD; Masterson et al., 2010) ($M_{Experimental}$ = 83.10, $SD_{Experimental}$= 103.57, $M_{Control}$= 70.00, $SD_{Control}$= 74.71, $t(18)= .351$, $p= .755$) and the Subtitle-based Word Frequencies for British-English (SUBTLEX-UK; van Heuven et al., 2014) ($M_{Experimental}$ = 4.26, $SD_{Experimental}$= 4.9, $M_{Control}$= 3.86, $SD_{Control}$= .72, $t(22)= 1.62$, $p= .119$).

Finally, the experimental and control words were matched in terms of AoA (Kuperman et al., 2012) with no statistically significant difference between the two groups ($M_{Control}$= 5.23 years, $SD_{Control}$= .99 years, $M_{Experimental}$ = 5.46 years; $SD_{Experimental}$=1.38 years; $t(19.91)=-.071$, $p=.944$). Care was taken to ensure metaphors with a range of AoAs (3-11 years) were included, reducing the likelihood of floor- or ceiling effects occurring. Stimuli can be seen in Appendix A.

5.2.2.1 Multiple-Choice Response Options

Four response options were created for each target word. For experimental words, two of the four response options were definitions of the target word, with one response option defining the first sense of the metaphoric word and the other response option defining the second sense. Conversely, for the control words, only one of the response options was a definition. All definitions were selected from the Wordsmyth online dictionary, available online at wordsmyth.net (Parks et al., 1998). The remaining response options were foils (i.e., incorrect definitions which were similar to correct responses). It was ensured that all response options, for both experimental and control trials, defined real concepts and were, thus, feasible. Consequently, the identification of the correct definitions of the target word would imply the participant's precise understanding of a word, rather than vague knowledge.

For the experimental trials, each incorrect response created was designed to be a foil for one of the two correct response options. For example, in the case of the metaphoric word "*snake*", the two correct response options were:

(1) a reptile with no legs

(2) a person who betrays their friends.

The two foils were:

(1) an animal with eight legs

(2) a person who is always late.

Similarly, all three foils in the control trials were created to be foils to the single correct response option. As the second sense of a number of the metaphoric words related to something you might call a person, where possible, at least one of the foils for the control items described a personal attribute or a name you may give to a person. For example, in the case of the control word "*tuba*," the correct response option was *a musical instrument you blow into*. The three foils were:

(1) a musical instrument you hit.

(2) a musical instrument you stand on.

(3) a person who marches in a band.

Finally, response options were matched on a number of parameters akin to the target words themselves. Namely, the number of words in each response option, the average number of letters of each word contained in the sentence and the average Kuperman's AoA for each word were shown to be statistically similar amongst all four response options. The means and standard deviations for these variables are reported in Table 38.

Table 38

Means and standard deviations (M (SD)) for the number of words in each response option, number of letters in each word, and Kuperman's AoA for the control and experimental trials, and correct response options and foils within each trial.

	Correct Response Option(/s)		Foils	
	Control	Experimental	Control	Experimental
Number of words in each response option	6.08 (1.00)	5.42 (1.61)	5.64 (1.10)	5.17 (1.09)
Number of letters in each word	4.14 (2.31)	4.22 (2.41)	3.93 (2.28)	4.08 (2.31)
Kuperman's AoA of each word (in years)	4.51 (1.29)	4.70 (1.59)	4.34 (1.17)	4.40 (1.44)

A series of t-tests found no significant differences of these criteria between the correct response options and foils, or between the experimental and control trials ($p > .05$).

5.2.3 Procedure

5.2.3.1 Language and Cognitive Measures

Language and cognitive measures were the same as those described in Chapter 4 (Experiment 3).

5.2.3.2 The Metaphoric Task

The experiment was created using PsychoPy version 3.0 (Pierce, 2007) and presented to the children on laptops. Trials were presented in a random order, as determined by PsychoPy, to prevent order effects.

Participants were first given instructions to the task. It was explained to the children that the task was about "words that looked the same, sounded the same, but sometimes meant more than one thing". Children were given the example of the word “*chicken*” which was sometimes used to refer to *the clucking farmyard animal*, but also used to denote *a person who is too scared to do something*; the word “*chicken*”, consequently, had two meanings. Researchers stated that the study was interested in the meanings the children knew of the words and, thus, they were required to think carefully about each word. Participants were to be tested on their knowledge of the words in two ways. Firstly, participants were told that they would be presented with a word and asked to define it. They were instructed to respond verbally to the experimenter. Secondly, the participants were informed that they would then see the word alongside four definitions and would be required to select the correct definition/s; sometimes only one of the response options would be the correct definition of the target word, sometimes two of the response options would be correct definitions. Children were instructed to respond using the mouse pad of the laptop. Instructions were presented visually on the screen, one paragraph at a time, lasting 57 seconds in total. After this time, a ‘Let’s Practice’ button was presented allowing participants to proceed to the first of the two practice trials.

Participants were given two practice trials before continuing onto the experimental trials. Each trial comprised of two components: (1) an open-ended question, and (2) a multiple-choice question. In the open-ended component of the trial, the participant was asked to define a word presented on the screen. All experimental trials were presented in the same format as the practice trial of “kettle” (a control word) which is outlined below:

“kettle”

What does this word mean?

Target words were presented in all lower case to aid recognition, and participants were able to respond verbally to the experimenter who would key in their answers. The format of the multiple-choice component was as follows:

"kettle"

Now, what does this word mean from the options below?

Be careful, it could mean more than one of these things!

(1) A person who drinks tea

(2) A device that boils water

(3) A way of cleaning the kitchen

(4) A device that freezes food

The presentation of the correct response options and foils were randomised by PsychoPy. Participants selected the definition/s of the target word using the touchpad of the laptop. Upon selection, response options would turn green. Participants were instructed to press the "Next" button in the bottom right corner of the screen when they believed they had selected all appropriate response options (whether this was after one or two selections). After a brief pause (0.2 seconds), during which the screen displayed no text, the next trial began.

After completing the two practice trials, a second set of short instructions were given. Children were told that the practice trials were over and were given an opportunity to ask any questions before the experimental trials began. These instructions lasted 11 seconds. After this point, the 'Let's Go!' button was presented allowing participants to proceed to the first trial.

In all cases, information was simultaneously presented visually and auditorily to account for differences caused by participants' reading abilities. The audio was recorded by a female, native British-English speaker with a neutral accent using Audacity (version 2.1.2).

The Metaphoric Task took approximately 20 minutes to administer. Accuracy of response and response time (in milliseconds) were recorded automatically by PsychoPy for both the open-ended and multiple-choice components of each trial.

Children were taken from their class and tested individually in a quiet classroom, onsite at school. Although ethical approval from parents had already been obtained, the testing procedure was first explained to the children before the experiment began, and they gave their verbal assent to continue. During the Metaphor and CNRep tasks, participants wore noise-cancelling headphones to listen to the audio and limit environmental distractions. Each child's assessment took a total of approximately 30 minutes to complete.

5.2.3.2.1 Scoring of the Metaphoric Task

5.2.3.2.1.1 Open-Ended Component

The open-ended component of the Metaphor Task was coded in two ways with responses being assigned both a score and a code. Firstly, a child could either receive a score of 0, 1, or 2. A child received a score of 0 if they failed to give the correct definition of the target word, a score of 1 if they gave one of the correct definitions of the target word, and a score of 2 if they were able to give two correct definitions for the target word.

Secondly, these scores were then coded. Scores of 0 were either coded as an '*Incorrect Definition'* (i.e., the definition they gave was not correct) or as '*Omission'*, when they did not give an answer. For correct scores (i.e., those given a score of either 1 or 2), the sense of the definition(s) given was coded for the later analysis of the most popular response. Accordingly, each of the senses of the target word was assigned a number. The literal interpretation of the word was assigned the code '*Sense1*', while the more metaphorical sense of the word was assigned the code '*Sense2'*. For example, a definition of '*lion'* as *a carnivorous member of the cat family*, would be coded as *Sense1*, while a definition of '*lion'* as *a person who possesses strength and courage*, would be coded as *Sense2*. If the child was to give a different, yet correct definition of the word, this was scored as '*Other'* (for example, if a child gave the definition of '*lion'* as *a type of chocolate bar*).

5.2.3.2.1.2 Multiple-Choice Component

Responses to the multiple-choice component of the task were also scored using a 0-2 scale. A child received a score of 0 if they failed to select either of the correct answers; a score of 1 if they identified one of the two correct responses for the target word, or a score of 2 if they were able to identify both correct definitions of the target word.

5.3 Results

Two t-tests were initially performed revealing accuracy rates to be significantly higher in the control condition than the experimental condition in both the open-ended component ($M_{Experimental}$= .13, $SD_{Experimental}$ = .34; $M_{Control}$= .69, $SD_{Control}$= .46; t(2807.19)= -38.53, p< .001) and the multiple-choice component of the task ($M_{Experimental}$= .18, $SD_{Experimental}$ = .38; $M_{Control}$= .67, $SD_{Control}$= .47; t(2954.05)= -31.85, p< .001). When considering only experimental trials, a final t-test revealed participants to perform significantly better on the Multiple-Choice component (M= .18, SD= .38) compared to the Open-Ended component (M= .13, SD= .34) (t(1535)= 4.67, p< .001) of the task.

Descriptive statistics were first conduced regarding the *type* of response participants gave. Firstly, this was considered at a group level (including all participants), then the Year 3 children were compared to the Year 6 children. Finally, the monolinguals were compared to their monolingual peers (using the binary classification of language experience). Prior to all descriptive statistics, Levene's tests for equality of variances were conducted. In cases where this assumption was violated, degrees of freedom were adjusted and a statistic not assuming the homogeneity of variance was reported.

Subsequently, accuracy and latency data were analysed using a series of Multilevel Linear Models, implemented in IMB SPSS Statistics (Version 26) using a Maximum Likelihood method. The dependent variable was either the accuracy or the latency of the response discussed separately below. The models included the predictors of KS1 SAT Reading score, KS1 SAT Writing score, Streamlined-CELF Receptive Language Score (RLS), Expressive Language Score (ELS), Backwards Digit Recall, Total LeBLEQ-T score, and language experience classification (monolingual, bilingual) as fixed factors, as well as an interaction of Year Group and Pupil Premium. The model was then run for the second time. Here, the same predictors were included except language experience classification (monolingual, bilingual) which was replaced with the child's Cumulative Language Input Score generated from the LeBLEQ-C. The Bayesian information criterion (BIC) of the two models was then compared in order to determine the model with the best fit to the data. In all models, random intercepts were included for each participant. Responses to Open-Ended questions and Multiple-Choice questions are described separately below.

5.3.1 Model Comparison

Bayesian information criterion (BIC) values were compared across the models including the binary (monolingual vs bilingual) and continuous (Cumulative Input Score) language experience classifications. BIC values for all models are outlined in Table 39.

Table 39

Bayesian information criterion values for Metaphor models including the binary (monolingual vs bilingual) and continuous (Cumulative Input Score) language experience classifications

		BIC	
		Binary	Continuous
Open-Ended Component			
	Accuracy	1853.99	1789.33
	Reaction Time	416.18	398.25
Multiple-Choice Component			
	Accuracy	955.54	903.98
	Reaction Time	514.96	482.96

Literature considers the model with the lowest BIC preferable (e.g., see Vrieze, 2012). As can be seen in Table 39, the models that include the continuous, Cumulative Language Input Score have lower BIC values, indicating better fitting models. Consequently, only the findings of the continuous models will be discussed below.

5.3.2 Open-Ended Component

5.3.2.1 Response Type Analyses in the Open-Ended Component

First, specific response types were analysed (see Table 40). Responses were coded as ‘0’ (no correct response), ‘1’ (only one correct response), or ‘2’ (two correct responses) depending on how many responses participants produced.

Table 40

Response types across experimental trials of the Open-Ended Component of the Metaphors Task, stratified by age of child and language experience classification

		Year 3		Year 6		
Response code	Response type	Monolingual	Bilingual	Monolingual	Bilingual	Total
0	Total	20.00%	29.58%	9.52%	12.80%	18.96%
	Incorrect	14.33%	18.13%	7.14%	10.37%	13.06%
	Omission	5.67%	11.46%	2.38%	2.44%	5.91%
1	Total	69.33%	65.21%	71.03%	68.50%	68.04%
	Sense1	67.33%	62.08%	70.24%	67.07%	66.08%
	Sense2	2.00%	2.92%	0.79%	1.42%	1.90%
	Other	0.00%	0.21%	0.00%	0.00%	0.07%
2	Total	10.67%	5.21%	19.44%	18.70%	12.99%
	Sense1 and Sense 2	6.33%	2.50%	13.89%	12.20%	8.27%

Sense1 and other	4.33%	2.71%	5.56%	6.50%	4.72%
Sense2 and other	0.00%	0.00%	0.00%	0.00%	0.00%

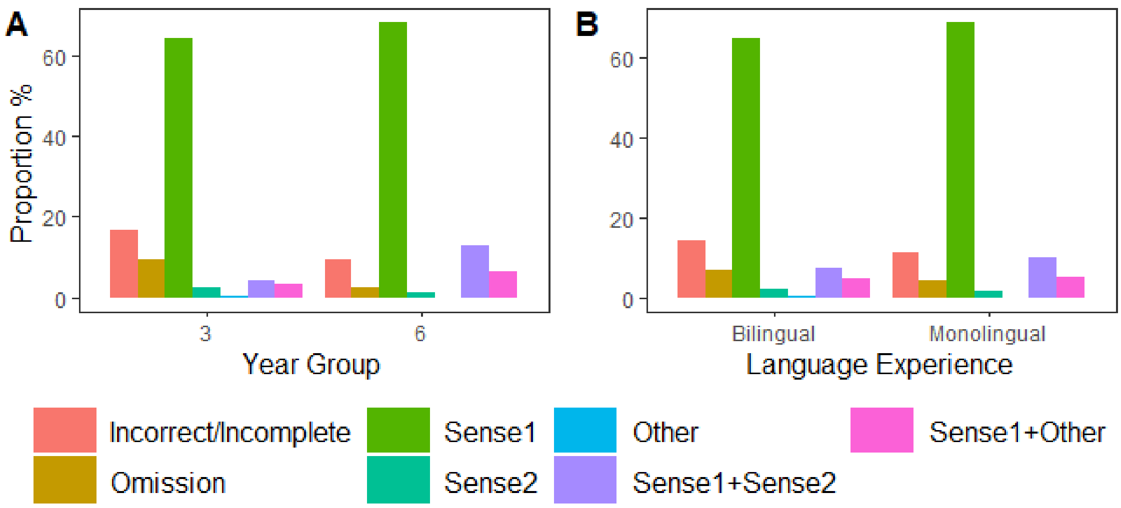
Overall, across all participants, the most common response was reporting only one of the senses of the metaphoric word (M= 68.04, SD= 16.12) rather than reporting two senses correctly (M= 12.99, SD= 13.88), or none at all (M= 18.96, SD= 16.08) [F(2, 380)= 488.96, $p < .001$). If only one response was given, participants were significantly more likely to state the first, more literal sense of the word (M= 66.08, SD= 15.86) compared to a common figurative sense (M= 1.90, SD= 3.51) or a less common figurative sense (M= .07, SD= .74), [F(2, 252)= 2019.08, $p < .001$]. When participants scored a ‘0’, this was more likely to be due to an incorrect or incomplete definition of the word (M= 13.06, SD= 13.10), rather than due to not responding at all (M= 5.91, SD= 8.47; t(126)= 5.34, $p < .001$). Finally, in the instances where participants scored ‘2’, this was significantly more likely to be due to giving the literal sense and a common figurative interpretation (M= 8.27, SD= 10.73) as opposed to a literal sense and a less common figurative sense (M= 4.72, SD= 6.77; $t(126)= 3.51$, $p < .01$). None of the children gave two figurative senses.

Children in Year 3 received significantly fewer ‘2’ scores on the open-ended component of the Metaphors task (M= 7.31, SD= 9.02) in comparison to the children in Year 6 (M=18.95, SD= 15.55; $t(96.95)= -5.13$, $p < .001$) and instead received proportionally more scores of ‘0’ ($M_{Year\text{-}3}$= 25.90, $SD_{Year\text{-}3}$= 17.12; $M_{Year\text{-}6}$= 11.69, $SD_{Year\text{-}6}$= 11.03; $t(109.97)= 5.58$, $p < .001$). Upon further inspection, this was due to Year 3 children having both significantly higher proportions of incorrect/incomplete answers (M= 16.67, SD= 14.51) and omissions (M= 9.23, SD= 9.10) than the children in Year 6 ($M_{Incorrect/Incomplete}$= 9.27, $SD_{Incorrect/Incomplete}$= 10.25; $t_{Incorrect/Incomplete}(115.34)= 3.33$, $p < .01$; $M_{Omission}$= 2.42, $SD_{Omission}$= 6.11; $t_{Omission}$ $(112.45)= 4.98$, $p < .001$). The younger children also gave significantly lower proportions of both Sense1/Sense2 answers (M= 3.97, SD= 6.61) and Sense1/other answers (M= 3.33, SD= 5.27) than the Year 6 children ($M_{Sense1/Sense2}$= 12.77, $SD_{Sense1/Sense2}$= 12.33; $t_{Sense1/Sense2}(92.40)= -4.98$, $p < .001$;

$M_{Sense1/Other}$= 6.18, $SD_{Sense1/Other}$= 7.81; $t_{Sense1/Other}$ (106.20)= -2.39, p< .05). Although there was an overall non-significant difference of giving only one correct definition of the target word (i.e., a score of '1') across the Year 3 and Year 6 pupils, the younger, Year 3 children gave a higher proportion of second sense, more figurative definitions (M= 2.56, SD= 3.88) compared to the Year 6 children (M= 1.21, SD= 2.96; t(119.36)= 2.22, p< .05). No significant difference was found between Year 3 and Year 6 in the proportions of times they reported the first, more literal sense of the word (M_{Year-3}= 64.10, SD_{Year-3}= 15.30; M_{Year-6}= 68.15, SD_{Year-6}= 16.29 respectively) and another, less frequent, figurative sense of the word (which was only reported once by a Year 3 child) (see Figure 11A).

Figure 11

Types of responses given in the Open-Ended component of the Metaphors Task, stratified by year (A) and language experience (B)



Finally, monolingual and bilingual children were equally likely to receive a score of '2' ($M_{Monolingual}$= 14.67, $SD_{Monolingual}$= 13.96; $M_{Bilingual}$= 12.04, $SD_{Bilingual}$= 13.82), and equally likely to receive a score of '1' ($M_{Monolingual}$= 70.11, $SD_{Monolingual}$= 14.86; $M_{Bilingual}$=66.87, $SD_{Bilingual}$= 16.77). However, bilinguals were more likely to receive a score of '0' than monolinguals ($M_{Monolingual}$= 15.22, $SD_{Monolingual}$= 14.09; $M_{Bilingual}$= 21.09, $SD_{Bilingual}$= 16.83; t(125)= -2.00, p< .05). There were no other significant differences between monolinguals and bilinguals (see Figure 11B).

5.3.2.2 Analysis of Accuracy Rates in the Open-Ended Component

When including individuals Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the open-ended trials, the relationship between scores on the metaphors task and predictors showed significant variance in intercepts across participants, Var(u_{0j})= .014, $\chi^2(1)$= 10.59, *p*< .01. Here, there was a significant effect of RLS [F(1, 88.99)= 6.43, *p*< .05] and an interaction between School Year and Pupil Premium status [F(3, 89.77)= 6.25, *p*< .01]. Contributions of all other fixed factors can be seen in Table 41. Conducting separate multilevel models on those who were in receipt of Pupil Premium and those who were not (as above), revealed that, for those who received Pupil Premium, there was no significant effect of year [F(1, 33)= 2.78, *p*= .105, *b*= -.14, *SE*= .08, *t*= -1.67], indicating that performance of children in Year 3 did not differ from performance of children in Year 6. There was, however, a significant effect of year in the children who were not in receipt of Pupil Premium [F(1, 56.41)= 16.74, *p*< .01, *b*= -.32, *SE*= .03, *t*= -4.09] with children in Year 6 showing more accurate performance than children in Year 3.This demonstrates that children in receipt of Pupil Premium do not advance in metaphorical skill between Year 3 and Year 6 as their peers without the subsidiary do.

5.3.2.3 Analysis of Reaction Times in the Open-Ended Component

Reaction time analyses excluded both errors and outliers (87.01% of trials Following recent literature investigating child language (e.g., Commissaire et al., 2019; Grundy & Keyvani Chahi, 2017; Levy & Hanulíková, 2019; Persici et al., 2019) responses that were 2.5 standard deviations above/below each participant’s mean were considered deviations. Errors amounted to a substantial 87.01% of trials (no outliers were detected) being excluded from RT analyses, indicating that the children performed considerably poorly on the current task. This was despite the task being specifically designed for this age group based on the AoA ratings obtained in Experiment 2 (Chapter 3). The remaining RTs were z-scored to further minimise the impact of potential outliers.

The relationship between scores on the metaphors task and predictors showed significant variance in intercepts across participants, Var(u_{0j})= .036, $\chi^2(10)$= 195.45, *p*< .01. Here, there were significant effects of ELS [F(1, 53.03)= 9.97, *p*<

.01), and RLS [$F(1, 68.53) = 12.29$, $p < .01$], demonstrating that response times on this task were influenced by both receptive and expressive skill. However, it must be noted that the direction of these effects were different. While an increased ability in expressive language was associated with a shorter response time, an increased ability in receptive language skill was linked with a child taking longer to respond to the trial. Contributions of all other fixed factors were non-significant and can be seen in Table 41.

Table 41

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Open-Ended component of the Metaphors Task when using a Continuous Measure of Language Experience

	Accuracy				Reaction Time			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	.58	.18	3.28	.002**	-.16	.85	-.18	.856
Cumulative Language Input Score	-.08	.08	-1.02	.311	-.44	.37	-1.16	.249
Gender (Male)	.05	.04	1.21	.230	-.32	.18	-1.81	.079
KS1 Reading SATs (below average)	.07	.07	1.06	.294	.18	.29	.60	.548
KS1 Writing SATs (below average)	-.06	.06	-.99	.327	-.21	.28	-.75	.456
Receptive Language Score	.01	.00	2.54	.013**	.03	.01	3.51	.001**
Expressive language Score	.00	.00	.97	.337	-.02	.01	-3.16	.003**
Working Memory Score	-.03	.02	-1.27	.208	-.14	.10	-1.37	.180
LeBLEQ-T	-.00	.00	-.25	.802	.01	.00	1.39	.170
Year*Pupil Premium (Year 3, without Pupil Premium)	-	-	-	-	.10	.30	.33	.742

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

5.3.3 Multiple-Choice Component

5.3.3.1 Response Type Analyses in the Multiple-Choice Component

Specific response types were first analysed in subsequent analyses and can be seen in Table 42. As with the Open-Ended component, responses were coded as ‘0’ (no response), ‘1’ (only one response), or ‘2’ (responses) depending on how many responses participants gave. However, in the Multiple-Choice component, participants could only either obtain a code which signified that they (1) selected an incorrect response, (2) selected only Sense1, (3) selected only Sense2, or (4) select both correct senses.

Table 42

Response types across experimental trials of the Multiple-Choice Component of the Metaphors Task, stratified by age of child and language experience classification

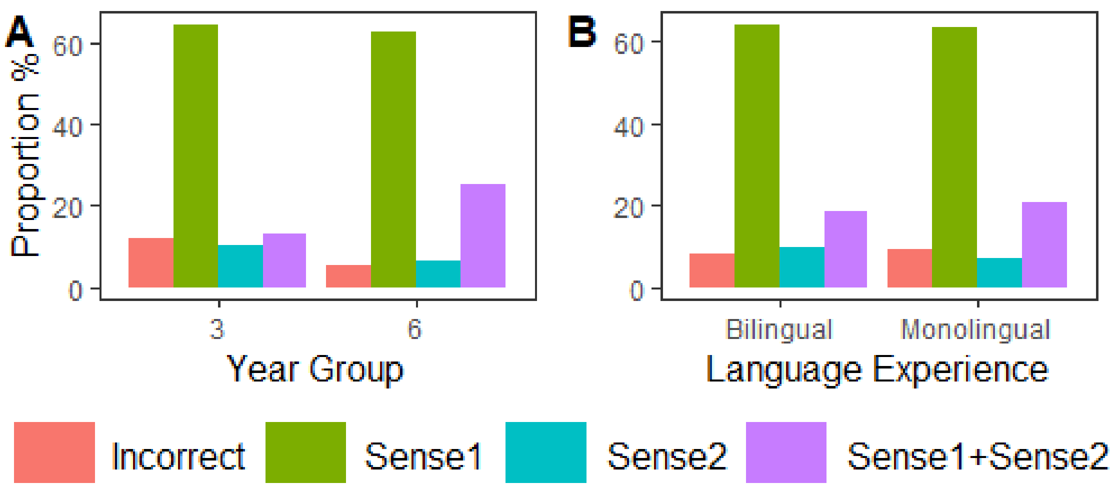
		Year 3		Year 6		Total
		Monolingual	Bilingual	Monolingual	Bilingual	
0	Incorrect	13.00%	11.04%	4.76%	5.49%	8.60%
1	Total	71.33%	77.71%	69.05%	69.51%	72.38%
	Sense1	63.00%	65.83%	63.89%	62.20%	63.78%
	Sense2	8.33%	11.88%	5.16%	7.32%	8.60%
2	Sense1 and Sense 2	15.67%	11.25%	26.19%	25.00%	19.03%

When looking across all participants, the most common response in the Multiple-Choice component of the Metaphors Task was reporting only one of the senses of the metaphoric word (M= 72.38, SD= 15.74) rather than reporting two senses correctly (M= 19.03, SD= 17.41), or none at all (M= 8.60, SD= 9.39), [F(2,380)= 698.06, *p*< .001]. If only one response was given, participants were more likely to give the first, more literal sense of the word (M= 63.78, SD= 15.55) than the second, more figurative sense (M= 8.60, SD= 7.78), (t(126)= 32.91, *p*< .001).

As can be seen in Figure 12A, children in Year 3 received significantly fewer ‘2’ scores on the Multiple-Choice component of the Metaphors task (M= 12.95, SD= 12.76) in comparison to children in Year 6 (M= 25.40, SD= 19.35; *t*(104.90)= -4.26, *p*< .001) and instead received proportionally more scores of ‘0’ (M_{Year-3}= 11.79, SD_{Year-3}= 10.71; M_{Year-6}= 5.24, SD_{Year-6}= 6.26; *t*(104.07)= 4.23, *p*< .001), and of ‘1’ (M_{Year-3}= 75.26, SD_{Year-3}= 13.34; M_{Year-6}= 69.35, SD_{Year-6}= 17.52; *t*(113.91)= 2.13, *p*< .05). Upon further inspection, this was due to Year 3 children giving significantly higher proportions of Sense2-Only answers (M= 10.51 SD= 7.69) than children in Year 6 (M= 6.59, SD= 7.41; *t*(125)= 2.93, *p*< .01). Children in Year 3 and Year 6 were equally as likely to give Sense1-Only responses (M_{Year-3}= 64.74, SD_{Year-3}= 14.27; M_{Year-6}= 62.77, SD_{Year-6}= 16.85).

Figure 12

Types of responses given in the Multiple-Choice component of the Metaphors Task, stratified by year (A) and language experience (B)



There were no significant differences between monolinguals and bilinguals in terms of types of responses given (see Figure 12B).

5.3.3.2 Analysis of Accuracy Rates in the Multiple-Choice Component

When including participants' Cumulative Language Input Score (generated from the LeBLEQ-C) as a fixed factor in the model, the relationship between scores on the metaphors task and predictors showed significant variance in intercepts across participants, $Var(u_{0j}) = .009$, $\chi^2(1) = 21.54$, $p < .01$. Although RLS approached significance ($p = .064$), only the interaction between School Year and Pupil Premium status was a significant predictor of performance on the Metaphoric Task [$F(3,89.88) = 3.34$, $p < .05$]. Contributions of all other fixed factors were non-significant and can be seen in Table 43.

When this interaction was explored further, analyses showed that, for those who received Pupil Premium, there was no significant contribution of School Year [$F(1,33) = 2.57$, $p = .119$, $b = -.10$, $SE = .06$, $t = -1.60$], indicating that performance of children in Year 3 did not differ from performance of children in Year 6. There was, however, a significant effect of School Year in the children who were not in receipt of Pupil Premium [$F(1,56.33) = 9.79$, $p < .01$, $b = -.17$, $SE = .05$, $t = -3.13$] with children in Year 6 showing more accurate performance than children in Year 3. Thus, mirroring the findings for the open-ended component of the task, children in receipt of Pupil Premium do not advance in metaphorical skill between Year 3 and Year 6 as their peers without the subsidiary do.

5.3.3.3 Analysis of Reaction Times in the Multiple-Choice Component

Reaction times (RTs) were again z-scored with the analyses excluding errors and outliers (i.e., responses that were 2.5 standard deviations above/below a participant's mean). This amounted to a substantial82.09% of trials being excluded from RT analyses, indicating the children, overall, performed considerably poorly on the current task. This was despite the task being specifically designed for this age group based on the AoA ratings obtained in Experiment 2 (Chapter 3).

Analyses showed significant variance in intercepts across participants, ($Var(u_{0j}) = .453$, $\chi^2(1) = 49.59$, $p < .01$). The only factor that predicted reaction times on the

Metaphors Task significantly was the child Total LeBLEQ-T Score [$F(1, 62.64) = 15.89$, $p < .001$]. The positive *b* -value indicated that the lower the score a teacher gave a pupil across all subsets of the LeBLEQ-T (which indicated better language skills), the quicker children were able to respond on the metaphors task. Contributions of all other fixed factors can be seen in Table 43. At no point did academic attainment, working memory or language experience make a significant contribution to either the accuracy or reaction time models.

Table 43

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Multiple-Choice component of the Metaphors Task when using a Continuous Measure of Language Experience

	Accuracy				*Reaction Time*			
	B	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	.08	.13	.59	.558	-1.09	.95	-1.15	.256
Cumulative Language Input Score	-.08	.06	-1.33	.188	.10	.47	.204	.839
Gender (Male)	-.01	.03	-.27	.788	-.09	.21	-.43	.666
KS1 Reading SATs (below average)	.08	.05	1.65	.102	-.18	.33	.55	.584
KS1 Writing SATs (below average)	-.08	.05	-1.83	.071	.34	.31	1.12	.267
Receptive Language Score	.00	.00	1.88	.064	.00	.01	.09	.932
Expressive language Score	.00	.00	.03	.977	.00	.01	.67	.509
Working Memory Score	-.00	.02	-.18	.857	.08	.12	.72	.478
LeBLEQ-T	.00	.00	.42	.674	.02	.01	3.99	.000**
Year*Pupil Premium (Year 3, without Pupil Premium)	-	-	-	-	.38	.36	1.06	.295

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

5.4 Discussion

The present research aimed to investigate the lexical ambiguity skills of primary school-aged children in order to understand the impact of bilingualism on the comprehension of metaphors and whether competence could be better predicted using a continuous score of bilingualism rather than a binary classification (monolingual vs bilingual). A novel metaphor task was administered alongside measures of general linguistic ability and working memory (Streamlined-CELF) to Year 3 (7-8 year-old) and Year 6 (10-11 year-old) pupils in schools of relatively low socioeconomic status associated with the Born in Bradford cohort. The Leeds-Bradford Language Exposure Questionnaire for Children was administered to record levels of bilingualism of the pupils on a continuous scale (rather than simply using the binary classification monolingual/bilingual). The Leeds-Bradford Language Exposure Questionnaire for Teachers was also administered to obtain an indication of teachers' perceptions of the child's linguistic behaviour in the classroom. Scores on KS1 Reading and Writing SATs were also obtained from the schools, as was information regarding Pupil Premium (which was taken as an indicator of disadvantage).

Primarily, it must be noted that the accuracy rates for both of these response formats were considerably low; 12.99% for the Open-Ended component and 17.91% for the Multiple-Choice component. A similar effect was observed in the Homonymy task (Chapter 4, Experiment 3); this remains to be surprising considering the stimuli were specifically designed for use with the current age groups, as determined by the age of acquisition ratings in Experiment 2 (Chapter 3). It is once again suggested that this effect is due to the overall low SES of the city of Bradford as a whole Index (IDACI; Department for Communities and Local Government, 2015), but could be further attenuated by the possible discrepancy between the SES of the children tested in the current study, and the general population sampled in the AoA norming experiment (see the discussion of Chapter 4 for a full explanation of this).

Accuracy rates across the two component-types demonstrated that children found the Open-Ended component more difficult than the Multiple-choice component of the Metaphors Task. This is in line both with predictions and previous literature showing that children find explaining metaphors harder than answering multiple-choice questions about them (Arcara et al., 2019; Perlini et al., 2018; Pouscoulous, 2014). These results highlight the importance of response format when studying ambiguities.

Finally, Linear Mixed Effects Models demonstrated various contributors to metaphoric understanding. Although older children outperformed younger children on this novel task, different effects were seen between high- and low-SES groups with those from poorer socio-economic strata being adversely impacted. Though there was no effect of language experience when either binary or continuous measures were taken, a model including the continuous measure of the Cumulative Language Input Score was shown to fit the data best. Binary language classification, however, was shown to affect the types of response children gave in the Open-Ended component of the task. Also in accordance with predictions was the finding that a child's response time was significantly predicted by both their Receptive Language Score and their Expressive Language Score. While an increased ability in expressive language was associated with a shorter response time (though only in the Open-Ended component), an increased ability in receptive language skill was linked with both an increased accuracy rate but, interestingly, also an increased reaction time. Children who demonstrated higher teacher-rated linguistic proficiency in the classroom environment were also better at identifying two senses of a metaphor (in the Multiple-Choice component). All these factors are discussed separately below.

5.4.1 SES and Age of child/School Year

In both the Open-Ended and Multiple-Choice components of the task, there were no effects of School Year in children who received Pupil Premium. However, for pupils who did not receive Pupil Premium (and are thus considered to be higher SES even within the disadvantaged areas where testing took place), Year 6 demonstrated greater metaphorical understanding than their younger peers. Such findings suggest age to be a clear predictor of metaphoric awareness in children but this to be highly moderated by SES, with those from higher socio-economic positions outperforming those from lower strata with increasing age. These findings oppose original predictions made for the current study which expected children in receipt of Pupil Premium to be outperformed by those without in Year 3, but this effect to be eliminated by Year 6 (as per the findings of

Hutchinson et al., 2016). However, the current study found the differences between children who received Pupil Premium and those who did not to be mostly apparent in Year 6. It is possible that this effect was observed due to children from lower SES being required to direct their attentional resources on the development of other, more widespread skills needed that are vital to enable them to progress within the classroom environment. Although focussing on such skills may enable children to reach a level of attainment on par with their higher SES peers in skills in general language use and performance on standardised academic examinations (Hoff & Tian, 2005), it may divert attention from the development of their higher order, metaphorical skills which become hindered in relation to their higher SES peers.

Perhaps most importantly, however, such a trend highlights the fact that children in Year 6 are not reaching the goals set out by the Department for Education. The learning objectives outlined by the government stipulate that by Year 6, children should have developed a more sophisticated understanding of figurative language (Department for Education, 2013a). However, the children in the present sample fail to demonstrate this. Though this is apparent for all pupils, it is more prominent for the children receiving Pupil Premium, i.e., those from the lower SES groups. Thus, interventions which aim to improve the abilities of metaphorical competence in primary-school-aged children should be predominantly focussed on children from lower socioeconomic strata who need more support than their higher-SES peers. These are the children who may benefit the most if such interventions are implemented.

Furthermore, not only did age significantly predict accuracy and reaction time of identifying more than one sense of a metaphoric word, but age also affected the type of responses provided by the children. Younger children were not only less likely to provide a correct answer but were also significantly more likely to provide either an incorrect answer or fail to respond to the question at all. However, they were equally as likely as older children to only be able to provide one definition of the target word. These findings demonstrate a clear shift in metaphorical understanding of children between the ages of 7 and 11 years-old where older children are more likely to know at least one sense of a metaphor in comparison to those who are younger. Such a shift between these ages is likely due to the introduction of figurative language in schooling. The national curriculum for England suggests children should demonstrate some form of understanding of figurative language from Year 3, but that more sophisticated understanding should become apparent by Year 6 (Department for Education, 2013a). Consequently, the increases in the number of single-definition responses and both-definition responses in the older years is likely due to the increase of metaphoric encounters that take place in the school environments of the older children. More encounters of such senses increase the likelihood that the meanings will be acquired into an individual's vocabulary.

5.4.2 Language Experience

Contrary to predictions, a child's level of exposure to English, as measured by either binary or continuous language classification, did not predict their performance on any component of the Metaphoric Task. Despite this, language experience was found to be a significant contributor to the *types* of responses a child gave. To this effect, bilingual children were significantly more likely than monolingual children to give an incorrect definition of the metaphorical word or be unable to give any definition at all in the Open-Ended component of the task. Bilinguals also gave fewer correct responses with two definitions and responses with only the one correct response (though this effect was not significant). These findings are indicative of differences in processing between these two groups of children and the slight advantage monolinguals have over bilinguals. The present results are in line with previous findings by Hessel and Murphy (2019) and Segal and Gollan (2018) who also demonstrated monolinguals and bilinguals to perform differently on tasks of metaphoric competence. Such differences between monolinguals and bilinguals are attributed to the increased exposure monolinguals receive in English, which has also been shown to impact many other aspects of language development, such as receptive and expressive vocabulary in bilingual children (e.g., Hoff et al., 2012; Paradis, 2011; Thordardottir, 2011; Unsworth, 2013, 2017).

Exposure to language is known to be a substantial contributor to the rate of monolingual and bilingual children's language growth (Hart & Risley, 1995; Hoff, 2003; Pearson et al., 1997) with the rate of vocabulary learning shown to be proportional to exposure (Chondrogianni & Marinis, 2011; De Houwer, 2009; De Houwer, 2007; Gathercole et al., 2013; Hoff et al., 2012; Thordardottir, 2011; Thordardottir et al., 2006). This is attributed to bilingual children experiencing less of each of their languages in comparison to monolingual children of the same age due to being required to divide their time between each of their languages. The present findings contribute to this body of literature by demonstrating that the same applies to metaphoric words; bilingual children's reduced exposure to English negatively impacts knowledge of the disparate senses of metaphoric words.

Most interesting, however, is the finding that the models shown to best fit the data were those that included the continuous measures of language experience; the Cumulative Input Score from the Leeds-Bradford Language Experience Questionnaire for Children (LeBLEQ-C). Though there was no significant contribution of either the binary or continuous measures of language experience, the continuous model had a lower Bayesian Information Criterion (BIC) value, indicating that language exposure is indeed likely to contribute to knowledge of metaphors, but a binary classification system may not be sensitive enough to capture such differences. This is in line with literature suggesting bilingualism to be better suited to being measured along a continuum, rather than simply considered to be a binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020).This could explain why previous studies that used binary classifications (e.g., Johnson & Rosano, 1993; Trosbrog, 1985) did not find significant contributions of binary language experience classification to metaphoric knowledge. Nevertheless, it must be noted that although more continuous measures of bilingualism may be preferable, this distinction remained non-significant in all models and thus findings must be interpreted with caution.

It must be noted that both Segal and Gollan (2018) and Hessel and Murphy (2019) observed clear differences in accuracy rates, which were not seen in the current study. It is possible, however, that such effects were observed due to the ages of the participants and the types of tasks in these studies. In particular, Segal and Gollan (2018) in a divided visual field study investigated the processing of metaphors in adult populations using a divided visual field study, not the explicit knowledge of metaphorical interpretations like the present study. On the other hand, Hessel and Murphy (2019) studied children in Year 1 and Year 2 of primary education. Thus, it is possible that many of the bilingual children recruited in their research had less English exposure than the children recruited in the current study. In the present study, even if children began school with limited English language ability, by Year 3, children would have had over three years of exposure to English, while the Year 6 children would have had over six years. In other words, the number of years of schooling and subsequent exposure to English might explain the differences observed.

5.4.3 Expressive Language

As predicted, children who scored higher on the expressive language tasks of the Streamlined-CELF were quicker in correctly identifying both meanings of a metaphor in the Open-Ended component of the task. Furthermore, a child's expressive language skill was not a significant predictor to any other aspects of metaphoric competence. This effect is attributed to the specific demands of the open-ended component of the Metaphorical Task. Obtaining a correct score required children to articulate sufficiently the definition of a word, which is highly reliant on expressive language ability. Consequently, those who had higher communicative abilities were able to verbalise their knowledge of the metaphors more easily and quickly than children who had lower expressive abilities.

Additionally, it could be argued that this added element of expressive language ability may have contributed to the fact that participants found the Open-Ended component of the Metaphoric Task to be harder than the Multiple-Choice component. This was a purposeful element of the study design, with the Open-Ended component of the task intended to capture different, more verbal skills, but also to ensure the task can be used across a wide age range. To make this possible, the Open-Ended component was added to allow for rich, detailed data to be gathered whilst limiting the task's susceptibility to ceiling effects. This design was in line with research showing increased difficulties for verbal-explanation tasks compared to multiple-choice tasks when assessing figurative language understanding (Arcara et al., 2019; Perlini et al., 2018; Pouscoulous, 2014). For example, Arcara et al. (2019) found adults with brain trauma had significantly higher accuracy rates on multiple-choice tasks compared to tasks that required verbal explanation, highlighting the importance of response format when studying metaphoric competence. The present study is the first, to the author's knowledge to extend these effects to children, highlighting once again that the choice of task plays an important role in the observed performance of participants.

5.4.4 Receptive Language

The ability of a child's Receptive Language Score to predict accuracy rates and response times on the Open-Ended component of the task and their accuracy on the Multiple-Choice component of the novel Metaphors Task is consistent with the view that receptive language ability is related to metaphorical competence (Deckert et al., 2019; Hessel & Murphy, 2019; Norbury, 2005; Rundblad & Annaz, 2010b). In particular, previous research showed that individuals who had high receptive language scores also had high metaphorical skills. This was shown to be the case for typically developing children at various ages (Deckert et al., 2019; Rundblad & Annaz, 2010), bilingual children (Hessel & Murphy, 2019) and children with ASD (Norbury, 2005). The present study is the first one, to the author's knowledge, to use receptive language directly as a predictor of knowledge of metaphorical words in bilingual populations. Though Hessel and Murphy (2019) demonstrated a link between receptive language and metaphoric comprehension, this relationship was associative only and was based on the use of metaphorical expressions which resemble idioms (e.g., "to be on cloud nine"), rather than metaphorical words. The present study built upon and further extended such past findings by explicitly using robust receptive language scores as a predictor in the models for both accuracy and reaction time of metaphorical understanding.

Finally, past literature employed the use of the British Picture Vocabulary Scale II (BPVS; Dunn et al., 1997) to capture receptive language ability. Despite the BPVS being used in all three English studies described above (Hessel & Murphy, 2018; Norbury, 2005; Rundblad & Annaz, 2010a), it is considered to be a rather simplistic measure of linguistic understanding. The BPVS requires participants to solely point to an image that best represents a statement heard; thus, it only provides information about whether a child can understand a word, in isolation, and select it from a small series of illustrations in front of them. The current study, however, used a much more robust measure of receptive language ability, namely three tasks of the Streamlined-CELF. The Receptive Language Score (RLS) of the Streamlined-CELF (described in Chapter 2) is a composite score encompassing three measures of receptive language skills: Concepts and Following Directions, Word Classes – Receptive, and Sentence Structure. The RLS of the Streamlined-CELF, thus, provides a more comprehensive view of a child's receptive language ability. It does not only assess the children's receptive vocabulary like the BPVS, but crucially it provides valuable information about the ability of the children to understand, and perform, ordered instructions, to group similar items together, and to understand full sentences. Such tasks are more akin to those used in daily life. As such, receptive scores gathered from the Streamlined-CELF are, thus, more robust and indicative of a child's understanding of daily conversation.

However, it must be noted that, although the findings are in the predicted direction for the accuracy rates (as discussed above) the same cannot be said for the relationship between children's receptive language ability and their reaction time on a task of metaphorical skill. As such, children who had higher receptive language skill were slower in the Open-Ended component of the metaphoric task than those with lower scores. However, this could be due to the method of employed to measure reaction time. In the current study, reaction time was measured from the onset of the trial (i.e., the moment the metaphor was displayed on the screen and the audio commenced) to the time the participant had finished their response. Consequently, it may not be that the children with higher receptive language skill were slower in deliberating the meanings of the metaphoric words, but rather they gave longer, more detailed responses, due to better understanding of the word, which affected their reaction times. Retrospectively, more informative data may have been gathered if the time in which a participant began their answer was recorded, as well as their rate of speech. This would have given a more reliable indicator of reaction time and an individual's knowledge of a metaphoric word.

5.4.5 Leeds-Bradford Language Experience Questionnaire for Teachers

The Total Score of the LeBLEQ-T was shown to significantly predict response times on the Multiple-Choice questions when considering the continuous Cumulative Input Score of the child. The LeBLEQ-T quantifies a teacher's perception of a child's linguistic ability taking into consideration aspects other than expressive and receptive language. In particular, the total score is comprised of listening, speaking, reading and writing components as well as more pragmatic abilities, providing a more comprehensive indication of a child's communicative competence. Thus, though still capturing language proficiency, the LeBLEQ-T is conceptually different from the measures gathered by the Streamlined-CELF and as such it is not surprising that its contributions to metaphoric competence differ.

The present findings are in line with Johnson and Rosano (1993) who also found no significant differences between monolingual and bilingual populations (i.e., the binary classification of language experience) but a significant positive relationship between metaphorical competence and communicative proficiency. Johnson and Rosano (1993) employed the Language Assessment Scales (De Avila & Duncan, 1983) which, like the LeBLEQ-T, assesses English communicative competence in relation to pronunciation, comprehension, vocabulary, syntax, pragmatics and general communicative competence. Interestingly, although this study was conducted with an adult population, the Language Assessment Scales are also rating-scales for teachers to assess their students, indicating that teachers' rating of their students skills are good predictors of metaphoric language skill, regardless of student age. Consequently, not only does this significant relationship between metaphoric ability and general language skill advance the knowledge of the links between oral and pragmatic language ability, but it also serves to aid in validating the novel measure of the LeBLEQ-T by producing results similar to those previously observed in the literature.

5.4.6 Conclusion

In conclusion, the current study found receptive language skills to be a substantial predictor of metaphorical understanding in children, both in terms of accuracy rates and processing speeds. This effect was apparent regardless of the child's age, language experience, working memory abilities, socio-economic status or mode of response. In addition, when children were required to verbalise their knowledge of a word, expressive language ability also became a significant contributor. Though language experience was shown to have no significant impact on accuracy rates of the Metaphoric Task, the quality/type of answer a child gave was significantly impacted by their language experience classification (i.e., if they were classed as monolingual or bilingual). Models including a continuous measure of language experience were also shown to fit the data better than models only including a binary language classification, suggesting that metaphoric ability might be associated with language experience to some extent, but not to the degree of other, more influential factors. In fact, the most prominent predictor was shown to be the age of an individual (and their School Year), but this relationship was heavily moderated by socio-economic status.

It must be noted, however, that despite only including in the task metaphors for which both meanings were reported to have been learnt before the age of 10 years-old (as determined by the Norming Study, Experiment 2, Chapter 3), there were still a substantial number of instances where children were unable to define both senses. As the ability to correctly define two senses has been shown to increase with age, this poses the question of whether such metaphorical knowledge would significantly increase in secondary school or adult populations and whether a continuous measure of language experience might still be more informative to a binary classification. The next chapter explores another trope of ambiguity, metonymy, and the extent it is impacted by the same variables as homonymy and metaphor.

6 The Effect of Language Experience on the Comprehension of Metonyms in Primary-School aged Children

Metonymy is considered of particular relevance to communication (Barcelona, 2003; Benczes et al., 2011). The process of metonymy is both linguistic and cognitive whereby individuals use a simple concept to refer to a more complex, more abstract idea (Littlemore, 2015). Like metaphors, metonymy can be used for various functions, including illocution, cohesion, euphemism, humour, and persuasion (Littlemore & Tagg, 2018). The disparate interpretations of metonyms also come in various forms (Copestake & Briscoe, 1995; Rabagliati et al., 2011; Srinivasan et al., 2019), including animal for meat (e.g., “horse”), place for people (e.g., “Downing Street”), instrument for action (e.g., “screw”), and producer for product alternations (e.g., “Scorsese”), with a body of research finding these shifts in interpretation to be relatively easy to comprehend (Clark & Gerrig, 1983; Frisson & Pickering, 2007; McElree et al., 2006; Murphy, 2006). As such, skills disambiguating metonyms are considered to be acquired relatively early in childhood with understanding of metonymy beginning to develop around the age of 3 years-old (Van Herwegen et al., 2013) and skills being associated with the ability to partake in daily conversations (Blasko, 1999; Shelestiuk, 2005) and organise new ideas (Glucksberg, 2001; Ortony, 1975).

In opposition to the meaning/sense of both homonyms and metaphors (discussed in Chapters 4 and 5 respectively), both the primary and secondary senses of metonyms are highly related and literal. In the example of the metonym “horse,” the second sense (i.e., the meat) is thought to have developed from the first (i.e., the animal). This is thought to occur via a process of regular and predictable rules that have been observed across numerous languages, referred to as sense extension (Clark & Clark, 1979; Klepousniotou, 2002; Srinivasan & Rabagliati, 2015).

Sense extension is believed to help simplify the process of word learning, providing children with clues about new senses of words already in their vocabulary (Smith et al., 2002; Srinivasan & Rabagliati, 2015). A body of research agrees that the knowledge of one sense of a polysemous word is likely to facilitate the acquisition of other polysemous senses (Rabagliati & Srinivasan, 2018; Srinivasan et al., 2017). For instance, when a child encounters a word for an animal that is used in a context denoting food, they may reason that the word must refer to the meat that is derived from that animal, facilitating their interpretation of second senses of metonymic words such as “chicken”, “rabbit” or “horse” to mean a type of food. Consistent with this view, children have demonstrated an ability to establish a reference for a new sense of a metonymic word (Rabagliati et al., 2010; Srinivasan et al., 2017; Srinivasan & Snedeker, 2011, 2014). For example, upon learning a novel word for an animal, 4-5-year-old children are able to attribute that same word to refer to the meat derived from the same animal reasonably effortlessly (Srinivasan & Snedeker, 2014). Likewise, children have demonstrated their ability to easily shift between the representational objects and their abstract content (Srinivasan & Snedeker, 2011), to use container for contents (Rabagliati et al., 2010), and instrument for action (Srinivasan et al., 2017).

The ability to use sense extension in such a way has also been found to be true in the case of learning artificial senses of words. For example, Rodd et al. (2012) found adults were able to learn new senses of truly unambiguous words (e.g., “sip” to mean “a small amount of computer data”). Regardless of whether participants received a short reading task or an intensive, four-day training period, participants’ lexical decision making was shorter to trained words in comparison to untrained words. Rodd et al. (2012) concluded that the new senses of the existing words had been sufficiently consolidated, even in such a task that did not require participants to access semantic knowledge. Maciejewski et al. (2020) found similar results, further suggesting that such acquisition can affect the processing of the original word, slowing comprehension, and thus mirroring the ambiguity disadvantage effect observed in studies using existing ambiguous words (e.g., Duffy et al., 1988; Hoffman & Woollams, 2015).

A similar effect or sense extension has also been found in children. Srinivasan et al. (2019) found 3-4-year-old children could extend a new object label to objects of the same shape, following the well-known heuristic of the shape-bias (Diesendruck et al., 2003). In fact, when presented with both a solid object and a word embedded in a count noun context (e.g., ‘This is a dax’), children as young as two-years-old were prone to the shape bias, extending novel words to other objects that match in shape whilst ignoring differences in other visual properties (Landau et al., 1988; Srinivasan et al., 2017; Srinivasan et al., 2019). Such experimental work consequently mirrors the impact of ambiguity in natural language and the way in which new senses for existing words are learnt. These findings are also consistent with the view that polysemy benefits word recognition (e.g., Armstrong & Plaut, 2008; Klepousniotou & Baum, 2007; Rodd et al., 2002) via a process that begins in childhood and continues into adulthood (Landau et al., 1988).

Such word learning is facilitated by the separate senses of metonyms showing no typical dominance effect or competition like other forms of ambiguities, such as homonyms and metaphors (e.g., Brocher et al., 2016, 2018; Frazier & Rayner, 1990; Frisson & Pickering, 1999; Klepousniotou et al., 2008; 2012). This non-interference between the disparate senses of metonyms is thought to be an indicator of metonyms only having one, rich semantic representation (MacGregor et al., 2015; Klepousniotou, 2002; Klepousniotou & Baum, 2007).

As such, eye-tracking studies have demonstrated reading times for late disambiguating sentences with metonyms in the first sense to be equivalent to those in the alternative sense (Frazier & Rayner, 1990; Frisson & Frazier, 2005; Frisson & Pickering, 2007; McElree et al., 2006; Pickering & Frisson, 2001). This suggests that readers do not commit to a specific interpretation of a metonym as soon as it is encountered; rather they wait until they are provided with additional context that supports one or the other sense. The one, single representation consequently acts as a “gateway” for a metonym’s multiple senses (Frisson, 2009).

However, there is much debate on the information contained within the representation of a metonym (Falkum & Vicente, 2015). One argument suggests a “common core” view, whereby the representation includes all features shared by the separate senses of the word (Copestake & Briscoe, 1995; Klepousniotou et al., 2008). As such, readers of a metonymic word access the dominant sense (e.g., “horse” to mean the animal) from the representation whenever the polysemic word is encountered. However, they must also derive the alternative sense (e.g., “horse” to mean the meat) via a productive rule (due to “horse” being in the class of producer-for-product) when that sense is intended. The second explanation refers to the same common core more abstractly in a position referred to the “underspecified view” (e.g., Caramazza & Grober, 1976; Frisson

& Pickering, 1999; Ruhl, 1989). In this alternative view, individuals encountering a metonymic word will access the general sense first, then use contextual information to decide which of the senses is most likely, the dominant or the alternate. It must be noted however, that although these two opposing views disagree on what the core comprises and how they are disambiguated, both views agree that metonymic words share only a single representation.

One of the areas of metonymic research that may be the most widely researched, however, is its relationship with receptive language. The association between higher receptive language and higher metonymic understanding has been demonstrated in typically developing children (Rundblad & Annaz, 2010a), in children with Williams Syndrome (Annaz et al., 2009) and children with ASD (Rundblad & Annaz, 2010b; Van Herwegen & Rundblad, 2018). However, to the author's knowledge, there has been little research to date exploring the relationship between metonymic knowledge and any other forms of language competence, such as reading, writing, or speaking ability. Additionally, in cases where the receptive language was explored, literature determined receptive language using the British Picture Vocabulary Scale II (BPVS; Dunn et al., 1997), a simplistic tool requiring participants to solely point to an image that best represents a statement heard. Finally, there has been no research exploring children's understanding of the different metonymic types (e.g., animal for meat. or place for people), or the differences in metonymic comprehension in monolingual and bilingual populations. This is despite the knowledge that exposure to language is known to strongly affect language growth (Berko Gleason, 2005; Hart & Risley, 1995; Hoff, 2003; Pearson et al., 1997), with the rate of such exposure also being proportional to vocabulary (Chondrogianni & Marinis, 2011; De Houwer, 2007, 2009; Duursma et al., 2007; Gathercole & Thomas, 2009; Hammer et al., 2008; Hoff et al., 2012; Pearson et al., 1997; Thordardottir, 2011; Thordardottir et al., 2006).

6.1 The Current Study (Experiment 5)

The research presented in this chapter aimed to investigate metonymic comprehension in both monolingual and bilingual, primary-school-aged children and the factors that contribute to better understanding. Given the lack of standardised measures, a novel metonymic task was developed whereby children were required to respond to a metonym in two ways. Firstly, children were asked to verbally explain a metonym, clarifying the intended sense contained within a sentence. Afterwards, the children were asked to match two sentences which used the same sense of a metonym. As such, the metonymic study had both an open-ended and multiple-choice component and maintained uniformity with other experiments contained within this thesis.

As previously mentioned, the national curriculum for England suggests children should demonstrate an understanding of figurative language from Year 3, with the development of more sophisticated understanding apparent by Year 6 (Department for Education, 2014b), therefore the current study was tested in these two age groups with older children expected to excel. Following on from findings from the Metaphors task however, the interaction between year group and SES was also investigated. Once again, whether the child was in receipt of Pupil Premium was used as a proxy for SES grouping.

In line with work discussed in previous chapters of this thesis, metonymic understanding was explored as a product of language experience. As such, bilingualism was measured using both a binary classification (monolingual/bilingual status) and a continuous scale (as per scores generated by the LeBLEQ-C described in Chapter 2) with the prediction that the latter would be a more robust predictor of metonymic ability due to bilingualism being better suited to measures along a continuum (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020).

Finally, associations with scores of academic attainment (Reading and Writing SATs) and teacher's perceptions of linguistic competence (LeBLEQ-T) were also investigated alongside scores of expressive language ability, Working Memory performance and gender. Of particular interest was the ability of a child's Receptive Language Score (RLS) to predict metonymic awareness. The RLS of the Streamlined-CELF is a composite score assessing receptive ability on more than one task, consequently allowing for a more detailed understanding of the link between metonymic comprehension and receptive ability.

6.2 Methods

6.2.1 Participants

Participants were the same as those in Chapter 4 (Experiment 3).

6.2.2 Materials

The experimental task had two components to test participants' knowledge of metonymic words. Firstly, participants were presented with a sentence with a target word in the sentence-final position. They were instructed to rephrase the sentence to show their understanding of the sense of the target word. Secondly, participants then saw the sentence alongside two sentences which also used the target word in the sentence final position. They were instructed to select the sentence that used the word in the same way as the target sentence. If they saw no difference in the word uses, they were instructed to select both sentences. Children were instructed to respond using the mouse pad of the laptop.

Twelve metonymic (i.e., experimental) words were selected from the list of words surveyed in Experiment 2 (Chapter 3). These experimental words had a single meaning (a single orthographic and phonological form) but multiple, related senses. Like the senses of the metaphoric stimuli included in the previous chapter, the senses of the metonymic stimuli were related. Unlike metaphors, however, both the basic and secondary senses of metonymic words are literal, and are highly related through sense extension (Shutova et al., 2013). Classification was based on the definitions of metonyms found in past literature (e.g., Klepousniotou et al., 2012). The experimental stimuli contained within the current research had also been used in previous literature investigating metonymic processing and understanding (e.g., Klepousniotou et al., 2012). Experimental words were selected on a number of parameters. Firstly, only words where *both senses* of the word were reported to have been known by the majority of the respondents were included. If a sense of a word was reported to have been unknown more than once, the word (i.e., both senses) was excluded from the list of potential stimuli. Furthermore, it was ensured that the ages at which both Sense-1 and Sense-2 of the words were known across a wide range (3-11 years) to limit the likelihood of floor- or ceiling effects being observed. All stimuli can be found in Appendix E.

The 12 metonyms were also selected to have similar characteristics as the 12 metaphors used in the Metaphors Task (Chapter 5). This was to ensure results from each of the polysemic tasks could be later compared. Firstly, while it was not feasible to match the AoA of *both* senses of the polysemic words (using norms gathered in Experiment 2, Chapter 3), the sense-specific AoA of the first senses of the words were statistically similar (p> .05) between the Metaphors Task (M= 4 year, 3 months, SD= 2 years, 3 months) and the Metonyms Task (M= 4 years, 8 months; SD= 1 year, 10 months). The Kuperman et al., (2012) non-sense-specific AoAs were likewise considered and metonyms had statistically similar (p> .05) ages of acquisition to the metaphoric stimuli ($M_{Metaphors}$= 5 years, 5 months, $SD_{Metaphors}$ = 1 year, 4 months; $M_{Metonyms}$= 5 years 4 months; $SD_{Metonyms}$= 1 year, 6 months).

There were no statistically significant differences (all ps>.05) between the polysemic items in terms of the number of letters, syllables, or phonemes within a word and all words matched on imageability, concreteness, and familiarity using the MRC Psycholinguistic Database (Coltheart, 1981a) as well as multiple measures of frequency. This included the Kučera and Francis (1967) and Thorndike-Lorge written frequency scores (Thorndike & Lorge, 1944) and Brown verbal frequency scores (Brown, 1984) in addition to the frequency scores from the Children's Printed Word Database scores (CPWD; Masterson et al., 2010), and the Subtitle-based Word Frequencies for British-English (SUBTLEX-UK; van Heuven et al., 2014). There were no statistically significant differences when employing any of these norms and all means and standard deviations can be seen in Table 44.

Table 44

Means and standard deviations of the experimental metonymic words in comparison to both the experimental metaphoric words (M (SD)) and the control metonymic words.

	Metonyms (Experimental)	Metaphors (Experimental)	Control (Metonyms)
Letters in word	4.25 (.97)	4.17 (.83)	4.75 (1.14)
Syllables in word	1.25 (.45)	1.08 (.29)	1.50 (.52)
Phonemes in word	3.58 (.79)	3.50 (1.08)	3.70 (.82)
Imaginability	582.73 (29.26)	598.60 (22.51)	601.00 (22.18)
Concreteness	590.45 (25.43)	602.30 (30.01)	604.10 (20.76)
Familiarity	549.55 (47.72)	520.30 (37.14)	518.91 (68.01)
Written Frequency (Kucera & Francis, 1967)	28.75 (25.09)	24.36 (24.11)	30.09 (63.52)
Written Frequency (Thorndike-Lorge, 1942)	380.55 (417.57)	157. 27 (174.41)	462.00 (195.66)
Verbal Frequency (Brown, 1989)	6.75 (5.45)	2.40 (2.61)	24.50 (27.58)
Frequency in Children's Books (CWPD, Masterson et al., 2010)	103.36 (120.25)	83.09 (103.57)	120.90 (170.44)
Frequency in TV (SUBTLEX-UK; Van Heuven et al., 2014)	4.34 (.48)	4.26 (.49)	4.06 (.52)

A list of 12 control (non-metonymic) words was created by selecting items from Armstrong et al. (2012) on the basis of having only one definition. Experimental and control words were matched in terms of AoA (Kuperman et al., 2012) with no statistically significant difference (p> .05) between the two groups ($M_{Experimental}$ = 5 years, 4 months, $SD_{Experimental}$= 1 year, 6 months; $M_{Control}$= 5 years 4 months, $SD_{Control}$= 1 year, 4 months).

As above, there were no statistically significant differences (all ps> .05) between the experimental and control items in terms of the number of letters, syllables, or phonemes within a word, or the word's imageability, concreteness, and familiarity using the MRC Psycholinguistic Database the Kucera and Francis (1967) and Thorndike-Lorge (1942) written frequency scores, the Brown verbal frequency scores (Brown, 1989), norms form the Children's Printed Word Database (CPWD; Masterson, Stuart, Dixon, & Lovejoy, 2010), and the Subtitle-based Word Frequencies for British-English (SUBTLEX-UK; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). All means and standard deviations can be seen in Table 44 and a list of stimuli can be seen in Appendix A.

6.2.2.1 Target Sentences

Sentences were then created to incorporate the target words. Sentences varied from five to nine (M= 7.54, SD= 1.57) words in length and all contained the target word in the sentence-final position.

While only one sentence was created for the control words, two sentences were created for each metonymic item; one biasing the first sense of the word, and one biasing the second sense of the word. The first sense of the word was always referred to the more physical sense (i.e., the glass of the bottle, or fruit that is the lemon), while the second sense referred to the more abstract sense of the word (i.e., the milk within the bottle, or the flavour of the lemon). The inferred sense of the word was unambiguous (i.e., was clearly the Sense-1 interpretation and could not be mistaken for the Sense-2 interpretation). For example, in the case of the metonymic word '*lemon*', the two target sentences created were:

Sentence 1: *At the market, she picked up a lemon*

Sentence 2: *The cake she made tasted like lemon*

However, in order for participants to not be exposed to both senses of a metonymic word and for the task to be kept relatively short, two versions of the task, were created. These two versions of the task were identical in all aspects, differing only in terms of stimuli. List-A and List-B contained the same 12 control words and the same 12 metonymic words, only in opposing senses. Half of the metonymic words in each list were presented in sentences which emphasised the first sense, while the other half of the sentences emphasised the second sense. In the example of '*lemon'* above, Sentence 1, which stressed the first sense of the word, was contained in List-A, while Sentence 2, which stressed the second sense of the word, was contained in List-B.

Target sentences were matched on three parameters; (1) the number of words contained within the sentence, (2) the average number of letters each word contained, and (3) the average Kuperman et al.'s AoA for each word. The means and standard deviations for these variables are reported in Table 45.

Table 45

Means and standard deviations (M (SD)) outlining characteristics of the Target Sentences in the Metonymic Task.

	Control	Experimental	
		Sense-1	Sense-2
Words in each target sentence	7.17 (.94)	7.42 (1.08)	7.67 (1.56)
Letters in each word	3.95 (1.96)	3.81 (1.77)	3.98 (1.62)
Kupermans AoA of each word (in years)	4.44 (1.41)	4.65 (1.52)	4.54 (1.38)

A series of t-tests revealed no significant differences of these criteria between the control and experimental target sentences, or between Sense-1 and Sense-2 sentences (all p-values> .05). There were no statistical differences regarding these three parameters between the target sentences assigned to List-A and List-B (p> .05).

6.2.2.2 Response Options

Two response options were created for each target word. These response options were sentences of six to nine words (M= 7.56, SD= 1.27) that also used the target word in the sentence-final position. In the case of the experimental words, one of the sentences used the first sense of the word, while the second sentence used the second sense of the word. The response options were the same for List-A and List-B. For example, in the case of the metonymic word '*lemon'*, the two response options were:

Response Option 1: *The orange was smaller than the lemon*

Response Option 2: *The new kitchen cleaner smelt of lemon*

If the participant had been exposed to List-A ("*At the market, she picked up a lemon*"), Response Option 1 would be the correct match, with both sentences using the word '*lemon'* to mean *the physical fruit*. If the participants were exposed to List-B, however, ("*The cake she made tasted of lemon*"), Response Option 2 would be the correct match, with both sentences using the word '*lemon'* to mean *properties of the fruit*. Consequently, only one of the response options in the experimental trials was correct as only one response option used the same sense of the target word that was used in the target sentence.

In the case of the control trials, there was no difference between the first and second sentence as, obviously, there was only one sense of the word. For example, in the case of the control word '*lamp'*, the target sentence was "I tripped over whilst holding the lamp". The two response options were:

Response Option 1: *I switched on the big red lamp*

Response Option 2: *Mum wanted to buy a new lamp*

Consequently, *both* of the response options in the control trials were correct as both used the same (and only) sense of lamp.

Response options were matched on a number of parameters akin to the target words themselves. Namely, the number of words in each response option, the average number of letters each word contained in the sentence and the average Kuperman et al.'s AoA for each word were statistically similar ($p > .05$) amongst all four response options. The means and standard deviations for these variables are reported in Table 46.

Table 46

Means and standard deviations (M (SD)) outlining characteristics of the Response Options the Metonymic Task.

	Control	Experimental	
		Sense-1	Sense-2
Words in each response option	7.33 (.70)	7.75 (1.21)	7.83 (1.27)
Letters in each word	3.89 (2.16)	3.90 (1.96)	4.45 (2.16)
Kupermans AoA of each word (in years)	4.50 (1.37)	4.56 (1.55)	4.63 (1.34)

A series of t-tests revealed no significant differences of these criteria between the control and experimental response options or between the Sense-1 and Sense-2 response options of the experimental trials (all p-values> .05).

6.2.3 Procedure

6.2.3.1 Language and Cognitive Measures

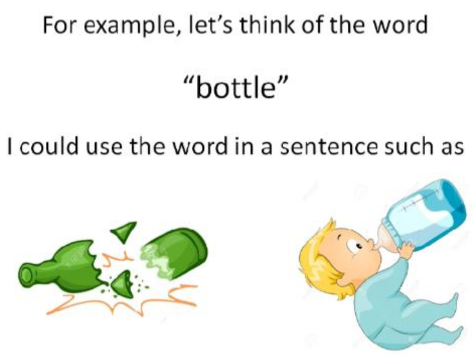
Language and cognitive measures were identical to those described in Chapter 4 (Experiment 3).

6.2.3.2 The Metonyms Task

The experiment was created using PsychoPy version 3.0 (Pierce, 2007) and presented to the children on laptops. Trials were presented in random order, determined by PsychoPy, to prevent order effects. Participants were first given instructions to the task. It was explained to the children that that task they were about to participate in regarded "words that looked the same, sounded the same, but sometimes meant more than one thing". Children were given the example of the word "*bottle*" and presented with the image displayed in Figure 13.

Figure 13

Image displayed in the example ("bottle") of the Metonymic Task



Researchers explained that the word bottle could be used in two different ways. In the case of the first picture, the baby is drinking the bottle, but what the baby actually drinking is whatever substance is inside of the bottle. Consequently, the intended sense of "bottle" of "the baby drank the bottle" is the same as the intended sense of "the baby drank the milk" or "the baby drank the liquid inside of the bottle."

In the case of the second picture, however, the sentence "I smashed the bottle", does not denote what is contained within the bottle, but rather the physical entity; the outside. Here, the sense of the word "bottle" in "I smashed the bottle" equates to "I smashed the container" or "I smashed the part of the bottle made of glass." Participants were informed that the current study was interested in the senses the children knew of the words and, thus, they were required to think carefully about each word.

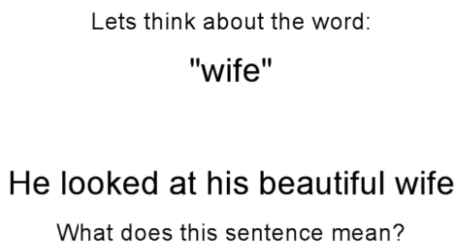
Participants were tested on their knowledge of the words in two ways. Firstly, they were told that they would be presented with a sentence and would be asked to rephrase the final (target) word (such as in the examples given above). They were instructed to respond verbally to the experimenter. Secondly, the participants were informed that they would then see the target sentence alongside two other sentences that contained the target word. They would then be required to select the sentence/s that used the target word in the same way. Sometimes one of the sentences would use the word in the same way (the experimental trials) while other times, both sentences would use the word in the same way (control trials). Children were given the example of matching “the baby drank his bottle” to “Mum spilt her milk” and not “I smashed my bottle”. Children were instructed to respond using the mouse pad of the laptop.

Instructions were presented verbally by the researcher so it could be ensured the child was aware what they were being required to do. Once the participants confirmed they understood the task, and what was expected from them, they were permitted to click a ‘Let’s Practice’ button displayed on the screen.

Participants were given two practice trials before continuing onto the experimental trials. Each trial comprised of two components: (1) an open-ended question, and (2) a multiple-choice question. In the open-ended component of the trial, the participant was asked to rephrase a target sentence to show their understanding of the intended sense of the target word. All trials were presented in the same format as the practice trial of ‘*wife*’ (a control word) which can be seen in Figure 14.

Figure 14

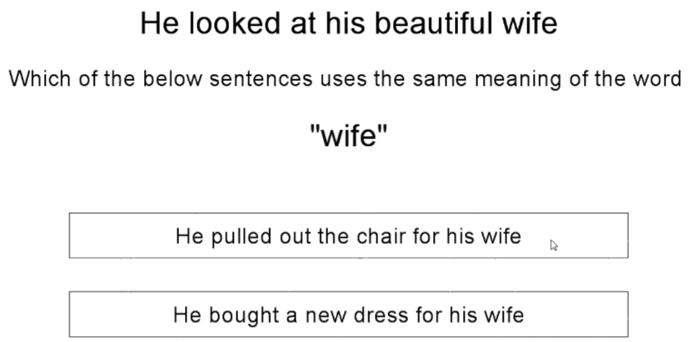
Presentation of the Open-ended question in the Metonyms Task.



Target words were presented in all lower case to aid recognition, and participants were able to respond verbally to the experimenter who would key in their answers. The format of the multiple-choice component of the trials can be seen in Figure 15.

Figure 15

Presentation of the Multiple-Choice question in the Metonyms Task.



The presentation of the correct and incorrect response options (i.e., foils) were randomised by PsychoPy. Participants selected the sentence with the same sense of the target word using the touchpad of the laptop. Once selected, the response options would turn green. When the participants had finished their selections, they were instructed to press the ‘Next’ button in the bottom right corner of the screen. After a brief pause (0.2 seconds) during which the screen displayed no text, the next trial began.

After completing the two practice trials, a second set of short instructions were given. Children were told that the practice trials were over and were given an opportunity to ask any questions before the experimental trials began. These instructions lasted 11 seconds. After this point, the ‘Let’s Go!’ button was presented allowing participants to proceed to the first trial.

In all cases, information was simultaneously presented visually and auditorily to account for differences caused by participants’ reading abilities. The audio was recorded by a female, native British-English speaker with a neutral accent using Audacity (version 2.1.2).

The Metonymic Task took approximately 20 minutes to administer. Accuracy of response and response time (in milliseconds) were recorded automatically by PsychoPy for both the open-ended and multiple-choice components of each trial. Children were taken from their class and tested individually in a quiet classroom, onsite at school. Although ethical approval from parents had already been obtained, the testing procedure was first explained to the children before the experiment began, and they gave their verbal assent to continue. During the

Metonym and CNRep tasks, participants wore noise-cancelling headphones to listen to the audio and limit environmental distractions. Each child's assessment took a total of approximately 30 minutes to complete.

6.2.3.2.1 Scoring of the Metonymic Task

6.2.3.2.1.1 Open-Ended Component

The open-ended component of the Metonymic Task was coded in two ways with responses being assigned both a score and a code. Firstly, a child could either receive a score of 0 or 1. A child received a score of 1 if they were able to rephrase the sentence to show understanding of the target word, in the correct sense. If they were unable to do so, they scored a 0.

Secondly, these scores were then coded. Scores of 0 were coded in one of four ways: (1) '*Flawed Definition'* (i.e., the way in which the child rephrased the sentence did not show understanding of the target word); (2) '*Omission'*, where the child refused to give an answer; (3) '*Incorrect Sense'*, the child was able to rephrase the sentence and show correct understanding of the word, but described the opposing, incorrect sense of the target word; or (4) '*Ambiguous'* whereby the child was able to rephrase the sentence to show understanding of the target word, but the way in which they did this did not emphasise a single sense of the word; it could be either correct or incorrect (i.e., the first or second sense).

For correct scores (i.e., those given a score of 1), the sense of the word described by the child was coded for the later analysis of the most popular response. Responses were coded as being the '*Correct Sense'* if the participant was able to rephrase the sentence in a way that shows they have identified the correct sense of the target word.

For example, if a participant was given the target sentence of "*At the market, she picked up a lemon,"* responding with "*She picked up a yellow, oval fruit*" would have been coded as '*Correct Sense'*. Responding "*She picked up a zesty flavour*" would have been coded '*Incorrect Sense'* (as this emphases the second sense of the word, not the first), and responding "*She picked up the object*" would have been coded '*Ambiguous'* as it cannot be clear whether the participant interpreted the target word to have the first or second sense.

6.2.3.2.1.2 Multiple-Choice Component

Responses to the multiple-choice component of the task were scored using a 0-1 scale. A child received scored a 1 if they selected both responses in the control trial (correct) or the single response option in the experimental trial that used the same sense of the metonymic word contained in the target sentence. Conversely, they received a score of 0 if they selected either both responses of an experimental trial, or only one response on a control trial.

6.3 Results

Two t-tests were initially performed revealing accuracy rates to be significantly higher in the control condition than the experimental condition in the open-ended component ($M_{Experimental}$= .35, $SD_{Experimental}$= .48; $M_{Control}$= .77, $SD_{Control}$= .42; *t*(3043.84)= 3.41, *p*< .01). However, the children performed significantly better in the experimental task than in the control in the multiple-choice component ($M_{Experimental}$= .43, $SD_{Experimental}$= .50; $M_{Control}$= .36, $SD_{Control}$= .48; *t*(2996.98)= -25.97, *p*< .001) of the task. This means that children were more likely to select only one of the response options matching the target sentence (which contained a non-metonymic word) than, correctly match both (as there were no differences between the sentences due to the target words only having a singular sense).

When considering experimental trials only, a t-test revealed participants to perform significantly better in the Multiple-Choice component (M= .43, SD= .50) compared to the Open-Ended component (M= .35, SD= .48; *t*(1500)= -4.81, *p*< .001) of the task. Finally, there was no statistically significant difference between the children who were exposed to List-A ($M_{Open\text{-}Ended}$= .34, $SD_{Open\text{-}Ended}$= .47; $M_{Multiple\text{-}Choice}$= .41, $SD_{Multiple\text{-}Choice}$= .49) compared to List-B ($M_{Open\text{-}Ended}$= .36, $SD_{Open\text{-}Ended}$= .48; $t_{Open\text{-}Ended}$(1520.67)= -1.10, *p*= .273; $M_{Multiple\text{-}Choice}$= .44, $SD_{Multiple\text{-}Choice}$= .50; $t_{Multiple\text{-}Choice}$(1519.69)= -1.07, *p*= .284).

Descriptive statistics were first conduced regarding the *type* of response participants gave. Firstly, this was considered at a group level (including all participants), then the Year 3 children were compared to the Year 6 children. Finally, the monolinguals were compared to their monolingual peers (using the binary classification of language experience). Prior to all descriptive statistics, Levene’s tests for equality of variances were conducted. In cases where this assumption was violated, degrees of freedom were adjusted and a statistic not assuming the homogeneity of variance was reported.

Subsequently, accuracy and latency data were analysed using a series of Multilevel Linear Models, implemented in IMB SPSS Statistics (Version 26) using a Maximum Likelihood method. The dependent variable was either the accuracy or the latency of the response (described separately below). The models included the predictors of KS1 SAT Reading score, KS1 SAT Writing score, Streamlined-CELF Receptive Language Score (RLS), Expressive Language Score (ELS), Backwards Digit Recall, Total LeBLEQ-T score, gender, sense of target and language experience classification (monolingual, bilingual) as fixed factors, as well as an interaction of Year Group and Pupil Premium. The model was then run for the second time. Here, the same predictors were included except language experience classification (monolingual, bilingual) which was replaced with the child's Cumulative Language Input Score generated from the LeBLEQ-C. The Bayesian information criterion (BIC) of the two models was then compared in order to determine the model with the best fit to the data. In all models, random intercepts were included for each participant. Responses to Open-Ended questions and Multiple-Choice questions are described separately below.

6.3.1 Model Comparison

Bayesian information criterion (BIC) values were compared across the models including the binary (monolingual vs bilingual) and continuous (Cumulative Input Score) language experience classifications. BIC values for all models are outlined in Table 47.

Table 47

Bayesian information criterion values for Metonym models including the binary (monolingual vs bilingual) and continuous (Cumulative Input Score) language experience classifications

		BIC	
		Binary	Continuous
Open-Ended Component			
	Accuracy	1576.13	1505.26
	Reaction Time	841.45	790.38
Multiple-Choice Component			
	Accuracy	1667.85	1598.12
	Reaction Time	1098.19	1051.33

Literature considers the model with the lowest BIC preferable (e.g., see Vrieze, 2012). As can be seen in Table 47, the models that include the continuous, Cumulative Language Input Score have lower BIC values, indicating better fitting models. Consequently, only the findings of the continuous models will be discussed below.

6.3.2 Open-Ended Component

6.3.2.1 Response Type Analyses in the Open-Ended Component

First, specific response types were analysed (see Table 48). Responses were coded as being either correct or incorrect. An incorrect response could be due to (1) not answering the question; (2) giving an flawed definition of the word (e.g., using the word to define the target or describing the target incorrectly); (3) giving a correct definition of the word but in the wrong sense; and (4) giving a correct definition of the word but such a definition being ambiguous in terms of the sense implied.

Table 48

Response types across experimental trials of the Open-Ended Component of the Metonyms Task, stratified by age of child and language experience classification.

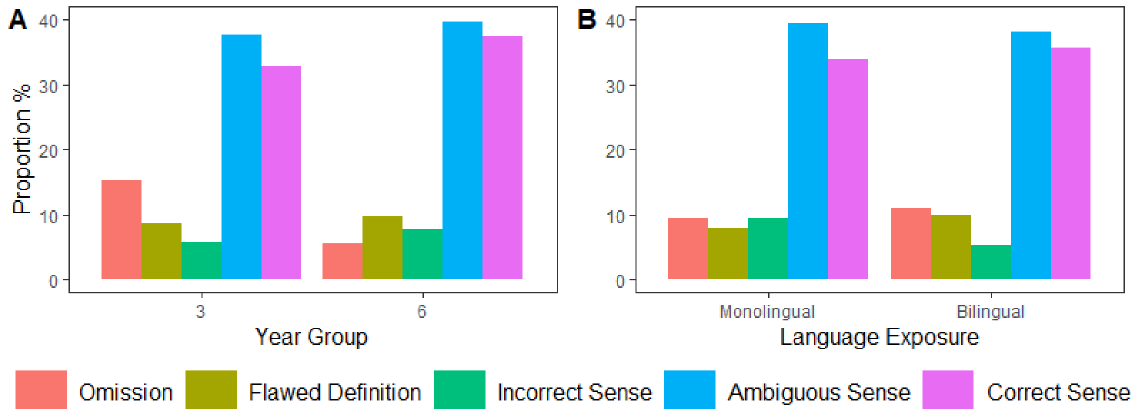
Response type	Year 3 Monolingual	Year 3 Bilingual	Year 6 Monolingual	Year 6 Bilingual	Total
Omission	11.00%	17.92%	7.54%	4.47%	10.50%
Flawed Definition	8.67%	8.75%	7.14%	11.18%	9.25%
Incorrect Sense	8.67%	3.96%	10.32%	6.50%	6.76%
Ambiguous Sense	39.00%	36.67%	39.68%	39.43%	38.52%
Correct Response	32.67%	32.71%	35.32%	38.42%	34.97%

Overall, across all participants, there was a significant difference in proportion of responses given [$F(4, 634) = 193.12$, $p < .001$]. Subsequent post-hoc analyses using the Bonferroni test indicated that this was driven by children providing significantly more ambiguous senses (M= .39, SD= .17) compared to omissions (M= .10, SD= .12), flawed definitions (M= .09, SD= .11) or wrong senses (M= .07, SD= .08) (all p-values< .001). Children also provided significantly more correct senses (M= .35, SD= .14) than omissions (M= .10, SD= .12), flawed definitions (M= .09, SD= .11) or wrong senses (M= .07, SD= .08) (all at p< .001).

Children across the two year groups performed similarly, with the only difference being that the younger children were more likely to not answer the question (M= .15, SD= .13) in comparison to their older peers (M= .06, SD= .08; $t(110.22) = 5.22$, $p < .001$). In addition, the children in Year 3 were marginally less likely to give a correct response (M= .33, SD= .13) than the Year 6 children (M= .37, SD= .14), though this only approached significance ($t(125) = -1.92$, $p = .057$). Responses can be seen in Figure 16A.

Figure 16

Types of responses given in the Open-Ended component of the Metonyms Task, stratified by year (A) and language experience (B).



Finally, monolingual children were significantly more likely to provide an interpretation of the target word but describe it in the wrong sense (M= .09, SD= .9) compared to their bilingual peers (M= .05, SD= .07; t(125)= 2.89, p< .01). Responses can be seen in Figure 16B.

6.3.2.2 Analysis of Accuracy Rates in the Open-Ended Component

When including individuals Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the open-ended trials, the relationship between scores on the Metonyms task and predictors showed no variance in intercepts across participants. Excluding this factor, however, demonstrated a significant model $\chi^2(1)= 70.22$, $p< .01$. Here, there was a significant effect of KS1 Reading score [$F(1, 1068)= 6.43$, $p< .05$], where participants performed better on Metonyms task when they performed at, or above average, in such compulsory scores of academic attainment. Furthermore, there was a significant contribution of the sense of the target [$F(3, 1068)= 16.03$, $p< .001$] where participants were more able to accurately define the metonym when presented in its second sense. Contributions of all other fixed factors can be seen in Table 49.

6.3.2.3 Analysis of Reaction Times in the Open-Ended Component

Reaction time analyses excluded both errors (65.07% of trials) and outliers. Following recent literature investigating child language (e.g., Commissaire et al., 2019; Grundy & Keyvani Chahi, 2017; Levy & Hanulíková, 2019; Persici et al., 2019) responses that were 2.5 standard deviations above/below each participant's mean were considered outliers (a further 1.51% of trials). This amounted to a substantial 66.58% of trials being excluded from RT analyses, indicating that children performed considerably poorly on the current task. This was despite the task being specifically designed for this age group based on the AoA ratings obtained in Experiment 2 (Chapter 3). The remaining RTs were z-scored to further minimise the impact of potential outliers. Prior to all analyses, Levene's tests for equality of variances were conducted. In cases where this assumption was violated, a t-statistic not assuming the homogeneity of variance was computed.

The relationship between scores on the Metonyms task and predictors showed significant variance in intercepts across participants, $Var(u_{0j})= .077$, $\chi^2(2)= 58.40$, $p< .01$. Here, there was a significant effect of ELS only [$F(1, 82.31)= 6.71$, $p< .05$] where increased expressive language skills contributed to a quicker time formulating a response on the Open-Ended component of the Metonyms task. Contributions of all other fixed factors were non-significant and can be seen in Table 49.

Table 49

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Open-Ended component of the Metonyms Task when using a Continuous Measure of Language Experience.

	Accuracy				Reaction Time			
	b	*SE*	*T*	*p*	*b*	*SE*	*t*	*p*
Intercept	.28	.13	2.22	.026*	-.05	.42	-.12	.904
Cumulative Language Input Score	.04	.06	.70	.484	.05	.18	.25	.805
Gender (Male)	.03	.03	.81	.418	.12	.10	1.21	.229
KS1 Reading SATs (below average)	-.10	.05	-2.11	.035*	.25	.17	1.51	.134
KS1 Writing SATs (below average)	.01	.05	-.321	.749	-.29	.15	-1.85	.067
Receptive Language Score	-.00	.00	-.45	.652	.01	.00	1.58	.119
Expressive Language Score	.00	.00	1.65	.099	-.01	.00	-2.59	.011*
Working Memory Score	-.01	.02	-.77	.439	.01	.06	.17	.869
LeBLEQ-T	.00	.00	.23	.815	.00	.00	1.80	.075
Year*Pupil Premium (Year 3, without Pupil Premium)	.01	.05	.14	.888	.04	.16	.24	.812
Sense of Target (Sense-1)	-11	.03	-4.00	.000**	.09	.07	1.22	.222

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

6.3.3 Multiple-Choice Component

6.3.3.1 Response Type Analyses in the Multiple-Choice Component

When providing an incorrect response (i.e., score of 0), children were statistically more likely to select just one, but incorrect sense (M= .60, SD= .31) rather than selecting both senses (M= .40, SD= .31; t(254)= 5.09, p< .001). This indicates that the children could see a difference between the two senses, yet had difficulty identifying the specific sense of the word. This effect was the same across both Year 3 ($M_{IncorrectSense}$= .56, $SD_{IncorrectSense}$= .33; $M_{BothCorrect}$= .44, $SD_{BothCorrect}$= .33; t(128)= 2.09, p< .05) and Year 6 ($M_{IncorrectSense}$= .64, $SD_{IncorrectSense}$= .29; $M_{BothCorrect}$= .36, $SD_{BothCorrect}$= .29; t(122)= 5.36, p< .001) children. Whilst the same effect was seen in bilinguals ($M_{IncorrectSense}$= .63, $SD_{IncorrectSense}$= .33; $M_{BothCorrect}$= .37, $SD_{BothCorrect}$= .33; t(160)= 4.90, p< .001), monolinguals were equally likely to select one incorrect sense or both correct senses ($M_{IncorrectSense}$= .55, $SD_{IncorrectSense}$= .28; $M_{BothCorrect}$= .45, $SD_{BothCorrect}$= .28).

6.3.3.2 Analysis of Accuracy Rates in the Multiple-Choice Component

When including participants' Cumulative Language Input Score (generated from the LeBLEQ-C) as a fixed factor in the model, the relationship between scores on the Metonyms task and predictors showed significant variance in intercepts across participants, Var(u_{0j})= .006, $\chi^2(2)$= 71.85, *p*< .01. Here, gender significantly contributed to a child's score on the Metonyms task [F(1, 88.79)= -4.62, *p*< .05) with males performing significantly better in comparison to females. A child's working memory ability was also a significant contributor [F(1, 87.93)= 4.45, *p*< .05] with those with a lower working memory score performing better in the current task. Contributions of all other fixed factors were non-significant and can be seen in Table 50.

6.3.3.3 Analysis of Reaction Times in the Multiple-Choice Component

Reaction times (RTs) were again z-scored with the analyses excluding both errors (57.93% of trials) and outliers (a further 1.51% of trials).This amounted to a substantial 59.44% of trials being excluded from RT analyses, indicating the children, overall, performed considerably poorly on the current task. This, again, was despite the task being specifically designed for this age group based on the AoA ratings obtained in Experiment 2 (Chapter 3).

Analyses showed significant variance in intercepts across participants, ($Var(u_{0j})$= .294, $\chi^2(2)$= 159.61, *p*< .01). Here, there was a contribution bordering significance of gender [F(1, 81)= 3.93, p= .051], with males responding quicker than females. Furthermore, there was also a significant interaction between school year and whether the child was in receipt of Pupil Premium [F(3, 83.59)= 3.32, p< .05]. Contributions of all other fixed factors were non-significant; all effects can be seen in Table 50.

The School Year by Pupil Premium interaction was then broken down by conducting separate multilevel models on those who were in receipt of Pupil Premium and those who were not. The models specified were the same as the main model, but excluding the interaction term of Pupil Premium and, instead, including simply a main effect of School Year. These models revealed that, for those who received the subsidiary, there was no significant effect of year [F(1, 26.58)= .52, *p*= .477, *b*= -.21, *SE*= .29, *t*= -.72]. There was a significant effect of year, however, in children who did not receive the subsidiary [F(1, 59.35)= 4.03, *p*< .05, *b*= -.51, *SE*= .26, *t*= 2.01] with children in Year 6 responding significantly faster than children in Year 3 to the Multiple-Choice component of the Metonyms task. This demonstrates an increase in the performance of metonyms between Year 3 and Year 6 (captured in the speed of processing), but only in children who are not in receipt of Pupil Premium.

Table 50

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Multiple-Choice component of the Metonyms Task when using a Continuous Measure of Language Experience.

	Accuracy				*Reaction Time*			
	b	*SE*	*T*	*p*	*b*	*SE*	*t*	*p*
Intercept	.43	.15	2.89	.005**	-.60	.58	-1.02	.308
Cumulative Language Input Score	.-.00	.07	-.02	.998	.37	.28	1.31	.193
Gender (Male)	.08	.04	2.15	.034*	-.29	.15	-1.98	.051
KS1 Reading SATs (below average)	-.10	.06	-1.70	.093	.16	.24	.66	.513
KS1 Writing SATs (below average)	.03	.05	.53	.596	-.19	.22	-.86	.394
Receptive Language Score	.00	.00	.52	.603	-.00	.01	-.32	.751
Expressive Language Score	.00	.00	.76	.449	.01	.00	1.48	.143
Working Memory Score	-.04	.02	-2.11	.038*	.13	.08	1.48	.143
LeBLEQ-T	-.00	.00	-1.02	.312	.01	.00	1.79	.077
Year*Pupil Premium (Year 3, without Pupil Premium)	-.09	.06	-1.54	.126	-.21	.23	-.93	.356
Sense of Target (Sense-1)	-.04	.03	-1.21	.226	.06	.06	.94	.348

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

6.4 Discussion

The present research aimed to investigate the metonymic skills of primary school-aged children in order to understand the impact of bilingualism on the comprehension of polysemy. The study also aimed to explore whether metonymic competence could be better predicted using a continuous score of bilingualism rather than a binary classification (monolingual/bilingual). A novel metonymic task was administered alongside measures of general linguistic ability and Working memory (Streamlined-CELF) to Year 3 (7-8-year-old) and Year 6 (10-11-year-old) pupils in schools of relatively low socioeconomic status associated with the Born in Bradford (BiB) cohort. The Leeds-Bradford Language Exposure Questionnaire for Children (LeBLEQ-C) was administered to record levels of bilingualism of the pupils on a continuous scale (rather than simply using the binary classification monolingual/bilingual). The Leeds-Bradford Language Exposure Questionnaire for Teachers (LeBLEQ-T) was also administered to obtain an indication of teachers' perceptions of the child's linguistic behaviour in the classroom. Scores on KS1 Reading and Writing SATs were also obtained from the schools, as was information regarding Pupil Premium (which was taken as an indicator of disadvantage).

Firstly, it must be noted that, although the children performed better on the control trials in comparison to the experimental trials when considering the Open-Ended component of the task, the children were significantly better in the experimental trials than the control on the Multiple-Choice component of the task. The findings revealed that children were more likely to select only one of the response options matching the control target sentence (which contained a non-metonymic word) than correctly matching it to both response options (as the target word only had a single sense, there was no difference between the response options in terms of the way in which the target word was used). This indicates that children were able to see subtle distinctions in the way the words were used, even when the word in question was non-metonymic. Such conclusions were likely subconsciously drawn by the children due to the prevalence of metonyms, and general ambiguity in the English language. Polysemy is considered the rule rather than the exception in most languages. In the case of English, a small 15.4% of the words in the dictionary are truly unambiguous and have only one sense (Parks et al., 1998). The remaining majority of the words (77.9%) are pure polysemes and have more than two senses. Consequently, children are accustomed to observing words being used in different ways, and thus possess the understanding that one word can have multiple interpretations.

Accuracy rates for both components of the Metonymic task were considerably low with an accuracy rate of 35% for the Open-Ended component and 43% for the Multiple-Choice component. Similar effects were seen for the Homonyms task (Experiment 3, Chapter 2) and the Metaphoric Task (Experiment 4, Chapter 5); this remains to be surprising considering the stimuli were specifically designed for use with the age groups used, as determined by the age of acquisition ratings in Experiment 2 (Chapter 3). It is once again suggested that this effect was due to the overall low SES of the city of Bradford as a whole Index (IDACI; Department for Communities and Local Government, 2015), but could be further attenuated by the possible discrepancy between the SES of the children tested in the current study, and the general population sampled in the AoA norming experiment (see the discussion of Chapter 4 for a full explanation of this).

Accuracy rates across the two component-types demonstrated that children found the Open-Ended component more difficult than the Multiple-choice component of the Metonyms Task. This is in line both with predictions and previous literature showing that children find explaining ambiguities harder than answering multiple-choice questions about them (Arcara et al., 2019; Perlini et al., 2018; Pouscoulous, 2014). The Open-Ended component of the task allowed for increasingly detailed, rich data to be collected, whilst limited the task's susceptibility to ceiling effects, and highlighted the importance of response format when studying metonymic competence. These results highlight the importance of response format when studying ambiguities.

Finally, Linear Mixed Effects Modelling demonstrated various contributors to metonymic understanding. Although older children (Year 6) performed on par to younger children (Year 3) on this novel task, different effects were seen between high- and low-SES groups with those from poorer socio-economic strata being slower in the Multiple-choice component of the task than those who did not receive the subsidiary. Though there was no effect of language experience when either binary or continuous measures were taken, a model including the continuous measure of the Cumulative Language Input Score was shown to fit the data best. The expressive language abilities of the child impacted the processing speed of the metonyms presented, with those with higher expressive ability being quicker to correctly clarify the intended sense of a metonym. A child's KS1 Reading Score also significantly predicted accuracy, with those with lower (below average) reading ability performing worse on the Metonyms task. Finally, the sense of the word contained within the target sentence also impacted accuracy rates, with children being more able to correctly clarify, or define, the word when presented with its second sense. This second sense was the interpretation of the metonym as a food substance (in most cases) and was disparate from the interpretation of the word as its physical entity. For instance, the second sense of the word "cup" meant the space enclosed by the physical entity, such as the liquid inside (e.g., "I was thirsty and drank the whole cup"). All significant factors are discussed separately below.

6.4.1 SES and Age of child/School Year

In the Multiple-Choice component of the task, there were no effects of School Year in children who received Pupil Premium but there was an emerging difference for pupils who did not receive Pupil Premium (and were, thus, considered to be higher SES even within the disadvantaged areas where testing took place). In particular, Year 6 children demonstrated faster processing of metonyms compared to younger Year 3 peers. Such findings suggest age to be an indicator of metonymic awareness in children but this to be highly moderated by SES, with those from higher socio-economic positions outperforming those from lower strata with increasing age. These findings oppose original predictions made for the current study which expected children in receipt of Pupil Premium to be outperformed by those without in Year 3, but this effect to be eliminated by Year 6 (similar to the findings of Hutchinson et al., 2016). In contrast, the current study found the differences between children who received Pupil Premium and those who did not to be mostly apparent in Year 6. As discussed in previous chapters, it is possible that this effect was observed due to children from lower SES being required to direct their attentional resources on the development of other, more widespread skills needed that are vital to enable them to progress within the classroom environment. As a result, this may have diverted attention from the development of their higher order, metonymic skills which become hindered in comparison to their higher SES peers.

Perhaps most importantly, however, is the fact that the present data highlights that children in Year 6 are not reaching the goals set of by the Department for Education. The learning objectives outlined by the government stipulate that by Year 6, children should have developed a more sophisticated understanding of figurative language (Department for Education, 2013a); yet, the children contained within this sample do not demonstrate this. Even though this is apparent for all pupils, it is more prominent for the children receiving Pupil Premium (i.e., those from the lower SES groups).

Furthermore, not only did children's age significantly predict reaction times when selecting sentences referring to the same sense of a metonymic word, age also affected the type of responses provided by the children. Younger children were not only less likely to correctly clarify the intended sense of a metonym but were also significantly more likely to not even attempt to do so. These findings demonstrate a slight shift in metonymic understanding of children between the ages of 7 and 11 years-old with older children being more likely to know the intended sense of a metonym and/or attempting to give a response. Such a shift between these ages is likely due to the introduction of figurative language in schooling. The national curriculum for England suggests children should demonstrate some form of understanding of figurative language from Year 3, but that more sophisticated understanding should become apparent by Year 6 (Department for Education, 2013a). Consequently, the increases in knowledge and processing in the older years is likely due to the increase of metonymic encounters that take place in the school environments of the older children, thus increasing the likelihood that such senses are acquired into an individual's vocabulary.

6.4.2 Language Experience

Contrary to predictions, a child's level of exposure to English, as measured by either binary or continuous language classification, did not predict their performance on any component of the Metonymic Task. Despite this, language experience was found to be a significant contributor to the *types* of responses a child gave. To this effect, monolingual children were significantly more likely than bilingual children to provide a correct interpretation of the target word but describe it in the wrong sense. On the other hand, monolingual and bilingual children were equally likely to provide incorrect senses, ambiguous answers, correct answers, or not provide a response at all. Language exposure is known to be a substantial contributor to the rate of monolingual and bilingual children's language growth (Hart & Risley, 1995; Hoff, 2003; Pearson et al., 1997) with the rate of vocabulary learning being proportional to exposure (Chondrogianni & Marinis, 2011; De Houwer, 2009; De Houwer, 2007; Gathercole et al., 2013; Hoff et al., 2012; Thordardottir, 2011; Thordardottir et al., 2006). This is attributed to bilingual children experiencing less of each of their languages in comparison to monolingual children of the same age due to being required to divide their time between each of their languages. The present findings contribute to this body of literature by demonstrating that the same applies to metonymic words, with the reduced exposure to English that bilingual children receive negatively impacting their abilities to disentangle disparate senses of metonymic words.

Most interesting, however, is the finding that the models shown to fit the data best were those that included the continuous measures of language experience; namely, the Cumulative Input Score from the Leeds-Bradford Language Experience Questionnaire for Children (LeBLEQ-C). Though there was no significant contribution of either the binary or continuous measures of language experience, the continuous model had a lower Bayesian Information Criterion (BIC) value, indicating that language exposure is indeed likely to contribute to knowledge of metonyms, but a binary classification system may not be sensitive enough to capture such differences. This is in line with literature suggesting bilingualism to be better suited as a measure along a continuum, rather than a simple binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020). However, although more continuous measures of bilingualism may be preferable, this distinction remained non-significant in all models and thus findings must be interpreted with caution.

6.4.3 Expressive Language

Children who scored higher on the expressive language tasks of the Streamlined-CELF were quicker in correctly clarifying the intended sense of a metonym on the Open-Ended component of the task. This effect is attributed to the specific demands of the task. In particular, obtaining a correct score required children to sufficiently articulate the intended sense of a metonymic word, which is highly reliant on expressive language ability. Consequently, those who had higher communicative abilities were able to verbalise their knowledge of the metonym more easily, and more quickly, than children who had lower expressive abilities. As in previous chapters, it is this additional demand for higher expressive language ability that could have contributed to the lower accuracy rates in the Open-Ended component of the Metonymic Task compared to the Multiple-Choice component.

6.4.4 Reading Ability

Children who scored below average on their KS1 Reading SATs performed significantly worse on the Open-Ended component of the Metonyms task. While this is a common finding in the polysemic literature in terms of metaphoric knowledge (e.g., Smith Cairns et al., 2004), the association between metonymic understanding and reading ability is a novel finding within the literature. Increased exposure to language, in general, results in increased comprehension (Bialystok et al., 2010). With the high prevalence of polysemes (including metonyms) in the English language (77.9% of words in the Wordsmyth Dictionary, Parks et al., 1998), it would be intuitive to assume a high prevalence of metonymic works in children's literature. Consequently, though children who read more are likely to be stronger readers from the offset, these children are likely to encounter metonyms more frequently compared to those who read less. Increased exposure to metonyms is likely to aid in the comprehension of such ambiguities. Subsequently, when a metonymic word is encountered, it is readily understood, thus making reading easier.

Furthermore, it must be noted that the KS1 Reading SAT scores were a simple binary measure (indicating whether the children performed as expected of their age, or below) and a more rigorous comprehension measure of reading ability might have been more able to predict metonymic ability. As such, it is suggested that additional research is conducted to further research in this relationship. More research is also needed into the prevalence of metonymic words, and polysemes as a whole, in reading materials to which primary-school aged children are likely to be exposed. This would increase understanding of how knowledge of metonyms may be increased, or conversely, whether metonymic teaching can be included in interventions to increase the reading abilities of poor-comprehenders.

6.4.5 Sense of Target

The sense of the word contained within the target sentence (sense-1/sense-2) was demonstrated to significantly impact the accuracy of scores in the Open-Ended component of the Metonyms task, with children being more able to correctly clarify, or define, the word when presented with its second sense. This second sense was the interpretation of the metonym that was different from the physical entity sense of the word. For instance, the second sense of the word "cup" described the space enclosed by the physical entity, such as the liquid inside (e.g., "I was thirsty and drank the whole cup"). Such a finding could imply that this second sense is better understood, or, more likely, it could be that this sense is easier to verbalise than the sense implying the physical entity.

In the example above, "I was thirsty and drank the whole cup," children were able to clarify their understanding of such a second sense by stating the thing which was drunk, for example, "I drank the cup of milk, water, of liquid." Clarification was more difficult, however, when the target sentence contained the first sense of the word. When attempting to clarify the sense used in the target sentence "I tripped and broke the cup" the children could only achieve a correct answer by describing the physical appearance or the make-up of the cup in some way; for example, by responding "the china that the cup was made form broke" or "the thing that you pour your drink in/that you drink from broke." Using a synonym would have been classed as ambiguous as it was unclear whether the child referred to the first or second sense of the metonym. Thus, it may be that the second sense was not necessarily easier to comprehend, but rather it was easier to clearly explain without ambiguity, leading to higher accuracy rates.

6.4.6 Conclusion

Overall, the most notable factors that contributed to a child's understanding of metonymic words included reading ability, expressive language ability and specific sense of the target metonym. Though the understanding of metonyms did not appear to substantially increase between the ages of 7-11 years, older children responded faster to the stimuli in than their younger peers. This was, however, highly moderated by a child's socio-economic position, a novel finding within research on children's metonymic understanding.

Although language experience did not have a significant impact on accuracy rates of the Metonymic Task, the type of answer a child gave was significantly impacted by their language experience classification (i.e., whether they were monolingual or bilingual). In line with findings of previous tasks of this thesis (the Homonyms task in Chapter 4 and the Metaphors task in Chapter 5), models including a continuous measure of language experience fit the data better than models only including a binary language classification. This suggests the propensity of metonymic ability to be associated with language experience to some extent, but not to the degree of other, more influential factors such as SES.

It must be noted, however, that despite only including in the task metonyms for which both senses were reported to have been learnt before the age of 10 years-old (as determined by the Norming Study, Experiment 2, Chapter 3), there were still a substantial number of instances where children were unable to clarify the specific sense of a metonymic word. This could be attributed to the low SES contained within the sample; higher scores might be expected in higher SES samples, or older children who have had more experience of language in general. In addition, past research with adults demonstrated that it is very difficult for participants to separate the two closely related senses of metonymous words (e.g., Williams, 1992). Thus, it is possible that although the children in the present study perceived the differences between the two senses, they had difficulty verbalising them and separating completely the two senses. In conclusion, then, the current study adds to existing literature and contributes substantially to the limited literature investigating children's understanding of metonymic words in bi/multi-lingual populations. The next chapter looks at the final trope of ambiguity to be explored within this thesis, idiom, and the extent it is impacted by the same variables as homonymy, metaphor and metonymy.

7 The Effect of Language Experience on the Comprehension of Idioms in Primary-School aged Children

Idioms are a form of figurative (non-literal) language where the overall meaning of the phrase cannot always be derived from the meaning of the individual words of which it is comprised. Essentially, they are “phrases that are more than the sum of their parts” (Nordmann & Jambazova, 2017, p.198). For example, the correct intended meaning of the common British-English idiom ‘*to kick the bucket’* is ‘*to die*.’ This meaning, however, is very different from the literal interpretation of a person physically kicking a cylindrical container.

Although idioms are present in all languages, the English language is particularly rife with these expressions (Brenner, 2011; Zyzik, 2011), with figurative utterances constituting approximately 36% of a child’s and 25% of an adult’s exposure to language (Lazar et al., 1989; Van Lancker-Sidtis & Rallon, 2004). Idioms are consequently deemed to be relatively ubiquitous in everyday interactions, present in many books, songs and quotations (Langacker, 1987). Their use is considered the most common expression of creativity in everyday communication with many of their forms also being included in poetry and creative writing (Carter, 2004).

Idioms are used mostly colloquially (Cacciari & Tabossi, 1988), produced with apparently little effort (Tabossi et al., 2008), and quickly understood (Tabossi et al., 2009). The presence of idioms in speech is deemed to be as numerous as the presence of individual words (Jackendoff, 1995), with the ability to interpret them being considered an essential skill in the use of language (De Caro, 2009; Sridhar & Karunakaran, 2013). Consequently, research into the comprehension of figurative language considers idiom comprehension a skill necessary in everyday communication, contributing to reading comprehension (Fusté-Herrmann, 2008), educational achievement (Cain et al., 2005), social participation (Laval, 2003; Swineford et al., 2014), and the ability to establish intimacy (Gerrig & Gibbs, 1988). Many also affirm awareness, interpretation and correct usage of idioms to be an indication of full language proficiency (Beloussova, 2015; Celce-Murcia, 2007; Ellis, 1997; Saleh & Zakaria, 2013; Yorio, 1989). However, there has been relatively little research regarding the knowledge of such units of language in child populations, especially the differences in understanding purely idiomatic constructs across monolingual and bilingual children. This is despite literature demonstrating idiomatic items to present significant challenges to both children (Cain et al., 2009) and adult second language learners (Barfield & Gyllstad, 2009; Cieślicka, 2015; Conklin & Schmitt, 2008; Steinel et al., 2007; Titone et al., 2015).

This is particularly worrisome as vocabulary is known to affect school outcomes (Hutchinson et al., 2003) and the knowledge that bilingual children are already comparatively disadvantaged to their monolingual peers upon entering the schooling system. While monolinguals have received 4-5 years of pure exposure in one language (in the UK, typically in English), this is not necessarily the case for bilingual children who may have limited, or even no exposure to English in the home environment prior to schooling. As idioms tend to not be explicitly taught within the educational system, those who do not understand such units of speech (such as bilinguals whose English exposure is likely to be less than that of their monolingual peers) which may be included in teachers' instructions are at risk of falling behind their peers. Consequently, greater understanding of idiomatic understanding may provide additional knowledge of how language is acquired in both monolingual and bilingual populations. The limited number of studies regarding idiomatic understanding in both monolingual and bilingual children at different developmental points, however, may be attributed to the lack of appropriate available measures.

Due to the lack of standardised assessments for lexical-pragmatics, many studies within the literature had to develop their own tasks for assessing such skills. However, due to limited methodological detail that is often included in this literature (as discussed in Chapter 3), these studies are not always possible to replicate outside the lab in which they were developed. The assessments that have been developed across the literature also employ vastly different methodologies, drawing, thus, on different stages of idiom competence. In particular, research suggests that there is a four-stage process in which idioms are fully integrated into an individual's repertoire; identification, interpretation, explanation and use (Caillies & Le Sourn-Bissaoui, 2013; Kerbel & Grunwell, 1998; Nippold & Taylor, 2002). While some studies may test the identification of idioms (Nikoalenko, 2004) other studies may test the use (e.g., Nesi et al., 2006), yet the findings are compared across research. This leads to inconsistencies in the ages in which individuals acquire idioms. For example, while there is a body of work suggesting idioms are acquired between the ages of 7-11 years (Cain et al., 2009; Gibbs, 1992; Levorato & Cacciari, 1999; Levorato & Cacciari, 1995; Nippold & Rudzinski, 1993; Nippold & Martin, 1989; Nippold & Taylor, 1995; Nippold et al., 2001), research by Nippold and Rudzinski (1993) and Nippold (2007) states acquisition may be even as young as 5-years-old.

Furthermore, it is not unusual for both assessment and research with idioms to include proverbs (e.g., "*Rome wasn't built in a day*" or "*the apple doesn't fall far from the tree*") within idiomatic assessments, without a clear distinction between the two when analysed (e.g., Nordmann et al., 2013). This compromises the construct validity of the results when conclusions are drawn by researchers using such measures. Many of the studies investigating idioms are also either heavily dependent on reading ability (e.g., Jaen, 2007), or focussed heavily on idiom explanation tasks which can place additional cognitive load on the participants due to the added need for sufficient communicative skill (e.g., Cain et al., 2009; Nippold et al., 2001). This becomes especially difficult when working with bilingual children; thus, such measures are not optimal for use with children who speak two languages.

Smith and Murphy (2015) aimed to develop a measure of English "Multi-word Phrase knowledge" (MWP). By Smith and Murphy's (2015) definition, MWPs range from entirely literal to completely non-literal units that regularly occur together. Consequently, the stimuli included a number of British-English idioms. Smith and Murphy (2015) assessed 108 children (68 monolinguals and 40 bilinguals) from 7-10 years-old using an adapted version of the Dutch CONTRIX measure (Revier, 2009). In this task, the children were required to finish a target sentence by creating a 2- or 3-word multi-word phrase from a 2×3 or 3×3 matrix of single words. For example, the participant might have seen the words 'break,' 'catch,' and 'pay' in the first column, and 'studies,' 'attention,' and 'work' in the second column and were required to create the semi-transparent MWP 'pay attention' by selecting its constituent parts. The findings of Smith and Murphy (2015) revealed the pattern of acquisition of MWPs to be similar to that of vocabulary knowledge in Year 4 and Year 5. In Year 4, the monolingual and bilingual children differed on all three components of the CONTRIX task: transparent, semi-transparent and non-transparent units. However, in Year 5 monolinguals outperformed their bilingual peers in all tasks except those regarding semi-transparent units. This was not the case in Year 3 pupils, however, with monolinguals and bilinguals performing statistically similar on all MWP tasks, but monolinguals outperforming bilinguals on administered vocabulary measures (such as the British Picture Vocabulary Scale; BPVS; Dunn et al., 1997). Smith and Murphy (2015) concluded that this was likely due to increased reading practices in this age group. This was in line with similar studies by Levorato and Cacciari (1995; 1999) who found children gradually improve in their understanding of both opaque and figurative language between 7-11 years old. More importantly, however, the findings reported by Smith and Murphy (2015) also seem to indicate that MWP knowledge develops qualitatively differently among monolingual and bilingual populations. In essence, while monolinguals' performance on the MWP task was correlated with other language measures, there was less association between the measures of bilingual children. They concluded that this is likely related to exposure, associations with reading ability, and/or the lack of instruction in academic settings alerting bilinguals to the presence of such units of speech.

These findings are in support of the literature showing greater MWP to be associated with faster language processing in both monolingual and bilingual adult populations (Conklin & Schmitt, 2008; Wray, 2008) with such units also being read quicker than novel phrases (Conklin & Schmitt, 2008; Siyanova-Chanturia et al., 2011). It would not be unreasonable to assume the same is true for idiomatic units, where the order of words is also fixed. For instance, the idiom "*to be a piece of cake*" is recognised to mean "*to be easy*", whereas the same cannot be said when similar words are used, for example, "*to be a slice of cake*" (Baker, 1992). As such, lack of idiomatic knowledge may slow both language processing and reading speed in much the same way as in fixed MWPs. However, the aims of this study by Smith and Murphy (2015) were predominantly focussed on MWPs and, thus, not all of the stimuli were pure idioms, leaving the body of literature focussing on both monolingual and bilingual child populations to be relatively limited.

Following the assumptions of the Graded Salience Hypothesis, individuals are more likely to know an idiom if they hear it frequently (Giora, 1997, 1999). However, frequency of exposure to English idioms is likely to be reduced in bilingual populations due to them receiving less exposure to each of their languages compared to monolingual children of the same age. Being unable to devote as much time to becoming proficient in each of their languages as they would if they were only learning one, bilingual children are required to divide their time between two languages. As a result, bi-/multilingual children acquire each of their languages slower than their monolingual peers (e.g., Hoff et al., 2012). It is well-established that this exposure to language impacts growth of single word vocabulary in bilingual child populations (Kimbrough Oller et al., 2007; Thordardottir et al., 2006), yet little is known about idiomatic phrases.

7.1 The Current Study (Experiment 6)

This chapter aimed to investigate idiomatic comprehension in both monolingual and bilingual, primary-school-aged children and the factors that contribute to better understanding of idioms. Given the lack of standardised measures, a novel idiomatic task, similar to that of the homonymic and metaphoric tasks, discussed in Chapters 4 and 5 respectively, was developed whereby children were required to respond to an idiom in two ways. Firstly, they were asked to verbally explain the idiom in the open-ended component of the task, then they were asked to select the correct meanings of the idiomatic phrases from a series of options. This design was chosen in order to ensure the idiomatic task was as comparable as possible to the homonymic and metaphoric tasks. In addition, it was informed by previous research demonstrating that children find explaining idioms more difficult than answering multiple-choice questions on them (Arcara et al., 2019; Papagno & Caporali, 2007; Perlini et al., 2018) and more difficult to explain than comprehend in general (Levorato & Cacciari, 1995). Both response methods were, thus, employed to reduce the likelihood of ceiling effects occurring and allow for rich linguistic data collection. Furthermore, using both methods attested in the literature with the same materials and participants would allow direct comparisons between methods, leading to better understanding of idiom processing and comprehension by monolingual and bilingual primary-school children. Including different response methods also allowed for investigation of the multiple stages of idiom acquisition (Caillies & Le Sourn-Bissaoui, 2013; Kerbel & Grunwell, 1998; Nippold & Taylor, 2002). In the current study all four stages of idiom comprehension (i.e., identification, interpretation, explanation and use) were assessed to an extent, though the focus was on the first three stages.

Assessing the various stages of idiom acquisition may allow for the disentanglement of the inconsistencies found across the current literature regarding the age in which idioms are first acquired. As previously mentioned, such disparities could be attributed to the stage of idiomatic acquisition captured in the different forms of idiom assessment employed.

However, although many aspects of the Idioms task were designed to be as consistent as possible with both the Homonymic and Metonymic tasks, a primary difference with the Idioms task was that the targets were presented in the context of a sentence. The study was designed in this way due to the literature stating context to be a key factor in idiom comprehension (Libben & Titone, 2008; Titone & Libben, 2014) and concerns that without such a vital indicator, comprehension would be low across all children. Furthermore, when idioms are presented in the classroom environment, though they are not often explained, they are seldom used in isolation (Ezell & Goldstein, 1991). Thus, including the idioms in the context of a sentence allowed for a more naturalistic design.

Due to the national curriculum for England stating that children should demonstrate an understanding of figurative language, including idioms, from Year 3, with the development of more sophisticated understanding apparent by Year 6 (Department for Education, 2014c), the current study was tested in these two age groups with older children expected to excel. The relationship between age and Socio-economic status was also investigated, as in previous chapters. Again, the receipt of Pupil Premium was used as a proxy for SES grouping.

The current study also investigated links between the purely idiomatic constructs and other measures of language. This included the Receptive and Expressive Language scores of the Streamlined-CELF and the Total Score of the LeBLEQ-T to measure teacher perceptions of a child’s linguistic ability. The relationship between oral language (receptive and expressive language) was attested by Smith and Murphy (2015) who found positive correlations between both expressive and receptive language across year groups 3, 4 and 5. The present study aimed to further these findings by exploring this relationship in both the monolingual and bilingual children.

As for the previous experiments, it was hypothesised that a child’s Cumulative Language Input (CLI) score, would be a more robust predictor of idiomatic ability, due to bilingualism being suited better to measures along a continuum, rather than a binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020), and research demonstrating bilinguals to struggle with figurative tasks more than their monolingual peers (Barfield & Gyllstad, 2009; Cieślicka, 2015; Conklin & Schmitt, 2008; Steinel et al., 2007; Titone et al., 2015). Associations between idiomatic understanding and gender, academic attainment and working memory were also considered. Finally, as with previous studies contained within this thesis, the stimuli were carefully controlled for, with considerations of AoA and decomposition being paramount to the idioms contained within the study.

7.2 Methods

7.2.1 Participants

Participants were the same as those in Chapter 4 (Experiment 3).

7.2.2 Materials

The experimental task had four components to test participants' knowledge of idiomatic phrases. Firstly, participants were presented with an idiomatic phrase, both in context and in isolation, and were instructed to explain the meaning of the idiom. Secondly, they partook in a series of three multiple-choice tasks. These tasks presented the child with a figurative (target) phrase and instructed them to pair it with (1) a literal phrase of the same meaning, or (2) another figurative phrase of the same meaning. The third multiple-choice task required the participants to match a *literal* (target) phrase with a figurative phrase of the same meaning. There were always four response options of which only one was correct. Details of how the further three response options were selected are described in Section 7.2.2.4 (Response Options). Children were instructed to respond using the mouse pad of the laptop.

Twenty-four idiomatic (i.e., experimental) phrases were selected from the list of words surveyed in Experiment 2 (Chapter 3). All were multi-word utterances that contained ambiguity at the phrase-level (rather than containing ambiguity at the word-level, as explored in the Homonymy, Metaphor, and Metonymy experiments of the previous three chapters). Idioms were selected for inclusion in the current study on a number of parameters. The first parameter considered was the Decomposition Survey (Experiment 1, Chapter 3). Employing the method of Koleva, Mon-Williams and Klepousniotou (2019), using a 7-point Likert scale, idioms with a decomposition score of four or above were considered decomposable, while those with a score below four were considered non-decomposable. Equal numbers of decomposable and non-decomposable idioms were selected (i.e., twelve of each).

The final list contained only idioms which were reported to have been known by the majority of the respondents. If a sense of an idiom was reported to have been unknown more than once, the idiom was excluded from the list of potential stimuli. Furthermore, it was ensured that the ages at which the idioms were learnt spanned a wide range (3-11 years) to ensure no floor- or ceiling-effects were obtained. All stimuli can be found in Appendix F.

To ensure results from the Idiomatic task would be comparable across idiom types, equal numbers of decomposable and non-decomposable idioms were selected. Firstly, the AoA of the decomposable and non-decomposable idioms were matched (M_{Decomp}= 8 years, 11 months, SD_{Decomp}= 8 months; $M_{Non\text{-}dec}$ = 8 years, 5 months, $SD_{Non\text{-}dec}$= 1 year, 0 months) as were the number of words in the experimental/control phrase (M_{Decomp}= 5.08, SD_{Decomp} = .90; $M_{Non\text{-}dec}$ = 5.17, $SD_{No\text{-}dec}$= 1.34) and the average number of letters in each word (M_{Decomp}= 3.19, SD_{Decomp} = .33; $M_{Non\text{-}dec}$ = 3.28, $SD_{No\text{-}dec}$= .46). Finally, there was no statistically significant difference between the decomposable and non-decomposable idioms in terms of familiarity, as determined by Koleva, Mon-Williams and Klepousniotou (2019), (M_{Decomp}= 4.54, SD_{Decomp} = 1.06; $M_{Non\text{-}dec}$ = 4.70, $SD_{No\text{-}dec}$= .76). All p-values were above .05.

7.2.2.1 Control (literal) Phrases

A list of 24 control phrases were created by taking idioms surveyed in Chapter 3 and creating literal versions of them. For example, '*to hit the hay'* became '*to go to bed*.' It was ensured that the control phrases and the experimental idioms were matched on both the number of words in the phrase ($M_{Experimental}$= 5.13, $SD_{Experimental}$ = 1.12; $M_{Control}$ = 4.79, $SD_{Control}$= 1.02) and the average number of letters in each word ($M_{Experimental}$= 3.29, $SD_{Experimental}$ = .47; $M_{Control}$ = 3.24, $SD_{Control}$= .96) with no statistical differences between the items (all ps > .05).

7.2.2.2 Stimuli Lists for the Idiomatic Task

Due to the large number of stimuli pooled, the words were split into two lists, List-A and List-B (a list of the stimuli can be seen in Appendix A.). The use of these two lists left the tasks identical in terms of format and instructions, while the target items differed. This allowed for a large number of idioms to be tested but for administration time to remain relatively short. This resulted in 24 items in each list; 12 experimental trials and 12 control trials. The two lists were matched on the parameters outlined above (all p-values above .05), namely, AoA of the idioms ($M_{List\text{-}A}$= 8 years, 8 months, $SD_{List\text{-}A}$ = 10 months; $M_{List\text{-}B}$ = 8 years, 9 months, $SD_{List\text{-}B}$= 9 months), number of words in the phrase ($M_{List\text{-}A}$= 4.63, $SD_{List\text{-}A}$ = 1.35; $M_{List\text{-}B}$ = 4.92, $SD_{List\text{-}B}$= .97) and the average number of letters in each word ($M_{List\text{-}A}$= 3.32, $SD_{List\text{-}A}$ = .63; $M_{List\text{-}B}$ = 3.21, $SD_{List\text{-}B}$= .87). There was no statistically significant difference between the familiarity of the idioms in the two lists, as determined by Koleva, Mon-Williams and Klepousniotou (2019), ($M_{List\text{-}A}$= 5.03, $SD_{List\text{-}A}$ = .79; $M_{List\text{-}B}$ = 4.18, $SD_{List\text{-}B}$= .91).

Finally, to allow for inter-rater reliability, and generalisation across lists, each list contained a 10% cross over. Consequently, three items (one decomposable idiom, one non-decomposable idiom and one control trial) were taken from List-A and added to List-B. Likewise, three trials from List-B were added to List-A, creating Version-A and Version-B so the same six items were seen across both versions. This resulted in participants being exposed to 27 trials; 13 controls, seven decomposable idioms and seven non- decomposable idioms.

7.2.2.3 Target Sentences

Sentences were then created to incorporate the experimental idioms/control phrases. Sentences varied from 6 to 14 (M= 11.50, SD= 2.26) words in length and all contained the experimental idiom/control phrase in the sentence-final position. The variance was high due to the variability in the number of words in the idiom itself.

Each target sentence placed the idiom in context. For example, the idiomatic expression ‘to get off the ground’ was placed in the context of ‘They couldn’t get the business off the ground’. Target sentences were matched on both the number of words contained within the sentence and the average number of letters each word contained. The means and standard deviations for these variables are reported in Table 51.

Table 51

Means and standard deviations (M; SD) outlining characteristics of the Target Sentences in the Idiomatic Task.

	Control	Experimental	
		Decomposable	Non-decomposable
Words in each target sentence	11.21; 2.26	11.25; 2.01	12.33; 2.06
Letters in each word	3.94; 1.85	4.17; 2.08	3.70; 1.82

A series of t-tests revealed no significant differences between the control and experimental target sentences or between decomposable and non-decomposable idiom trials (all ps > .05).

7.2.2.4 Response Options

Four response options were created for each target sentence. These response options were sentences of two to nine words (M= 5.41, SD= 1.04). The large variability was the result of ensuring all response options made sense within the context of the target sentence. Furthermore, the figurative response options (unlike the literal or control options) were idioms and, thus, had a set number of words that could not be changed. For example, adding or subtracting words from an idiom such as “to be sitting on a nest egg” changes the meaning of the phrase entirely.

There were three types of response options due to the three different components of the Idioms task: Literal/control, Figurative and Inverse. Although all trials had a set of response options which were literal, half of the trials also had a second response option that was figurative (i.e., all response options were idioms). If the target was figurative, the response option set was referred to as Figurative; if the target phrase was literal, it was referred to as Inverse. All six of the crossover items previously mentioned when stratifying the stimuli lists into Version-A and Version-B had both literal and figurative response option trials.

In all trials, only one of the response options was correct. For the literal response option trials, the correct response was the sentence that conveyed the correct interpretation of the idiom. The other three response options fit within the context of the target (idiomatic) sentence (even though some expressions were more likely than others) but were foils. One of the foils was always the opposite meaning of the idiom. Another one of the foil options related to the literal meaning of the idiom in question but was not its literal interpretation. The final foil was simply plausible within the context of the sentence but was unrelated to the target idiom. For example, in the case of the idiomatic trial '*After a fizzy drink, Mum said I had "ants in my pants"'*, the four response options in the literal trials were:

Correct Response Option: *I was full of energy*

Opposing Foil: *I spilt my drink*

Literal Foil: *I had bugs all over me*

Unrelated Foil: *She could see my pants*

For the figurative response-option trials, the correct response was the idiomatic sentence that conveyed the same interpretation as the target idiom. Again, the other three response options were idiomatic expressions that fit within the general context of the target sentence (in some form) but were foils. All the foils were chosen from the Idiomatic Survey (Chapter 3) but had not been used as target idioms but were reported to be known by the majority of the survey respondents. If more than two respondents had reported to not know the idiom, it was not included as a response option. For example, in the case of the idiomatic trial '*After a fizzy drink, Mum said I had "ants in my pants"'*, the four response options in the figurative trials were:

Correct Response Option: *I was full of beans*

Foil 1: *I had pins and needles*

Foil 2: *I was taking the biscuit*

Foil 3: *My heart wasn't in it*

For the control trials, there was only one correct response option which was a rephrasing of the non-idiomatic phrase. As with the literal foils, there was an opposing foil and an unrelated foil. However, as the target sentence was already literal, there was no literal foil, but rather a related foil which re-used a key word in the non-idiomatic phrase. For example, in the case of the control trial '*After a long day, he couldn't wait to "go to bed"'* (taken form the literal interpretation of '*to hit the hay')*, the four response options were:

Correct Response Option: *He was really tired*

Opposing Foil: *He was wide awake*

Related Foil: *He hadn't made his bed this morning*

Unrelated Foil: *He was excited to go to the park*

Finally, in the Inverse response options, only one of the response options was correct again, namely the idiom that had the same interpretation as the literal (target) sentence. The other three response options fit generally within the context of the target sentence (in some form). All the incorrect responses were selected from the Idiomatic Survey (Chapter 3) but had been unused as target idioms. It was ensured that all idioms used were known by the majority of the survey respondents. For example, in the case of the control trial '*After a long day, he couldn't wait to "go to bed"'*, the four response options were:

Correct Response Option: *He couldn't wait to hit the hay*

Foil 1: *He couldn't wait to chew the fat*

Foil 2: *He had an axe to grind*

Foil 3: *He was getting itchy feet*

Response options were matched for the number of words in each response option and the average number of letters each word in the sentence contained. The means and standard deviations for these variables are reported in Table 52.

Table 52

*Means and standard deviations (M: SD) outlining characteristics of the Response Options the Idiomatic Task***.**

	Control	Experimental	
		Decomposable	Non-decomposable
Words in each response option	5.95; .93	6.01; .91	5.64; .90
Letters in each word	3.85; .52	3.99; .57	3.84; .32

A series of t-test showed no significant differences between the control and experimental response options or between the decomposable and non-decomposable response options ($p < .05$).

7.2.3 Procedure

7.2.3.1 Language and Cognitive Measures

Language and cognitive measures were the same as those described in Chapter 4 (Experiment 3).

7.2.3.2 The Idiomatic Task

The experiment was created using PsychoPy version 3.0 (Pierce, 2007) and presented to the children on laptops. Trials were presented in a random order, determined by PsychoPy, to prevent order effects. Participants were first given instructions to the task. It was explained to the children that they were about to participate in a game with idioms. They were told "*idioms are phrases that don't always mean the same thing as what you think, based on the words in the phrase*". Children were given the example of the idiom "*to give a hand*" and it was explained that when you ask someone for 'a hand' what you are really asking for is 'help'.

Participants were told they were to be tested on their knowledge of the idioms in three ways. Firstly, they were told that they would be presented with a sentence and would be asked what they think the sentence means (see examples above). They were instructed to respond verbally to the experimenter; this was the Open-Ended version of the task. Secondly, the participants were informed that they would then see four options of what the target sentence could mean and would be required to select the sentence that was correct in its meaning; this was the first of two Multiple-Choice questions. For some of the idioms, children were told they would also be asked to match the target idiom to one of four other idioms seen on the screen; this was the second, (figurative) multiple-choice question. In the case of the two multiple-choice trials, children were instructed to respond using the mouse pad of the laptop. Instructions lasted 57 seconds. Once the participants stated that they understood the task, and what was expected from them, they were permitted to click a "*Let's Practice*" button displayed on the screen. This allowed them to proceed to the first of the two practice trials.

In the open-ended component of the trial, participants were asked to rephrase a target sentence to show their understanding of the meaning of the idiom. All trials were presented in the same format as the practice trial of "*to give a hand"* which can be seen in Figure 17.

Figure 17

Presentation of the Open-ended question in the Idiomatic Task

"After struggling with the shopping, she asked her husband to give her a hand"

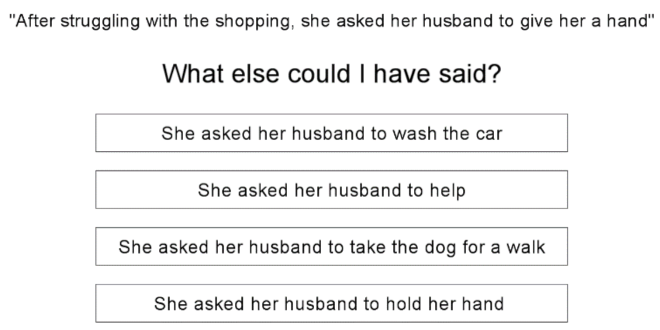
What does this mean?

"to give a hand"

Target sentences were presented in lower case to aid recognition, and the control phrase/idiom was always presented in isolation again to ensure participants knew which part to focus on. Participants were asked to respond verbally to the experimenter. The format of the multiple-choice component of the literal trials can be seen in Figure 18.

Figure 18

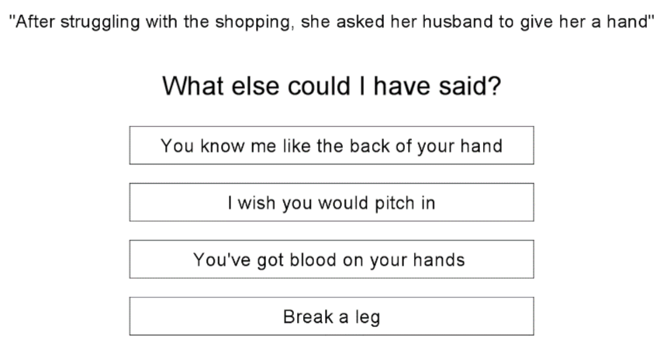
Presentation of the Multiple-Choice question in the Literal trials of the Idiomatic Task



Participants selected the response option with the same meaning as the idiom/control phrase. After a short pause (0.2 seconds) during which the screen displayed no text, the next practice trial began. In addition to the open-ended question and the literal multiple-choice question, this second practice trial also contained a figurative multiple-choice question. The format of the multiple-choice component of the figurative trials can be seen in Figure 19.

Figure 19

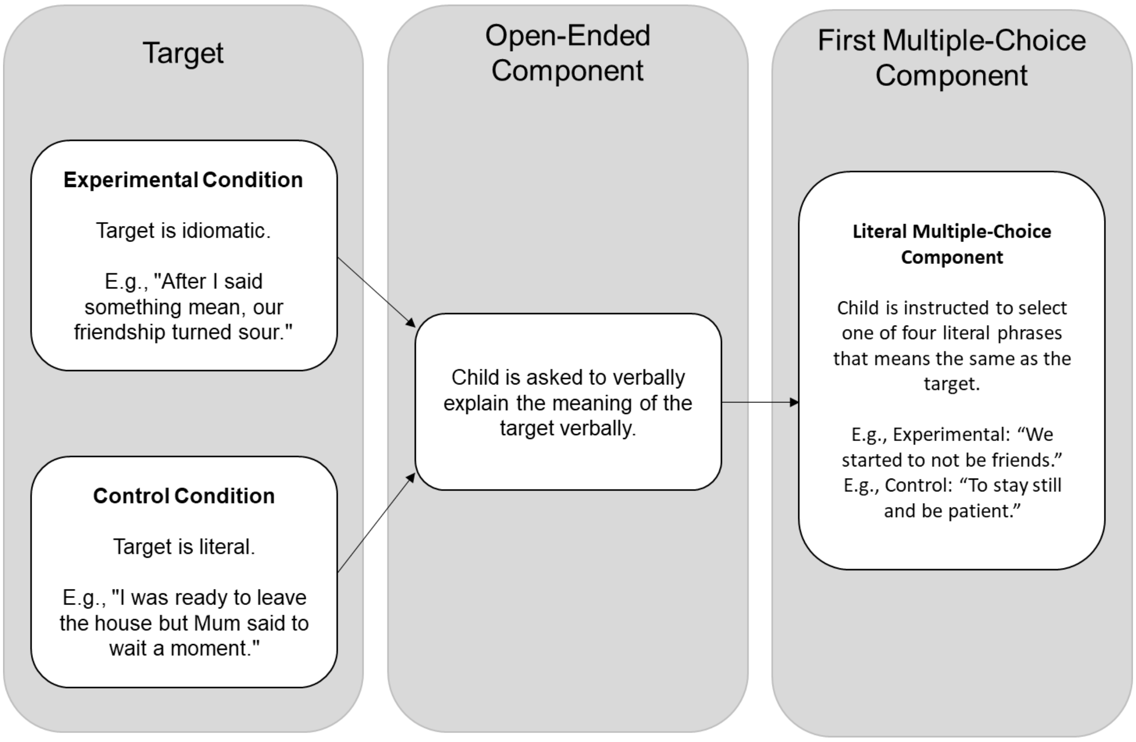
Presentation of the Multiple-Choice question in the Figurative trials of the Idiomatic Task



After completing the two practice trials, a second set of short instructions were given. Children were told that the practice trials were over, and were given an opportunity to ask any questions before the experimental trials began. These instructions lasted 11 seconds. After this point, the 'Let's Go!" button was presented allowing participants to proceed to the first 12 trials (comprised of six control trials, three decompositional trials and three non-decompositional trials). In the first half of the experiment, the children were required to make one multiple-choice selection after they completed the open-ended component (Figure 20).

Figure 20

Order of presentation of the components in the first half of the Idioms task

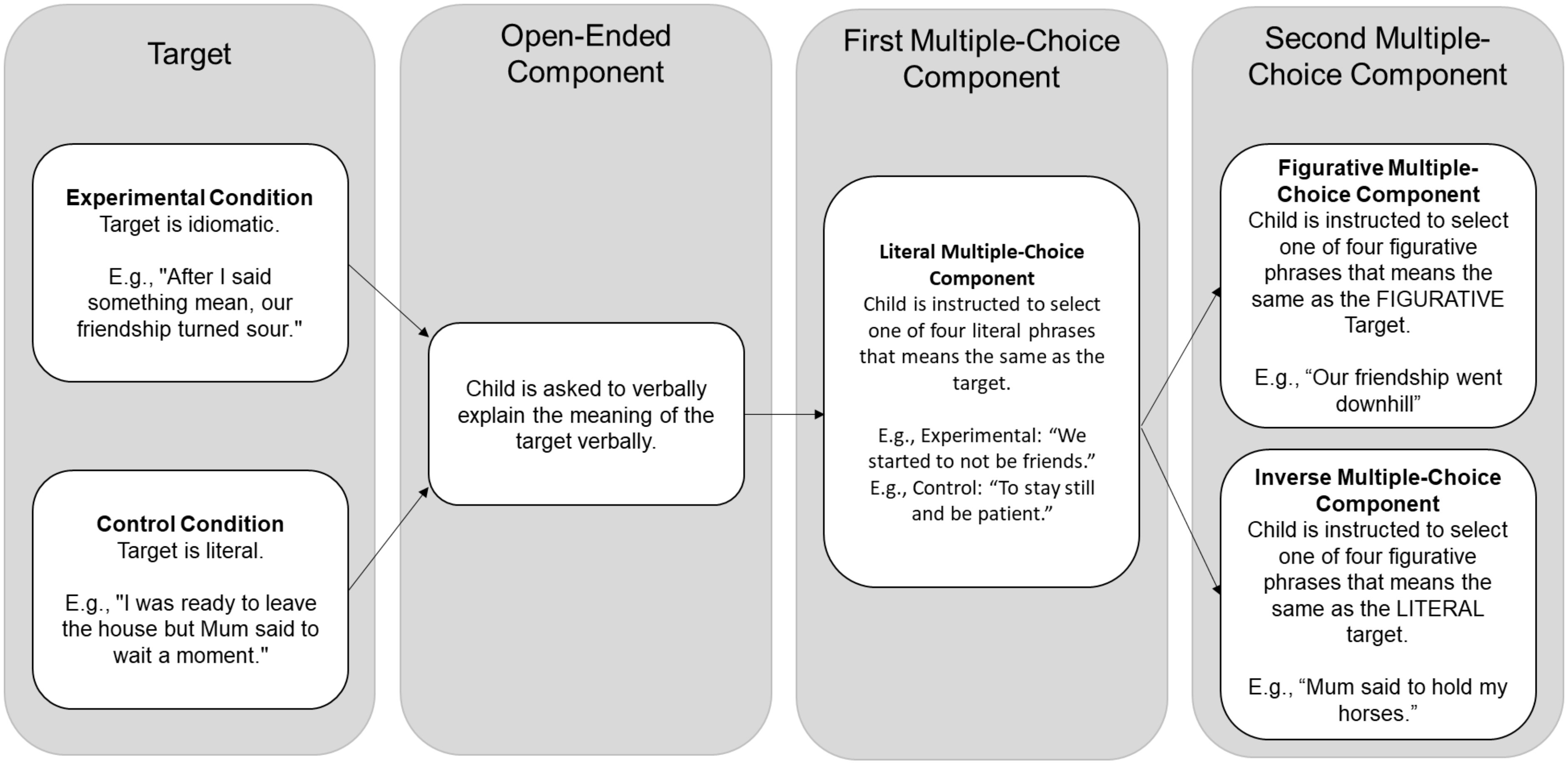


After these 12 trials, participants were then presented with a second set of instructions. These instructions marked the beginning of the trials that, in addition to the in the Open-Ended and Literal Response Option trials that were seen in the first half of the experiment, included the Figurative Response Option trials. The children were told they were again going to match the idioms to literal sentence, but additionally, they would also be asked to "*match the idiom to another idiom that means the same thing – just like we practiced at the beginning*". To ensure the children were not affected by the order in which they were exposed to the ‘Figurative’ and ‘Literal’ response types in this second half of the experiment, two further versions of the tasks were created. These were referred to as Order-1 and Order-2 (see Figure 21a and Figure 21b respectively). In Order-1, the children were given the literal response option trials first, followed by the figurative response option trials; the converse was true for Order-2.

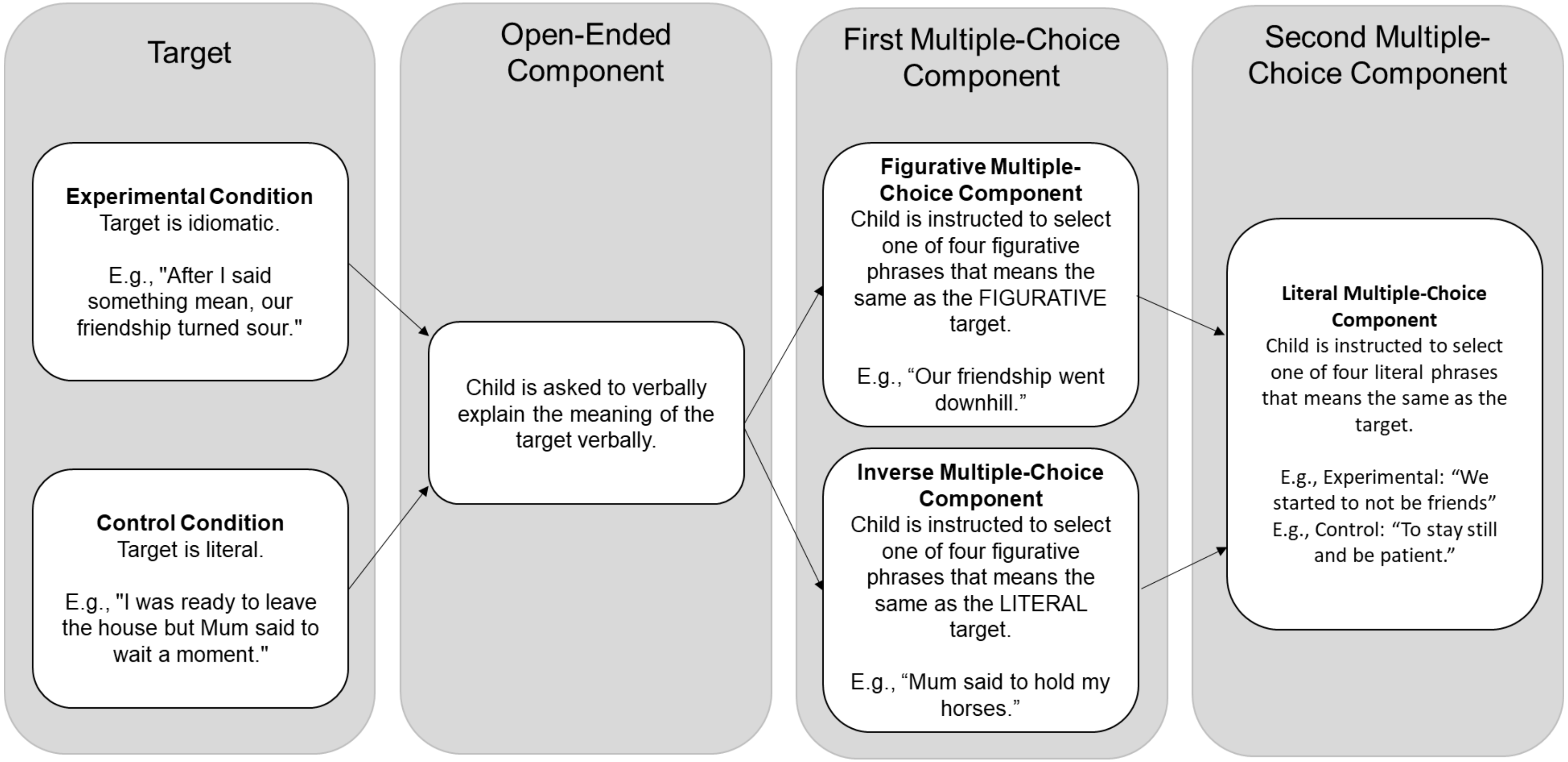
Figure 21

Order of presentation of the components in the second half of the Idioms task

(a) Order 1



(b) Order 2


Target
Open-Ended Component
First Multiple-Choice Component
Second Multiple-Choice Component
Experimental Condition
Target is idiomatic.
E.g., "After I said something mean, our friendship turned sour."
Control Condition
Target is literal.
E.g., "I was ready to leave the house but Mum said to wait a moment."
Child is asked to verbally explain the meaning of the target verbally.
Figurative Multiple-Choice Component
Child is instructed to select one of four figurative phrases that means the same as the FIGURATIVE target.
E.g., “Our friendship went downhill.”
Inverse Multiple-Choice Component
Child is instructed to select one of four figurative phrases that means the same as the LITERAL target.
E.g., “Mum said to hold my horses.”
Literal Multiple-Choice Component
Child is instructed to select one of four literal phrases that means the same as the target.
E.g., Experimental: “We started to not be friends”
E.g., Control: “To stay still and be patient.”

This second half of the experiment contained 15 trials; the six control trials, three decomposable trials and three non-decomposable trials, plus the three items from the opposing stimuli list which would allow for inter-rater reliability. In all cases, information was simultaneously presented orthographically and auditorily to account for differences in participants' reading abilities. The audio was recorded by a female, native British-English speaker using Audacity (version 2.1.2). The speaker had a neutral accent. The Idiomatic Task took approximately 30 minutes to administer. Accuracy of response and response time (in milliseconds) were recorded automatically by PsychoPy for both the open-ended and multiple-choice components of each trial. Children were taken from their class and tested individually in a quiet classroom, onsite at school. Although ethical approval from parents had already been obtained, the testing procedure was first explained to the children before the experiment began, and they gave their verbal assent to continue. During the Idiom and CNRep tasks, participants wore noise-cancelling headphones to listen to the audio and limit environmental distractions. Participants were assigned to one of the four experimental tasks, Idioms-A1, Idioms-A2, Idioms-B1, or Idioms-B2. Each child's assessment took a total of approximately 30 minutes to complete.

7.2.3.2.1 Scoring of the Idiomatic Task

7.2.3.2.1.1 Open-Ended Component

The Open-Ended component of the Idiomatic Task was coded in two ways with responses being assigned both a score and a code. Firstly, a child could either receive a score of 0 or 1. A child received a score of 1 if they were able to rephrase the idiomatic/control phrase to show understanding of its meaning. If they were unable to do so, they scored a 0.

Secondly, these scores were then coded. Scores of 0 were either coded as a '*Literal Interpretation'* (where the child did not understand the phrase to be figurative and gave a meaning for the sentence that implied they understood the term to be literal), as '*Incorrect/Incomplete'* (where the answer they gave was unrelated to either the literal or figurative interpretation or did not give enough information to imply understanding), or as '*Omission/Repetition'*, (where they refused to give an answer or repeated the sentence). For example, if a participant was given the idiom “We couldn’t get the new business off the ground” and responded with “*The new business was on the floor,*” this would have been coded as ‘*Literal Interpretation’*. Alternatively, responding “The new business was boring” would have been coded as ‘*Incorrect’.* Scores of 1 were simply coded as ‘*Correct’*.

7.2.3.2.1.2 Multiple-Choice Components

Responses to the Multiple-Choice component of the task were also scored using a 0-1 scale. A child received a score of 1 if they selected the response option which correctly rephrased the target sentence, or a 0 if they selected a response option that did not mean the same thing as the target sentence.

In the case of the Literal Multiple-Choice component, the scores were then coded to account for instances whereby children selected the foil which gave the literal interpretation for the idiomatic expression, As such, scores of 1 were coded as ‘*correct’* and scores of 0 were coded as either ‘*incorrect’* or ‘*literal*.’

Responses in the Figurative and Invers Multiple-Choice components were coded as either correct (a score of 1) or incorrect (a score of 0), as there was no literal option.

7.3 Results

Initially, a series of t-tests were conducted to determine whether there was an experimental effect. Analysis revealed accuracy rates to be significantly higher in the control condition than the experimental condition in the Open-Ended Component ($M_{Experimental}$= .20, $SD_{Experimental}$ = .40; $M_{Control}$= .60, $SD_{Control}$= .49; *t*(3186.87)= -25.79, *p*< .001), as well as between the control of the Multiple-Choice component (M= .76, SD= .43) and experimental conditions of the Multiple-Choice Component. This includes the Literal Multiple-Choice Component (M= .49, SD= .50; *t*(3422.12)= -17.15, *p*< .001), the Figurative Multiple-Choice Component (M= .43, SD= .50; *t*(1922.77)= -17.865, *p*< .001), and the Inverse Multiple-Choice Component of the task (M= .48, SD= .50; *t*(16.1.49)= -14.41, *p*<

.001).[1] Consequently, all analysis will henceforth only consider experimental trials.

As with previous experiments contained within this thesis, differences in accuracy rates across the two different components (Open-Ended/Multiple-Choice) were explored in the current Idioms Task. A series of t-tests revealed participants performed significantly better on all Multiple-Choice components, with the average score for the Open-Ended component (M= .20, SD= .40) being statistically lower than the Literal (M= .49, SD= .50; $t(3417.14) = -19.07$, $p < .001$), the Figurative (M= .43, SD= .50; $t(1785.80) = -12.68$, $p < .001$), and the Inverse Multiple-Choice Component of the task (M= .48, SD= .50; $t(1483.88) = -14.49$, $p < .001$)[1]. Additionally, an ANOVA revealed significant differences among the three Multiple-Choice versions [F(2, 3711]= 4.65, $p = .01$). A post-hoc Bonferroni analysis determined participants performed significantly better on the Literal-Multiple-Choice trials in comparison to the Figurative-Multiple-Choice trials ($p < .01$); there were no other significant differences.

Subsequently, a series of t-tests were conducted to investigate the differences between the two lists (List-A and List-B) that created the four different versions of the Idiomatic Task (i.e., Idioms-A1 and Idioms-A2 from List-A, and Idioms-B1 and Idioms-B2 from List-B). Despite the lists being controlled for AoA (as determined in Experiment 2, Chapter 3), familiarity (Koleva et al., 2019), number of letters in each word, and number of words in the idiomatic expression, participants performed significantly better across all tasks when exposed to List-B (M= .40, SD= .49) compared to List-A (M= .37, SD= .48; t(5490.85)= -2.10, p< .05). When stratifying this in the different response types, this pattern was observed in the Literal Multiple-Choice component ($M_{List\text{-}A}$= .26, $SD_{List\text{-}A}$= .50; $M_{List\text{-}B}$= .51, $SD_{List\text{-}B}$= .50; t(1790)= -2.11, p< .05) and the Open-Ended component ($M_{List\text{-}A}$= .15, $SD_{List\text{-}A}$= .40; $M_{List\text{-}B}$= .25, $SD_{List\text{-}B}$= .43; t(1710.11)= -5.20, p< .001). However, the effects were in the opposite direction for the Inverse component ($M_{List\text{-}A}$= .53, $SD_{List\text{-}A}$= .50; $M_{List\text{-}B}$= .42, $SD_{List\text{-}B}$= .50; t(893.63)= 3.45, p< .01) and non-

[1] The same effects were found when analyses were conducted separately for the four different versions of the Idioms Task: IdiomsA1, IdiomsA2, IdiomsB1 and IdiomsB2. In all cases, participants had statistically higher scores on the control trials in comparison to the experimental trials.

significant in the Figurative Multiple-Choice component ($M_{List\text{-}A}$= .43, $SD_{List\text{-}A}$= .50; $M_{List\text{-}B}$= .42, $SD_{List\text{-}B}$= .50; t(1022)= .32, p= .747).

Nevertheless, when focusing on the four items that spanned across both stimuli lists, there were no significant differences between the accuracy scores of the participants on three out of four instances. These items included two decomposable items which were “to kick the bucket” ($M_{List\text{-}A}$= .14, $SD_{List\text{-}A}$= .35; $M_{List\text{-}B}$= .17, $SD_{List\text{-}B}$= .37; *t*(382)= -.84, *p*= .403) and “to take the bull by the horns” ($M_{List\text{-}A}$= .38, $SD_{List\text{-}A}$= .49; $M_{List\text{-}B}$= .32, $SD_{List\text{-}B}$= .47; *t*(381.96)= 1.28, *p*= .203), as well as two non-decomposable items which were “to get off the ground” ($M_{List\text{-}A}$= .45, $SD_{List\text{-}A}$ = .50; $M_{List\text{-}B}$= .40, $SD_{List\text{-}B}$= .49; *t*(382)= .97, *p*= .331) and “ to have a mental block”. For this last idiom only, differences were observed between the two lists with children exposed to List-B scoring significantly higher ($M_{List\text{-}A}$= .30, $SD_{List\text{-}A}$= .46; $M_{List\text{-}B}$= .40, $SD_{List\text{-}B}$= .49; *t*(378.28)= -2.05, *p*= .041). Items were consequently interpreted as demonstrating good inter-rater reliability and all individuals were included in the same analyses, regardless of which specific version of the idioms task the children completed.

An effect of order of presentation of the trials in the accuracy rates of the participants was also observed. Participants’ accuracy rates in the Literal-Multiple-Choice component were statistically similar regardless of whether they had to match the idiom to a literal phrase first (i.e., Order-1; M= .50, SD= .50) or to a figurative phrase first (i.e.,Order-2; M= .46, SD= .50; *t*(1022)= 1.35, *p*< .177). However, in the figurative trials, participants were more accurate when they were required to match the idiom to a literal phrase first (M= .48, SD= .50) than to a figurative phrase first (M= .37, SD= .48; *t*(997.60)= 3.37, *p*< .005).

Descriptive statistics were first conducted regarding the *type* of response participants gave. Firstly, this was considered at a group level (including all participants), then the Year 3 children were compared to the Year 6 children. Finally, the monolinguals were compared to their bilingual peers (using the binary classification of language experience). Prior to all descriptive statistics, Levene’s test for equality of variances was conducted. In cases where this assumption was violated, degrees of freedom were adjusted and a statistic not assuming the homogeneity of variance was reported.

Accuracy and latency data were analysed using a series of Multilevel Linear Models, implemented in IMB SPSS Statistics (Version 26) using a Maximum Likelihood method. The dependent variable was either the accuracy or the latency of the response discussed separately below. As with previous chapters, the models included predictors of KS1 SAT Reading score, KS1 SAT Writing score, Streamlined-CELF Receptive Language Score (RLS), Expressive Language Score (ELS), Backwards Digit Recall, Total LeBLEQ-T score, gender and language experience classification (monolingual, bilingual) as fixed factors, as well as an interaction of Year Group and Pupil Premium. In the current experiment, list type (List-A or List-B) and decomposition type (decomposable/non-decomposable) of the target idiom were also included and, when exploring the trials where the participants were required to match the idiom to both a literal and figurate expression (i.e., the full trials), the order of presentation (Order-1, where the literal options were presented first, or Order-2, where the figurative options were presented first) was also included as a fixed factor.

The model was then run for the second time. Here, the same predictors were included except language experience classification (monolingual, bilingual) which was replaced with the child’s Cumulative Language Input Score generated from the LeBLEQ-C. The Bayesian information criterion (BIC) of the two models was then compared to determine the model with the best fit to the data. In all models, random intercepts were included for each participant. Responses to Open-Ended questions and Multiple-Choice questions are described separately below.

7.3.1 Model Comparison

Bayesian information criterion (BIC) values were compared across the models including the binary (monolingual vs bilingual) and continuous (Cumulative Language Input Score) language experience classifications. BIC values for all models are outlined in Table 53.

Table 53

Bayesian information criterion values for Idiom models including the binary (monolingual vs bilingual) and continuous (Cumulative Language Input Score) language experience classifications

		BIC	
		Binary	Continuous
Open-Ended Component			
	Accuracy	1144.27	1049.50
	Reaction Time	571.50	539.12
Literal Multiple-Choice Component			
	Accuracy	1835.21	1750.59
	Reaction Time	1183.86	1136.08
Figurative Multiple-Choice Component			
	Accuracy	1068.52	1019.99
	Reaction Time	762.19	718.644
Inverse Multiple-Choice Component			
	Accuracy	932.14	926.70
	Reaction Time	599.86	586.25

Literature considers the model with the lowest BIC preferable (e.g., see Vrieze, 2012). As can be seen in Table 53, the models that include the continuous, Cumulative Language Input Score have lower BIC values, indicating better fitting models. Consequently, only the findings of the continuous models will be discussed below.

7.3.2 Open-Ended Component

7.3.2.1 Response Type Analyses in the Open-Ended Component

First, specific response types were analysed (see Table 54). Responses were coded as incorrect when the child gave either an Omission or Repetition, gave an incorrect or incomplete response, or gave the literal interpretation of the idiomatic phrase. Responses were only recorded as correct when children provided an accurate interpretation that aligned with the figurative meaning of the idiom.

Table 54

Response types across experimental trials of the Open-Ended Component of the Idioms Task, stratified by age of child and language experience classification

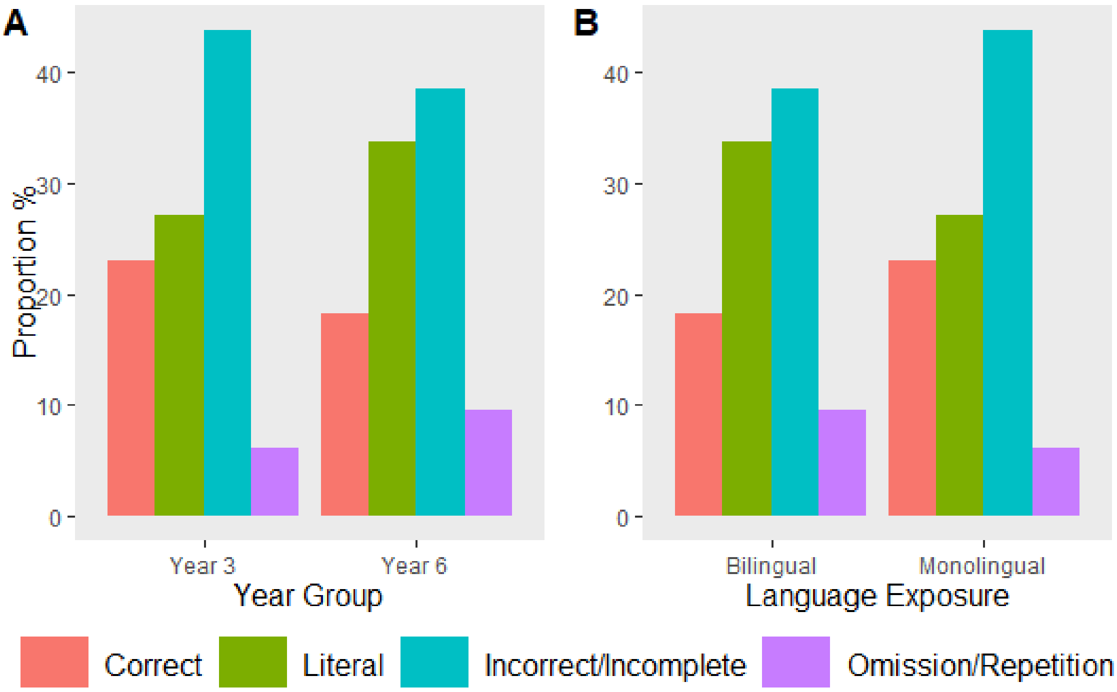
Response Type	Year 3		Year 6		Total
	Monolingual	Bilingual	Monolingual	Bilingual	
Correct	11.43%	10.54%	36.73%	25.68%	19.98%
Literal Interpretation	37.71%	45.36%	14.29%	22.62%	31.31%
Incorrect/Incomplete	42.29%	30.00%	45.58%	46.60%	40.40%
Omission/Repetition	8.57%	14.11%	3.40%	5.10%	8.31%

Overall, an ANOVA revealed significant differences in the types of responses given across all participants [$F(3, 509)= 76.61$, $p< .001$]. A post-hoc Bonferroni analysis determined participants to provide more incorrect or incomplete responses ($M= .40$, $SD= .19$) in comparison to literally interpreting the figurative idiom ($M= .31$, $SD= .21$, $p< .001$) to correctly interpreting the idiom ($M= .20$, $SD= .18$, $p< .001$) and to not responding to the question ($M= .08$, $SD= .13$, $p<. 001$). Children also provided more literal interpretations than correct responses, literal interpretations than omissions, and correct responses than omissions (all at the $p < .001$ level).

Furthermore, a series of t-tests revealed a number of significant differences between children in Year 3 and children in Year 6. Children in Year 6 were more likely to provide correct responses ($M = .29$, $SD = .17$) and incorrect responses ($M = .46$, $SD = .19$) compared to the younger children in Year 3 ($M_{Correct} = .11$, $SD_{Correct} = .11$, $t_{Correct}(100.38) = -6.73$, $p < .001$; $M_{Incorrect} = .35$, $SD_{Incorrect} = .18$, $t_{Incorrect}(126) = -3.51$, $p < .01$). However, children in Year 3 were more likely to interpret the figurative expression literally ($M = .42$, $SD = .19$), or not respond at all ($M = .12$, $SD = .15$), in comparison to children in Year 6 ($M_{Literal} = .20$, $SD_{Literal} = .16$, $t_{Literal}(126) = 7.32$, $p < .001$; $M_{Omission} = .05$, $SD_{Omission} = .08$, $t_{Omission}(95.64) = 3.46$, $p < .01$).This suggests that children develop increased idiomatic understanding with age and older children are more aware that the literal interpretation of a figurative item is not always plausible. Consequently, they will attempt to make sense of the figurative phrase, using other information such as context, even if this leads to formulating an incorrect interpretation. Percentages of the types of responses given can be seen in Figure 22A. Finally, there were no significant differences between the types of responses monolingual and bilingual children provided (see Figure 22B).

Figure 22

Types of responses given in the Multiple-Choice component of the Idioms Task, stratified by year (A) and language experience (B)



7.3.2.2 Analysis of Accuracy Rates in the Open-Ended Component

When including individuals' Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the Open-Ended trials of the Idioms task, the relationship between predictors and scores showed significant variance in intercepts across participants, $Var(u_{0j})=$ .003, $\chi^2(1)= 17.06$, $p< .01$. There was a significant effect of RLS [$F(1, 89)= 9.63$, $p< .01$] with children with higher receptive language abilities also demonstrating increased idiomatic knowledge. The stimuli list the child was exposed to also impacted accuracy rates, with children performing better when they were assigned to List-B rather than to List-A [$F(1, 89)= 17.25$, $p< .001$]. Children who received a lower Total LeBLEQ-T score (indicative of good linguistic ability) also achieved higher scores on this component of the Idioms task [$F(1, 89)= 5.45$, $p< .05$], as did children with lower Cumulative Language Input Scores [$F(1, 89)= 4.51$, $p< .05$] (indicative of being more akin to monolinguals than bilinguals). Interestingly, children with lower Working Memory score [$F(1, 89)= 10.17$, $p< .01$] also performed better on the idioms task. Finally, the decomposition type of the idiom significant impacted children's knowledge of the phrase's meaning with children showing greater understanding of decomposable idioms in comparison to non-decomposable idioms [$F(1, 1157)= 17.18$, $p< .001$]. There was also a significant interaction between School Year and Pupil Premium status [$F(3, 89)= 16.69$, $p< .001$] (further discussed below). Contributions of all other fixed factors can be seen in Table 55.

The interaction between School Year and Pupil Premium was then broken down by conducting separate multilevel models on those who were in receipt of Pupil Premium and those who were not. The models specified were the same as the main model, but excluding the interaction term of Pupil Premium and, instead, including simply a main effect of year. These models revealed that, for those who received the subsidiary, there was a significant effect of year [$F(1, 462)= 12.09$, $p< .01$, $b= -.27$, $SE= .05$, $t= -3.48$], indicating that children in Year 6 outperformed Year 3. The same effect was found for children who were not in receipt of the subsidiary, but to a greater extent; in particular, Year 6 children outperformed Year 3 children to a larger degree [$F(1, 56)= 37.01$, $p< .001$, $b= -.27$, $SE= .05$, $t= -6.08$]. This demonstrated that children in receipt of Pupil Premium do not advance in idiomatic skill between Year 3 and Year 6 to the same extent as their peers without the subsidiary do.

7.3.2.3 Analysis of Reaction Times in the Open-Ended Component

Analyses excluded errors (80.01% of trials) and responses that were 2.5 standard deviations (following recent literature also investigating child language, such as Commissaire et al., 2019; Grundy & Keyvani Chahi, 2017; Levy & Hanulíková, 2019; Persici et al., 2019) above/below each participant's mean (a further 1%, resulting in a total 81.01% trials excluded). .

The relationship between scores on the idioms task and predictors showed significant variance in intercepts across participants, $Var(u_{0j})= .014$, $\chi^2(1)= 7.79$, $p< .01$. Here, there was a significant effect of decomposition type only [$F(1, 192.162)= 8.09$, $p< .01$] with children responding significantly faster to non-decomposable items in comparison to decomposable items. Contributions of all other fixed factors can be seen in Table 55.

Table 55

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Open-Ended component of the Idioms Task when using a Continuous Measure of Language Experience

	Accuracy				*Reaction Time*			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	.22	.10	2.12	.029*	.25	.55	.46	.650
Cumulative Language Input Score	-.10	.05	-4.15	.000**	-.37	.25	-.15	.148
Gender (Male)	.02	.03	.88	.380	-.15	.12	-1.25	.219
KS1 Reading SATs (below average)	.05	.04	1.29	.199	.15	.21	.74	.462
KS1 Writing SATs (below average)	-.00	.04	-.04	.968	-.11	.19	-.57	.565
Receptive Language Score	.00	.00	3.10	.003**	.01	.01	1.04	.300
Expressive Language Score	.00	.00	.50	.621	-.00	.00	-1.35	.182
Working Memory Score	-.04	.01	-3.19	.002**	-.03	.06	-.53	.598
LeBLEQ-T	-.00	.00	-2.34	.022*	-.00	.00	-.14	.886
Year*Pupil Premium (Yr3, without Pupil Premium)	-.19	.04	-4.85	.000**	.12	.20	.62	.535

Stimuli List (List-A)	-.10	.02	-4.15	.000**	.08	.12	.69	.490
Decomposition Type (non-decomposable)	-.08	.02	-4.14	.000**	-.30	.10	-2.84	.005*

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

7.3.3 Literal Multiple-Choice

7.3.3.1 Response Type Analyses in the Literal Multiple-Choice Component

First, specific response types were analysed (see Table 56). There were two types of incorrect responses: when a child selected the literal interpretation of the idiom, or when they selected either of the other two incorrect options.

Table 56

Response types across experimental trials of the Literal Multiple-Choice Component of the Idioms Task, stratified by age of child and language experience classification

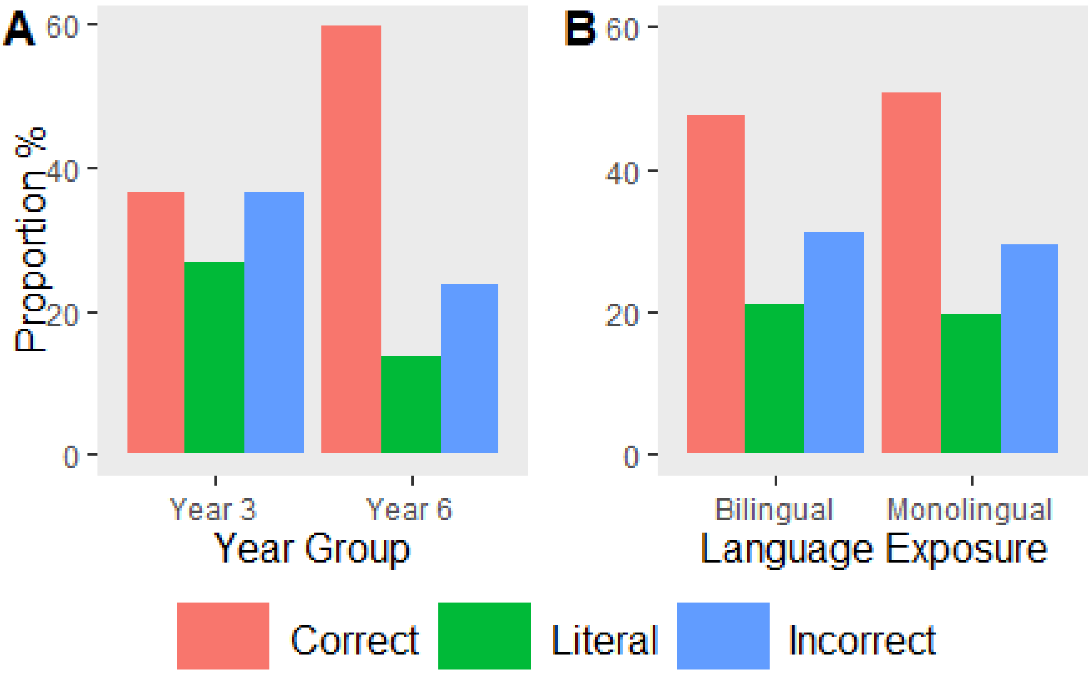
Response Type	Year 3		Year 6		Total
	Monolingual	Bilingual	Monolingual	Bilingual	
Correct	37.43%	36.07%	66.67%	58.84%	49.75%
Literal Option	25.71%	25.50%	12.59%	14.97%	20.19%
Incorrect Option	36.86%	36.43%	20.75%	26.19%	30.06%

Overall, an ANOVA revealed significant differences in the types of responses given across all participants [$F(2, 583)= 88.24$, $p< .001$]. A post-hoc Bonferroni analysis determined participants selected more correct options ($M= .49$, $SD= .22$, $p< .001$) than incorrect options ($M= .31$, $SD= .16$), and more incorrect options than literal options ($M= .21$, $SD= .13$, $p< .001$). Participants were also more likely to select the correct option than the literal ($p< .001$).

Error analysis, once again, revealed a number of significant differences between the types of responses between children in Year 3 and Year 6. A series of t-tests revealed children in Year 6 selected the correct response option more often than children in Year 3 ($M_{Year3}= .37$, $SD_{Year3}= .15$; $M_{Year6}= .61$, $SD_{Year6}= .21$; $t(113.57)= -7.73$, $p< .001$). In addition, Year 6 pupils were less likely to select literal interpretations (M_{Year3}= .27, SD_{Year3}= .13; M_{Year6}= .14, SD_{Year6}= .08; t(109.96)= 6.58, p< .001) and incorrect options (M_{Year3}= .37, SD_{Year3}= .14; M_{Year6}= .24, SD_{Year6}= .16; t(126)= 4.70, p< .001). This can be seen in Figure 23A. There were no statistical differences between the monolinguals and bilinguals in terms of the options selected (see Figure 23B).

Figure 23

Types of responses given in the Literal Multiple-Choice component of the Idioms Task, stratified by year (A) and language experience (B)



7.3.3.2 Analysis of Accuracy Rates in the Literal Multiple-Choice Component

When including individuals' Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the Literal Multiple Choice component of the Idioms task, the relationship between predictors and scores showed significant variance in intercepts across participants, $Var(u_{0j})$= .003, $\chi^2(3)$= 25.99, *p*< .01. Here, there was a significant effect of decomposition [F(1, 1155.65)= 25.97, *p*< .001] with children's accuracy rates being higher for decomposable items compared to non-decomposable items. Cumulative Language Input Score also made a significant contribution [$F(1, 87.67) = 3.94$, $p = .05$] with children with lower CLI scores (i.e., children who were less bilingual and more akin to monolinguals) performing better on this component of the Idioms task. Finally, there was a significant interaction between School Year and Pupil Premium status [$F(3, 88.76) = 14.67$, $p < .001$]. Contributions of all other fixed factors can be seen in Table 57.

When the interaction between School Year and Pupil Premium was explored further (as above), analyses revealed that, for those who received the subsidiary, there was a significant effect of Year [$F(1, 33.91) = 10.54$, $p < .01$, $b = -.21$, $SE = .06$, $t = -3.25$], indicating that children in Year 6 outperformed children in Year 3. The same effect was found for children who were not in receipt of the subsidiary, but to a greater extent; Year 6 children outperformed Year 3 children to a larger degree [$F(1, 56.34) = 44.37$, $p < .001$, $b = -.34$, $SE = .05$, $t = -6.66$]. This demonstrated that children in receipt of Pupil Premium do not advance in idiomatic skill between Year 3 and Year 6 to the same extent as their peers without the subsidiary do.

7.3.3.3 Analysis of Reaction Times in the Literal Multiple-Choice Component

Analyses excluded errors (51.40% of trials) and responses that were 2.5 standard deviations above/below each participant's mean (a further 1.22%, resulting in a total 52.62% trials excluded). The relationship between scores on the idioms task and predictors showed significant variance in intercepts across participants, $Var(u_{0j}) = .300$, $\chi^2(3) = 48.36$, $p < .01$. Here, there was a significant effect of decomposition type [$F(1, 482.94) = 10.69$, $p < .01$], with children being faster to respond to decomposable items in comparison to non-decomposable items. There was also a significant effect of order of presentation [$F(2, 532.02) = 19.27$, $p < .001$] with participants performing faster when literal trials were presented second, after figurative trials (i.e., Order-2; b= -.40, SE= .07, t= -6.03, p< .001) than when the literal trials were presented first, before the figurative trials. Contributions of all other fixed factors can be seen in Table 57.

Table 57

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Literal Multiple-Choice component of the Idioms Task when using a Continuous Measure of Language Experience

	Accuracy				*Reaction Time*			
	B	*SE*	*T*	*p*	*b*	*SE*	*t*	*p*
Intercept	.58	.13	4.62	.000**	-.46	.56	-.82	.413
Cumulative Language Input Score	-.11	.06	-1.98	.050	.11	.26	.43	.668
Gender (Male)	-02	.03	-.49	.623	-.27	.14	-1.92	.059
KS1 Reading SATs (below average)	-.08	.05	-1.67	.099	.10	.21	.48	.632
KS1 Writing SATs (below average)	.03	.04	.56	.574	.11	.20	.53	.596
Receptive Language Score	.00	.00	1.03	.307	-.00	.01	-.69	.495
Expressive Language Score	.00	.00	1.14	.257	.00	.00	1.16	.248
Working Memory Score	-.01	.02	-.69	.491	.13	.08	1.67	.099
LeBLEQ-T	-.00	.00	-.76	.450	.00	.00	1.40	.165
Year*Pupil Premium (Yr3, without Pupil Premium)	-.25	.05	-5.10	.000**	.21	.22	.97	.337

Stimuli List (List-A)	-.03	.03	-1.05	.298	.20	.13	1.58	.117
Decomposition Type (non-decomposable)	-.13	.03	-5.10	.000**	.15	.05	3.27	.001**
Order of Presentation (Literal first)	.02	.03	-.54	.592	-.11	.06	-1.85	.065

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

7.3.4 Figurative Multiple-Choice

7.3.4.1 Analysis of Accuracy Rates in the Figurative Multiple-Choice Component

When including individuals' Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the Figurative Multiple-Choice component of the Idioms task, the relationship between predictors and scores showed significant variance in intercepts across participants, Var(u_{0j})= .000, $\chi^2(1)$= 6.71, *p*< .01. Here, there was a significant effect of Total LeBLEQ-T score [F(1, 712)= 11.83, *p*< .01] with children whose teachers deemed them to possess superior linguistic ability (i.e., a lower score on the LeBLEQ-T) to have objectively better idiomatic knowledge in this component of the task. There was also a significant effect of Decomposition Type [F(1, 712)= 30.97, *p*< .001] with children being more accurate in selecting a matching figurative phrase when the target idiom was decomposable. Furthermore, Order of Presentation was significant [F(1, 712)= 6.74, *p*< .05] with children being more accurate when they were exposed to the literal trials first and the figurative trials second (i.e., Order-1). Finally, there was a significant interaction between School Year and Pupil Premium status [F(3, 712)= 2.99, *p*< .001]. Contributions of all other fixed factors included in the model can be seen in Table 58.

When the interaction between School Year and Pupil Premium was explored further, models revealed that, for those in receipt of the subsidiary, there was no significant effect of Year [F(1, 264)= .76, *p*= .384, *b*= -.07, *SE*= .09, *t*= -.87], indicating that performance of children in Year 3 did not differ from performance of children in Year 6. There was, however, a significant effect of Year in the children who were not in receipt of Pupil Premium [F(1, 448)= 7.56, *p*< .01, *b*= -.19, *SE*= .07, *t*= -2.75] with children in Year 6 showing more accurate performance than children in Year 3. This demonstrates that children in receipt of Pupil Premium do not advance in idiomatic skill between Year 3 and Year 6 as their peers without the subsidiary do.

7.3.4.2 Analysis of Reaction Times in the Figurative Multiple-Choice Component

Analyses excluded errors (57.03% of trials) and responses that were 2.5 standard deviations above/below each participant's mean (a further 1.27%, resulting in a total 58.30% trials excluded). The relationship between scores on the Figurative component of the idioms task and predictors showed significant variance in intercepts across participants, $Var(u_{0j})= .304$ $\chi^2(2)= 14.37$, $p< .01$. Here, there was a significant effect of Gender [$F(1, 84.96)= 6.44$, $p< .05$], with males performing significantly faster than females, and ELS [$F(1, 80.15)= 4.06$, $p< .05$], with those with higher expressive language abilities performing faster on the Figurative Multiple-Choice Component of the Idioms Task. Furthermore, Total LeBLEQ-T Score was a significant contributor to speed of response [$F(1, 99.21)= 5.27$, $p< .05$] with those scoring higher on the Teacher-report measure (an indication that their teacher perceived their general language skills to be poor) performing faster on the task, revealing a speed-accuracy trade-off for these children. Finally, there was a significant effect of Decomposition Type [$F(1, 219.48)= 12.22$, $p< .01$], demonstrating that children performed faster when presented with decomposable items. Contributions of all other fixed factors can be seen in Table 58.

Table 58

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Figurative Multiple-Choice component of the Idioms Task when using a Continuous Measure of Language Experience

	Accuracy				Reaction Time			
	B	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	.40	.16	2.58	.010*	-.90	.71	-1.26	.210
Cumulative Language Input Score	.04	.07	.62	.539	-.14	.31	-.46	.646
Gender (Male)	-.01	.04	-.27	.785	-.43	.17	-2.54	.013*
KS1 Reading SATs (below average)	-.05	.06	-.87	.380	.31	.25	1.25	.215
KS1 Writing SATs (below average)	.08	.05	1.47	.143	.06	.24	.26	.798
Receptive Language Score	.00	.00	1.37	.172	-.01	.01	-.91	.364
Expressive Language Score	.00	.00	1.02	.307	.01	.01	2.02	.047*
Working Memory Score	-.03	.02	-1.67	.095	.16	.09	1.74	.085
LeBLEQ-T	-.00	.00	-3.44	.001**	.01	.00	2.30	.024*
Year*Pupil Premium (Yr3, without Pupil Premium)	-.15	.06	-2.40	.017*	.11	.26	.41	.683

Stimuli List (List-A)	-.01	.04	-.36	.719	.23	.16	1.49	.141
Decomposition Type (non-decomposable)	-.19	.03	-5.57	.000**	.28	.08	3.50	.001**
Order of Presentation (Literal first)	.10	.04	2.60	.010*	-2.7	.16	-1.69	.094

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

7.3.5 Inverse Multiple-Choice

7.3.5.1 Analysis of Accuracy Rates in the Inverse Multiple-Choice Component

When including individuals' Cumulative Language Input Score (as generated from the LeBLEQ-C) as a fixed factor in the model predicting accuracy on the Inverse Multiple-Choice component of the Idioms task, the relationship between predictors and scores showed significant variance in intercepts across participants, Var(u_{0j})= .006, $\chi^2(2)$= 44.54, *p*< .01. Here, there was a significant effect of ELS [F(1, 89)= 8.15, *p*< .01] with children with improved expressive language skills possessing superior knowledge on this component of the idiomatic task. Cumulative Language Input Score [F(1, 89)= 4.13, *p*< .05] was also a significant contributor with children with a lower CLI score (indicative of less exposure to a language other than English) demonstrating higher accuracy rates. Finally, there was also a significant interaction between School Year and Pupil Premium status [F(3, 89)= 3.82, *p*< .05]. Contributions of all other fixed factors included in the model can be seen in Table 59.

When the interaction between School Year and Pupil Premium was explored further, models revealed that for those who were in receipt of the subsidiary, there was a significant effect of Year [F(1, 231)= 11.98, *p*< .01, *b*= -.30, *SE*= .09, *t*= -3.46], indicating that children in Year 6 outperformed children in Year 3 in terms of accuracy on the Inverse portion of the Idioms Task. Though the same pattern was observed in the children who were not in receipt of Pupil Premium, the difference was not statistically significant [F(1, 56)= 3.43, *p*= .07, *b*= -.15, *SE*= .08, *t*= -1.85]. This demonstrates that although the ability to match a figurative expression to a literal phrase does not improve with age in children who do not receive Pupil Premium (possibly because they are already at a higher level), abilities do advance in those from lower SES groups that are in receipt of the subsidiary.

7.3.5.2 Analysis of Reaction Times in the Inverse Multiple-Choice Component

Analyses excluded errors (52.17% of trials) and responses that were 2.5 standard deviations above/below each participant's mean (a further 2.12%, resulting in a total 54.29% trials excluded). Here, although the relationship between scores on the Inverse component of the idioms task and predictors showed significant variance in intercepts across participants ($Var(u_{0j})$= .298 $\chi^2(2)$= 95.58, p< .01), there was no significant contribution of any one particular predictor. Contributions of all fixed factors can be seen in Table 59.

Table 59

Fixed factors included in the Multi-Level Model for Accuracy and Reaction Time of the Inverse Multiple-Choice component of the Idioms Task when using a Continuous Measure of Language Experience

	Accuracy				Reaction Time			
	B	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	.36	.18	2.01	.047	-.27	.64	-.42	.679
Cumulative Language Input Score	-.17	.08	-2.03	.045*	-.04	.28	-.16	.875
Gender (Male)	.01	.05	.18	.859	-.20	.15	-1.31	.194
KS1 Reading SATs (below average)	.00	.07	.03	.976	.20	.23	.86	.393
KS1 Writing SATs (below average)	.03	.06	.41	.681	-.07	.22	-.31	.755
Receptive Language Score	-.00	.00	-1.01	.313	-.01	.01	-1.06	.293
Expressive Language Score	.00	.00	2.86	.005**	.00	.00	1.05	.295
Working Memory Score	-.02	.02	-.79	.434	.09	.08	1.11	.269
LeBLEQ-T	-.00	.00	-.531	.597	.00	.00	1.06	.291
Year*Pupil Premium (Yr3, without Pupil Premium)	-.17	.08	-2.03	.045*	.25	.23	1.08	.283
Stimuli List (List-A)	-.08	.04	1.84	.069	.04	.14	.26	.795

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

7.4 Discussion

The present research aimed to investigate the idiomatic skills of primary school-aged children in order to understand the impact of bilingualism on the comprehension of idioms. The study also aimed to explore whether idiomatic competence could be predicted better using a continuous score of bilingualism rather than a binary classification (monolingual/bilingual). A novel idioms task was administered alongside measures of general linguistic ability and working memory (Streamlined-CELF) to Year 3 (7-8-year-old) and Year 6 (10-11-year-old) pupils in schools of relatively low socioeconomic status associated with the Born in Bradford cohort. The Leeds-Bradford Language Exposure Questionnaire for Children (LeBLEQ-C) was administered to record levels of bilingualism of the pupils on a continuous scale (rather than simply using the binary classification monolingual/bilingual). The Leeds-Bradford Language Exposure Questionnaire for Teachers (LeBLEQ-T) was also administered to obtain an indication of teachers' perceptions of children's linguistic behaviour in the classroom. Scores on KS1 Reading and Writing SATs were also obtained from the schools, as was information regarding Pupil Premium (which was taken as an indicator of disadvantage).

Primarily, it must be noted that the accuracy rates for all response formats were considerably low; 19.99% for the Open-Ended component, 48.60% for the Literal Multiple-Choice component, 42.97% for the Figurative Multiple-Choice component and 47.82% for the Inverse Multiple-Choice component. This contradicts original predictions considering that the stimuli were specifically selected for use with the particular age groups used in this study, as determined by the age of acquisition ratings in Chapter 3 of this thesis. However, similar findings were observed in the Homonyms task (Experiment 3, Chapter 4), Metaphors Task (Experiment 4, Chapter 5) and Metonyms Task (Experiment 5, Chapter 6) where children also achieved low levels of accuracy across tasks. Similar to the performance patterns observed in the previous Chapters of the thesis, this could be attributed to the lower SES of the participants included in the current research (compared to existing literature) and the well-known impact of deprivation on language development (Hoff, 2003; Johnson & Kossykh, 2008; Pan et al., 2005; Rowe, 2008). All three schools included in the current sample were considered to be in low SES neighbourhoods, as indicated by the Income

Deprivation Affecting Children Index (IDACI; Department for Communities and Local Government, 2015), with Bradford, as a whole, considered to be one of the most deprived cities in the UK (Ministry of Housing, Communities & Local Government, 2015). However, the findings could be further attenuated by the possible discrepancy between the SES of the children tested in the current study, and the general population sampled in the AoA norming experiment (see the discussion of Chapter 4 for a full explanation of this).

Accuracy rates across the two components of the Idioms task demonstrated that children found the Open-Ended component more difficult than any of the Multiple-choice components. This is in line both with predictions and past literature showing increased difficulties for verbal-explanation tasks compared to multiple-choice tasks when assessing figurative language understanding both in adults and children (Arcara et al., 2019; Perlini et al., 2018). For example, Arcara et al. (2019) found adults with brain trauma had significantly higher accuracy rates on multiple-choice tasks compared to tasks that required verbal explanation, highlighting the importance of response format when studying idiomatic competence. Additionally, the current study also demonstrated a significant difference in scores across Multiple-Choice components where participants performed significantly better on the Literal-Multiple-Choice trials in comparison to the Figurative-Multiple-Choice trials. The varying difficulty levels across the tasks was a purposeful element of the study design, with the different components of the task intended to capture different skills. Research (e.g., Caillies & Le Sourn-Bissaoui, 2013, Nippold & Taylor 2002, Grunwell & Kerbel 1998) suggests that there is a four-stage process in which idioms are fully integrated into an individual’s repertoire: identification, interpretation, explanation and use; all these stages were assessed (to some degree) within the current experiment.

Whilst all components in the current study draw upon the ability to identify idiomatic language (Stage 1), the Open-Ended and Literal- and Figurative-Multiple-Choice components also tap into the ability to interpret idioms (Stage 2). The ability to explain idioms (Stage 3) is assessed in the Open-Ended component and, finally, the ability to use idioms (Stage 4) is indirectly assessed in both the Figurative- and Inverse-Multiple-Choice components. Although the *spontaneous* use of idioms is not directly captured in the current methodology, children's ability to match literal phrases to figurative expressions could arguably be indicative of their ability to produce the idiom themselves, or at least possess the capacity to do so. In particular, when the children are presented with a literal expression knowing they will be required to match it to a figurative phrase, the highly proficient learners might be able to automatically think of the idiomatic equivalent without needing to see the response options (rather, seeing the figurative phrase they believed to be correct in the response options may simply confirm their prior knowledge). As such, if learners can respond to a literal expression with a figurative equivalent with ease, they may be more likely to be able to spontaneously use the idiomatic expression itself in everyday communication.

Thus, the inclusion of the different components assessing the separate processes of idiom acquisition allowed for rich, detailed data to be gathered whilst limiting the task's susceptibility to ceiling effects. In addition, it ensured the ability of the task to be used with a wide age-range. Consequently, the present study supports previous literature that found children show poorer performance on open-ended idiom tasks in comparison to multiple-choice tasks (Arcara et al., 2019; Perlini et al., 2018). However, this is the first study to the author's knowledge that extends such effects to both monolingual and bilingual populations and shows performance can differ depending on the type of multiple-choice question asked. This demonstrates that method of testing plays a significant role in the observed performance of idiomatic understanding. Thus, the importance of response format must be considered when studying figurative language such as idioms.

Finally, Linear Mixed Effects Models demonstrated various contributors to idiomatic understanding. Overall, there was a significant interaction between the age of the child/School Year and their SES as indicated by their entitlement to Pupil Premium where the effect of age was highly moderated by SES. In particular, children in Year 6 demonstrated increased understanding of idioms compared to the children in Year 3; however, the difference was greater for children who did not receive Pupil Premium, indicating that children from lower SES backgrounds do not advance in idiomatic understanding to the same degree as their peers from higher socio-economic strata. Furthermore, a lower Cumulative Language Input Scores (indicative of increased exposure to the

English language, and less exposure to any other languages, as measured using the LeBLEQ-C) predicted higher scores on the idioms task. Including this CLI score in the model was also shown to fit the data better than when including a simple binary classification of language experience (monolingual/bilingual). Children's expressive and receptive language abilities also significantly impacted their scores on the Idioms task, with higher language skill indicating better performance. This related not only to the objective scores children received on the Streamlined-CELF but also the subjective, teacher generated scores from the LeBLEQ-T. Idiomatic accuracy rates could also be predicted by the child's Total LeBLEQ-T Score, with lower scores on the teacher-reported measure (indicating better general language proficiency) predicting a higher score on the Idioms Task. The target idiom itself also impacted response rates, with decomposition type and stimuli list predicting performance. Overall, children were quicker, and more accurate, when exposed to decomposable rather than non-decomposable idioms. Finally, accuracy on the Figurative Multiple-Choice component on the Idioms task was predicted by the order of presentation; children were more accurate in correctly selecting a matching idiom when they were exposed to the literal component first and the figurative component second (i.e., Order-1). All impacting factors (e.g., age and language experience) are discussed separately below.

7.4.1 SES and Age of child/School Year

The effect of SES (i.e., the interaction between School Year and Pupil Premium) was evident in all components of the idioms task. In both the Open-Ended and the Literal Multiple-Choice components, a significant effect of School Year was found in both the children who received Pupil Premium and those who did not with children in Year 6 outperforming those in Year 3. Those who were not in receipt of the subsidiary, however, showed a larger increase in idiomatic understanding between the two year groups. Similarly, in the Figurative Multiple-Choice component, those in Year 6 outperformed children in Year 3; however, this increase was only significant in the children who did not receive the subsidiary. These findings demonstrated that the children in receipt of Pupil Premium do not advance in idiomatic skill between Year 3 and Year 6 to the same extent as their peers without the subsidiary. The findings suggest age to be a clear predictor of idiomatic awareness in children but this to be highly moderated by SES, with those from higher socio-economic positions outperforming those from lower strata with increasing age.

These findings oppose original predictions made for the current study which expected children in receipt of Pupil Premium to be outperformed by those without in Year 3, but this effect to be eliminated by Year 6 (as per the findings of Hutchinson et al., 2016). In contrast, the current study found the gap between children who received Pupil Premium and those who did not to increase in Year 6. Although these findings do not corroborate previous literature, they are compatible with the findings in the Metaphors Task (Chapter 5 of this thesis). This strengthens the position that these effects are potentially observed due to children from lower SES being required to direct their attentional resources on the development of other, more widespread skills needed that are vital to enable them to progress within the classroom environment. While the focus of such general skills may allow children to perform akin to their higher SES peers on a multitude of other tasks (e.g., Hoff & Tian, 2005; Hutchinson et al., 2016), it may divert attention from the development of their higher order, idiomatic skills which become hindered in relation to their higher SES peers.

It must be noted however, that the opposite effect was observed in the Inverse Multiple-Choice component of the task. Although there was an increase in accuracy rates between the year groups both in the children who received Pupil Premium and those that did not, only the difference between children who did receive the subsidiary was significant. As such, the children from the lower SES homes improved to a greater extent in comparison to their peers from slightly higher Socio-economic strata.

Furthermore, not only did idiomatic knowledge increase with age across all components, on average, but the age of a child also affected the type of responses provided. Older children were more likely to provide a correct response in the Open-Ended component compared to children in Year 3 and were also more likely to attempt to explain the idiom rather than not respond at all. However, this increased the likelihood of older children also providing an incorrect answer. Finally, older children were less likely to interpret the phrase literally. Thus, not only do older children have increased understanding of idiomatic expressions, but they also make more attempts to guess the meaning of idioms when they are unsure. Furthermore, older children were more aware that the literal interpretation of an idiom is not always plausible. In contrast, younger children did not make this distinction, and were more likely to interpret an idiom literally, even when that interpretation did not make sense within the context of a sentence. Similar findings were observed in the response-type analysis of the Literal Multiple-Response component of the task where older children were more likely to select the correct interpretation of the idiom but were also less likely to select the literal interpretation or a completely incorrect interpretation. Such response patterns support existing literature consistently finding that younger children give more literal responses in comparison to older children (Ackerman, 1982; Cacciari & Levorato, 1989; Levorato & Cacciari, 1995). Ackerman (1892) suggested children under the age of 9-10 years to be less aware of the intended, figurative meaning of an idiomatic phrase and incorrectly taking the literal interpretation of the phrase to be true; he described them to be "biased to the literal interpretations of utterances" (p.451). On the other hand, typically developing older children and adults are less predisposed to such preferences.

The current findings clearly demonstrate a shift in idiomatic understanding of children between the ages of 7 and 11 years-old with older children being more likely to correctly interpret an idiom and know when a direct literal interpretation is not plausible. This was in line with similar literature by Levorato (1999) and Levorato and Cacciari (1995) who found children 7-11 years old to be gradually improving in their understanding of figurative language. Such a shift between these ages is likely due to the introduction of figurative language in schooling (DfE, 2013) where the increase in exposure to idioms in the school environments of the older children likely cause an increase in idiomatic knowledge. The instances in which an individual encounters such an idiomatic unit increases the likelihood that the meaning will be acquired into an individual's vocabulary (Levorato & Cacciari, 1992).

Perhaps most importantly, however, the findings highlight the fact that children in Year 6 are not reaching the goals set out by the Department for Education. The learning objectives outlined by the government stipulate that by Year 6, children should have developed a more sophisticated understanding of figurative language (Department for Education, 2013a) of which idiomatic language is a part. However, the children in the present sample failed to demonstrate this. Though this was apparent for all pupils, it was more prominent for the children receiving Pupil Premium (i.e., those from the lower SES groups) who were slower and less accurate on the tasks. Thus, interventions which aim to improve idiomatic competence in primary-school-aged children should be predominantly focussed on children from lower socioeconomic strata who need more support than their higher-SES peers. These are the children who may benefit the most if such interventions are implemented.

7.4.2 Language Experience

Another interesting finding, was that, models that best fit the data included the continuous measures of language experience, the Cumulative Language Input Score (rather than a binary measure). This was determined using the Bayesian Information Criterion (BIC) value, where lower scores indicate better fit of the models. Lower BIC scores were found across all models containing the Cumulative Language Input Score compared to models containing the binary measure of language experience (monolingual/bilingual). These results were corroborated by the finding that there was no significant contribution of language exposure when including the binary measure of language experience in the model.

There was, however, a significant contribution of language experience when a continuous measure was included in the model. Here, a child’s level of exposure to English, as measured by their Cumulative Language Input Score from the LeBLEQ-C was able to predict accuracy in the Open-Ended, Literal Multiple-Choice and Inverse Multiple-Choice components of the Idioms task. In all instances, a lower CLI score, which was indicative of more exposure to the English language, and less to another language, was predictive of an increased ability in explaining the correct meaning of an idiomatic phrase. This is in line with literature suggesting bilingualism to be better suited to being measured along a continuum, rather than simply considered to be a binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020).

It must be noted that, contrary to the current findings, Smith and Murphy (2015) observed clear differences in accuracy rates between monolingual and bilingual children in their sample, in regards to transparent and semi-transparent idioms.

The different effects seen between Smith and Murphy (2015) and the current study may be explained, however, by pre-existing differences between the monolinguals and bilinguals in their sample. Namely, Smith and Murphy (2015) found bilinguals were also outperformed by monolinguals in all other administered tests of language ability, namely the Assessment of Comprehension and Expression Phrasal Verbs subtest (ACE; Adams et al., 2001), the British Picture Vocabulary Scale-II (BPVS; Dunn et al., 1997), and receptive and expressive sections of the Test of Word Knowledge (TOWK; Wiig & Secord, 1991). In contrast, there were no differences in the general language tasks between the monolingual and bilingual children in the current sample. Consequently, the differences attributed to language exposure in the research by Smith and Murphy (2015), could have been simple differences in the degree of their general linguistic ability which, in turn, impacted performance on the pragmatics tasks.

Given the significant influence a child's Cumulative Language Input Score has on their accuracy in three of the four components of the Idioms task and the finding that this continuous score of language better predicts idiom competency, it can be concluded that language exposure is indeed likely to contribute to knowledge of idioms. As a result, a continuous score of exposure, such as the one used in the present study, has proved to be more informative and better suited to capture any differences than the traditional binary classification (i.e., monolingual/bilingual) that may not be sensitive enough.

7.4.3 Expressive Language

Contrary to predictions, a child's expressive language skill, measured using the Expressive Language Score from the Streamlined-CELF, was not a significant contributor to accuracy of response on the Open-Ended component of the task. However, expressive language ability was a significant predictor in terms of accuracy rate in the Inverse Multiple-Choice component and response time in the Figurative Multiple-Choice component. These findings were in line with previous literature by Smith and Murphy (2015) who also found links between expressive language and Multi-Word Phrases (MWPs), many of which included idiomatic constructs. This is the first piece of research to the author's knowledge, however, to use expressive language ability directly as a predictor of idiomatic knowledge in bilingual child populations, using purely idiomatic units.

The relationship between higher expressive language skills and the increased accuracy rate observed in the Inverse Multiple-Choice component can likely be attributed to the notion that this component is underpinned by the ability of the child to produce idiomatic expressions themselves. This is the only component that provides the participant with a literal phrase and asks them to match it to one of four figurative expressions; in all other Multiple-Choice components, the child is provided with a figurative target. The ability to do this is likely linked to the ability to produce an idiom in everyday communication, a task that requires higher-order language skills (Levorato & Cacciari, 1995). Consequently, these results suggest that children who are more proficient in understanding idiomatic expressions are also better equipped to produce idioms themselves, a task which also relies on the child’s ability to coherently produce speech in general.

The reduced reaction time in the Figurative Multiple-Choice component can likely be explained using the same rationale. Matching two figurative expressions in one single process was the task in which participants demonstrated the highest accuracy rates, a finding that was more pronounced in the participants who were exposed to Order-2 of the Idioms task (i.e., where they were required to first match an idiom to another idiom with the same meaning, and then to the correct, dictionary (literal) definition of the phrase). Such a task requires the participants to first correctly interpret the idiom, and then to select a figurative expression that maps onto that interpretation, all in a single step. Consequently, the contribution of expressive language ability in predicting the speed of reaction of the Figurative Multiple-Choice component may stem from this component also tapping into the likelihood of a child being able to produce an idiomatic phrase themselves. As with the Inverse Multiple-Choice component, the child is still required to conjure a figurative phrase from a literal one, drawing on knowledge of language use. Consequently, it could be concluded that those who have higher communicative abilities may be able to produce idioms faster and more accurately than those who have lower expressive abilities.

7.4.4 Receptive Language

The ability of a child's Receptive Language Score to predict their accuracy on the Open-Ended component of the task is consistent with the view that receptive language ability is related to idiomatic competence. Smith and Murphy (2015) demonstrated individuals who had high receptive language scores better understood Multi-word phrases (MWP) than their peers with lower receptive skills in children from Years 3, 4 and 5. Though the current research supports such previous literature, it is the first, to the author's knowledge, to use receptive language directly as a predictor of idiomatic knowledge in bilingual child populations. Furthermore, though Smith and Murphy (2015) demonstrated a link between receptive language and MWP comprehension, this relationship was associative only and was based on the use of MWPs ranging from entirely literal (e.g., "*to catch mice*"), to those including metaphorical words (e.g., "*to catch a cold*"), to completely non-literal units that regularly occur together (e.g., "*to break the ice*"; i.e., idioms). Consequently, though the stimuli included a number of British-English idioms, the primary focus was not in figurative expressions. The present study, thus, extended past findings by using only highly controlled, purely idiomatic units.

Importantly, the present study employed a more robust receptive language measure compared to the relatively simplistic British Picture Vocabulary Scale (Dunn et al., 1997) that was used in Smith and Murphy (2015). Results were generated using the Receptive Language Score (RLS) of the Streamlined-CELF (described in Chapter 2) which provides a more comprehensive view of a child's receptive language ability, and is more indicative of speech encountered in a child's typical interactions. As a result, the present findings demonstrated that idiomatic understanding develops alongside a child's understanding of daily conversation.

7.4.5 Leeds-Bradford Language Experience Questionnaire for Teachers

The Total Score of the LeBLEQ-T significantly predicted children's accuracy rates in the Open-Ended component of the Idioms task, as well as both accuracy and speed of response in the Figurative Multiple-Choice component. In these cases, lower Total LeBLEQ-T scores, which indicate better general linguistic ability as evaluated by classroom teachers, predicted higher accuracy rates and faster response times.

The LeBLEQ-T quantifies a teacher’s perception of a child's linguistic ability, taking into consideration listening, speaking, reading and writing skills of children as well as more pragmatic abilities, providing, thus, a comprehensive indication of a child’s communicative competence within the classroom environment. Though the LeBLEQ-T is conceptually different from measures normally used within the literature which directly assess oral proficiency of children (such as the BPVS or various measures of the CELF), the present findings demonstrate that teachers’ perceptions of a child’s linguistic ability are a reliable alternative and support previous research demonstrating language skills to be a substantial predictor of idiomatic competence (e.g., Caillies, & Le Sourn–Bissaoui, 2012; Smith & Murphy, 2015). The current research highlights how teacher-report methods can be used in place of these more direct measurements whilst still providing rich information on the language skills of a particular child. While a direct assessment of oral language proficiency can take anywhere from 10-40 minutes, the LeBLEQ-T can be completed by classroom teachers in less than three minutes. Consequently, using teacher-report methods, which have been shown to be reliable tools for gathering information regarding child experience (e.g., Rimfeld et al., 2019), can be advantageous when testing idiomatic comprehension in school environments, whilst still being predictive of a child’s linguistic ability.

7.4.6 Decomposition

Children were more likely to correctly interpret an idiom in the Open-Ended, the Literal- and the Figurative- Multiple-Choice components of the task when the idiom was decomposable, rather than non-decomposable. In the latter two components, they were also faster at interpreting decomposable idioms. These findings are in line with previous literature demonstrating decomposable idioms to be more easily and more quickly processed (e.g., Gibbs et al., 1989; Gibbs, 1991). In their seminal work, Gibbs et al. (1989) demonstrated that individuals take less time in concluding decomposable idioms to be meaningful phrases in English than non-decomposable idioms. Gibbs (1989) proposed the idiom decomposition hypothesis according to which decomposable idioms are easier to process than non-decomposable idioms because individuals will always attempt to decompose an idiomatic phrase to make sense of it.

In contrast to the above findings, in the Open-Ended component, children had faster reaction times to non-decomposable in comparison to decomposable idioms. The direction of this result, however, may be influenced by the way in which the Open-Ended responses were timed. Reaction times were measured from the onset of the trial (i.e., the moment the idiom was displayed on the screen and the audio commenced) to the time the participant had finished their response. Consequently, it may not be that the children were slower in arriving at an answer for decomposable idioms, but rather that they gave longer, more detailed responses due to increased knowledge of the phrases. This, in turn, artificially increased their response rates. In retrospect, more informative data may have been gathered if the time in which a participant began their answer was recorded, as well as their rate of speech. This would have provided a more reliable indicator both of reaction times and the individual's knowledge of an idiomatic phrase.

7.4.7 Order of Presentation

As hypothesised, accuracy rates differed across the two orders of presentation. Though participants' accuracy rates in the Literal-Multiple-Choice component were statistically similar regardless of the order of presentation, a significant difference was found in accuracy rates in the Figurative Multiple-Choice component. Higher accuracy rates in the Figurative Multiple-Choice Task were observed when participants were required to first match the idiom to a literal phrase and then to a figurative phrase (rather than first match the idiom to a figurative phrase and then to a literal phrase). This finding could be explained by a bootstrapping effect where the Literal Multiple-Choice component bridges the gap in knowledge between the two figurative expressions. When this bridge in knowledge is not provided, children are required to correctly interpret the target idiom, then select another idiom that has the same meaning within a single process. Conversely, when the Literal Multiple-Choice option is presented first, this enables children to break the process down into two separate steps, thus minimising cognitive load and enabling the use of knowledge of the literal component to aid selection of the figurative idiom.

Regardless of the order of presentation the children were exposed to, children were first presented with an Open-Ended question asking "*what does this idiom mean?*". Although actual feedback from the experimenter/experiment is not provided after the Open-Ended component (nor any other component) of the Idioms task, children's prior assumptions of the correct interpretation of the idioms may be confirmed if they saw this interpretation listed in the options of the Literal Multiple-Choice component, thus, acting as a type of 'pseudo-feedback'. The children can then proceed to the Figurative Multiple-Choice component with more confidence in their interpretation of the target idiom. This may aid the selection of another figurative expression with the same meaning. As children who are exposed to the Figurative Multiple-Choice component first do not get this same 'pseudo-feedback', their knowledge depends solely on their detailed understanding of both the target idiom, and all the figurative response options contained within the Figurative Multiple-Choice component. Consequently, being presented with the Figurative Multiple-Choice option first involves more higher-order cognitive resources, thus making the task more difficult. Based on these findings, real world conclusions can be drawn on how to best explain idioms to a language learner, whether they are children, or second language learners of English. In particular, the present findings suggest that it is more constructive and effective to explain an idiomatic phrase using a literal interpretation first before attempting to link it to another form of figurative phrasing. For example, if an individual questions the meaning of a figurative statement such as "*he kicked the bucket*", it would be less effective to clarify the statement with another figurative phrase, such as "he bit the dust, he popped his clogs."

7.4.8 Conclusion

In conclusion, the present study demonstrated that receptive and expressive language abilities in addition to general language ability as reported by the classroom teacher were factors that contributed to the understanding of idiomatic expressions of primary school children. The most prominent predictors of idiomatic understanding, however, were the age of an individual, which was heavily moderated by socio-economic status. These findings support past literature which shows that children's idiomatic understanding vastly improves between the ages of 7-11 years (Cain et al., 2009; Gibbs, 1991; Levorato, 1999; Levorato & Cacciari, 1995; Nippold & Martin, 1989; Nippold, Moran & Schwatz, 2001; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995), and, importantly, further add to this body of research with evidence that SES is a large driving force behind these effects.

Furthermore, language experience also highly impacted idiomatic knowledge. Not only did the type of answer given by a child differed depending on their language experience classification (i.e., whether they were monolingual or bilingual) but linear effect models including a continuous measure of language experience fitted the data better than models only including a binary language classification. These models demonstrated that a child's cumulative exposure to language directly contributed to their degree of idiomatic knowledge. This finding supports predictions that idiomatic knowledge is dependent on general exposure to language. Importantly, this is the first study to use a continuous score of bilingualism to predict this ability, supporting the notion that bilingualism is better conceptualised along a scale, rather than as a simple binary variable.

Finally, as with previous chapters, it must, once again, be noted that, despite only including idioms in the task which were reported to have been learnt before the age of 10 years-old (as determined in Chapter 3, Experiment 2), there were still a substantial number of instances where children were unable to identify the meaning of an idiomatic phrase. This is attributed to the low SES contained within the sample; higher scores might be expected in higher SES samples, or older children who have had more language experience overall. Finally, the next chapter explores children's understanding of idioms alongside homonyms, metaphors and metonyms (as determined in Chapters 4, 5 and 6, respectively) to investigate further the development of pragmatic language and map an 'order of accuracy' in children's understanding of pragmatic (and ambiguous) language.

8 Patterns of Pragmatic Language Development

The present thesis set out to investigate pragmatic language abilities of both monolingual and bilingual primary-school aged children. The previous chapters focussed on children's skills regarding specific aspects of pragmatic language (ambiguity and figurative speech) in isolation. Collectively, however, these studies allow for determination of the order of acquisition of these tropes in comparison to each other. Literature often makes conclusions about which tropes are acquired first, or what children excel in, based on comparisons across studies. Such conclusions are flawed, however, due to comparisons being made across populations, and the different methods in which the tropes are assessed across studies. As discussed in both the past literature and current findings of the previous chapters, the way in which ambiguities are assessed can have substantial implications on the conclusions drawn. For example, differences in performance have been found depending on whether assessment involved verbal explanation or selection from a series of options in research concerning idioms (Arcara et al., 2019; Perlini et al., 2018) and metaphors (Pouscoulous, 2014), alike.

Furthermore, although there is research attempting to disentangle the order in which different ambiguity tropes are acquired, the stimuli contained in such research are often not controlled in comparable ways. For example, Gibbs (1990) found understanding of metaphors to be better than metonyms. In contrast, Rundblad and Annaz (2010) found the opposite with the understanding of metonyms being better than that of metaphors. It is possible that such directional differences are observed due to the way in which stimuli were (or were not) controlled. While Rundblad and Annaz (2010) controlled for both the metonymic and metaphoric words across numerous shared parameters (such as frequency and age of acquisition), Gibbs (1990) only reported controlling for the number of words contained in each of the target sentences. It is consequently possible that failure to control for parameters that are known to influence language processing may have resulted in large differences in the types of words used across studies. Hence, the accuracy rates across metonymic and metaphoric conditions could have been a result of the stimuli selection parameters, rather than the difference in the understanding of these two different tropes. Consequently, in order to arrive at reliable conclusions regarding the order of accuracy across different ambiguity tropes, studies must include the same participants, using similar methodologies and similar methods of stimuli selection.

8.1 The Current Analyses

The current analyses aimed to compare children's results across all four experimental studies described in the previous chapters of this thesis, namely the Homonyms Task (Chapter 4), the Metaphors Task (Chapter 5), the Metonyms Task (Chapter 6), and the Idioms Task (Chapter 7), to determine whether there is a form of pragmatic language children struggle with most when stimuli are highly controlled (for details of the methods for each of these experiments, see Chapters 4-7), or whether children experience similar levels of difficulty across all pragmatic language types. An additional aim was to determine if this order of difficulty was the same regardless of the age of the child (Year 3/6) or their language experience (monolingual/bilingual).

Finally, the current analyses compared the order of difficulty, as ascertained from the experimental studies, to results from the Age of Acquisition (AoA) norming survey (Experiment 2, Chapter 3). This aimed to determine whether children's knowledge of different pragmatic language types is aligned with the ages in which adults reported to have learnt such words/phrases (for details of the methods of the Age of Acquisition norming survey, see Chapter 3).

8.2 Analyses

A series of analyses were conducted in order to determine the ambiguity trope participants struggled with most (homonyms, metaphors, metonyms or idioms). This included further exploration of the patterns of difficulty across age groups (school year) and binary language experience (monolingual/bilingual). Accuracy rates were computed by combining the accuracy percentage scores of both the Open-Ended and Multiple-Choice components of the task in question.

A final ANOVA was conducted in order to determine whether the level of difficulty suggested from the accuracy rates across the experimental stimuli mirrored what would be predicted by the AoA ratings (Chapter 3). For these analyses, Sense-1 and Sense-2 of the ambiguous words were combined, as were the decompositional and non-decompositional idioms. The goal was to better understand the differences in accuracy rates across the different types of figurative language (rather than focus on differences within each trope which are explored in the previous experimental chapters).

8.2.1 Order of Accuracy as determined by the Experimental Studies

A three-way mixed 4 (Task: Homonyms, Metaphors, Metonyms, Idioms) x 2 (Year Group: Year 3, Year 6) x 2 (Language Experience: Monolingual, Bilingual) ANOVA, with a Greenhouse-Geisser correction, revealed a significant main effect of Task [$F(2.57, 316.45) = 116.08$, $p < .001$]. This implies that children do not experience the same level of difficulty with all types of pragmatic language, but, rather, understand some tropes better than others. Post hoc tests using the Bonferroni correction revealed participants (the whole sample) obtained the highest scores on the Metonyms Task ($M = .39$, $SD = .12$), followed by the Idioms Task ($M = .33$, $SD = .17$), the Homonyms Task ($M = .21$, $SD = .17$) and, finally, the Metaphors Task ($M = .15$, $SD = .14$) which children appeared to find the most difficult (all differences were significant at the $p < .01$ level). There was also a significant interaction of Task and Year Group [$F(2.57, 316.45) = 11.28$, $p < .001$], with children in Year 6 performing significantly better than Year 3 children on each of the four pragmatic tasks. However, the strength of the relationship between the two age groups varied across tasks. Four independent measures t-tests revealed the difference between the children in Year 3 and the children in Year 6 to be greatest for Idioms ($MD = .20$; $M_{Year3} = .23$, $SD_{Year3} = .09$; $M_{Year6} = .44$, $SD_{Year6} = .17$; $t(94.84) = -8.34$, $p < .001$), followed by Homonyms ($MD = .18$; $M_{Year3} = .12$, $SD_{Year3} = .10$; $M_{Year6} = .30$, $SD_{Year6} = .17$; $t(98.23) = -7.25$, $p < .001$), Metaphors ($MD = .12$; $M_{Year3} = .09$, $SD_{Year3} = .08$; $M_{Year6} = .22$, $SD_{Year6} = .15$; $t(94.13) = -5.59$, $p < .001$), and, finally, Metonyms ($MD = .06$; $M_{Year3} = .36$, $SD_{Year3} = .12$; $M_{Year6} = .42$, $SD_{Year6} = .12$; $t(125) = -3.06$, $p < .005$). There was no significant main effect of language experience with similar orders of accuracy for the pragmatic types found when analysing monolinguals and bilinguals independently,[1] with both monolinguals

[1] Two one-way ANOVAS demonstrated that overall both monolinguals and bilinguals had the highest understanding of metonyms, ($M_{Monolingual} = .37$, $SD_{Monolingual} = .12$; $M_{Bilingual} = .40$, $SD_{Bilingual} = .13$), followed by idioms ($M_{Monolingual} = .36$, $SD_{Monolingual} = .20$; $M_{Bilingual} = .32$, $SD_{Bilingual} = .15$), homonyms ($M_{Monolingual} = .22$, $SD_{Monolingual} = .14$; $M_{Bilingual} = .20$, $SD_{Bilingual} = .16$) and, finally, metaphors ($M_{Monolingual} = .17$, $SD_{Monolingual} = .14$; $M_{Bilingual} = .15$, $SD_{Bilingual} = .13$) [$F_{Monolingual}(3, 183) = 17.54$, $p < .001$; $F_{Bilingual}(3, 323) = 51.54$, $p < .001$]. All differences for the monolinguals were at the $p < .001$ level apart from between metonyms and idioms, and between homonyms and metaphors where the differences were non-significant. All differences for the bilinguals were at the $p < .01$ level apart from the difference between homonyms and metaphors which was non-significant.

and bilinguals following the same pattern outlined above (metonyms first, followed by idioms, homonyms and, finally, metaphors). No additional significant interactions were shown.

In order to explore further the effects within each Year Group, separate analyses were conducted for Year 3 and Year 6 children. Firstly, a 4 (Task: Homonyms, Metaphors, Metonyms, Idioms) x 2 (Language Experience: Monolingual, Bilingual) ANOVA for Year 3 revealed a significant main effect of Task [F(2.37, 149.55)= 106.17, p< .001]. Post-hoc Bonferroni tests showed that the highest accuracy rates in Year 3 overall were in the Metonyms Task (M= .36, SD= .12), followed by the Idioms Task (M= .23, SD= .09), the Homonyms Task (M= .12, SD= .10) and the Metaphors Task (M= .09, SD= .09) which did not differ statistically from the Homonyms (all other differences between the tasks were significant at the p< .001 level). For Year 3 children, there was no main effect of Language Experience[2], demonstrating both monolinguals and bilinguals in this year group experience the same pattern of difficulty. The pattern of difficulty for both monolinguals and bilinguals in Year 3 can be seen in Figure 24. There was no interaction between Task and Language Experience.

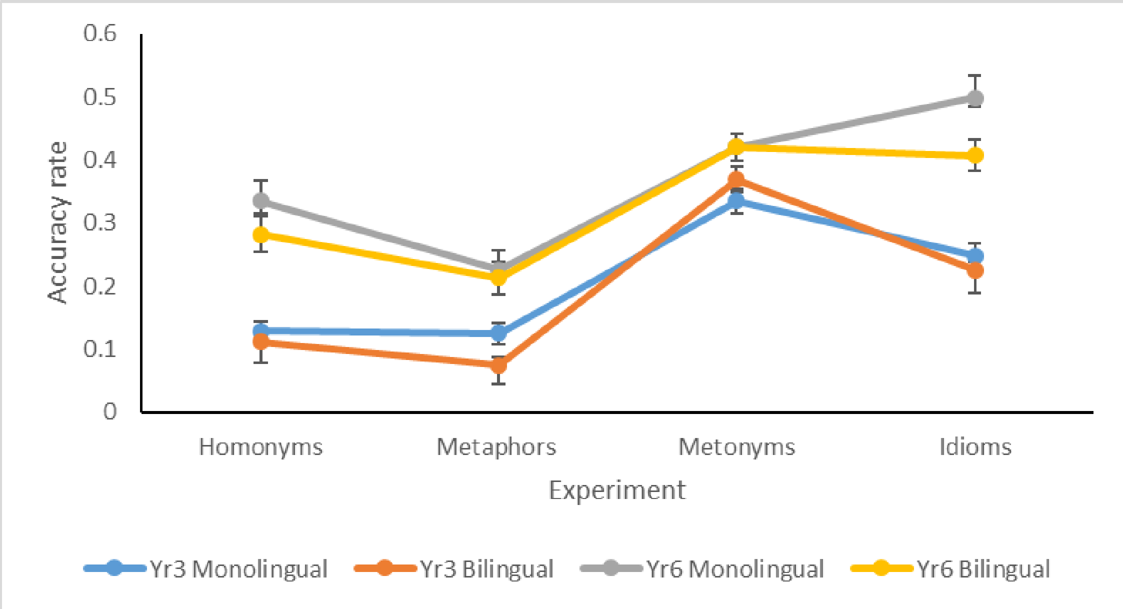
A 4 (Task: Homonyms, Metaphors, Metonyms, Idioms) x 2 (Language Experience: Monolingual, Bilingual) ANOVA for Year 6 revealed a significant main effect of Task [F(2.57, 155.81)= 41.64, p< .001]. Bonferroni post-hoc tests revealed that the highest accuracy rates in the Year 6 group were in the Idioms Task (M= .44, SD= .17), followed by the Metonyms Task (M= .42, SD= .12), the Homonyms Task (M= .30, SD= .17), and finally the Metaphors Task (M= .22, SD= .15). All differences between tasks were significant at the p< .001 level apart from the difference between Idioms and Metonyms which was non-significant. The order of difficultly revealed in the Year 6 children was, thus, different to the order of difficult revealed in the Year 3 children (metonyms understood the best, followed by idioms, homonyms and, finally, metaphors). There was no main effect

[2] Two one-way ANOVAS demonstrated both monolinguals and bilinguals in Year 3 had the highest understanding of metonyms, ($M_{Monolingual}$= .33, $SD_{Monolingual}$= .02; $M_{Bilingual}$= .37, $SD_{Bilingual}$= .12), followed by idioms ($M_{Monolingual}$= .25, $SD_{Monolingual}$= .01; $M_{Bilingual}$= .23, $SD_{Bilingual}$= .08), homonyms ($M_{Monolingual}$= .13, $SD_{Monolingual}$= .02; $M_{Bilingual}$= .11, $SD_{Bilingual}$= .11) and, finally, metaphors ($M_{Monolingual}$= 12., $SD_{Monolingual}$= .10; $M_{Bilingual}$= .08, $SD_{Bilingual}$= .07) [$F_{Monolingual}$(3, 99)= 25.45, p< .001; $F_{Bilingual}$(3, 159)= 70.71, p< .001]. All differences for the monolinguals were at the p< .001 level apart from between metonyms and idioms, where the difference was at the p< .05 level, and between homonyms and metaphors where the differences were non-significant. All differences for the bilinguals were at the p< .01 level apart from the difference between homonyms and metaphors which was non-significant.

of Language Experience and no significant interaction between Task and Language Experience.

Figure 24

Average accuracy scores across all four experimental tasks (homonyms, metaphors, metonyms, idioms), stratified by year and language experience



However, follow-up analyses revealed differences in the patterns of difficulty between monolingual and bilingual Year 6 children. For the monolingual Year 6 children, a one-way ANOVA revealed the highest accuracy rates were in the Idioms Task (M= .50, SD= .20), followed by the Metonyms Task (M= .42, SD= .12), the Homonyms Task (M= .34, SD= .19), and, finally, the Metaphors Task (M= .23, SD= .17) (differences were significant between all tasks at the $p< .05$ level apart from the difference between the Idioms and Metonyms and between Metonyms and Homonyms tasks which did not differ). For bilingual Year 6 children, however, a one-way ANOVA showed the pattern of accuracy rates mirrored the pattern observed in the Year 3 children. In particular, the highest accuracy rates were observed in the Metonyms Task (M= .42, SD= .12), closely followed by the Idioms Task (M= .41, SD= .14), the Homonyms Task (M= .28, SD= .16), and, finally, the Metaphors Task (M= .21, SD= .15) [$F(2.52, 100.96)= 28.24$, $p< .001$]. Post hoc tests using the Bonferroni correction revealed participants to obtain significantly higher scores between all tasks at the $p< .001$

level apart from the difference between the Homonyms and Metaphors tasks which was at the $p < .05$ level, and the difference between Metonyms and Idioms tasks which was not significant. Consequently, although the performance of the monolinguals and bilinguals on each individual task was not significantly different, the *pattern* of accuracy rates differed depending on Language Experience. The pattern of difficulty for both monolinguals and bilinguals in Year 6 can be seen in Figure 24.

8.2.2 Order of Understanding as determined by the AoA Survey

A one way ANOVA demonstrated a statistically significant difference between the ages of acquisition of the four different types of figurative language (homonyms, metaphors, metonyms, idioms), as reported in the AoA Survey (Chapter 3) [$F(3, 583) = 58.68$, $p < .001$]. A Bonferroni post-hoc analysis revealed individuals reported learning the meaning of metonyms first ($M = 4.50$, $SD = 2.07$), then metaphors ($M = 5.60$, $SD = 2.84$), homonyms ($M = 6.93$, $SD = 2.46$) and, finally, idioms ($M = 8.75$, $SD = 1.54$). All differences were at the $p < .001$ level apart from the difference between metaphors and metonyms which was at the $p < .05$ level. This order of understanding of the four pragmatic language types as determined by the AoA survey (metonyms first, followed by metaphors, homonyms and, finally, idioms) was different to the overall order of understanding as determined by the experimental studies (metonyms first, followed by idioms, homonyms and, finally, metaphors).

8.3 Discussion

The current research is the first to compare the order of acquisition of four tropes of ambiguity (homonyms, metaphors, metonyms and idioms), in both monolingual and bilingual, primary-school-aged children in order to determine whether children's understanding of pragmatic language differed across the four tropes. Four novel tasks were developed, as described in previous chapters of this thesis, to assess Homonymous (Chapter 4), Metaphoric (Chapter 5), Metonymic (Chapter 6), and Idiomatic understanding (Chapter 7). Comparing the four tasks in this way was possible because the stimuli selection criteria were consistent across all tasks. The procedures across all tasks were also kept as consistent as possible. Patterns of difficulty experienced in the experimental tasks were then compared with the average AoA ratings gathered for each trope in the norming studies of Experiment 2 (Chapter 3).

8.3.1 Findings across the Experimental Studies

Results revealed that, overall, children had higher comprehension of metonyms, followed by idioms, homonyms and, finally, metaphors. Though this general pattern remained in the Year 3 group when the two age groups (Year 3 and Year 6) were stratified, children in Year 6 performed slightly better in the Idioms Task compared to the Metonyms task (though this difference was non-significant). The patterns of performance between monolingual and bilingual children did not differ when the group was considered as a whole, but differences did emerge when focussing solely on the Year 6 children. While Year 6 monolinguals demonstrated greater knowledge of idioms, followed by metonyms, homonyms and, finally, metaphors (the same pattern that was observed when the Year 6 group was considered as a whole), the bilinguals in Year 6 mirrored the pattern that was seen in Year 3 monolingual and bilingual children.

The finding that metonyms are easier than other tropes is consistent with literature stating metonymic words are relatively easy to learn (Goossens, 1995; Taylor, 1995). Additionally, research shows that understanding of metonymic words is higher than metaphoric words across a range of ages (e.g., Rundblad & Annaz, 2010), with advantages in the processing of metonymies both in terms of comprehension and production (Annaz et al., 2009). These results are likely due to the high prevalence of metonymy within the English language (Armstrong, 2016), but also due to the fact that speakers do not necessarily perceive a clear distinction between the senses of metonyms. Unlike other forms of ambiguity, the senses of metonymic words are intrinsically linked, demonstrating no typical effects of dominance (Brocher et al., 2016, 2018; Frazier & Rayner, 1990; Frisson & Pickering, 1999; Klepousniotou et al., 2008, 2012). As a result, a processing advantage has been shown for metonymic words, demonstrating metonyms to be processed easier than even non-ambiguous words with a single meaning (e.g., Klepousniotou & Baum, 2007). Consequently, the finding that primary-school aged children deem metonymic ambiguity one of the easiest forms of pragmatic language, when considering both ambiguity at the word-level and idiomatic expressions, is consistent with existing literature.

Results also demonstrated a substantial increase in idiomatic understanding between Year 3 and Year 6 (discussed in Chapter 7). Such findings align with research by Levorato and Cacciari (1995) who showed idiomatic competence developed between 7 and 11 years of age. The present research is also corroborated by literature showing increases in idiomatic understanding occur alongside increased general language processing skills that also develop during that time (Cain et al., 2009). This effect is more pronounced for monolinguals than bilinguals, leading to idioms being the best understood trope of the four that were tested in the current thesis for monolingual children in Year 6. It is possible that the larger leap for the monolingual children can be attributed to the higher level of exposure to English which monolinguals receive in comparison to their bilingual peers (Chondrogianni & Marinis, 2011; De Houwer, 2007, 2009; Duursma et al., 2007; Gathercole & Thomas, 2009; Hammer et al., 2008b; Hoff et al., 2012; Pearson et al., 1997; Thordardottir, 2011; Thordardottir et al., 2006). Although bilingual children also experience this sizable increase in the knowledge of idiomatic phrases between the ages of 7-11 years, they do not improve to the extent that their monolingual peers do who receive more exposure to the English language in general, and, by proxy, are likely to be exposed to more figurative language. Consequently, the current research supports previous research (e.g., Smith & Murphy, 2015) which argues that the effect of language exposure is not limited to vocabulary only but also extends to other aspects of language development.

Furthermore, it poses the question of whether the observed increase in idiomatic knowledge between the two year groups (Year 3 and Year 6) would continue to be larger than for the other tropes tested (homonyms, metaphors and metonyms) as children get older. Alternatively, it may be that the rate of development for different types of pragmatic language shift throughout childhood. For example, the development of the other tropes may "catch-up" to the rate of idiomatic learning, or another trope may exceed the rate of idiomatic learning. This would alter considerably the order of difficulty suggested in the current thesis. Consequently, further research should be conducted in older age groups to determine whether the rate of development of each trope remains consistent throughout childhood, or whether there are ages in which certain tropes develop considerably faster until they potentially reach some kind of a plateau.

Finally, the finding that metaphors appear to be the most difficult of the ambiguity tropes assessed in the present thesis is particularly interesting. Though this appears to be the first research of its kind to suggest such a hierarchy of difficulty, this pattern is not completely unexpected as metaphors require a higher degree of conscious abstract thinking. The Conceptual Metaphor Theory (Gibbs, 1994; Lakoff & Johnson, 1980; 1999) states that metaphors represent highly abstract concepts by tying concepts to more concrete representations. As such, metaphors are considered to be representational goals, rather than simply a linguistic phenomenon (Lakoff & Johnson, 1980; 1999). Consequently, children require more than simple vocabulary knowledge to understand the second, more abstract sense of a metaphor. They additionally require the ability to pair real concepts to concepts which are not concrete, or immediately present in their world. However, according to the seminal research conducted by Piaget (1963), children do not develop the ability to think in abstract terms until the age of 12 years old which is beyond the age tested in the current research. Consequently, it may be determined that although primary-school-aged children do show some understanding of metaphorical meanings, this understanding is reasonably underdeveloped, with the potential of higher accuracy rates being seen in older children.

It must be noted, however, that the data analysed within this chapter is cross-sectional, and not longitudinal. Therefore, the current research does not account for both inter- and intra-individual variability. Further research should be conducted to attempt to map the timeline of development of these tropes throughout the course of KS2. This would provide further insight into children's development of various forms of ambiguity (both at the word- and phrase-level) and ensure educational resources are appropriately targeted.

8.3.2 Findings from the AoA Norming Surveys

Results from the AoA survey, however, reveal a different, more predictable pattern of results. In the AoA studies, individuals reported learning metonyms first, followed by metaphors, homonyms and, finally, idioms. Not surprisingly, metonyms are rated as the first to be acquired. This corroborates the findings from the comparison of the present experimental tasks discussed above, as well as past literature that has shown that metonyms are viewed as relatively easy to learn (Goossens 1995a, 1995b; Taylor 1995), are highly prevalent within the English language (Armstrong, 2016), and are processed the fastest compared to the other ambiguity types (Klepousniotou & Baum, 2007).

The notion of metaphor lying between homonymy and metonymy has also already been proposed in the literature (e.g., Klepousniotou, 2002; Klepousniotou & Baum, 2007). For instance, in a cross-modal sentence-priming lexical decision task, Klepousniotou (2002) found metonymic words to be processed faster than metaphors but metaphors to be processed faster than homonyms. As such, they concluded homonymy and metonymy are at polar ends of the lexical ambiguity continuum, ranging from the separately stored homonyms to the rule generated metonyms. Metaphors, however, possess both characteristics (though to lessening extents), thus lying somewhere in the middle of this continuum. Consequently, reports that metaphors are acquired between the ages of the acquisition of metonyms and homonyms add further support to this body of literature.

The finding that idioms are learnt last is also in accordance with previous literature that considered them to be complicated tropes of ambiguity (e.g., Adkins, 1968; Cedar, 2008; Cooper, 1998; Gibbs, 1994; Irujo, 1986; Nippold, 1991). Homonyms, metaphors and metonyms are all considered to be ambiguous at the word-level. In contrast, idioms are ambiguous at the sentence-level, requiring increased processing, thus, making them more difficult to comprehend. For example, research has demonstrated that even when all words in an idiom are understood in isolation, the meaning of the idiomatic phrase can still be misconstrued (Lodge & Leach, 1975). This is further complicated as the figurative meaning of an idiom can be unpredictable, with no standard rules of interpretation (Cooper, 1998). Idiomatic understanding is, thus, not exclusively dependent on vocabulary knowledge, but is influenced by other factors (such as decomposition or concreteness for example). Furthermore, there are no concrete referents for the meaning of idiomatic phrases. Instead, idiomatic expressions refer to increasingly abstract concepts which children generally find difficult to understand (Flavell & Ross, 1981). It is not surprising then, that idioms were rated as the last trope to be acquired.

8.3.3 Differences between the Experimental Studies and the AoA Survey

The differences between the findings of the AoA norming study and the experimental research may be explained, in part, by the presence of context. Participants in the AoA rating survey were exposed to all targets both in isolation, and within the context of a sentence. This differed from the experimental studies where only the Idioms task presented the targets in context. Context is known to be influential in the understanding of homonyms (Duffy et al., 1988; Yasushi Hino et al., 2002; Jager & Cleland, 2016; Pexman et al., 2004; Piercey & Joordens, 2000; Rayner & Duffy, 1986) metaphors (Giora, 1997, 2002, 2003; Giora & Fein, 1999), metonyms (Caramazza & Grober, 1976; Frisson, 2009; Frisson & Pickering, 1999; Ruhl, 1989) and idioms (Libben & Titone, 2008; Titone & Libben, 2014), increasing the likelihood of such tropes being understood. Consequently, it is possible that the children demonstrated higher levels of understanding in the Idioms task due to the presence of context. Future research comparing the four novel tasks developed as part of this thesis should either omit the context in the Idioms task, or place the target homonyms, metaphors and metonyms within the context of a sentence. This would enable improved comparisons of the tasks.

8.3.4 Conclusions

In conclusion, although differences remain between children depending on their age and language experience in terms of their understanding of different tropes of ambiguity (as discussed in depth in previous chapters), the patterns of understanding are largely consistent across these groups of individuals. As such, the current research clearly demonstrates that children do not find all types of pragmatic language of equal difficulty, but rather there is a gradient of difficulty from metonyms which are deemed the easiest form of figurative language to comprehend to metaphors which remained the most difficult.

9 General Discussion

The extent of an individual's bilingualism is known to affect various aspects of language development (e.g., Hoff et al., 2012; Paradis, 2011; Thordardottir, 2011; Unsworth, 2013b, 2017). Yet, until now, little research has been conducted on how the degree of exposure to a language impacts pragmatic awareness. The chapters contained in this thesis explored the relationship between specific aspects of pragmatic competence and language exposure in addition to the degree in which this relationship may be moderated by factors associated with bilingualism, such as academic attainment, general language development, age and working memory. Research also aimed to explore the order in which different figurative language tropes were acquired. The present research complements past literature and further highlights the tropes children struggle most with, thus, indicating where further attention and intervention is required.

Delving into the present research highlighted gaps in the literature and the requirement for additional materials to be developed for comparative purposes. The first of these necessary measures that were developed or modified and adjusted was the Streamlined-CELF which aimed to gather detailed information about a child's receptive and expressive language ability in the school setting, in a relatively short time frame. The second series of materials that were developed were two questionnaires. The first of these questionnaires, the Leeds-Bradford Language Exposure Questionnaire for Children (LeBLEQ-C), aimed to enable researchers to quantify a child's language experience whilst mitigating the need for parental report, a factor which cannot always be relied upon in school settings, or in certain populations. The second questionnaire, the Leeds-Bradford Language Exposure Questionnaire for Teachers (LeBLEQ-T) is complementary to the LeBLEQ-C in gathering detailed information about a child's language and pragmatic abilities from the child's teacher. This provided detailed linguistic information of a child's classroom-based competencies, in a fraction of the time it would take to administer standardised linguistic batteries. These purpose-developed materials provided receptive and expressive language scores, an indicator of working memory, an indicator of classroom communicative ability and a cumulative score of bilingualism that could be used as benchmarks when assessing a child's pragmatic language ability.

Finally, the current research developed AoA questionnaires and collected meaning-specific AoA norms for homonyms, metaphors, metonyms and idioms for use with British-English populations. Not only was such information vital in the stimuli selection of the experimental chapters contained within this thesis, but it also has the potential to aid other research that investigates ambiguity in child populations.

9.1 Summary of the Findings

Despite the predictions of the current research that a child's degree of exposure of a language other than English would significantly predict their level of pragmatic understanding, all studies demonstrated only a minimal effect of language experience on pragmatic awareness. While a child's idiomatic knowledge was predicted by their cumulative language score (as expected), idiomatic understanding was the only trope where accuracy or reaction times could be predicted by language exposure. A child's level of exposure to English, as measured using either binary classification or continuous language input score, did not predict performance on either component of the Homonymous, Metaphoric or Metonymic Tasks. Despite the non-significant differences in these three tasks, however, it is important to note that a child's continuous language input score was shown to be a better predictor of all types of pragmatic ability when included in the linear mixed effects models compared to the traditional binary (monolingual/bilingual) classification. Consequently, it can be concluded that language exposure is indeed likely to contribute to the understanding of pragmatic language, but a binary classification system may not be sensitive enough to capture such differences. This is in line with literature suggesting bilingualism is better suited to being measured along a continuum, rather than a simple binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020). Nevertheless, language experience was a significant contributor to the *types* of responses a child gave, with monolinguals tending to provide more sophisticated answers than their bilingual peers in these three tasks, indicating that the quality of the responses was mediated by language exposure.

Additionally, children's age was shown to make a significant contribution to pragmatic language both in accuracy rates and response types across all four studies, with more sophisticated understanding in older children. Importantly, this was significantly moderated by the child's SES, with those from higher socio-economic positions outperforming those from lower strata with increasing age. This is likely due to reduced quality and quantity of language that is often observed in low-SES households (Hoff, 2006) attributed to a series of parenting behaviours and child-rearing practices (Conger & Donnellan, 2007). Research has shown that mothers from lower socio-economic strata (comparable to those who receive Pupil Premium in the present sample) speak markedly less to their children than mothers from higher socio-economic groups, with the quality of their speech also found to be conceptually different (Hoff, 2006). While mothers in higher SES households engage in more conversational speech that would elicit a response from the child, mothers from lower SES use speech more to direct their child's attention to objects (Hoff-Ginsberg, 1998; Hoff-Ginsberg, 1991; Hoff, 2003, 2006; Hoff et al., 2002; Huttenlocher et al., 2010). The current research has demonstrated that the impact of SES on language development is not only observable in terms of vocabulary (Fernald et al., 2013; Hart & Risley, 1995; Hoff-Ginsberg, 1998; Hoff, 2003; Hoff & Tian, 2005; Huttenlocher et al., 2010; Oller & Eilers, 2002; Pan et al., 2005; Rescorla & Alley, 2001; Rowe, 2008), syntax (Huttenlocher et al., 2002) and grammar (Dollaghan et al., 1999), but also in the more subtle rules of spoken language, namely, pragmatics. These factors (based on SES) appear to have impacted the language development of the current sample as a whole, due to the disadvantaged conditions in the city of Bradford generally, not just the children in the sample that were in receipt of Pupil Premium.

Receptive language was also one of the most consistent predictors of the understanding of pragmatic language, with higher vocabulary knowledge being associated with higher pragmatic understanding. The present findings advance previous literature which demonstrated that receptive language ability is indicative of metaphoric (Hessel & Murphy, 2019), idiomatic (Smith & Murphy, 2015) and metonymic competence (Annaz et al., 2009; Rundblad & Annaz, 2010a, 2010b; Van Herwegen et al., 2013). The current research is the first, however, to use a comprehensive assessment of receptive language ability to predict the understanding of metaphoric words, (with Hessel and Murphy, 2019, investigating metaphoric phrases) and homonymic awareness. It is also the first to suggest that the understanding of language in general is intrinsically linked to multiple aspects of pragmatic understanding, not just specific tropes, with the two appearing to develop cohesively.

Performance on the LeBLEQ-T also significantly predicted scores on the pragmatic components, including both accuracy rates (Homonyms and Idioms Tasks) and response times (Metaphors Tasks). This ability of the LeBLEQ-T to predict pragmatic competence advances the knowledge of the links between oral and pragmatic language ability. These findings also assist in validating the novel measure of the LeBLEQ-T (developed in Chapter 2) by demonstrating that teachers are indeed knowledgeable and reliable sources of information of a child's language ability (Stone et al., 2010), including metaphoric (Johnson & Rosano, 1993), but also homonymous and idiomatic ability.

Expressive language was additionally shown to be a contributor to the Open-Ended components of both polysemic tasks (Metaphors in Chapter 5 and Metonyms in Chapter 6), and two Multiple-Choice components of the Idiomatic Task. Both of these Multiple-Choice tasks (Figurative and Inverse Multiple-Choice components) were considered a precursor of the child's ability to produce idiomatic units themselves. Therefore, the association found between expressive language and these particular components is attributed to the specific demands of verbalising information, or the tendency of the children to verbalise information especially in the case of idioms. Children with higher communicative abilities are better able to express their knowledge of ambiguities or figurative speech in comparison to children who have lower expressive abilities.

Interestingly, males were shown to outperform females on the Multiple-Choice components of three of the four pragmatic tasks. In particular, males in the Metonyms Task, were shown to have higher accuracy rates than their female peers, with males also being faster at selecting the correct responses in both the Homonyms Task and the Idioms Task. While the finding of gender contributing to Metonyms and Idioms is novel, such a finding is not unprecedented in homonymic understanding. In a study investigating the comprehension of homonymous words, (Corthals, 2010) found gender differences in 12-year-olds, with males demonstrating higher accuracy rates than their female peers. However, further research is needed to determine the underlying mechanisms that could be driving this gender difference in the pragmatic understanding of primary-school-aged children, and why this difference is only evident in some (but not all) of the pragmatic tasks.

In accordance with expectations, form of presentation was also shown to affect accuracy rates across all tasks of pragmatic language, as was the method of response. Children had higher accuracy rates when asked to pick from a series of options (i.e., the Multiple-Choice component) in comparison to when asked to verbalise their understanding (i.e., the Open-Ended component) of the pragmatic items in all four pragmatic tasks. They were also more accurate when response options were presented pictorially rather than orthographically. This builds on knowledge suggesting that the methods in which individuals are tested highly influence the results gathered (Arcara et al., 2019; Perlini et al., 2018). These findings, then, further highlight the importance of uniformity in methodology when making comparisons across studies.

A child's reading ability scores were predictive of their metonymic ability. While such a relationship with reading is well documented in idiomatic research (Fusté-Herrmann, 2008) and research investigating ambiguity in general (Smith Cairns et al., 2004), the association between metonymic understanding and reading ability is a novel finding. It must be noted, however, that children's reading ability scores were only predictive of their metonymic ability and not homonymic or idiomatic ability (where there is literature of such a link) or metaphoric ambiguity. It is suggested that the reading ability of the child contributed little to the pragmatic-language scores due to the nature of measurement used for reading ability. The KS1 SAT scores were used as an indicator of reading skill across the current research. However, this was a dichotomous scoring system that was simply used to indicate whether the child was performing at the expected level or not. While the KS1 SATS were easily obtainable measures of reading ability that negated the need to add another lengthy assessment to the already long list of tasks, a more detailed and targeted measure of reading ability (such as the YARC for example; Snowling et al., 2009) might have been able to provide more detailed information of the degree in which reading ability is associated with pragmatic ability.

Working memory was also shown to play a role, but the relationship between pragmatic understanding and working memory in the multiple-choice components of the Homonyms and Metonyms and the open-ended component of the Idioms tasks was in an unexpected direction Findings of the current thesis indicated that children with lower working memory performance possessed increased pragmatic knowledge. This contradicts previous literature which suggests linguistic knowledge and memory are positively correlated (e.g., Baddeley et al., 1998; Gathercole & Baddeley, 1990). However, most of the research determining this link has assessed phonological short-term memory, *not* working memory. For instance, Baddeley et al. (1998), claimed phonological short-term memory is vital in the construction of representations of the phonological form of new words both in one's native language and in a foreign language. Consequently, though working memory and pragmatic knowledge might be related to some extent, it may be that pragmatic knowledge is differentially associated with short-term memory, a factor which is more consistently linked with general linguistic skill. Further research should therefore be conducted to identify more precisely the link between different memory types (e.g., working memory, short-term memory etc.) and different tropes of pragmatics.

9.2 Implications of the Present Findings

The present thesis developed four measures that can be used to investigate monolingual and bi-/multilingual children's understanding of pragmatic language to provide detailed information about children who might be underperforming. These measures are contained in Chapters 4-7 and assess homonymic, metaphoric, metonymic and idiomatic comprehension, respectively. Identifying children with poor knowledge of ambiguity and figurative expressions is particularly important as these units of speech are highly prevalent in everyday life (Lazar et al., 1989; Parks et al., 1998; Van Lancker-Sidtis & Rallon, 2004) and inability to understand pragmatic language is known to impact a multitude of other skills. For example, the literature demonstrates a link between children's pragmatic knowledge and their social skills, including the ability to establish intimacy (Gerrig & Gibbs, 1988), understand humour (Hirsh-Pasek et al., 1978) and participate in social interactions (Kerbel & Grunwell, 1997; Laval, 2003; Swineford et al., 2014). Research has also shown that children with enhanced pragmatic skills tend to be more liked by their peers (Gertner et al., 1994; Place & Becker, 1991). Thus, pragmatic ability is linked to social inclusion and well-being.

Pragmatic language abilities are also associated with children's reading ability, with research demonstrating poor pragmatic language skills are linked with poor reading comprehension (e.g., Doherty, 2000; Fusté-Herrmann, 2008). This has been demonstrated even in the early years of schooling (Tunmer & Hoover, 2018; Wankoff, 1983) with children's ability to report the dual meaning of ambiguous words being predictive of future reading skill even in children who have not yet learned to read (Smith Cairns et al., 2004). Consequently, using the measures developed within the current thesis to identify children with poor pragmatic language skills also has the potential to highlight other areas of development that children may struggle with that is likely to negatively impact their school experience.

Importantly, the ability to identify children with poor pragmatic language skills using the novel methods developed within this thesis allows for resources to be appropriately distributed to children who may be disproportionately challenged. This includes targeted interventions. Such interventions may be aimed solely at improving children's pragmatic language skills, or, alternatively, may be employed with the aims to improve other skills that are correlated with pragmatic language competence, such as social skills or reading comprehension. For example, Nelson and Stage (2007) developed an intervention that aimed to improve the reading capabilities of primary-school aged children by increasing their understanding of words with multiple meanings. This research demonstrated that increasing ambiguity knowledge significantly improved reading comprehension in both Year 3 and Year 5 children. Consequently, assessing pragmatic language competence can not only identify those who are disproportionately challenged, but may also lead to targeted interventions that will improve not only said the targeted pragmatic skills, but also, by proxy, social skills and academic outcomes of primary-school aged children.

Interventions may be particularly useful as lower than expected skills were observed in all the children in the current sample. This highlights that children still struggle with pragmatic language even at the age of 11 years old, a finding that is particularly concerning as the school curriculum states that children of this age should be able to understand figurative language (Colston & Kuiper, 2002; Department for Education, 2013b). Therefore, the present findings suggests that ambiguities and figurative language should not continue to be included in academic resources (for example, textbooks or key reading materials) without explicit teaching. Increased explicit instruction of pragmatic language in the curriculum may, thus, be required to enable disadvantaged children to catch up with their more advantaged peers.

There are also substantial implications for the tools developed and modified as part of the current thesis that do not relate solely to pragmatics. These tools, the Streamlined CELF and the LeBLEQ, were developed out of necessity but have the potential to improve efficiency when testing in environments where access to parental information, and time testing the children themselves is limited. Firstly, the CELF-4 UK was truncated in a methodological way to considerably reduce the number of items in the assessment by approximately a third without compromising the reliability of the measures. Most importantly in terms of practicality, however, this reduced the administration time from 80 minutes (in the CELF-4 UK) to 30-40 minutes (in the Streamlined-CELF). This allows more children to be tested within any given time frame, increasing the amount of data gathered by researchers or practitioners visiting a school. The reduced administration time also reduces the likelihood of children exhibiting boredom effects or cognitive fatigue from an overly long testing period. Consequently, the modified battery is a less resource- and time-expensive method of gaining reliable information of children's receptive, expressive, and core language skills. When testing time within the classroom environment is limited, the Streamlined-CELF can improve the efficiency of the testing experience for researchers and practitioners alike.

The LeBLEQ and its two components were also tools that were developed within this thesis due to necessity. Namely, the low response rates from parents seen both within this project (see Chapter 2), and across other studies within Bradford more generally, and the lack of existing measures that capture children's language background highlighted the need to develop novel methods to obtain this information without relying on parental reports. Consequently, two questionnaires were developed to gather information of a child's language background: a questionnaire for children (the LeBLEQ-C) and a questionnaire for teachers (the LeBLEQ-T). The LeBLEQ-C enables researchers, teachers or practitioners to quantify a bilingual child's language experience. While such continuous scores are usually gathered from extensive interview techniques or paperwork procedures that rely on parental involvement, the current research developed a method of obtaining such detailed information during a relatively short, conversational procedure with the children themselves. The LeBLEQ-C, thus, allows for vital information to be gathered without the need for parental involvement, easing the processes of data collection in environments where parental response rates are low, and providing detail that would otherwise be missed if parental response alone was relied upon.

Finally, while the LeBLEQ-T complements the LeBLEQ-C, its value is unique. The teacher's component of the LeBLEQ has been demonstrated to reliably predict aspects of children's classroom-based language skills in accordance with what would be expected of a child their age. Skills predicted include receptive, expressive and core language ability (as measured using the objective Streamlined-CELF) in addition to pragmatic language ability, as well as some aspects of academic attainment (SATs). Furthermore, this is a measure that can be used with both monolingual and bilingual populations and can be completed in less than three minutes. Consequently, if the time spent with a child is limited, the LeBLEQ-T can be employed as a quick and effective method of gaining information regarding children's general language proficiency.

9.3 Limitations and Future Directions

It is important to note that although SES was controlled for across all experimental studies contained within the current thesis, SES is a complex, multidimensional construct with many different identifiers (Cheng & Goodman, 2015). Consequently, more than one indicator was employed to indicate SES. Whilst the entire study took part in areas of deprivation (as identified by the IDACI profiles of the schools in which testing took place, Department for Communities and Local Government, 2015), information on the SES of the children contained within the study was additionally collected in the form of receipt of Pupil Premium. However, while some language research uses Pupil Premium as an indicator of SES (e.g., Babayiğit, 2015), others use maternal education level (e.g., Fernald et al., 2013), or a combination of many identifiers including those mentioned above, household income or occupation (e.g., Dixon et al., 2012) with the use of different identifiers often producing differing results. For example, Qi et al. (2006) used monthly household income, teenage mother status, maternal education level, marital status, and number of children in the family as indicators of SES, finding only the final three were associated with children's performance on the Peabody Picture Vocabulary Test-Third Edition (PPVT-III; Dunn & Dunn, 1997). Such effects are not unique to language research, but present in other developmental domains also. For example, Cools et al. (2011) found a significant relationship between motor competence and parental education level, but this effect disappeared when they used parental occupation as an indicator of SES. This highlights the need for further research to be conducted into the relationship between pragmatic language ability and SES using a wider range of SES indicators or more sophisticated measures. Although prominent links were found between SES and pragmatic competence when using Pupil Premium, it is possible that the relationship between SES and pragmatic language may differ (both in strength and significance) when employing other indicators of SES such as maternal education, household income or parental occupation.

Subsequent research should also attempt to assess children from both higher, and more moderate socio-economic strata. The current research was only administered in low SES neighbourhoods, with the low accuracy rates across all language assessments (receptive, expressive, and pragmatic alike) being attributed to the overall underperformance. Thus, it is important to investigate whether children from higher SES groups would perform better on the tasks contained in the current thesis. Furthermore, there has been research to suggest that SES may differ depending on culture (Hoff & Tain, 2005) and ethnic group (Fairley et al., 2014). When investigating ethnic-specific latent classes of SES within the Born in Bradford (BiB) cohort, Fairley et al. (2014) found differences in the clustering of socioeconomic indicators across White British and Pakistani British samples. Thus, that the way in which SES is conceptualised in populations such as those in Bradford, which is known to have communities consisting predominantly of ethnic minorities, has the potential to impact the conclusions drawn regarding the effect of SES on language development. Future research should explore in more depth the links between the conceptualisation of SES and language development generally, as well as pragmatic language development more specifically.

The current research has also highlighted the need for norms that are more current and extensive. For instance, the Kucera and Francis (1967) written frequency norms were employed to match the stimuli within all the experimental studies. Whilst these norms remain among the most consistently used written frequency lists (e.g., Boscolo Nale, 2020; Borghesani et al., 2020; Guan, & Fraundorf, 2020) they were established more than 50 years ago. Subsequently, the frequency values may not be wholly reflective of language use in the 21st century. Furthermore, this need for updated, more extensive norms is particularly relevant for types of language such as pragmatic language (including homonyms, metaphors, metonyms and idioms). In the current research, the only parameters that could be matched across the ambiguous stimuli at the word-level (homonyms, metaphors and metonyms), and sentence-level (i.e., idioms) were either those which were not specific to ambiguous stimuli (e.g., the AoA ratings of Kuperman et al., 2012), or were gathered in the scope of the current thesis. Updated, more sophisticated norms should consequently be generated to allow researchers to control better for confounding effects of words. For example, frequencies of pragmatic language in children's reading materials is an area that appears to be underreported. Counting the occurrences of pragmatic language in the books that track the development of children across the national curriculum (i.e., the books set by the school that determine reading bands), would be highly informative when conducting research with pragmatic language in child populations. Such norms could be used in much the same way as the non-meaning-specific norms generated in the Children's Printed Word Database (Masterson et al., 2010). Likewise, similar frequencies could be generated for the number of times in which such units are encountered in media directed at children. Uses of this information would be akin to the uses of the SUBTLEX-UK (van Heuven et al., 2014) which counted the frequencies of occurrences of non-meaning-specific words on children's channels such as CBBC and CBeebies. Finally, the generation of such frequencies need not apply merely to the ambiguities contained within the current thesis (homonyms, polysemes and idioms) but also for instances of simile, irony, sarcasm which are also considered pragmatic language (e.g., Carrow-Woolfolk, 1999).

It is vital to note, however, that the current format of all the measures within this thesis are only appropriate for British-English speakers. This is due to the pragmatic stimuli in all four experimental studies being selected based, in part, on British-English frequency norms and on the meaning-specific AoA ratings gathered only from native British-English speakers in the four AoA norming studies contained within this thesis. With some alterations, this method might be applicable for speakers of other dialects of English. Consequently, while the assessment in its current form can only be used in populations where one of the languages spoken is British-English, there is potential for it to be used in parts of the world that speak a different dialectal variation of English with only minor reconsiderations and alterations. This could include other parts of the UK (e.g., Wales, Ireland or Scotland), but also other countries where English is one of the official languages such as (e.g., the US, Canada, Australia, or Uganda).

Furthermore, the pragmatic tasks contained within the present thesis have yet to be standardised. While standardisation of the tasks is possible, the sample size of the current research would have to increase vastly and psychometric norms would need to be developed. Though no difficulties with the tasks would be expected in age groups from 7-11 years (as both children at the upper and lower bounds of these ages groups were able to complete the task), the measure may be unsuitable for use with both younger populations who may find the task too difficult, as well as older children who are more likely to be prone to ceiling effects (though the overall low performance of the children within the sample must be considered). Consequently, the extent to which the current stimuli are appropriate for other age groups outside of the realms of the current research is an avenue for future enquiry.

Standardising the tasks developed within the thesis, using participants similar to those included in the current sample, would be highly advantageous as few language batteries are standardised for both monolingual and bi-/multi-lingual populations. This includes the CELF-4 UK (Semel et al., 2003), which, as discussed in Chapter 2, was norm-referenced for monolingual children only. Resultantly, some researchers have called into question the suitability of the CELF-4 UK for bi-/multi-lingual populations (e.g., ASHA, 2004; Oxley et al., 2019). Whilst the current research set out to highlight the necessity for a *pragmatic* language assessment that could be administered to both monolingual and multi-lingual children, as a by-product, it also highlighted the need for more language assessments to be standardised for diverse populations. This is particularly important as a rising number of the world's population speaks more than one language (Department for Education, 2020). While the current study did not use the CELF-4 UK for diagnostic purposes, a comprehensive, core language assessment that was norm-referenced with both monolingual and multilingual children would have been a more appropriate tool to validate the current research. However, there is no such battery currently in distribution within the UK for primary-school-aged children. When such a battery is developed, additional research should be conducted to further validate the tasks developed within the current thesis.

Finally, it must be noted that the data collected and analysed within this thesis are cross-sectional, and not longitudinal. Whilst conducting this research longitudinally would account for both inter- and intra-individual variability within each of the experiments, this was not feasible within the relatively short timeframe of the PhD (3-4 years). The issue of feasibility is particularly pertinent considering the age groups of interest in the current research and the point the children were in education (i.e., at the beginning (Year 3) and the end (Year 6) of KS2, respectively). Further research conducted using a within-participants design would allow for a timeline of development of the tropes studied (homonyms, metaphors, metonyms and idioms) to be mapped across the course of KS2. Such findings would provide further insight into children's development of various forms of ambiguity (both at the word- and phrase-level) and ensure educational resources are appropriately targeted.

9.4 Conclusions

The current research is the first to assess both monolingual and bilingual, primary-school-aged children on four tropes of ambiguity: homonyms, metaphors, metonyms and idioms, employing similar methodologies and stimuli selection processes across tasks. Outcomes of this work have produced tools which can be used in the school setting to quickly and efficiently assess children's receptive, expressive and core language skills, their degree of exposure to languages other than English, and their understanding of ambiguous and figurative language.

Although previous research has made tentative conclusions regarding the order in which different tropes of ambiguity are acquired (e.g., Gibbs, 1990; Rundblad & Annaz, 2010a), the current study is the first to conduct research on the AoAs of homonyms, metaphors, metonyms and idioms across the same group of participants. The present research concluded that the developmental trajectories of different types of pragmatic language vary, with some being learnt with more ease than others. Namely, while metonyms are acquired first (followed by idioms, homonyms and metaphors), the development of idiomatic expressions is considerably faster between the ages of 7-11 years compared to the other three tropes.

The current research also demonstrated age, socio-economic position and receptive language skills of children to be the most reliable predictors of how well they understand pragmatic language. The work in this thesis also provides increased understanding of the effects of bilingual language experience and in particular, how children's exposure to language can impact their development of pragmatic knowledge. The present findings support emerging research suggesting bilingualism is a continuous measure, rather than a binary classification (e.g., Baum & Titone, 2014; Kašćelan et al., 2020; Serratrice & De Cat, 2020).

Finally, the low overall performance of the children in the current research has highlighted that children with limited exposure to language, either due to low SES or exposure to other languages within the home, may experience increased challenges in understanding pragmatic language. Understanding pragmatic speech is particularly vital, however, as a child's competence in understanding pragmatic language is closely linked with many skills that allow them not only to be academically successful within the school setting but also affect their successful social inclusion and overall well-being (Gerrig & Gibbs, 1988; Gertner et al., 1994; Hirsh-Pasek et al., 1978; Kerbel & Grunwell, 1997; Laval, 2003; Place & Becker, 1991; Swineford et al., 2014). The present research, thus, is original in that it allows the early identification of children whose understanding of pragmatic language is below expected. This knowledge is vital in assisting to direct resources, such as targeted interventions, to these children so they can improve their pragmatic skills and, in turn, their overall school experience, including their academic attainment, social interactions and well-being.

References

Abbott, I., Middlewood, D., & Robinson, S. (2015). It's not just about value for money: A case study of values-led implementation of the Pupil Premium in outstanding schools. *Management in Education*, *29*(4), 178–184. https://doi.org/10.1177/0892020615593789

Abel, B. (2003). English idioms in the first language and second language lexicon: A dual representation approach. *Second Language Research*, *19*(4), 329–358. https://doi.org/10.1191/0267658303sr226oa

Ackerman, B. P. (1982). On comprehending idioms: Do children get the picture? *Journal of Experimental Child Psychology*, *33*, 439–454. https://doi.org/10.1016/0022-0965(82)90058-3

Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of comprehension and expression 6-11. GL assessment*.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 206–245. https://doi.org/10.3102/0034654310368803

Adkins, P. G. (1968). Teaching idioms and figures of speech to non-native speakers of English. *The Modern Language Journal*, *52*(3), 148–152. https://doi.org/10.1111/j.1540-4781.1968.tb01884.x

Adler, S. (1991). Assessment of language proficiency of limited English proficient speakers: Implications for the speech-language specialist. *Language, Speech, and Hearing Services in Schools*, *22*(2), 12–18.

Agirdag, O., & Vanlaar, G. (2018). Does more exposure to the language of instruction lead to higher academic achievement? A cross-national examination. *International Journal of Bilingualism*, *22*(1), 123–137. https://doi.org/10.1177/1367006916658711

American Speech-Language-Hearing Association. (2004). Knowledge and skills needed by speech-language pathologists and audiologists to provide culturally and linguistically appropriate services. *ASHA Supplement 24*.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), Comprehension and teaching: Research Reviews (pp. 77- 117). Newark, Del.: International Reading Association.

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Andrich, D., & Sheridan, B. (2011). *RUMM2030*. RUMM Laboratory.

Annaz, D., Van Herwegen, J., Thomas, M., Fishman, R., Karmiloffsmith, A., & Rundblad, G. (2009). Comprehension of metaphor and metonymy in children with Williams syndrome. *International Journal of Language and Communication Disorders*, *44*(6), 962–978. https://doi.org/10.3109/13682820802525005

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics*, *5*, 395–415. https://doi.org/10.1146/annurev-linguistics-011718-011820

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism*, *18*(4), 683–695. https://doi.org/10.1017/S1366728914000777

Apresjan, J. D. (1974). Regular polysemy. *Linguistics*, *12*(142), 5–32. https://doi.org/10.1515/ling.1974.12.142.5

Arcara, G., Tonini, E., Muriago, G., Mondin, E., Sgarabottolo, E., Bertagnoni, G., Semenza, C., & Bambini, V. (2019). Pragmatics and figurative language in individuals with traumatic brain injury: fine-grained assessment and relevance-theoretic considerations. *Aphasiology*, 1–30. https://doi.org/10.1080/02687038.2019.1615033

Armstrong, B. (2016). *Number of meanings, number of senses, and part-of-speech frequencies from the Wordsmyth Online Dictionary [data file]*. https://cnbc.cmu.edu/~armstrong/tools/

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

Armstrong, B. C., Tokowicz, N., & Plaut, D. C. (2012). eDom: Norming software and relative meaning frequencies for 544 English homonyms. *Behavior Research Methods*, *44*(4), 1015-1027. https://doi.org/10.3758/s13428-012-0199-8

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from lowand middle-income families. *Applied Psycholinguistics*, *19*(2), 209–223. https://doi.org/10.1017/s0142716400010043

Ashby, M. (2009). *The Oxford Dictionary for Learners of English*.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (CD-ROM). Linguistic data consortium*.

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

Babayiğit, S. (2014). The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of English language. *Journal of Research in Reading*, *37*(1), 22–47. https://doi.org/10.1111/j.1467-9817.2012.01538.x

Babayiğit, S., & Shapiro, L. (2020). Component skills that underpin listening comprehension and reading comprehension in learners with English as first and additional language. *Journal of Research in Reading*, *43*(1), 78–97. https://doi.org/10.1111/1467-9817.12291

Backscheider, A. G., & Gelman, S. A. (1995). Children’s understanding of homonyms. *Journal of Child Language*, *22*(1), 107–127. https://doi.org/10.1017/S030500090000965X

Baddeley, A. (1986). *Working Memory*. Oxford University Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158. https://doi.org/10.1037/0033-295X.105.1.158

Bain, B., & Yu, A. (1980). Cognitive consequences of raising children bilingually: One parent, one language. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *34*(4), 304. https://doi.org/10.1037/h0081106

Baird, G., & Norbury, C. F. (2016). Social (pragmatic) communication disorders and autism spectrum disorder. *Archives of Disease in Childhood*, *101*(8), 745–751.

Baker, C. (2011). Foundations of Bilingual Education and Bilingualism. In *Multilingual Matters*.

Baker, M. (1992). *In Other Words: A Coursebook on Translation*. Routledge. https://doi.org/10.2307/329719

Balladares, J., Marshall, C., & Griffiths, Y. (2016). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean preschoolers. *First Language*, *36*(3), 338–351. https://doi.org/10.1177/0142723715626067

Balota, D. A., & Faust, M. (2001). Attention in dementia of the Alzheimer's type. *Handbook of Neuropsychology*, *6*, 51–80.

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science*, *18*(2), 89–94. https://doi.org/10.1111/j.1467-8721.2009.01615.x

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*, *83*(2), 413–422. https://doi.org/10.1111/j.1467-8624.2011.01707.x

Barcelona, A. (2003). The case for a metonymic basis of pragmatic inferencing. In *Metonymy and Pragmatic Inferencing* (81–102). John Benjamins Publishing Company.

Barfield, A., & Gyllstad, H. (2009). Introduction: Researching L2 collocation knowledge and development. In *Researching Collocations in Another Language: Multiple Interpretations* (1–18). Palgravw Macmillan. https://doi.org/10.1057/9780230245327

Baron, A., Bedore, L. M., Peña, E. D., Lovgren-Uribe, S. D., López, A. A., & Villagran, E. (2018). Production of Spanish grammatical forms in U.S. bilingual children. *American Journal of Speech-Language Pathology*, *27*(3), 975–987. https://doi.org/10.1044/2018_ajslp-17-0074

Barre, N., Morgan, A., Doyle, L. W., & Anderson, P. J. (2011). Language abilities in children who were very preterm and/or very low birth weight: A meta-analysis. *Journal of Pediatrics*, *158*(5), 766–774. https://doi.org/10.1016/j.jpeds.2010.10.032

Barron-Hauwaert, S. (2004). Language strategies for bilingual families: The one-parent-one-language approach. In *Multilingualism Matters*. https://doi.org/10.1017/s0272263105250283

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, *35*(5), 857–894. https://doi.org/10.1017/S0142716414000174

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1–28. https://doi.org/10.2167/beb392.0

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T. M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism*, *15*(3), 616. https://doi.org/10.1017/S1366728912000090

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In *Right hemisphere language comprehension: Perspectives from cognitive neuroscience.* (pp. 225–284).

Beloussova, V. (2015). *Idiom learning materials for Estonian Secondary School students*.

Ben-Zeev, S. (1984). Bilingualism and cognitive development. In *Bilingualism and language disability*. Springer, Boston, MA.

Ben-Zeev, S. (1977). The Influence of Bilingualism on cognitive strategy and cognitive development. *Child Development*, 1009=1018. https://doi.org/10.2307/1128353

Benczes, R., Barcelona, A., & Ibanez, F. (2011). Defining Metonymy in Cognitive Linguistics. Towards a consensus view. In *America*. John Benjamins Publishing.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, *6*(2), 183–200. https://doi.org/10.1017/S0305000900002245

Benjamin, L., Newton, C., & Ebbels, S. (2020). Investigating the effectiveness of idiom intervention for 9–16-year-olds with developmental language disorder. *International Journal of Language and Communication Disorders*, *55*(2), 266–286. https://doi.org/10.1111/1460-6984.12519

Bercow, J. (2008). The Bercow Report: A review of the Services for Children and Young People (0-19) with Speech, Language and Communication Needs. *Defs*. https://doi.org/ISBN: 978-1-84775-211-6

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research*, *24*(1), 57–65. https://doi.org/10.1016/j.cogbrainres.2004.12.006

Berko Gleason. (2005). *The Development of Language*. Pearson.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology*, *94*(2), 291–304. https://doi.org/10.1037/0022-0663.94.2.291

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences2*, *16*(4), 240–250.

Bialystok, E. (2001). Bilingualism in Development. In *Bilingualism in*

Development. Cambridge University Press. https://doi.org/10.1017/cbo9780511605963

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism*, *12*(1), 3. https://doi.org/10.1017/S1366728908003477

Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, *11*(4), 485–508. https://doi.org/10.1080/15248372.2010.516420

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, Supplement*, *10*(3), 89–129. https://doi.org/10.1177/1529100610387084

Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language*, *109*(2–3), 93–100. https://doi.org/10.1016/j.bandl.2008.09.001

Bialystok, E., & Herman, J. (1999). Does bilingualism matter for early literacy? *Bilingualism: Language and Cognition*, *2*(1), 35–44. https://doi.org/10.1017/s1366728999000139

Bialystok, E., Klein, R., Craik, F. I. M., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*, 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. M. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging*, *29*(3), 696. https://doi.org/10.1037/a0037254

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*(3), 494–500. https://doi.org/10.1016/j.cognition.2009.06.014

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism (Cambridge, England)*, *13*(4), 525.

Birdsong, D. (2005). Interpreting Age Effects in Second Language Acquisition. *Handbook of Bilingualism: Psycholinguistic Approaches*, *109*, 127. https://doi.org/10.1353/lan.2008.0062

Bishop, D. V. M. (2000). Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In D. V. M. Bishop & L. Leonard (Eds.), *Speech and language impairment in children - causes, characteristics, intervention and outcome*. Psychology Press.

Bishop, D. V. M. (2010). Overlaps between autism and language impairment: Phenomimicry or shared etiology? *Behavior Genetics*, *40*(5), 618–629. https://doi.org/10.1007/s10519-010-9381-x

Bishop, D. V. M., & Baird, G. (2001). Parent and teacher report of pragmatic aspects of communication: Use of the Children's Communication Checklist in a clinical setting. *Developmental Medicine and Child Neurology*, *43*(12), 809–818. https://doi.org/10.1111/j.1469-8749.2001.tb00168.x

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *37*(4), 391–403. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In *Formulaic sequences: Acquisition, processing, and use* (227–244). https://doi.org/10.1075/lllt.9.12bis

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology*, *28*(2), 561–571. https://doi.org/10.1207/s15326942dn2802_1

Blasko, D. G. (1999). Only the tip of the iceberg: Who understands what about metaphor? *Journal of Pragmatics*, *31*(12), 1675–1683. https://doi.org/10.1016/S0378-2166(99)00009-0

Blasko, D. G., & Connine, C. M. (1993). Effects of Familiarity and Aptness on Metaphor Processing. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *19*(2), 295–308. https://doi.org/10.1037/0278-7393.19.2.295

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476. https://doi.org/10.1017/S0142716416000060

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. https://doi.org/10.1016/j.jecp.2014.06.007

Blumenfeld, H., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660. https://doi.org/10.1080/01690960601000746

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*(3), 343–346. https://doi.org/10.3758/BF03198118

Boers, F., Demecheleer, M., & Eyckmans, J. (2004). Etymological elaboration as a strategy for learning idioms. In *Vocabulary in a second language: Selection, acquisition and testing* (pp. 53–78). https://doi.org/10.1075/lllt.10.07boe

Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish-English bilinguals. *International Journal of Bilingual Education and Bilingualism*, *13*(3), 325–344. https://doi.org/10.1080/13670050903342019

Bond, T. G. (2003). Validity and assessment : a Rasch measurement perspective. *Metodologia de Las Ciencias Del Comportamiento*, *5*, 179–194.

Bond, T. G., & Fox, C. M. (2007). Applying the rasch model: Fundamental measurement in the human sciences: Second edition. In *Applying the Rasch Model: Fundamental Measurement in the Human Sciences: Second Edition*. https://doi.org/10.4324/9781410614575

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science*, *14*(2), 256–269. https://doi.org/10.1111/j.1467-7687.2010.00974.x

Bonnell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom.* Washington, DC: George Washington Press.

Booth, J. R., Harasaki, Y., & Burman, D. D. (2006). Development of lexical and sentence level context effects for dominant and subordinate word meanings of homonyms. *Journal of Psycholinguistic Research*, *35*(6), 531–554. https://doi.org/10.1007/s10936-006-9028-5

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(1), 63–85. https://doi.org/10.1037/0278-7393.22.1.63

Boscolo Nale, V. (2020). Late talkers and typically developing children at 30 months: psycholinguistic variables as a lexical semantic predictor for developmental language disorder.

Bosma, E., Blom, E., Hoekstra, E., & Versloot, A. (2019). A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive vocabulary. *International Journal of Bilingual Education and Bilingualism*, *22*(4), 371–385. https://doi.org/10.1080/13670050.2016.1254152

Bovolenta, F., Goldoni, M., Clerici, P., Agosti, M., & Franceschini, M. (2009). Robot therapy for functional recovery of the upper limbs: a pilot study on patients after stroke. *Journal of Rehabilitation Medicine*, *41*(12), 971–975.

Bowes, A., & Katz, A. (2015). Metaphor creates intimacy and temporarily enhances theory of mind. *Memory and Cognition*, *43*(6), 953–963. https://doi.org/10.3758/s13421-015-0508-4

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. https://doi.org/10.1007/s11145-016-9699-8

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*(1), 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers: Performance on nonword repetition and lexical tasks. *American Journal of Speech-Language Pathology*, *24*(4), 126–138. https://doi.org/10.1044/2015_AJSLP-13-0106

Bréal, M. (1924). *Essai de sémantique: science des significations*. Gérard Monfort.

Brenner, G. (2011). *Webster's New World: American Idioms Handbook*. Houghton Mifflin Harcourt.

Brocher, A., Foraker, S., & Koenig, J. P. (2016). Processing of irregular polysemes in sentence reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(11), 1798. https://doi.org/10.1037/xlm0000271

Brocher, A., Koenig, J. P., Mauner, G., & Foraker, S. (2018). About sharing and commitment: the retrieval of biased and balanced irregular polysemes. *Language, Cognition and Neuroscience*, *33*(4), 443–466. https://doi.org/10.1080/23273798.2017.1381748

Brown, G. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behavior Research Methods, Instruments, & Computers*, *16*(6), 502–532. https://doi.org/10.3758/BF03200836

Borghesani, V., Hinkley, L. B., Ranasinghe, K. G., Thompson, M. M., Shwe, W., Mizuiri, D., ... & Nagarajan, S. S. (2020). Taking the sublexical route: brain dynamics of reading in the semantic variant of primary progressive aphasia. *Brain*, *143*(8), 2545-2560.

Cacciari, C., Bolognini, N., Senna, I., Pellicciari, M. C., Miniussi, C., & Papagno, C. (2011). Literal, fictive and metaphorical motion sentences preserve the motion component of the verb: A TMS study. *Brain and Language*, *119*(3), 149–157. https://doi.org/10.1016/j.bandl.2011.05.004

Cacciari, C., & Levorato, M. (1989). How children understand idioms in discourse. *Journal of Child Language*, *16*(2), 387–405.

Cacciari, C., & Glucksberg, S. (1995). Understanding Idioms: Do Visual Images Reflect Figurative Meanings? *European Journal of Cognitive Psychology*, *7*(3), 283–305. https://doi.org/10.1080/09541449508402450

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683. https://doi.org/10.1016/0749-596X(88)90014-9

Caillies, S., & Le Sourn-Bissaoui, S. (2013). Nondecomposable idiom understanding in children: Recursive theory of mind and working memory. *Canadian Journal of Experimental Psychology*, *67*(2), 108–116. https://doi.org/10.1037/a0028606

Cain, K., Oakhill, J., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, *9*(1), 65–87. https://doi.org/10.1016/j.jecp.2004.09.003

Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, *102*(3), 280–298. https://doi.org/10.1016/j.jecp.2008.08.001

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288. https://doi.org/10.1016/j.cognition.2013.11.015

Camaioni, L., Castelli, M. C., Longobardi, E., & Volterra, V. (1991). A parent report instrument for early language assessment. *First Language*, *11*(33), 345–358. https://doi.org/10.1177/014272379101103303

Caramazza, A., & Grober, E. (1976). Polysemy and the structure of the subjective lexicon. In *Georgetown University roundtable on languages and linguistics. Semantics: Theory and application* (pp. 181–206).

Carnell, S., Edwards, C., Croker, H., Boniface, D., & Wardle, J. (2005). Parental perceptions of overweight in 3-5 y olds. *International Journal of Obesity*, *29*(4), 353–355. https://doi.org/10.1038/sj.ijo.0802889

Carrow-Woolfolk, E. (1999). *CASL: Comprehensive assessment of spoken language*. American Guidance Services.

Carter, R. (2004). *Language and creativity: The art of common talk*. Routledge. https://doi.org/10.4324/9781315658971

Castarlenas, E., Jensen, M. P., Von Baeyer, C. L., & Miró, J. (2017). Psychometric properties of the numerical rating scale to assess self-reported pain intensity in children and adolescents. *Clinical Journal of Pain*, *33*(4), 376–383. https://doi.org/10.1097/AJP.0000000000000406

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language and Communication Disorders*, *49*(6), 649–671. https://doi.org/10.1111/1460-6984.12082

Cedar, P. (2008). Learner’s recognition of Thai-English idioms counterparts. *Linguistic Journal*, *3*(3), 1–13.

Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In *Intercultural Language Use and Language Learning* (pp. 41–57). Springer. https://doi.org/10.1007/978-1-4020-5639-0_3

Charteris-Black, J. (2002). Second language figurative proficiency: A comparative study of Malay and English. *Applied Linguistics*, *32*(1), 104–133. https://doi.org/10.1093/applin/23.1.104

Cheng, T., & Goodman, E. (2015). Race, ethnicity, and socioeconomic status in research on child health. *Pediatrics*, *135*(1), 225-237. https://doi.org/10.1542/peds.2014-3109

Chondrogianni, V., & Kwon, H. (2019). The development of English tense and agreement morphology in Welsh-English bilingual children with and without specific language impairment. *Applied Psycholinguistics*, *40*(4), 821–852. https://doi.org/10.1017/S0142716418000772

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, *1*(3), 318–348. https://doi.org/10.1075/lab.1.3.05cho

Cieślicka, A. (2015). *Idiom acquisition and processing by second/foreign language learners*. In R. R. Heredia & A. B. Cieślicka (Eds.), Bilingual figurative language processing (pp. 208– 244). New York, NY: Cambridge University Press.

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, *22*(2), 115–144. https://doi.org/10.1191/0267658306sr263oa

Cilibrasi, L., Stojanovik, V., Loucas, T., & Riddell, P. (2018). The role of noninitial clusters in the Children's Test of Nonword Repetition: Evidence from children with language impairment and typically developing children. *Dyslexia*, *24*(4), 322–335. https://doi.org/10.1002/dys.1599

Clark, H. H., & Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 591–608. https://doi.org/10.1016/S0022-5371(83)90364-X

Clark, E. V., & Clark, H. H. (1979). When Nouns Surface as Verbs. *Language*, *55*, 767–811. https://doi.org/10.2307/412745

Colston, H. L., & Gibbs, R. (2002). Are Irony and Metaphor Understood Differently? *Metaphor and Symbol*, *17*(1), 57–80. https://doi.org/10.1207/s15327868ms1701_5

Colston, H. L., & Kuiper, M. S. (2002). Figurative Language Development Research and Popular Children's Literature: Why We Should Know, "Where the Wild Things Are." *Metaphor and Symbol*, *17*(1), 27–43. https://doi.org/10.1207/s15327868ms1701_3

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Cometa, M. S., & Eson, M. E. (1978). Logical operations and metaphor interpretation: A Piagetian model. *Child Development*, *49*, 649–659. https://doi.org/10.2307/1128232

Commissaire, E., Duncan, L. G., & Casalis, S. (2019). Investigating pseudohomophone interference effects in young second-language learners. *Journal of Experimental Child Psychology*, *180*, 1–18. https://doi.org/10.1016/j.jecp.2018.11.010

Coney, J., & Lange, A. (2006). Automatic and attentional processes in the comprehension of unfamiliar metaphors. *Current Psychology*, *25*(2), 93. https://doi.org/10.1007/s12144-006-1006-4

Conger, R. D., & Donnellan, M. B. (2007). An Interactionist Perspective on the Socioeconomic Context of Human Development. *Annual Review of Psychology*, *58*, 175–199. https://doi.org/10.1146/annurev.psych.58.110405.085551

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89. https://doi.org/10.1093/applin/amm022

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition*, *2*(1), 130–138.

Cools, W., de Kristine, M., Samaey, C., & Andries, C. (2011). Fundamental movement skill performance of preschool children in relation to family context. *Journal of Sports Sciences*, *29*(7), 649–660. https://doi.org/10.1080/02640414.2010.551540

Cooper, T. (1998). Teaching Idioms. *Foreign Langauge Annuls*, *31*(2), 255–266.

Copestake, A., & Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics*, *12*(1), 15–67.

https://doi.org/10.1093/jos/12.1.15

Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, *24*(3), 267–295. https://doi.org/10.1177/0267658308090182

Corthals, P. (2010). Nine- to twelve-year olds' metalinguistic awareness of homonymy. *International Journal of Language and Communication Disorders*, *45*(1), 121–128. https://doi.org/10.3109/13682820902745446

Couper, M. P., & Groves, R. M. (1996). Household-level determinants of survey nonresponse. *New Directions for Evaluation*, *70*, 63–70. https://doi.org/10.1002/ev.1035

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Currie, J., & Thomas, D. (2001). Early test scores, school quality and SES: Longrun effects on wage and employment outcomes. *Research in Labor Economics*, *20*, 103–132. https://doi.org/10.1016/S0147-9121(01)20039-9

Currie, N. K., Lew, A. R., Palmer, T. M., Basu, H., De Goede, C., Iyer, A., & Cain, K. (2018). Reading comprehension difficulties in children with rolandic epilepsy. *Developmental Medicine and Child Neurology*, *60*(3), 275–282. https://doi.org/10.1111/dmcn.13628

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, *64*(4), 413–428. https://doi.org/10.1086/318638

Davison, M. D., & Hammer, C. S. (2012). Development of 14 English grammatical morphemes in SpanishEnglish preschoolers. *Clinical Linguistics and Phonetics*, *28*(8), 728–742. https://doi.org/10.3109/02699206.2012.700679

De Avila, E. A., & Duncan, S. E. (1983). *Language Assessment Scales. Monterey*. San Rafeal, CA: Linguametrics Group.

deBot, K. (2014). The effectiveness of early foreign language in the

Netherlands. *Studies in Second Language Learning and Teaching*, *4*(3), 409-418.

de Bruijn, E. R. A., Dijkstra, T., Chwilla, D. J., & Schriefers, H. J. (2001). Language context effects on interlingual homograph recognition: evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition*, *4*(2), 155–168. https://doi.org/10.1017/s1366728901000256

De Caro, E. E. R. (2009). The Advantages and Importance of Learning and Using Idioms in English. *Cuadernos de Lingüística Hispánica*, *14*, 121–136.

De Cat, C. (2020). Predicting Language Proficiency in Bilingual Children. *Cambridge University Press*. https://doi.org/https://doi.org/10.1017/S0272263119000597

De Houwer, A. (2013). Current and cumulative exposure in bilingual acquisition. In *Bilingualism: Language and cognition*.

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, *28*(3), 411–424. https://doi.org/10.1017/S0142716407070221

De Houwer, A. (2009). *Bilingual first language acquisition*.

Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language*, *46*(2), 334–367. https://doi.org/10.1017/S0305000918000491

Demie, F. (2003). Raising the achievement of Caribbean pupils in British schools: Unacknowledged problems and challenges for policy makers. *London Review of Education*, *1*(3), 229–249. https://doi.org/10.1080/1474846032000146785

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

Department for Communities and Local Government. (2015). The English Index of Multiple Deprivation (IMD) 2015 – Guidance. https://doi.org/https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015

Department for Education. (2013a). *National curriculum in England: primary curriculum*.

Department for Education. (2013b). *The national curriculum in England*.

Department for Education. (2014a). *Early years foundation stage profile attainment by pupil characteristics: 2013 to 2014*.

Department for Education. (2014b). *National curriculum assessments at key stage 2, 2014*.

Department for Education. (2014c). *Phonics screening check and key stage 1 assessments: England 2014*.

Department for Education. (2016). *Social Mobility: A local plan to improve opportunities for Bradford's children and young people*.

Department for Education. (2017). *Schools, Pupils and Their Characteristics: January 2017*.

Department for Education. (2020). *Schools, pupils and their characteristics*. Retrieved from https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics

Diaz, V., & Farrar, M. J. (2018). The missing explanation of the false-belief advantage in bilingual children: a longitudinal study. *Developmental Science*, *21*(4). https://doi.org/10.1111/desc.12594

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science*, *14*(2), 164–168. https://doi.org/10.1111/1467-9280.t01-1-01436

Dijkstra, T., De Bruijn, E., Schriefers, H., & Ten Brinke, S. (2000). More on interlingual homograph recognition: language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition*, *3*(1), 69–78. https://doi.org/10.1017/s1366728900000146

Dijkstra, T., & van Hell, J. G. (2003). Testing the Language Mode Hypothesis Using Trilinguals. *International Journal of Bilingual Education and Bilingualism*, *6*(1), 2–16. https://doi.org/10.1080/13670050308667769

Dixon, L. Q., Wu, S., & Daraghmeh, A. (2012). Profiles in Bilingualism: Factors Influencing Kindergartners' Language Proficiency. *Early Childhood Education Journal*, *40*(1), 25–34. https://doi.org/10.1007/s10643-011-0491-8

Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, *27*(2), 367–392. https://doi.org/10.1017/S0305000900004153

Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*(1), 203–214. https://doi.org/10.1017/S030500090300583X

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1432–1443. https://doi.org/10.1044/jslhr.4206.1432

Döpke, S. (2000). The Interplay Between Language-Specific Development and Crosslinguistic Influence. In *Cross-linguistic structures in simultaneous bilingualism* (pp. 79–103). https://doi.org/10.1075/sibil.21.07dop

Douglas, J. D., & Peel, B. (1979). The development of metaphor and proverb translation in children grades 1 through 7. *Journal of Educational Research*, *73*(2), 116–119. https://doi.org/10.1080/00220671.1979.10885218

Dowdy, E., Dever, B. V., DiStefano, C., & Chin, J. K. (2011). Screening for Emotional and Behavioral Risk Among Students With Limited English Proficiency. *School Psychology Quarterly*, *26*(1), 14. https://doi.org/10.1037/a0022072

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 56(8),* 848–856.

https://doi.org/10.1111/jcpp.12378

Duffy, S. A., Kambe, G., & Rayner, K. (2001). *The effect of prior disambiguating context on the comprehension of ambiguous words: Evidence from eye movements.* https://doi.org/10.1037/10459-002

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446. https://doi.org/10.1016/0749-596X(88)90066-6

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(6), 1428-1446. https://doi.org/10.1037/0012-1649.43.6.1428

Duncan, L. G., & Seymour, P. H. K. (2000). Socio-economic differences in foundation-level literacy. *British Journal of Psychology*, *91*(2), 145–166. https://doi.org/10.1348/000712600161736

Dunn, L., & Dunn, D. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Clinical Assessments.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). British Picture Vocabulary Scale (BPVS-II) 2nd ed. In *NFER-Nelson Publishing Company*.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188. https://doi.org/10.1177/0267658309349431

Duursma, E., Romero-Contreras, S., Szuber, A., Proctor, P., Snow, C., August, D., & Calderón, M. (2007). The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Applied Psycholinguistics*, *28*(1), 171–190. https://doi.org/10.1017/S0142716407070099

Eaker, S., Bergström, R., Bergström, A., Adami, H. O., & Nyren, O. (1998). Response rate to mailed epidemiologic questionnaires: A population-based randomized trial of variations in design and mailing routines.

American Journal of Epidemiology, *147*(1), 74–82. https://doi.org/10.1093/oxfordjournals.aje.a009370

Eberhardt, M., & Nadig, A. (2018). Reduced sensitivity to context in language comprehension: A characteristic of Autism Spectrum Disorders or of poor structural language ability? *Research in Developmental Disabilities*, *72*, 284–296. https://doi.org/10.1016/j.ridd.2016.01.017

Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: the current state of the literature. *Psychonomic Bulletin and Review*, *22*(1), 13–37. https://doi.org/10.3758/s13423-014-0665-7

Egilson, S. T., Ólafsdóttir, L. B., Leósdóttir, T., & Saemundsen, E. (2017). Quality of life of high-functioning children and youth with autism spectrum disorder and typically developing peers: Self- and proxy-reports. *Autism*, *21*(2), 133–141. https://doi.org/10.1177/1362361316630881

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the national reading panel's meta-analysis. *Review of Educational Research*, *71*(3), 393–447. https://doi.org/10.3102/00346543071003393

Eigsti, I., de Marchena, A., Schuh, J., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681–691.

Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In *Vocabulary description, acquisition and pedagogy* (pp. 122–139).

Elston-Gúttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language*, *52*(2), 256–283. https://doi.org/10.1016/j.jml.2004.11.002

Emmott, A. S., West, N., Zhou, G., Dunsmuir, D., Montgomery, C. J., Lauder, G. R., & von Baeyer, C. L. (2017). Validity of simplified versus standard self-report measures of pain intensity in preschool-aged children undergoing venipuncture. *Journal of Pain*, *18*(5), 564–573.

https://doi.org/10.1016/j.jpain.2016.12.015

Engel de Abreu, P. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, *19*(5), 529–537. https://doi.org/10.1080/09658211.2011.590504

Engel de Abreu, P. M. J., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science*, *23*(11), 1364–1371. https://doi.org/10.1177/0956797612443836

Ensminger, M. E., & Fothergill, K. E. (2003). A decade of measuring SES: What it tells us and where to go from here. In *Socioeconomic Status, Parenting, and Child Development* (pp. 13–27). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410607027-9

Espinosa, L. M., & López, M. L. (2007). Assessment Considerations for Young English Language Learners Across Different Levels of Accountability. In *The National Early Childhood Accountability Task Force and First 5 LA*.

Evans, G. W., Maxwell, L. E., & Hart, B. (1999). Parental language and verbal responsiveness to children in crowded homes. *Developmental Psychology*, *35*(4), 1020–1027. https://doi.org/10.1037/0012-1649.35.4.1020

Evans, M. A., & Gamble, D. L. (1988). Attribute saliency and metaphor interpretation in school-age children. *Journal of Child Language*, *15*(2), 435–449. https://doi.org/10.1017/S0305000900012447

Ezell, H. K., & Goldstein, H. (1991). Comparison of idiom comprehension of normal children and children with mental retardation. *Journal of Speech and Hearing Research*, *34*(4), 812–819. https://doi.org/10.1044/jshr.3404.812

Fairley, L., Cabieses, B., Small, N., Petherick, E., Lawlor, D., Pickett, K., & Wright, J. (2014). Using latent class analysis to develop a model of the relationship between socioeconomic position and ethnicity: Cross-sectional analyses from a multi-ethnic birth cohort study. *BMC Public Health*, *14*(1), 835. https://doi.org/10.1186/1471-2458-14-835

Falkum, I., Recasens, M., & Clark, E. V. (2017). “The moustache sits down first”: On the acquisition of metonymy. *Journal of Child Language*, *44*(1), 87–119. https://doi.org/10.1017/S0305000915000720

Falkum, I., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua, 157, 1-16*. https://doi.org/10.1016/j.lingua.2015.02.002

Farmer, M. (1997). Exploring the links between communication skills and social competence. *Educational and Child Psychology*, *14*, 38–44.

Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics*, *32*(4), 711.

Faust, M. E., Balota, D. A., Duchek, J. M., Gernsbacher, M. A., & Smith, S. (1997). Inhibitory control during sentence comprehension in individuals with dementia of the Alzheimer type. *Brain and Language*, *57*(2), 225–253. https://doi.org/10.1006/brln.1997.1747

Faust, M., & Mashal, N. (2007). The role of the right cerebral hemisphere in processing novel metaphoric expressions taken from poetry: A divided visual field study. *Neuropsychologia*, *45*(4), 860–870. https://doi.org/10.1016/j.neuropsychologia.2006.08.010

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*. https://doi.org/10.2307/1166093

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Fillmore, L. W. (2000). Loss of family languages: Should educators be concerned? *Theory into Practice*, *39*(4), 203–210. https://doi.org/10.1207/s15430421tip3904_3

Flavell, J., & Ross, L. (1981). *Social cognitive development: Frontiers and possible futures*. Cambridge University Press.

Fletcher, K. L., & Reese, E. (2005). Picture book reading with young children: A conceptual framework. *Developmental Review*, *25*(1), 64–103. https://doi.org/10.1016/j.dr.2004.08.009

Flexner, S.B. (Ed.) (2003). Random House Webster's Unabridged Dictionary (2nd Ed). New York: Random House

Fontes, A. B. A. D. L., & Schwartz, A. I. (2011). Working memory influences on cross-language activation during bilingual lexical disambiguation. *Bilingualism*, *14*(3), 360–370. https://doi.org/10.1017/S1366728910000374

Fontes, A. B. A. D. L., & Schwartz, A. I. (2015). Bilingual access of homonym meanings: Individual differences in bilingual access of homonym meanings. *Bilingualism*, *18*(4), 639–656. https://doi.org/10.1017/S1366728914000509

Forster, K. I., & Bednall, E. S. (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition*, *4*(1), 53–61. https://doi.org/10.3758/BF03213255

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*(2), 181–200. https://doi.org/10.1016/0749-596X(90)90071-7

Friedrich, M., & Friederici, A. (2005). Semantic sentence processing reflected in the event-related potentials of one-and two-year-old children. *Neuroreport*, *16*(16), 1801–1804.

Frisson, S. (2009). Semantic underspecification in language processing. *Linguistics and Language Compass*, *3*(1), 111–127. https://doi.org/10.1111/j.1749-818X.2008.00104.x

Frisson, S., & Frazier, L. (2005). Carving up word meaning: Portioning and grinding. *Journal of Memory and Language*, *53*(2), 277–291. https://doi.org/10.1016/j.jml.2005.03.004

Frisson, S., & Pickering, M. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, *22*(4), 595–613. https://doi.org/10.1080/01690960601017013

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. https://doi.org/10.1037/0278-7393.25.6.1366

Fuste-Herrmann, B. (2009). Idiom comprehension in bilingual and monolingual adolescents. In *Dissertation Abstracts International Section A: Humanities and Social Sciences*. University of South Florida.

Fuster, J. M. (1997). Network memory. *Trends in Neurosciences*, *20*(10), 451–459. https://doi.org/10.1016/S0166-2236(97)01128-4

Garnham, W. A., Brooks, J., Garnham, A., & Ostenfeld, A. M. (2000). From synonyms to homonyms: Exploring the role of metarepresentation in language understanding. *Developmental Science*, *3*(4), 428–441. https://doi.org/10.1111/1467-7687.00137

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*(5), 613–656.

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, *22*(8), 1019–1037. https://doi.org/10.1002/acp.1407

Gathercole, S., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177. https://doi.org/10.1037/0012-1649.40.2.177

Gathercole, S, & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336–360. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

Gathercole, V. C. M., Thomas, E. M., Roberts, E., Hughes, C., & Hughes, E. K. (2013). Why assessment needs to take exposure into account: Vocabulary and grammatical abilities in bilingual children. *Issues in the Assessment of Bilinguals*, *12*(2), 20–55.

Gathercole, V., Kennedy, I., & Thomas, E. M. (2016). Socioeconomic level and bilinguals' performance on language and cognitive measures. *Bilingualism: Language and Cognition*, *19*(5), 1057–1078. https://doi.org/10.1017/S1366728915000504

Gathercole, V., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism, 12(2,) 213*. https://doi.org/10.1017/S1366728909004015

Gatt, D., Baldacchino, R., & Dodd, B. (2020). Which measure of socioeconomic status best predicts bilingual lexical abilities and how? A focus on four-year-olds exposed to two majority languages. *Journal of Child Language*, *1*, 29. https://doi.org/10.1017/S0305000919000886

Gernsbacher, M. (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *Psychology of Learning and Motivation - Advances in Research and Theory*, *27*, 217–263. https://doi.org/10.1016/S0079-7421(08)60125-5

Gernsbacher, M. (1997a). Coherence cues mapping during comprehension. In *Processing interclausal relationships. Studies in the production and comprehension of text* (pp. 3–22).

Gernsbacher, M. (1997b). Two decades of structure building. *Discourse Processes*, *23*(3), 265–304. https://doi.org/10.1080/01638539709544994

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281. https://doi.org/10.1037/0096-3445.113.2.256

Gernsbacher, M., Robertson, R. R., & Werner, N. K. (2001). The costs and benefits of meaning. In *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 119–137).

Gernsbacher, M., St John, M. F., & Gorfein, D. S. (2001). On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. In *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity.* https://doi.org/10.1037/10459-000

Gerrig, R., & Gibbs, R. (1988). Beyond the lexicon: Creativity in language production. *Metaphor and Symbol*, *3*(3), 1–19.

Gertner, B. L., Rice, M. L., & Hadley, P. A. (1994). Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech and Hearing Research*, *37*(4), 913–923. https://doi.org/10.1044/jshr.3704.913

Geurts, H. M., Verté, S., Oosterlaan, J., Roeyers, H., Hartman, C. A., Mulder, E. J., van Berckelaer-Onnes, I. A., & Sergeant, J. A. (2004). Can the Children’s Communication Checklist differentiate between children with autism, children with ADHD, and normal controls? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(8), 1437–1453. https://doi.org/10.1111/j.1469-7610.2004.00326.x

Gibbs, R. (1982). *A critical examination of the contribution of literal meaning to understanding nonliteral discourse* (pp. 9–27).

Gibbs, R. (1989). Understanding and literal meaning. *Cognitive Science*, *13*, 243–251. https://doi.org/10.1016/0364-0213(89)90006-2

Gibbs, R. (1992). What do idioms really mean? *Journal of Memory and Language*, *31*(4), 485. https://doi.org/10.1016/0749-596X(92)90025-S

Gibbs, R. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.

Gibbs, R., Bogdanovich, J. M., Sykes, J. R., & Barr, D. J. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language*, *37*(2), 141–154. https://doi.org/10.1006/jmla.1996.2506

Gibbs, R., & Colston, H. L. (2012). Interpreting Figurative Meaning. In *Interpreting Figurative Meaning*. Cambridge University Press. https://doi.org/10.1017/cbo9781139168779

Gibbs, R., & Gerrig, R. J. (1989). How Context Makes Metaphor Comprehension Seem “Special.” *Metaphor and Symbolic Activity*, *4*(3), 145–158. https://doi.org/10.1207/s15327868ms0403_3

Gibbs, R., & Gonzales, G. P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition*, *20*(3), 243–259. https://doi.org/10.1016/0010-0277(85)90010-1

Gibbs, R., & Nayak, N. P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology*, *21*(1), 100–138. https://doi.org/10.1016/0010-0285(89)90004-2

Gibbs, R., Nayak, N. P., Bolton, J. L., & Keppel, M. E. (1989). Speakers’ assumptions about the lexical flexibility of idioms. *Memory & Cognition*, *17*(1), 58–68. https://doi.org/10.3758/BF03199557

Gibbs, R., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*(5), 576–593. https://doi.org/10.1016/0749-596X(89)90014-4

Gibbs, R. W. (1991). Semantic analyzability in children’s understanding of idioms. *Journal of Speech and Hearing Research*, *34*(3), 613–620. https://doi.org/10.1044/jshr.3403.613

Gibbs, Raymond W. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 56–66. https://doi.org/10.1037/0278-7393.16.1.56

Gibson, T. A., Oller, D. K., Jarmulowicz, L., & Ethington, C. A. (2012). The receptive-expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism*, *15*(1), 102. https://doi.org/10.1017/S1366728910000490

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014). The relation between language experience and receptive-expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism*, *17*(1), 90–110. https://doi.org/10.1080/13670050.2012.743960

Gillberg, C., & Rasmussen, P. (1994). Brief report: Four case histories and a literature review of williams syndrome and autistic behavior. *Journal of Autism and Developmental Disorders*, *24*(3), 381–393. https://doi.org/10.1007/BF02172235

Gillett, A. (2004). *Speak English Like an American: You Already Speak English--Now Speak it Even Better!* Language Success Press.

Gilmour, J., Hill, B., Place, M., & Skuse, D. H. (2004). Social communication deficits in conduct disorder: A clinical and community survey. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(5), 967–978. https://doi.org/10.1111/j.1469-7610.2004.t01-1-00289.x

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, *31*(7), 919–929. https://doi.org/10.1016/s0378-2166(98)00100-3

Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics*, *34*(4), 487–506. https://doi.org/10.1016/S0378-2166(01)00045-5

Giora, R. (2003). On Our Mind: Salience, Context, and Figurative Language. In *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780195136166.001.0001

Giora, R., & Fein, O. (1999). On understanding familiar and less-familiar figurative language. *Journal of Pragmatics*, *31*(12), 1601–1618. https://doi.org/10.1016/s0378-2166(99)00006-5

Glucksberg, S. (2001). Understanding Figurative Language: From Metaphor to Idioms. In *Understanding Figurative Language: From Metaphor to Idioms*. Oxford University Press on Demand. https://doi.org/10.1093/acprof:oso/9780195111095.001.0001

Glucksberg, S., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and*

Verbal Behavior, *21*(1), 85–98. https://doi.org/10.1016/S0022-5371(82)90467-4

Glucksberg, S., Kreuz, R. J., & Rho, S. H. (1986). Context Can Constrain Lexical Access. Implications for Models of Language Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 323–335. https://doi.org/10.1037/0278-7393.12.3.323

Godbee, K., & Porter, M. A. (2013). Attribution of negative intention in Williams syndrome. *Research in Developmental Disabilities*, *34*(5), 1602–1612. https://doi.org/10.1016/j.ridd.2013.01.019

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics*, *29*, 1–25. https://doi.org/10.1017/S014271640808003X

Goldstein, B. A., Bunta, F., Lange, J., Rodriguez, J., & Burrows, L. (2010). The effects of measures of language experience and language ability on segmental accuracy in bilingual children. *American Journal of Speech-Language Pathology*, *19*(3), 238–247. https://doi.org/10.1044/1058-0360(2010/08-0086)

Gollan, T. H., & Acenas, L. A. R. (2004). What Is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalog-English bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(1), 246–269. https://doi.org/10.1037/0278-7393.30.1.246

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, *13*(2), 197–208. https://doi.org/10.1017/S1355617707070038

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition*, *33*(7), 1220–1234. https://doi.org/10.3758/BF03193224

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562. https://doi.org/10.1037/0894-4105.16.4.562

Goossens, L. (1995). From three respectable horses' mouths: Metonymy and conventionalization in a diachronically differentiated data base. In *Pragmatics and Beyond New Series* (pp. 175–204).

Gorman, S. (2018). *What are Pragmatic Language Skills?* Sensational Kids.

Goswami, U., & Bryant, P. (1990). Phonological skills and learning to read. In *Phonological Skills and Learning to Read*. Psychology Press. https://doi.org/10.4324/9781315785110

Grice, H. (1989). *Studies in the Way of Words*. Harvard University Press. https://doi.org/10.5840/gps19913934

Grimm, A., & Schulz, P. (2014). Specific language impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indicators Research*, *7*(4), 821–841. https://doi.org/10.1007/s12187-013-9230-6

Grosjean, F. (2010). *Bilingual*. Harvard University Press.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, *64*(3), 299–308. https://doi.org/10.1086/317990

Grundy, J. G., & Keyvani Chahi, A. (2017). Post-conflict slowing effects in monolingual and bilingual children. *Developmental Science,* .20(1), 1-12. https://doi.org/10.1111/desc.12488

Grunwell, D., & Kerbel, D. (1996). A study of idiom comprehension in children with semantic-pragmatic difficulties. Part I: Task effects on the assessment of idiom comprehension in children. *International Journal of Language and Communication Disorders.*, *33*(1), 1–22.

Grüter, T., & Paradis, J. (2014). *Input and experience in bilingual development*.

John Benjamins Publishing Company.

Guan, C. Q., & Fraundorf, S. H. (2020). Cross-linguistic word recognition development among Chinese children: A multilevel linear mixed-effects modeling approach. *Frontiers in psychology*, *11*.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, *24*(2), 267. https://doi.org/10.1017/S0142716403000158

Gutiérrez-Clellen, V. F., Restrepo, M., & Simón-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research*, *49*, 1209–1223. https://doi.org/10.1044/1092-4388(2006/087)

Guttman, L. (1950). The principal components of scale analysis. In *Measurement of prediction* (pp. 312–361).

Halle, T., Hair, E., Wandner, L., McNamara, M., & Chien, N. (2012). Predictors and outcomes of early versus later English language proficiency among English language learners. *Early Childhood Research Quarterly*, *27*(1), 1–20. https://doi.org/10.1016/j.ecresq.2011.07.004

Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting Spanish-English bilingual children's language abilities. *Journal of Speech, Language, and Hearing Research*, *55*, 1251–1264. https://doi.org/10.1044/1092-4388(2012/11-0016)

Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2008a). Exposure to english before and after entry into head start: Bilingual children's receptive language growth in Spanish and English. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 30–56. https://doi.org/10.2167/beb376.0

Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2008b). The effect of summer vacation on bilingual preschoolers' language development. *Clinical Linguistics and Phonetics*, *22*(9), 686–702. https://doi.org/10.1080/02699200802028033

Hammer, C. S., Miccio, A. W., & Wagstaff, D. A. (2003). Home literacy experiences and their relationship to bilingual preschoolers' developing english literacy abilities: An initial investigation. *Language, Speech, and Hearing Services in Schools*, 34, 20-30. https://doi.org/10.1044/0161-1461(2003/003)

Hands, P., Walter, E., & Woodford, K. (2012). *Collins COBUILD idioms dictionary*. HarperCollins Publishers.

Hansen, P., Simonsen, H. G., Łuniewska, M., & Haman, E. (2017). Validating the psycholinguistic aspects of LITMUS-CLT: Evidence from Polish and Norwegian. *Clinical Linguistics & Phonetics*, *31*(11), 910–930. https://doi.org/10.1080/02699206.2017.1307455

Hardin, B. J., Scott-Little, C., & Mereoiu, M. (2013). Developing the bio questionnaire: A bilingual parent report tool for prekindergarten english learners of latino heritage. *Journal of Research in Childhood Education*, *27*(4), 485–509. https://doi.org/10.1080/02568543.2013.824940

Harris, R. J., & Nelson, E. M. (1992). Bilingualism: Not the Exception Any More. *Advances in Psychology*, *83*, 3–14. https://doi.org/10.1016/S0166-4115(08)61485-5

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Harzing, A. W. (1997). Response rates in international mail surveys: Results of a 22-country study. *International Business Review*, *6*(6), 641–665. https://doi.org/10.1016/s0969-5931(97)00040-1

Henderson, L. M., Clarke, P. J., & Snowling, M. J. (2011). Accessing and selecting word meaning in autism spectrum disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(9), 964–973. https://doi.org/10.1111/j.1469-7610.2011.02393.x

Hermanto, N., Moreno, S., & Bialystok, E. (2012). Linguistic and metalinguistic outcomes of intense immersion education: How bilingual? *International Journal of Bilingual Education and Bilingualism*, *15*(2), 131–145. https://doi.org/10.1080/13670050.2011.652591

Herschensohn, J. (2007). Language development and age. In *Language Development and Age*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486487

Hessel, A. K., & Murphy, V. A. (2019). 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*, *46*(2), 265–291. https://doi.org/10.1017/S0305000918000399

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language, 55*(2), 247–273.

Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(4), 686. https://doi.org/10.1037/0278-7393.28.4.686

Hirsh-Pasek, K., Gleitman, L. R., & Gleitman, H. (1978). What did the brain say to the mind? A study of the detection and report of ambiguity by young children. In *The child's conception of language* (pp. 97–132). Springer. https://doi.org/10.1007/978-3-642-67155-5_7

Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Katz, E. S., Kheirandish-Gozal, L., Neubauer, D. N., O'Donnell, A. E., Ohayon, M., Peever, J., Rawding, R., Sachdeva, R. C., Setters, B., Vitiello, M. V., Ware, J. C., & Adams Hillard, P. J. (2015). National sleep foundation's sleep time duration recommendations: Methodology and results summary. *Sleep Health*, *1*(4), 233–243. https://doi.org/10.1016/j.sleh.2014.12.010

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, *19*(4), 603–629. https://doi.org/10.1017/s0142716400010389

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development*, *62*(4), 782–796. https://doi.org/10.1111/j.1467-8624.1991.tb01569.x

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental Psychology*, *49*(1), 4–14. https://doi.org/10.1037/a0027238

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, E., & Laursen, B. (2019). Socioeconomic Status and Parenting. In *Handbook of Parenting*. https://doi.org/10.4324/9780429401459-13

Hoff, E., Laursen, B., & Tardiff, T. (2002). Socioeconomic status and parenting. In *Handbook of Parenting, Biology and Ecology of Parenting*. https://doi.org/10.1007/978-3-211-99131-2_903

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders*, *38*(4), 271–278. https://doi.org/10.1016/j.jcomdis.2005.02.003

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 385. https://doi.org/10.1037/a0038995

Hoffman, R. R., & Kemper, S. (1987). What could reaction-time studies be telling us about metaphor comprehension? *Metaphor and Symbolic Activity*, *2*(3), 149–186. https://doi.org/10.1207/s15327868ms0203_1

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 265–274.

Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education*, *39*(5), 304–312. https://doi.org/10.1177/0741932518773154

Howlin, P., Mawhood, L., & Rutter, M. (2000). Autism and developmental receptive language disorder—a follow-up comparison in early adult life. II: Social, behavioural, and psychiatric outcomes. *Journal of Child Psychology and Psychiatry*, *41*(5), 561-578.

Hsu, C. F. (2013). Contextual integration of causal coherence in people with Williams syndrome. *Research in Developmental Disabilities*, *34*(10), 332–3342. https://doi.org/10.1016/j.ridd.2013.06.031

Hu, Y.-H., & Fong, Y.-Y. (2010). Obstacles to conceptual-metaphor-guided L2 idiom interpretation. In *Fostering Language Teaching Efficiency through Cognitive Linguistics* (pp. 293–317). https://doi.org/10.1515/9783110245837.293

Hung, P. F. (2010). Mental imagery and idiom understanding in adults: Examining dual coding theory. In *ProQuest Dissertations and Theses*. University of Oregon.

Hutchinson, E., Bavin, E., Efron, D., & Sciberras, E. (2012). A comparison of working memory profiles in school-aged children with Specific Language Impairment, Attention Deficit/Hyperactivity Disorder, Comorbid SLI and ADHD and their typically developing peers. *Child Neuropsychology*, *18*(2), 190–207. https://doi.org/10.1080/09297049.2011.601288

Hutchinson, J., Dunford, J., & Treadaway, M. (2016). Divergent Pathways: the disadvantage gap, accountability and the pupil premium. *Education Policy Institute*. https://doi.org/10.1016/j.steroids.2006.05.001

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), 19–32.

https://doi.org/10.1111/1467-9817.261003

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*(3), 337–374. https://doi.org/10.1016/S0010-0285(02)00500-5

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Iluz-Cohen, P., & Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Bilingualism*, *16(4)*, 884.

Irujo, S. (1986). A piece of cake: Earning and teaching idioms. *ELT Journal*, *40*(3), 236–242. https://doi.org/10.1093/elt/40.3.236

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288. https://doi.org/10.1016/j.actpsy.2007.06.003

Jackendoff, R. (1995). The Boundaries of the Lexicon. In *Idioms: Structural and Psychological Perspectives*.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, *20*(3), 523–549. https://doi.org/10.1017/S0305000900008461

Jaen, M. (2007). A corpus-driven design of a test for assessing the EAL collocational competence of university students. *International Journal of English Studies*, *7*(2), 127–147.

Jager, B., & Cleland, A. A. (2016). Polysemy advantage with abstract but not concrete words. *Journal of Psycholinguistic Research*, *43*(1), 143–156. https://doi.org/10.1007/s10936-014-9337-z

Jang, J., Jeong, M., & Hwang, M. (2014). Comprehending and defining homonyms in school-aged children from multicultural families. *Communication Sciences and Disorders*, *19*(1), 71–79. https://doi.org/10.12963/csd.14121

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, *21*(12), 2434–2444. https://doi.org/10.1162/jocn.2008.21179

Janz, K. F., Lutuchy, E. M., Wenthe, P., & Levy, S. M. (2008). Measuring activity in children and adolescents using self-report: PAQ-C and PAQ-A. *Medicine and Science in Sports and Exercise*, *40*(4), 767–772. https://doi.org/10.1249/MSS.0b013e3181620ed1

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning Vocabulary Through Reading. *American Educational Research Journal*, *21*(4), 767–787. https://doi.org/10.3102/00028312021004767

Jia, G, & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*(1), 131. https://doi.org/10.1017/s0142716403000079

Jia, G., Aaronson, D., & Wu, Y. (2002). Long-term language attainment of bilingual immigrants: Predictive variables and language group differences. *Applied Psycholinguistics*, *23*(4), 599. https://doi.org/10.1017/S0142716402004058

Jia, G., & Fuse, A. (2007). Acquisition of english grammatical morphology by native mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2007/090)

Johnson, J., & Rosano, T. (1993). Relation of cognitive style to metaphor interpretation and second language proficiency. *Applied Psycholinguistics*, *14*(2), 159–175. https://doi.org/10.1017/S014271640000953X

Johnson, P., & Kossykh, Y. (2008). *Early years, life chances and equality: a literature review*.

Jones, H. R. (1973). The use of visual and verbal memory processes by three-year-old children. *Journal of Experimental Child Psychology*, *15*(2), 340–351.

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1051–1062. https://doi.org/10.1037//0278-7393.20.5.1051

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–518. https://doi.org/10.1016/j.tics.2005.09.009

Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory and Cognition*, *29*(2), 363–372. https://doi.org/10.3758/BF03194931

Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English: Lexical-semantic skills in L1 and L2. *Journal of Speech, Language, and Hearing Research*, *48*, 372–383. https://doi.org/10.1044/1092-4388(2005/026)

Karoly, L. A., & Gonzalez, G. C. (2011). Early care and education for children in immigrant families. In *Future of Children* (pp. 71–101). https://doi.org/10.1353/foc.2011.0005

Kašćelan, D., Prevost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2020). A Review of questionnaires quantifying bilingual experience in children: Do we document the same constructs?.

Katz, A., & Ferretti, T. (2001). Moment-by-moment reading of proverbs in literal and nonliteral contexts. *Metaphor and Symbol*, *16*(3), 193–221. https://doi.org/10.1207/s15327868ms1603&4_5

Katz, A. (2017). Psycholinguistic approaches to metaphor acquisition and use. In *The Routledge handbook of metaphor and language* (pp. 472–481).

Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism*, *18*(1), 27–28. https://doi.org/10.1017/S1366728914000510

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, *32*(4), 474–516. https://doi.org/10.1006/jmla.1993.1026

Kerbel, D, & Grunwell, P. (1998). A study of idiom comprehension in children with semantic-pragmatic difficulties. Part I: Task effects on the assessment of idiom comprehension in children. *International Journal of Language & Communication Disorders*, *33*(1), 1–22.

Kerbel, D., & Grunwell, P. (1997). Idioms in the classroom: An investigation of language unit and mainstream teachers' use of idioms. *Child Language Teaching and Therapy*, *13*(2), 113–123. https://doi.org/10.1177/026565909701300201

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, *28*(4), 375–385. https://doi.org/10.1016/0749-596X(89)90017-X

Kidd, E., & Holler, J. (2009). Children's use of gesture to resolve lexical ambiguity. *Developmental Science*, *12*(6), 903–913. https://doi.org/10.1111/j.1467-7687.2009.00830.x

Kilgarriff, A. (1992). Dictionary word sense distinctions: An enquiry into their nature. *Computers and the Humanities*, *26*(5), 365–387. https://doi.org/10.1007/BF00136981

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*(2), 259–282. https://doi.org/10.1006/jmla.2001.2779

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*(1), 205–223. https://doi.org/10.1006/brln.2001.2518

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*(1), 1–24. https://doi.org/10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Gracco, V. L., & Pike, G. B. (2014). Pathways to lexical ambiguity: fMRI evidence for bilateral fronto-parietal involvement in language processing. *Brain and Language*, *131*, 56–64. https://doi.org/10.1016/j.bandl.2013.06.002

Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language*, *123*(1), 11–21. https://doi.org/10.1016/j.bandl.2012.06.007

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1534–1543. https://doi.org/10.1037/a0013012

Kohn, M. L. (1963). Social Class and Parent-Child Relationships: An Interpretation. *American Journal of Sociology*, *68*(4), 471–480. https://doi.org/10.1086/223403

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science*, *13*(3), 441–453. https://doi.org/10.1111/j.1467-7687.2009.00907.x

Kousaie, S., Laliberté, C., Zunini, R. L., & Taler, V. (2015). A behavioral and electrophysiological investigation of the effect of bilingualism on lexical ambiguity resolution in young adults. *Frontiers in Human Neuroscience*, *9*, 682. https://doi.org/10.3389/fnhum.2015.00682

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English, Dartmouth Publishing Group*.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kwak, S., Byeong-Yong, H., & Jeong, K. (2015). Understanding homophones of children from multicultural families according to project presentation and social and cultural context conditions. *Speech Therapy Research*, *24*(4), 379–390.

Kwok, E. Y. L., Brown, H. M., Smyth, R. E., & Oram Cardy, J. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *9*, 202–222. https://doi.org/10.1016/j.rasd.2014.10.008

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208. https://doi.org/10.1016/S0364-0213(80)80017-6

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Larrañaga, P., & Guijarro-Fuentes, P. (2013). The linguistic competence of early Basque-Spanish bilingual children and a Spanish monolingual child. *International Journal of Bilingualism*, *17*(5), 577–601. https://doi.org/10.1177/1367006911435704

Laval, V. (2003). Idiom comprehension and metapragmatic knowledge in French children. *Journal of Pragmatics*, *35*(5), 723–739. https://doi.org/10.1016/S0378-2166(02)00137-6

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2009/08-0142)

Laws, G., & Bishop, D. V. M. (2004). Pragmatic language impairment and social deficits in Williams syndrome: A comparison with Down's syndrome and specific language impairment. *International Journal of Language and Communication Disorders*, *39*(1), 45–64. https://doi.org/10.1080/13682820310001615797

Lazar, R. T., Warr-Leeper, G. A., Nicholson, C. B., & Johnson, S. (1989). Elementary school teachers' use of multiple meaning expressions. *Language, Speech, and Hearing Services in Schools*, *20*(4), 420–430.

https://doi.org/10.1044/0161-1461.2004.420

Le Sourn-Bissaoui, S., Caillies, S., Bernard, S., Deleau, M., & Brulé, L. (2012). Children's understanding of ambiguous idioms and conversational perspective-taking. *Journal of Experimental Child Psychology*, *112*(4), 437–451. https://doi.org/10.1016/j.jecp.2012.02.003

LeBlanc, A. G., Broyles, S. T., Chaput, J. P., Leduc, G., Boyer, C., Borghese, M. M., & Tremblay, M. S. (2015). Correlates of objectively measured sedentary time and self-reported screen time in Canadian children. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 38. https://doi.org/10.1186/s12966-015-0197-1

Leinonen, E., Ryder, N., Ellis, M., & Hammond, C. (2003). The use of context in pragmatic comprehension by specifically language-impaired and control children. *Linguistics*, *41*, 407–423. https://doi.org/10.1515/ling.2003.014

Leonard, L. B., & Bortolini, U. (1998). Grammatical morphology and the role of weak syllables in the speech of Italian-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1363–1374. https://doi.org/10.1044/jslhr.4106.1363

Levorato, M. C., & Cacciari, C. (1992). Children's comprehension and production of idioms: the role of context and familiarity. *Journal of Child Language*, *19*(2), 415–433.

Levorato, M. C., & Cacciari, C. (1999). Idiom comprehension in children: Are the effects of semantic analysability and context separable? *European Journal of Cognitive Psychology*, *11*(1), 51–66. https://doi.org/10.1080/713752299

Levorato, M. C., & Cacciari, C. (1995). The effects of different tasks on the comprehension and production of idioms in children. *Journal of Experimental Child Psychology*, *60*(2), 261–283. https://doi.org/10.1006/jecp.1995.1041

Levy, H., & Hanulíková, A. (2019). Variation in children's vowel production: Effects of language exposure and lexical frequency. *Laboratory Phonology*, *10*(1). https://doi.org/10.5334/labphon.131

Lewis, F. M., Murdoch, B. E., & Woodyatt, G. C. (2007). Linguistic abilities in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *1*(1), 85–100. https://doi.org/10.1016/j.rasd.2006.08.001

Lezak, M. (1995). Executive functions and motor performance. In *In Neuropsychological Assessment* (pp. 650–685).

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210. https://doi.org/10.3758/BF03192770

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory and Cognition*, *36*(6), 1103–1121. https://doi.org/10.3758/MC.36.6.1103

Lindsay, G., Soloff, N., Law, J., Band, S., Peacey, N., Gascoigne, M., & Radford, J. (2002). Speech and language therapy services to education in England and Wales. *International Journal of Language and Communication Disorders*, *37*(3), 273–288.

Liontas, J. (2017). Why teach idioms? A challenge to the profession. *Iranian Journal of Language Teaching Research*, *5*(3), 5–25.

Liontas, J. (2002). Exploring second language learners' notions of idiomaticity. *System*, *30*(3), 289–313. https://doi.org/10.1016/S0346-251X(02)00016-7

Littlemore, J. (2015). Metonymy. In *The Cambridge Handbook of Cognitive Linguistics*. Cambridge University Press. https://doi.org/10.1017/9781316339732.026

Littlemore, J., & Low, G. (2006). Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics*, *27*(2), 268–294. https://doi.org/10.1093/applin/aml004

Littlemore, J., & Tagg, C. (2018). Metonymy and text messaging: A framework for understanding creative uses of metonymy. *Applied Linguistics*, *39*(4), 481–507. https://doi.org/10.1093/applin/amw018

Lloyd, K. M., Tienda, M., & Zajacova, A. (2001). rends in the Educational Achievement of Minority Students since Brown v. Board of Education.

Achieving High Educational Standards for All, 147–182.

Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language and Communication Disorders*, *37*(1), 3–15. https://doi.org/10.1080/13682820110089911

Lodge, D., & Leach, E. (1975). Children’s acquisition of idioms in the English language. *Journal of Speech and Hearing Research*, *18*(3), 521–529.

Love, T., Maas, E., & Swinney, D. (2003). The influence of language exposure on lexical and syntactic language processing. *Experimental Psychology*, *50*(3), 204. https://doi.org/10.1026//1617-3169.50.3.204

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Lyons, J. (1981). *Language and linguistics*. Cambrige University Press.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676.

MacGregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia*, *68*, 126–138. https://doi.org/10.1016/j.neuropsychologia.2015.01.008

Maciejewski, G. (2018). *Representation and processing of semantic ambiguity*. The University of Leeds.

Maciejewski, Greg, Rodd, J. M., Mon-Williams, M., & Klepousniotou, E. (2020). The cost of learning new meanings for familiar words. *Language, Cognition and Neuroscience*, *35*(2), 188–210. https://doi.org/10.1080/23273798.2019.1642500

Macintyre, P. D., Dornyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal*, *82*(4), 545–562.

https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

Mantyla, K. (2004). Idioms and Language Users: The Effect of the Characteristics of Idioms on Their Recognition and Interpretation by Native and Nonnative Speakers of English. In *Dissertation Abstracts International, C: Worldwide*. Jyväskylän yliopisto.

Marcel, A. (1980). Explaining selective effects of prior context on perception: The need to distinguish conscious and preconscious processes in word recognition. In *Attention and performance*.

Marinis, T., & Armon-Lotem, S. (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Miltilingual Matters.

Marmurek, H. H. C., & Rossi, M. (1993). The development of strategic processing of ambiguous words: Riddles versus neutral context. *Journal of Genetic Psychology*, *154*(4), 475–486. https://doi.org/10.1080/00221325.1993.9914746

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism*, *11*(1), 81. https://doi.org/10.1017/S1366728907003227

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language*, *85*(3), 451–466. https://doi.org/10.1016/S0093-934X(03)00070-1

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

Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. C. (2009). Peer effects on children’s language achievement during pre-kindergarten. *Child Development*, *80*(3), 686–702. https://doi.org/10.1111/j.1467-8624.2009.01291.x

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children’s printed word database: Continuities and changes over time in children’s early reading vocabulary. *British Journal of Psychology*, *101*(2), 221–242. https://doi.org/10.1348/000712608X371744

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, *14*(3), 186–223. https://doi.org/10.1080/15475441.2018.1455584

Mazzocco, M. M. M. (1997). Children's interpretations of homonyms: A developmental study. *Journal of Child Language*, *24*(2), 441–467. https://doi.org/10.1017/S0305000997003103

McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science*, *30*(1), 181–192. https://doi.org/10.1207/s15516709cog0000_49

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders*, *42*(1), 35–47. https://doi.org/10.1007/s10803-011-1210-4

McKown, C. (2007). Concurrent validity and clinical usefulness of several individually administered tests of children's social–emotional cognition. *Journal of Clinical Child and Adolescent Psychology*, *36*(1), 29–41.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift Fur Sprachwissenschaft*, *28*(1), 5–34. https://doi.org/10.1515/ZFSW.2009.002

Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation*, *70*(4), 308–312. https://doi.org/10.5555/uri:pii:0003999389901512

Miccio, A., Tabors, P. O., Paez, M., Hammer, C. S., & Wagstaff, D. (2003). Vocabulary development in Spanish-speaking Head Start children of Puerto Rican descent. *Proceedings of the 4th International Symposium on Bilingualism*, 1614–1617.

Miles, C., Morgan, M. J., Milne, A. B., & Morris, E. D. M. (1996). Developmental and individual differences in visual memory span. *Current Psychology*, *15*(1), 53–67. https://doi.org/10.1007/bf02686934

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, *33*(2), 175–202.

Miyake, A., & Shah, P. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*(2), 187–202. https://doi.org/10.1016/j.jecp.2012.09.002

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 716–726. https://doi.org/10.1111/j.1467-7687.2007.00623.x

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116.

Mount-Weitz, J. (1996). Vocabulary development and disorders in African American children. In *Communication development and disorders in African American children: Research, assessment, and intervention* (pp. 189–226).

Muñoz, M. L., & Marquardt, T. P. (2003). Picture naming and identification in bilingual speakers of Spanish and English with and without aphasia. *Aphasiology*, *17*(12), 1115–1132. https://doi.org/10.1080/02687030344000427

Murphy, M. (2006). The Language of Word Meaning. *Language*, *82*(3), 672–673.

Nagy, W. E. (1988). Teaching vocabulary to improve reading comprehension. Illinois: National Council of Teachers of English

Nagy, W. E., & Anderson, R. C. (1995). Metalinguistic awareness and literacy acquisition in different languages. In *Center for the Study of Reading Technical Report*.

NALDIC. (2015). http://www.naldic.org.uk/research-and-information/eal-statistics/eal-pupils/. Accessed: 20/10/2020

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *Npj Science of Learning*, *2*(1), 1–4. https://doi.org/10.1038/s41539-017-0004-7

Nekrasova, T. M. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning*, *59*(3), 647–686. https://doi.org/10.1111/j.1467-9922.2009.00520.x

Nelson, J. R., & Stage, S. A. (2007). Fostering the development of vocabulary knowledge and reading comprehension though contextually-based multiple meaning vocabulary instruction. In *Education and Treatment of Children* (pp. 1–22). https://doi.org/10.1353/etc.2007.0003

Nelson, K., Rescorla, L., Gruendel, J., & Benedict, H. (1978). Early lexicons: What do they mean? *Child Development*, *49*, 960–968.

Nerlich, B., Clarke, D. D., & Todd, Z. (1999). "Mummy, I like being a sandwich." In *Metonymy in Language and Thought* (pp. 361–383). https://doi.org/10.1075/hcp.4.21ner

Nesi, B., Levorato, M. C., Roch, M., & Cacciari, C. (2006). To break the embarrassment: Text comprehension skills and figurative competence in skilled and less-skilled text comprehenders. *European Psychologist*, *11*(2), 128–136. https://doi.org/10.1027/1016-9040.11.2.128

Netten, A., Luyten, H., Droop, M., & Verhoeven, L. (2016). Role of linguistic and sociocultural diversity in reading literacy achievement: A multilevel approach. *Journal of Research in Reading*, *39*(2), 189–208. https://doi.org/10.1111/1467-9817.12032

Neuman, S. B., Kaefer, T., & Pinkham, A. M. (2018). Double dose of disadvantage: Language experiences for low-income children in home and school. *Journal of Educational Psychology*, *110*, 102–118.

https://doi.org/10.1037/edu0000201

NICHD Early Child Care Research Network. (2005). *Child care and child development: Results from the NICHD study of early child care and youth development*.

Nikoalenko, N. (2004). Metaphorical and associative thinking in healthy children and in children with Asperger's syndrome at different ages. *Human Physiology*, *30*(5), 532–536.

Nikolaenko, N. N. (2004). Metaphorical and associative thinking in healthy children and in children with Asperger's Syndrome at different ages. *Human Physiology*, *30*(5), 532–536. https://doi.org/10.1023/b:hump.0000042608.36581.3b

Nippold, M. (1991). Evaluating and enhancing idiom comprehension in language-disordered students. *Language, Speech, and Hearing Services in Schools*, *22*(3), 100–106. https://doi.org/10.1044/0161-1461.2203.100

Nippold, M. (2007). Later Language Development: School-Age Children. In *Adolescents, and Young Adults*. PRO-ED.

Nippold, M. A., & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech and Hearing Research*, *36*(4), 728–737. https://doi.org/10.1044/jshr.3604.728

Nippold, M. A., & Tarrant Martin, S. (1989). Idiom interpretation in isolation versus context: A developmental study with adolescents. *Journal of Speech and Hearing Research*, *32*(1), 59–66. https://doi.org/10.1044/jshr.3201.59

Nippold, M. A., & Taylor, C. L. (1995). Idiom understanding in youth: Further examination of familiarity and transparency. *Journal of Speech and Hearing Research*, *38*(2), 426–433. https://doi.org/10.1044/jshr.3802.426

Nippold, M. A. (2016). Later language development : school-age children, adolescents, and young adults. In *PRO-ED, Inc.*

Nippold, M. A., & Duthie, J. K. (2003). Mental imagery and idiom comprehension: A comparison of school-age children and adults. *Journal of*

Speech, Language, and Hearing Research, *46*, 788–799. https://doi.org/10.1044/1092-4388(2003/062)

Nippold, M. A., Moran, C., & Schwarz, L. E. (2001). Idiom understanding in preadolescents: Synergy in action. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/1058-0360(2001/016)

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*, *24*, 384–391. https://doi.org/10.1044/1092-4388(2002/030)

Norbury, C. F., & Sparks, A. (2014). Language impairment within the autism spectrum. In *The Routledge International Handbook of Young Children's Thinking and Understanding*.

Norbury, C. F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research*, *47*, 1179–1193. https://doi.org/10.1044/1092-4388(2004/087)

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, *90*(2), 142–171. https://doi.org/10.1016/j.jecp.2004.11.003

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. M. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist - 2. *International Journal of Language and Communication Disorders*, *39*(3), 345–365. https://doi.org/10.1080/13682820410001654883

Nordmann, E., Cleland, A. A., & Bull, R. (2013). Cat got your tongue? Using the tip-of-the-tongue state to investigate fixed expressions. *Cognitive Science*, *37*(8), 1553–1564. https://doi.org/10.1111/cogs.12060

Nordmann, E., & Jambazova, A. A. (2017). Normative data for idiomatic expressions. *Behavior Research Methods*, *49*(1), 198–215. https://doi.org/10.3758/s13428-016-0705-5

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, *26*(6), 527–567. https://doi.org/10.1080/02643290903519367

O’Connor, M., O’Connor, E., Tarasuik, J., Gray, S., Kvalsvig, A., & Goldfeld, S. (2018). Academic outcomes of multilingual children in Australia. *International Journal of Speech-Language Pathology*, *20*(4), 393–405. https://doi.org/10.1080/17549507.2017.1292546

O’Neill, D. K. (2007). The language use inventory for young children: A parent-report measure of pragmatic language development for 18- to 47-month-old children. *Journal of Speech, Language, and Hearing Research*, *50*, 214–228. https://doi.org/10.1044/1092-4388(2007/017)

Oden, G. C., & Spira, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *The Quarterly Journal of Experimental Psychology Section A*, *35*(1), 51–64. https://doi.org/10.1080/14640748308402116

Oller, D. K., & Eilers, R. E. (2002). Language and Literacy in Bilingual Children. In *Language and Literacy in Bilingual Children* (2nd ed.). Multilingual Matters. https://doi.org/10.21832/9781853595721

Oller, D. Kimbrough, Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, *28*(2), 191–230. https://doi.org/10.1017.S0142716407070117

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*(3), 225–236. https://doi.org/10.3758/BF03196957

Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, *25*(1), 45–53. https://doi.org/10.1111/j.1741-5446.1975.tb00666.x

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. https://doi.org/10.1037/0022-0663.98.3.554

Oxley, E., Cattani, A., Chondrogianni, V., White, L., & De Cat, C. (2019). *Assessing bilingual children: a survey of professional practices in the UK.*

Ozcaliskan, S. (2014). Development of metaphor. In *Encyclopedia of language development* (pp. 374–375).

Ozcaliskan, S. (2007). Metaphors we move by: Children’s developing understanding of metaphorical motion in typologically distinct languages. *Metaphor and Symbol*, *22*(2), 147–168.

Ozdener, N., & Satar, H. M. (2008). Computer-mediated communication in foreign language education: Use of target language and learner perceptions. *Turkish Online Journal of Distance Education*, *9*(2). https://doi.org/10.17718/tojde.21536

Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly*, *46*(112–125). https://doi.org/10.1016/j.ecresq.2018.04.001

Paivio, A. (1971). Imagery and language. In *Imagery* (pp. 7–32).

Paivio, A. (1986). *Mental Representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Palmer, M. B. (1972). Effects of categorization, degree of bilingualism and language upon recall of select monolinguals and bilinguals. *Journal of Educational Psychology*, *63*(2), 160. https://doi.org/10.1037/h0032443

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782. https://doi.org/10.1111/j.1467-8624.2005.00876.x

Papagno, C., & Caporali, A. (2007). Testing idiom comprehension in aphasic patients: The effects of task and idiom type. *Brain and Language*, *100*(2), 208–220. https://doi.org/10.1016/j.bandl.2006.01.002

Paradis, J. (2009). *Oral language development in French and English and the role of home input factors.*

Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). Bilingual children with specific language impairment: How do they compare with their monolingual peers. *Journal of Speech, Language and Hearing Research*, *46*, 1-15.

Paradis, J. (2011a). Individual differences in child English second language acquisition. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J. (2011b). The impact of Input factors on bilingual development. *Linguistic Approaches to Bilingualism*, *1*(1), 67–70. https://doi.org/10.1075/lab.1.1.09par

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, *43*(6), 474–497. https://doi.org/10.1016/j.jcomdis.2010.01.002

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children. *Studies in Second Language Acquisition*, *18*, 1–25. https://doi.org/10.1017/s0272263100014662

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among english-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2012/12-0050)

Parks, R., Ray, J., & Bland, S. (1998). *Wordsmyth English Dictionary-Thesaurus*. Online.

Patterson, J. L. (1998). Expressive vocabulary development and word combinations of Spanish-English bilingual toddlers. *American Journal of Speech-Language Pathology*, *7*(4), 46–56. https://doi.org/10.1044/1058-0360.0704.46

Patterson, J. L. (2000). Observed and Reported Expressive vocabulary and word combinations in bilingual toddlers. *Journal of Speech, Language, and*

Hearing Research, *43*(1), 121–128. https://doi.org/10.1044/jslhr.4301.121

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58. https://doi.org/10.1017/s0142716400009863

Pearson, B. Z, & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, *44*(4), 617–653. https://doi.org/10.1111/j.1467-1770.1994.tb00633.x

Pennington, B. F., Bennetto, L., McAleer, O., & Roberts Jr., R. J. (1996). Executive functions and working memory: Theoretical and measurement issues. In *Attention, memory, and executive function*.

Pérez-Tattam, R., Ezeizabarrena, M. J., Stadthagen-González, H., & Gathercole, V. C. (2019). Gender assignment to Spanish pseudowords by monolingual and Basque-Spanish bilingual children. *Languages*, *4*(3), 58.

Perfetti, C. A., & Goodman, D. (1970). Semantic constraint on the decoding of ambiguous words. *Journal of Experimental Psychology*, *86*(3), 420–427.

Perkins, S. C., Finegood, E. D., & Swain, J. E. (2013). Poverty and language development: Roles of parenting and stress. *Innovations in Clinical Neuroscience*, *10*(4), 10–19.

Perlini, C., Bellani, M., Finos, L., Lasalvia, A., Bonetto, C., Scocco, P., D'Agostino, A., Torresani, S., Imbesi, M., Bellini, F., Konze, A., Veronese, A., Ruggeri, M., & Brambilla, P. (2018). Non literal language comprehension in a large sample of first episode psychosis patients in adulthood. *Psychiatry Research*, *260*, 78–89. https://doi.org/10.1016/j.psychres.2017.11.032

Persici, V., Vihman, M., Burro, R., & Majorano, M. (2019). Lexical access and competition in bilingual children: The role of proficiency and the lexical similarity of the two languages. *Journal of Experimental Child Psychology*, *179*, 103–125. https://doi.org/10.1016/j.jecp.2018.10.002

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(6), 1252. https://doi.org/10.1037/0278-7393.30.6.1252

Pfenninger, S. (2014). The missunderstood variable: Age effects as a function of type of instruction. *Studies in second language learning and teaching*, *4*(3), 529-556.

Phelps-Terasaki, D., & Phelps-Gunn, T. (2007). *TOPL-2: Test of Pragmatic Language*. Pro-Ed.

Piaget, J. (1963). *The origins of intelligence in children.* Norton.

Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(2), 556. https://doi.org/10.1037/0278-7393.27.2.556

Pierce, J. . (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory and Cognition*, *28*(4), 657–666. https://doi.org/10.3758/BF03201255

Pigdon, L., Willmott, C., Reilly, S., Conti-Ramsden, G., & Morgan, A. T. (2020). What predicts nonword repetition performance? *Child Neuropsychology*, *26*(4), 518–533. https://doi.org/10.1080/09297049.2019.1674799

Place, K., & Becker, J. (1991). The influence of pragmatic competence on the likeability of grade-school children. *Discourse Processes*, *14*(2), 227–241.

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-Year-Olds' bilingual proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and*

Cognitive Processes, *12*(5), 765–806.

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 5353–5551. https://doi.org/10.1016/j.jecp.2012.06.013

Portes, A., & Hao, L. (1998). Bilingualism and loss of language in the second generation. *Sociology of Education*, *71*(4), 269–294. https://doi.org/10.2307/2673171

Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics*, *25*(1), 51–79. https://doi.org/10.1075/bjl.25.04pou

Pouscoulous, N. (2014). “The elevator’s buttocks”: Metaphorical abilities in children. In *Pragmatic development in first language acquisition* (pp. 239–259). https://doi.org/10.1075/tilar.10.14pou

Powell, J. L., Furlong, J., de Bézenac, C. E., O'Sullivan, N., & Corcoran, R. (2019). The pragmatics of pragmatic language and the curse of ambiguity: An fMRI study. *Neuroscience*, *418*, 96-109.

Pring, T., Flood, E., Dodd, B., & Joffe, V. (2012). The working practices and clinical experiences of paediatric speech and language therapists: A national UK survey. *International Journal of Language and Communication Disorders*, *47*, 696–708. https://doi.org/10.1111/j.1460-6984.2012.00177.x

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism*, *13*(2), 253. https://doi.org/10.1017/S1366728909990526

Qi, C., Kaiser, A., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Language, Speech, and Hearing Services in Schools*. https://doi.org/10.1044/0161-1461(2006/002)

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2010). Shifting senses in lexical semantic development. *Cognition*, *117*(1), 17–37. https://doi.org/10.1016/j.cognition.2010.06.007

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2011). Rules, radical pragmatics and restrictions on regular polysemy. *Journal of Semantics*, *28*(4), 485–512. https://doi.org/10.1093/jos/ffr005

Rabagliati, H., & Srinivasan, M. (2018). *Word Senses*. In C. Cummins (Ed.), *Oxford Handbook of Experimental Semantics and Pragmatics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198791768.013.20

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. In *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.*

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Raynor, P., Duley, L., Small, N., Tuffnell, D., Wild, C., Wright, J., Airey, R., Akhtar, S., Brown, H., Crompton, J., Davies, J., Emsley, S., Farrar, D., Hepworth, S., Jankowicz, D., Johnson, W., Lodge, V., McKinney, T., Oddie, S., … McKinney, P. (2008). Born in Bradford, a cohort study of babies born in Bradford, and their parents: Protocol for the recruitment phase. *BMC Public Health*, *8*, 327–340. https://doi.org/10.1186/1471-2458-8-327

Rescorla, L., & Achenbach, T. M. (2002). Use of the language development survey (LDS) in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language, and Hearing Research*, *45*, 733–743. https://doi.org/10.1044/1092-4388(2002/059)

Rescorla, L., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language, and Hearing Research*, *44*, 434–445. https://doi.org/10.1044/1092-4388(2001/035)

Restrepo, M. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1398–1411. https://doi.org/10.1044/jslhr.4106.1398

Restrepo, M., & Kruth, K. (2000). Grammatical Characteristics of a Spanish-English Bilingual Child with Specific Language Impairment. *Communication*

Disorders Quarterly, *21*(1), 66–76. https://doi.org/10.1177/152574010002100201

Revier, R. L. (2009). Evaluating a new test of whole english collocations. In *Researching Collocations in Another Language: Multiple Interpretations* (pp. 125–138). Palgravw Macmillan. https://doi.org/10.1057/9780230245327

Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language Use Contributes to Expressive Language Growth: Evidence From Bilingual Children. *Child Development*, *89*(3), 929–940. https://doi.org/10.1111/cdev.12770

Richdale, A. L., & Baglin, C. L. (2015). Self-report and caregiver-report of sleep and psychopathology in children with high-functioning autism spectrum disorder: A pilot study. *Developmental Neurorehabilitation*, *18*(4), 272–279. https://doi.org/10.3109/17518423.2013.829534

Riley, A. W., Forrest, C. B., Rebok, G. W., Starfield, B., Green, B. F., Robertson, J. A., & Friello, P. (2004). The child report form of the CHIP-child edition reliability and validity. *Medical Care*, *42*, 221–231. https://doi.org/10.1097/01.mlr.0000114910.46921.73

Rimfeld, K., Malanchini, M., Hannigan, L. J., Dale, P. S., Allen, R., Hart, S. A., & Plomin, R. (2019). Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *60*(12), 1278–1288. https://doi.org/10.1111/jcpp.13070

Roberts, E., Bornstein, M. H., Slater, A. M., & Barrett, J. (1999). Early Cognitive Development and Parental Education. *Infant and Child Development*, *8*(1), 49–62. https://doi.org/10.1002/(SICI)1522-7219(199903)8:1<49::AID-ICD188>3.0.CO;2-1

Roberts, J. E., Burchinal, M., & Durham, M. (1999). Parents' report of vocabulary and grammatical development of African American preschoolers: Child and environmental associations. *Child Development*, *70*(1), 92–106. https://doi.org/10.1111/1467-8624.00008

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245–266. https://doi.org/10.1006/jmla.2001.2810

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory and Cognition*, *40*(7), 1095–1108. https://doi.org/10.3758/s13421-012-0209-1

Rodriguez-Fornells, A., Rotte, M., Helnze, H. J., Nösseit, T., & Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature*, *415*(6875), 1026–1029. https://doi.org/10.1038/4151026a

Romaine, S. (1995). *Bilingualism*. Wiley-Blackwell.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research and Evaluation*, *11*(1), 10-13.

Rossen, M., Bihrle, A., Kuma, E. S., Bellugi, U., & Johns, W. (1996). Interaction between langauge and cognition: Evidence from Williams syndrome. In H. J. Beitchman, N. Cohen, M. Konstantareas, & R. Tannock (Eds.), *Langauge learning and behaviour*. Cambridge University Press.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205. https://doi.org/10.1017/S0305000907008343

Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, *19*(6), 736-755.

Rucklidge, J. J., McLean, A. P., & Bateup, P. (2013). Criminal offending and learning disabilities in New Zealand youth: Does reading comprehension predict recidivism?. *Crime & Delinquency*, *59*(8), 1263-1286.

Ruhl, C. (1989). *Monosemy: A study in lexical semantics*. SUNY Press.

Rundblad, G., & Annaz, D. (2010a). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*(3), 547–563. https://doi.org/10.1348/026151009X454373

Rundblad, G., & Annaz, D. (2010b). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism*, *14*(1), 29–46. https://doi.org/10.1177/1362361309340667

Runnymede Trust. (1998). *Biennial Report*.

Salameh, E. K., Nettelbladt, U., Hakånsson, G., & Gullberg, B. (2002). Language impairment in Swedish bilingual children: A comparison between bilingual and monolingual children in Malmö. *Acta Paediatrica, International Journal of Paediatrics*, *91*(2), 229–234. https://doi.org/10.1080/080352502317285261

Saleh, N., & Zakaria, N. (2013). Investigating the difficulties faced in understanding, and strategies used in processing, English idioms by the Libyan students. *International Journal of English Language and Translation Studies*, *1*(2), 69–90.

Scheele, A. F. (2010). *Home language and mono-and bilingual children's emergent academic language: A longitudinal study of Dutch, Moroccan-Dutch, and Turkish-Dutch 3-to 6-*. Utrecht University.

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31*(1), 117–140. https://doi.org/10.1017/S0142716409990191

Schmidt, G. L., DeBuse, C. J., & Seger, C. A. (2007). Right hemisphere metaphor processing? Characterizing the lateralization of semantic processes. *Brain and Language*, *100*(2), 127–141. https://doi.org/10.1016/j.bandl.2005.03.002

Schmidt, G. L., & Seger, C. A. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition*, *71*(3), 375–386. https://doi.org/10.1016/j.bandc.2009.06.001

Schmitterer, A. M. A., & Schroeder, S. (2019). Effects of reading and spelling predictors before and after school entry: Evidence from a German longitudinal study. *Learning and Instruction*, *59*, 46–53. https://doi.org/10.1016/j.learninstruc.2018.09.005

Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, *49*, 157–165. https://doi.org/10.1016/j.learninstruc.2017.01.005

Schvaneveldt, R., Ackerman, B. P., & Semlear, T. (1977). The effect of semantic context on children's word recognition. *Child Development*, *48*, 612–616. https://doi.org/10.2307/1128660

Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(2), 243–256. https://doi.org/10.1037//0096-1523.2.2.243

Schwantes, F. M. (1981). Locus of the context effect in children's word recognition. *Child Development*, *52*, 895–903. https://doi.org/10.2307/1129092

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, *22*(1), 106–129. https://doi.org/10.1080/01690960500463920

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*, *32*(7), 866. https://doi.org/10.1037/neu0000467

Segal, D., Shalev, L., & Mashal, N. (2017). Attenuated hemispheric asymmetry in metaphor processing among adults with ADHD. *Neuropsychology*, *31*(6), 636. https://doi.org/10.1037/neu0000370

Semel, E. M., Wiig, E. H., & Secord, W. (2003). *CELF: Clinical Evaluation of Language Fundamentals. (CELF-4).*

Sereno, S. C., O'Donnell, P. J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 335. https://doi.org/10.1037/0096-1523.32.2.335

Serratrice, L., & De Cat, C. (2020). Individual differences in the production of referential expressions: The effect of language proficiency, language exposure and executive function in bilingual and monolingual children. *Bilingualism*, *40*(1), 119–151. https://doi.org/10.1017/S1366728918000962

Shelestiuk, H. V. (2005). Metonymy as a tool of cognition and representation: A natural language analysis. *Semiotica*, 125–144. https://doi.org/10.1515/semi.2005.2005.155.1-4.125

Shimahara, N. (1991). Social mobility and education: Burakumin in Japan. *Minority Status and Schooling: A Comparative Study of Immigrant and Involuntary Minorities*, 327–357.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1461–1481. https://doi.org/10.1044/jslhr.4206.1461

Shutova, E., Kaplan, J., Teufel, S., & Korhonen, A. (2013). A computational model of logical metonymy. *ACM Transactions on Speech and Language Processing*, *10*(3), 1–28. https://doi.org/10.1145/2483969.2483973

Simpson, G. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 120–136.

Simpson, G. (1994). Context and the processing of ambiguous words. *Handbook of Psycholinguistics*, *22*, 359–374.

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, *30*(6), 627–643.

Simpson, G., & Foster, M. (1986). Lexical ambiguity and children’s word recognition. *Developmental Psychology*, *22*(2), 147.

Simpson, G., & Lorsbach, T. (1983). The development of automatic and conscious components of contextual facilitation. *Child Development*, *54*, 760–772.

Simpson, Greg B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin*, *96*(2), 316–340. https://doi.org/10.1037/0033-2909.96.2.316

Siren, U. (1995). Minority language transmission in early childhood. *International Journal of Early Years Education*, *3*(2), 75–84.

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. https://doi.org/10.1177/0267658310382068

Smith Cairns, H., Waltzman, D., & Schlisselberg, G. (2004). Detecting the Ambiguity of Sentences: Relationship to Early Reading Skill. *Communication Disorders Quarterly*, *25*(2), 68–78. https://doi.org/10.1177/15257401040250020401

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*(1), 13–19. https://doi.org/10.1111/1467-9280.00403

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. https://doi.org/10.1007/s11145-014-9527-y

Snow, C., Arlman-Rupp, A., Hassing, Y., Jobse, J., Joosten, J., & Vorster, J. (1976). Mothers' speech in three social classes. *Journal of Psycholinguistic Research*, *5*, 1–20. https://doi.org/10.1007/BF01067944

Snow, C. E. (1999). Social perspectives on the emergence of language. In *The emergence of language* (pp. 257–276).

Snow, C. E., & Kang, J. Y. (2007). Becoming Bilingual, Biliterate, and Bicultural. In *Handbook of Child Psychology* (pp. 75–102). https://doi.org/10.1002/9780470147658.chpsy0403

Snowling, M. J., Bishop, D. V. M., Stothard, S. E., Chipchase, B., & Kaplan, C.

(2006). Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *Journal of Child Psychology and Psychiatry*, *47*(8), 759-765.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Hulme, C. (2009). *YARC York Assessment of Reading for Comprehension Passage Reading.*

Sohr-Preston, S. L., Scaramella, L. V., Martin, M. J., Neppl, T. K., Ontai, L., & Conger, R. (2013). Parental socioeconomic status, communication, and children's vocabulary development: A third-generation test of the family investment model. *Child Development*, *84*(3), 1046–1062. https://doi.org/10.1111/cdev.12023

Sorenson Duncan. T., & Paradis, J. (2020). How does maternal education influence the linguistic environment supporting bilingual language development in child second language learners of English? *International Journal of Bilingualism*, *24*(1), 46–61. https://doi.org/10.1177/1367006918768366

Sornig, K. (1988). Idioms in language teaching. In *Understanding the lexicon: Meaning, sense and world knowledge in lexical semantics* (pp. 280–290).

Spector, C. (1996). Children's comprehension of idioms in the context of humor. *Language, Speech, and Hearing Services in Schools*, *27*, 307–313. https://doi.org/10.1044/0161-1461.2704.307

Spencer, S., et al. (2017). Contribution of spoken language and socio-economic background to adolescents' educational achievement. *International journal of language & communication disorders, 52(2*), 184-196

Squires, J., Bricker, D., & Potter, L. W. (1997). Revision of a parent-completed developmental screening tool: Ages and stages questionnaires. *Journal of Pediatric Psychology*, 313–328. https://doi.org/10.1093/jpepsy/22.3.313

Sridhar, M., & Karunakaran, T. (2013). Idioms and importance of teaching idioms to ESL students: a study on teacher beliefs. *Asian Journal of Humanities and Social Sciences (AJHSS)*, *1*(1), 110–122.

Srinivasan, M, Berner, C., & Rabagliati, H. (2017). Children's use of lexical flexibility to structure new noun categories. In *CogSci*.

Srinivasan, M., Al-Mughairy, S., Foushee, R., & Barner, D. (2017). Learning language from within: Children use semantic generalizations to infer word meanings. *Cognition*, *159*, 11–24. https://doi.org/10.1016/j.cognition.2016.10.019

Srinivasan, M., Berner, C., & Rabagliati, H. (2019). Children use polysemy to structure new word meanings. *Journal of Experimental Psychology: General*, *148*(5), 926–948. https://doi.org/10.1037/xge0000454

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, *157*, 124–152. https://doi.org/10.1016/j.lingua.2014.12.004

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, *62*(4), 245–272. https://doi.org/10.1016/j.cogpsych.2011.03.002

Srinivasan, M., & Snedeker, J. (2014). Polysemy and the taxonomic constraint: Children's representation of words that label multiple kinds. *Language Learning and Development*, *10*(2), 97–128. https://doi.org/10.1080/15475441.2013.820121

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32–71.

Stanovich, K., West, R., & Feeman, D. (1981). A longitudinal study of sentence context effects in second-grade children: Tests of an interactive-compensatory model. *Journal of Experimental Child Psychology*, *32*, 185–199.

Stathi, K. (2006). Corpus linguistics meets cognitive linguistics: A frame work of the analysis of idioms. *German Cognitive Linguistics Association*, *28*, 9–27.

Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, *29*, 449–484. https://doi.org/10.1017/S0272263107070271

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-Year-olds: A review. *Clinical Child and Family Psychology Review*, *13*(3), 254–274. https://doi.org/10.1007/s10567-010-0071-2

Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, *34*(2), 827. https://doi.org/10.1017/S0305000905007099

Strand, S, Malmberg, L., & Hall, J. (2015). *English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database*.

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

Streiner, D. L., & Norman, G. R. (2003). *Health Measurement Scales*. Oxford University Press.

Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development*, *47*, 82–95. https://doi.org/10.1016/j.cogdev.2018.04.005

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., Suskind, E., Leininger, L., Goldin-Meadow, S., & LeVine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of Child Language*, *43*(2), 366–406. https://doi.org/10.1017/S0305000915000033

Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language*, *86*(2), 326–343.

Swineford, L. B., Thurm, A., Baird, G., Wetherby, A. M., & Swedo, S. (2014). Social (pragmatic) communication disorder: A research review of this new DSM-5 diagnostic category. *Journal of Neurodevelopmental Disorders*, *6*(1), 1–8. https://doi.org/10.1186/1866-1955-6-41

Swinney, D. A., & Hakes, D. T. (1976). Effects of prior context upon lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *15*, 681–689.

Swinney, D., & Prather, P. (1989). On the comprehension of lexical ambiguity by young children: Investigations into the development of mental modularity. In *Resolving semantic ambiguity* (pp. 225–238).

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. https://doi.org/10.1016/S0022-5371(79)90355-4

Syssau, A., Brouillet, D., & Groen, S. (2000). The study of inhibitory processes in aging with the Faust et al. paradigm (1997): Methodological issues. *Brain and Language*, *72*(2), 150–157. https://doi.org/10.1006/brln.1999.2282

Tabossi, P., Wolf, K., & Koterle, S. (2008). Idiom syntax: Idiosyncratic or principled? *International Journal of Psychology*, *43*(3), 415–415.

Tabossi, P, & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language*, *32*(3), 359–372.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, *27*(3), 324–340. https://doi.org/10.1016/0749-596X(88)90058-7

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, *49*(2), 161–167. https://doi.org/10.1007/BF00308682

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory and Cognition*, *37*(4), 529–540. https://doi.org/10.3758/MC.37.4.529

Tao, L., Cai, Q., & Gollan, T. H. (2019). Effects of cumulative language exposure on heritage and majority language skills: Spanish and Manadrin heritage speakers in the USA. *Linguistic Approaches to Bilingualism*, 1–24. https://doi.org/10.1075/lab.18044.tao

Taylor, J. (1995). *Linguistic Categorization: Prototypes in Linguistic Theory*. Blackwell.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care and Research*, *57*(8), 1358–1362. https://doi.org/10.1002/art.23108

Tesser, P. T. M., Merens, J. G. F., Praag, C. V., & Iedema, J. (1999). Rapportage minderheden 1999: Positie in het onderwijs en op de arbeidsmarkt. In *Sociaal en Cultureel Planbureau.*

Teubner-Rhodes, S. E., Mishler, A., Corbett, R., Andreu, L., Sanz-Torrent, M., Trueswell, J. C., & Novick, J. M. (2016). The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery. *Cognition*, *150*, 213–231. https://doi.org/10.1016/j.cognition.2016.02.011

The Bell Foundation. (2018). *Educational Outcomes of Children with English as an Additional Language*.

Thierry, G., & Yan, J. W. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Thompson. (1997). Children talking: The development of pragmatic competence. In *Multilingual Matters*.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In *Assessing Multilingual Children:*

Disentangling Bilingualism From Language Impairment. https://doi.org/10.21832/9781783093137-014

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, *4*(1), 1–21. https://doi.org/10.1080/14769670500215647

Thordardottir, E. T., & Juliusdottir, A. G. (2013). Icelandic as a second language: A longitudinal study of language knowledge and processing by school-age children. *International Journal of Bilingual Education and Bilingualism*, *16*(4), 411–435. https://doi.org/10.1080/13670050.2012.693062

Thorndike, E., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers Col. *Bur. Publ., Columbia University.*

Titone, D, & Libben, M. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom priming: A cross-modal priming investigation. *The Mental Lexicon*, *9*(3), 473–496.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In *Bilingual Figurative Language Processing*. https://doi.org/10.1017/CBO9781139342100.011

Tomlinson, R. M., Keyfitz, L., Rawana, J. S., & Lumley, M. N. (2017). Unique contributions of positive schemas for understanding child and adolescent life satisfaction and happiness. *Journal of Happiness Studies*, *18*(5), 1255–1274. https://doi.org/10.1007/s10902-016-9776-3

Trosbrog, A. (1985). Metaphoric productions and preferences in second language learners. In *The Ubiquity of Metaphor. Metaphor in Language and Thought* (pp. 525–557). https://doi.org/10.1075/cilt.29.20tro

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134. https://doi.org/10.1016/S0010-0277(99)00032-3

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*.

Tunmer, W. E., & Hoover, W. A. (2018). Cognitive and Linguistic Factors in Learning to Read. In *Reading Acquisition* (pp. 175–214). https://doi.org/10.4324/9781351236904-7

Twilley, L. C, & Dixon, P. (2000). Meaning resolution processess for words; A parallel independent model. *Psychonomic Bulletin and Review*, *7*(1), 49–82.

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, *22*(1), 111–126. https://doi.org/10.3758/BF03202766

Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, *38*, 225–236. https://doi.org/10.1044/0161-1461(2007/024)

Unsworth, N., & Engle, R. W. (2007). On the division of short-Term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038. https://doi.org/10.1037/0033-2909.133.6.1038

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2011). *Bilingual acquisition of Greek voice morphology and Dutch gender:* 590–602.

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism*, *16*(1), 86–110. https://doi.org/10.1017/S1366728912000284

Unsworth, S. (2015). Amount of exposure as a proxy for dominance in bilingual language acquisition. In *Language Dominance in Bilinguals*. https://doi.org/10.1017/cbo9781107375345.008

Unsworth, S. (2017). *Quantifying language experience in HL development*. The Oxford handbook of first language attrition.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, *35*(4), 765–805. https://doi.org/10.1017/S0142716412000574

Valdes, G. & Figueroa, R. A. (1996). *Bilingualism and testing: A special case of bias.* ABlex Publishing.

Van Herwegen, J., Dimitriou, D., & Rundblad, G. (2013). Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities*, *34*(4), 1300–1311. https://doi.org/10.1016/j.ridd.2013.01.017

Van Herwegen, J., & Rundblad, G. (2018). A cross-sectional and longitudinal study of novel metaphor and metonymy comprehension in children, adolescents, and adults with autism spectrum disorder. *Frontiers in Psychology*, *9*, 945. https://doi.org/10.3389/fpsyg.2018.00945

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language and Communication*, *24*(3), 207–240. https://doi.org/10.1016/j.langcom.2004.02.003

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2010). Children with SLI: who is at risk for dyslexia? *In III. All-European Dyslexia Conference, Date: 2010/04/22-2010/04/24, Location: Bruges, Belgium. European Dyslexia Association.*

Vernon-Roberts, A., Lopez, R. N., Frampton, C., Gearry, R. B., Lemberg, D. A., & Day, A. S. (2019). A symptom self-report tool for children with Inflammatory Bowel Disease (IBDnow). *Journal of Pediatric*

Gastroenterology and Nutrition, *69*(1), 7–11. https://doi.org/10.1097/MPG.0000000000002329

Voigt, L. F., Koepsell, T. D., & Daling, J. R. (2003). Characteristics of telephone survey respondents according to willingness to participate. *American Journal of Epidemiology*, *157*(1), 66–73. https://doi.org/10.1093/aje/kwf185

Vosniadou, S. (1987). Children and metaphors. *Child Development*, 870–885.

Vosniadou, S., Ortony, A., Reynolds, R. E., & Wilson, P. T. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development*, *55*, 1588–1606. https://doi.org/10.2307/1130028

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *7*(2), 228. https://doi.org/10.1037/a0027127

Wankoff, L. S. (1983). *Selected metalinguistic variables as they relate to conservation and reading achievement in normally-achieving and some learning-disabled children*. City UNiversity New York.

Wankoff, L. S., & Cairns, H. S. (2009). Why ambiguity detection is a predictor of early reading skill. *Communication Disorders Quarterly*, *30*(3), 183–192. https://doi.org/10.1177/1525740108324096

Warren, R., & Warren, N. (1976). Dual semantic encoding of homographs and homophones. *Memory & Cognition*, *4*, 586–592.

Warriner, K., Goyder, J., & Miller, S. (2002). Evaluating socio-economic status (SES) bias in survey nonresponse. *Journal of Official Statistics*, *18*(1), 1–12.

Weinreich, U. (1966). Exploration in semantic theory. In T. Sebeok (Ed.), *Current trends in linguistic* (3rd ed.). Mouton.

Weismer, S. E., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1249–1260. https://doi.org/10.1044/jslhr.4205.1249

Wechsler, D. (1992). Wechsler individual achievement test.

West, R., & Stanovich, K. (1978). Automatic contextual facilitation in readers of three ages. *Child Development*, *49*, 717–727.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461. https://doi.org/10.1037/0033-2909.91.3.461

White, L. J., & Greenfield, D. B. (2017). Executive functioning in Spanish- and English-speaking Head Start preschoolers. *Developmental Science*. https://doi.org/10.1111/desc.12502

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*(3), 848–872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x

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. https://doi.org/10.1111/cdev.12615

Whitford, H. C., & Dixson, R. J. (1953). *Handbook of American idioms and idiomatic usage*. Regents Publishing Company.

Whyte, E. M., Nelson, K. E., & Khan, K. S. (2013). Learning of idiomatic language expressions in a group intervention for children with autism. *Autism*, *17*(3), 449–464. https://doi.org/10.1177/1362361311422530

Wiig, E. H., & Secord, W. (1991). *Test of word knowledge*. Toronto, Ontario, Canada: Psychological Corporation.

Williams, J. N. (1992). Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research*, *21*, 193–218. https://doi.org/10.1007/BF01068072

Willinger, U., Deckert, M., Schmöger, M., Schaunig-Busch, I., Formann, A. K., & Auff, E. (2019). Developmental steps in metaphorical language abilities: The influence of age, gender, cognitive flexibility, information processing speed, and analogical reasoning. *Language and Speech*, *62*(2), 207–228.

https://doi.org/10.1177/0023830917746552

Winner, E. (1997). *The Point of Words: Children's Understanding of Metaphor and Irony*. Harvard University Press. https://doi.org/10.2307/4611932

Winner, E., Engel, M., & Gardner, H. (1980). Misunderstanding metaphor: What's the problem? *Journal of Experimental Child Psychology*, *30*(1), 22–32. https://doi.org/10.1016/0022-0965(80)90072-7

Winsler, A., Kim, Y. K., & Richard, E. R. (2014). Socio-emotional skills, behavior problems, and Spanish competence predict the acquisition of English among English language learners in poverty. *Developmental Psychology*, *50*(9), 2242. https://doi.org/10.1037/a0037161

Woodard, E., & Rodman, A. (2007). Socio-economic status as a mediating factor in the relationship between bilingualism and the cognitive development of pre-school children. *Annual Meeting of the National Communication Association 93rd Annual Convention*.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford University Press.

Wright, J., Small, N., Raynor, P., Tuffnell, D., Bhopal, R., Cameron, N., Fairley, L., A Lawlor, D., Parslow, R., Petherick, E. S., Pickett, K. E., Waiblinger, D., & West, J. (2013). Cohort profile: The born in bradford multi-ethnic family cohort study. *International Journal of Epidemiology*, *42*(4), 978–991. https://doi.org/10.1093/ije/dys112

Yan, S., & Nicoladis, E. (2009). Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism*, *12*(3), 323–335. https://doi.org/10.1017/S1366728909990101

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews*, *33*(8), 1168–1177. https://doi.org/10.1016/j.neubiorev.2009.03.003

Yeomans, F. E., Clarkin, J. F., Altschul, E., & Hull, J. W. (1992). The role of figurative language in the induction phase of an expressive psychotherapy.

The Journal of Psychotherapy Practice and Research, *1*(3), 270–279.

Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In *Bilingualism axross the lifespan* (pp. 55–72).

Young, E. C., Diehl, J. J., Morris, D., Hyman, S. L., & Bennetto, L. (2005). The use of two language tests to identify pragmatic language problems in children with Autism spectrum disorders. *Language, Speech, and Hearing Services in Schools*, *36*, 62–72. https://doi.org/10.1044/0161-1461(2005/006)

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*(1), 1–29. https://doi.org/10.1006/jmla.2001.2834

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory and Cognition*, *32*(1), 31–38. https://doi.org/10.3758/BF03195818

Zillmer, E., & Spiers, M. (2001). Neuropsychological assessment. In *Principles of Neuropsychology* (pp. 440–466).

Zipke, M. (2007). *Metalinguistic instruction improves third graders' reading comprehension*. City University of New York.

Zipke, M. (2011). First graders receive instruction in homonym detection and meaning articulation: The effect of explicit metalinguistic awareness practice on beginning readers. *Reading Psychology*, *32*(4), 349–371. https://doi.org/10.1080/02702711.2010.495312

Zipke, M., Ehri, L. C., & Cairns, H. S. (2009). Using semantic ambiguity instruction to improve third graders' metalinguistic awareness and reading comprehension: An experimental study. *Reading Research Quarterly*, *44*(3), 300–321. https://doi.org/10.1598/rrq.44.3.4

Zyzik, E. (2011). Second language idiom learning: The effects of lexical knowledge and pedagogical sequencing. *Language Teaching Research*, *15*(4), 413–433. https://doi.org/10.1177/1362168811412025

Appendix A: Stimuli List for Decomposition Survey

Idiom	Context
to back the wrong horse	In all his years as a publisher, he never 'backed the wrong horse.'
to bark up the wrong tree	If you think I want to go to church with you, you're 'barking up the wrong tree.'
to be a piece of cake	Suzie got an A+ on her maths test, she said it was 'a piece of cake.'
to be as fit as a fiddle	Gran reassured us she was 'fit as a fiddle.'
to be at death's door	Despite being discharged form hospital earlier in the week, Paul still looked as though he was 'on death's door.'
to be full of beans	Even after running around the park all day, the children were still 'full of beans.'
to be given the cold shoulder	Since their argument, she had been giving him 'the cold shoulder.'
to be head over heels	After spending the evening laughing at all his terrible jokes, she realised she was 'head over heels.'
to be in hot water	He came home late again to a note saying he was 'in hot water.'
to be in the same boat	Me and you, we're 'in the same boat.'
to be kept on one's toes	The new teacher certainly kept all the students 'on their toes.'
to be let off the hook	Luckily the party was cancelled so he was 'off the hook.'
to be on the ball	When it came to her job, Jane was always 'on the ball.'
to be on the fence	When asked of his political opinions, he always replied he was 'on the fence.'
to be on the same page	Tim and Pete discussed the party arrangements to ensure they were 'on the same page.'
to be right as rain	After finishing his antibiotics, Grandpa said he felt 'right as rain.'
to be saved by the bell	Thankfully, my colleague then came into the room and I was 'saved by the bell.'
to be the apple of one's eye	After having three boisterous sons, his delicate daughter really was the 'apple of his eye.'
to be the elephant in the room	They decided not to address the 'elephant in the room.'
to be the icing on the cake	After my terrible day, losing my keys was just the 'icing on the cake.'
to be the last straw	When she discovered someone at work had used all her milk again, she decided it was the 'last straw.'
to be the tip of the iceberg	He said she was boring and that was just the 'tip of the iceberg.'
to be under one's thumb	His friends joked he was 'under the thumb.'

to be under the weather	After catching a cold the week before, he still looked a little 'under the weather.'
to beat around the bush	She voices her opinions and doesn't 'beat around the bush.'
to bite the bullet	If he wanted the raise, he decided he needed to 'bite the bullet.'
to blow one's mind	The new Taylor Swift song really 'blew my mind.'
to break a leg	Before Mary walked on to the stage, her sister told her to 'break a leg.'
to burn the candle at both ends	Trying to find the work-life balance often resulted in Steve 'burning the candle at both ends.'
to butter one up	If I need to ask someone a favour, I'll first try to 'butter them up.'
to call the shots	While his parents were away, his grandparents were 'calling the shots.'
to chew the fat	They made a pot of tea so they could 'chew the fat.'
to clam up	As soon as he got on stage he 'clammed up.'
to drive one up the wall	She loved her children but they could sometimes 'drive her up the wall.'
to get away by the skin of one's teeth	He escaped the bear 'by the skin of his teeth.'
to get off on the wrong foot	After initially calling him the wrong name, Penny worried that her and her date had 'gotten off on the wrong foot.'
to get out of hand	Her behaviour is really getting 'out of hand.'
to give the kiss of death	Rain at a barbecue is the 'kiss of death.'
to go against the grain	Her parents were both doctors but, as she hated the sight of blood, she decided to 'go against the grain.'
to go around in circles	After arguing about where to eat for an hour, Abbie and Tom felt they were 'going around in circles.'
to go down the drain	If the business fails, that's all of the investment money 'down the drain.'
to go on a wild goose chase	I feel like my manager purposely sent me 'on a wild goose chase.'
to go out on a limb	If I didn't have the proof to justify this, I wouldn't 'go out on a limb.'
to go overboard	She had wanted her birthday party to be special, but her mother worried she had 'gone overboard.'
to go the whole nine yards	If we're going to throw a party, we've got to go 'the whole nine yards.'
to happen once in a blue moon	She only drank wine 'once in a blue moon.'
to have a chip on one's shoulder	He used to be the star player and now walks around with 'a chip on his shoulder.'
to have ants in one's pants	He couldn't keep still and was moving like he 'had ants in his pants.'
to have bigger fish to fry	I don't need this, I've 'got bigger fish to fry.'
to have in the bag	The interview went well, I think I've got the job 'in the bag.'

to have itchy feet	Although he only recently returned from Spain, Bill already had 'itchy feet.'
to have on the cards	They had been happily married for years, but children were never 'on the cards.'
to have pins and needles	She had been sitting down so long, she had 'pins and needles.'
to have the ball in one's court	I couldn't deal with his indecisiveness any longer and told him 'the ball was in his court.'
to hit the hay	After a long day, he couldn't wait to 'hit the hay.'
to hold all the cards	When it comes to pay cuts, management 'holds all the cards.'
to hold one's horses	I was anxious to leave the house but my mum kept telling me to 'hold my horses.'
to kick the bucket	While we were at school, the fish 'kicked the bucket.'
to know the ropes	The manager promised the apprentice she would 'show him the ropes.'
to let the cat out of the bag	Anne wanted to tell her friend about the surprise party, but had promised everyone she wouldn't 'let the cat out of the bag.'
to lose one's head	When yet another customer complained, the manager tried not to 'lose his head.'
to mince one's words	Barry said I looked terrible today, he certainly doesn't 'mince his words.'
to pop the question	He got down on one knee and 'popped the question.'
to preach to the choir	Teaching IT skills to millennials is 'preaching to the choir.'
to pull one's leg	Dad said the dog ran away but he was only 'pulling my leg.'
to pull the plug	The project was going nowhere and so they decided to 'pull the plug.'
to push the envelope	Her death-defying stunts always 'pushed the envelope.'
to put a sock in it	He couldn't listen to his sister complain any longer and told her to 'put a sock in it.'
to put all one's eggs in a single basket	The problem with dating is that people often don't want to 'put all their eggs in one basket.'
to rain cats and dogs	There was a storm outside and it was 'raining cats and dogs.'
to run out of steam	Working two jobs meant he was sure to soon 'run out of steam.'
to save one's skin	He lied and blamed the accident on his friend 'to save his skin.'
to shake a leg	If he was going to make it into work on time, he would have to 'shake a leg.'
to sit on a nest egg	Thanks to opening a savings account when she was young, Molly was now 'sitting on a nest egg.'
to spill the beans	What happened at the party last night? 'Spill the beans!'
to steal the show	With her extravagant dress, she really 'stole the show.'
to take the biscuit	She has always been spoilt but her latest outburst really 'takes the biscuit.'

to take the bull by the horns	If she wanted to be successful, she decided she would have to 'take the bull by the horns.'
to throw in the towel	He was getting too old and decided it was time to finally 'throw in the towel.'
to tie the knot	After finally buying the dress and ordering the flowers, she couldn't wait to 'tie the knot.'
to wear one's heart on one's sleeve	She fell in love too easily and was known to 'wear her heart on her sleeve.'

Appendix B: Stimuli List for AoA surveys

Ambiguity type	Target	Sentence-1	Sentence-2
Balanced Homonym	band	My friend bought us tickets to see my favourite band.	She put her hair into a ponytail using a band.
Balanced Homonym	bark	The boy and the girl carved their names into the bark.	The dog kept everyone awake with its loud bark.
Balanced Homonym	bat	It is hard to play cricket without a bat.	If you go into a cave at night you might see a bat.
Balanced Homonym	beam	Many gymnasts perform on a balancing beam.	The sunlight was shining through the clouds in a beam.
Balanced Homonym	blow	During a tackle in rugby, Barry took a blow.	The wind knocked the tiles off the roof with a strong blow.
Balanced Homonym	bowl	She ate her cereal out of a bowl.	Once the skittles were positioned she was able to bowl.
Balanced Homonym	box	The postman delivered a box.	His brother was eager to learn how to box.
Balanced Homonym	brush	Her hair was a mess as she couldn't find her brush.	The explorer got lost in the brush.
Balanced Homonym	buffer	The car’s bumper worked as a buffer.	Her ring was tarnished, so Penny bought a buffer.
Balanced Homonym	calf	The cow just gave birth to her new born calf.	The boy could no longer play football as he had injured his calf.
Balanced Homonym	cane	At the factory they were making sugar from cane.	To help him walk, the old man uses a cane.
Balanced Homonym	cell	The man committed a crime so he was put in a prison cell.	In biology we learnt about the cell.
Balanced Homonym	charm	The bracelet contained a beautiful diamond charm.	The lady fell for the man's charm.
Balanced Homonym	chink	The two floorboards were separated by a chink.	When the metal touched, you could hear a loud chink.

Balanced Homonym	cricket	If you listen hard on a quiet night, you may hear a cricket.	My brother's favourite sport is cricket.
Balanced Homonym	date	The boy and the girl had a lovely time together on their date.	Her birthday and exam were both on the same date.
Balanced Homonym	fan	It was so hot, he turned on the fan.	She was the musician's number one fan.
Balanced Homonym	fly	I always dream that I can fly.	In the darkness, all I could hear was the annoying fly.
Balanced Homonym	forge	The metal was heated and shaped in the forge.	Although he was exhausted, he continued with forge.
Balanced Homonym	glare	The sun hit the car windscreen causing a glare.	The girl looked angry as she gave her friend a sharp glare.
Balanced Homonym	hail	The subjects welcomed the queen with a hearty hail.	The castle sat at the top of a large mount.
Balanced Homonym	jam	Fitting four of us into one compartment was a jam.	Sophie's favourite sandwich is filled with jam.
Balanced Homonym	lean	To ensure he didn't hit his head, Jerry had to lean.	The steak had no fat, it was lean.
Balanced Homonym	limp	The man was old and walked with a limp.	The lettuce was out of date and slightly limp.
Balanced Homonym	log	The walk was long and she had to rest on a log.	To ensure they didn't go over budget the company kept an expense log.
Balanced Homonym	mass	There was a good offer so she ordered the stock in mass.	The family attended the church for Sunday mass.
Balanced Homonym	match	The man and his wife are a perfect match.	The camper lit the fire using a match.
Balanced Homonym	mortar	He ground the ingredients using a pestle and mortar.	The house was made using bricks and mortar.
Balanced Homonym	mould	The walls of the house were damp and covered in mould.	The artist used clay and a bowl to create a mould.
Balanced Homonym	nail	To hang the picture, the builder used a hammer and a nail.	After just having a manicure, the girl was sad when she chipped a nail.

Balanced Homonym	net	To avoid being bitten, she bought a mosquito net.	He presented his gross profit, not his net.
Balanced Homonym	note	On Valentine's day John sent Mary a special note.	She sang her song beautifully and reached every high note.
Balanced Homonym	organ	The brain is the most important organ.	Hymns in church are accompanied by an organ.
Balanced Homonym	panel	Their gardens were separated by a long wooden panel.	The school talent show was judged by the panel.
Balanced Homonym	parrot	When I was younger, I had a pet parrot.	My sister always copies what I say, she is a parrot.
Balanced Homonym	peep	He couldn't wait to see her wedding dress and had a quick peep.	He would be angry if the kids made even a peep.
Balanced Homonym	peer	Her friendship group was made up of her peers.	He looked over the edge of the table to peer.
Balanced Homonym	picket	They went to the town hall to picket.	They separated their gardens with a picket.
Balanced Homonym	pitcher	The iced water was put in a pitcher.	His favourite position to play in cricket is the pitcher.
Balanced Homonym	plane	After levelling it out, the garden was on a plane.	The carpenter smoothed the wooden surface with the plane.
Balanced Homonym	plot	The horror story had a very scary plot.	The house was built on a large plot.
Balanced Homonym	prop	The book was used to keep open the door as a prop.	In the show, the old phone was used as a stage prop.
Balanced Homonym	prune	If you preserve a plum, it becomes a prune.	The roses were growing so fast they needed another prune.
Balanced Homonym	pupil	The teacher gave a sticker to the pupil.	The doctor examined the patient's pupil.
Balanced Homonym	seal	The postman stamped the letter with a seal.	The boy was playing on the beach when he saw a seal.
Balanced Homonym	spade	The children built sandcastles using a spade.	During a game of cards, the man picked up a spade.

Balanced Homonym	stable	The house had a good foundation, it was very stable.	The donkey lived in the stable.
Balanced Homonym	stake	The vampire was killed with a stake.	The company's new venture was very high stake.
Balanced Homonym	stalk	Plants draw water up through their stalk.	To provide food, lions must learn to stalk.
Balanced Homonym	stall	They took the donkey into it's stall.	The amount of information on the PC caused it to stall.
Balanced Homonym	steer	At the market, the farmer sold his prized steer.	The roads were very icy which made it difficult to steer.
Balanced Homonym	tap	To get the ketchup out of the bottle you must give it a small tap.	He filled the bathtub by turning on the tap.
Balanced Homonym	tick	In the still of the night he could hear the clock tick.	Caressing his dog, he removed the last tick.
Balanced Homonym	tie	He looked very smart for his interview in a suit and tie.	The present was beautifully wrapped and held together with a pink tie.
Balanced Homonym	utter	My students write every word I utter.	After the earthquake, the shock and devastation was utter.
Balanced Homonym	vault	Only one person had the key to the vault.	During gymnastics, he was always afraid to vault.
Balanced Homonym	yard	She couldn't throw the ball more than a yard.	The new house had a big back yard.
Unbalanced Homonym	angle	You have to get the compass at the right angle.	To fish with a hook, is to angle.
Unbalanced Homonym	arch	To get into the church, you go through the arch.	She described his look as arch.
Unbalanced Homonym	ash	The paper burnt and all that was left was ash.	Some of the trees were willow, others were ash.
Unbalanced Homonym	ball	As it came towards him, he kicked the ball.	She couldn't wait to dance at the ball.
Unbalanced Homonym	bay	The holidaymakers drove the boat slowly into the bay.	The bed was positioned by the window, in the bay.

Unbalanced Homonym	bear	The children were scared of the big grizzly bear.	The sofa was too heavy a load to bear.
Unbalanced Homonym	bluff	When playing poker, it's important to know how to bluff.	He stood on the edge of a bluff.
Unbalanced Homonym	bolt	The plumber said the problem was a lose bolt.	He was worried the horse would bolt.
Unbalanced Homonym	bush	They spotted two robins in the holly bush.	To reduce friction in machinery, you need to use a bush.
Unbalanced Homonym	bust	To ensure the dress fitted, the tailor measured her bust.	The riverbank was about to bust.
Unbalanced Homonym	camp	At the weekend, the family went to the lake to camp.	The clothes that the man wore were rather camp.
Unbalanced Homonym	cape	The superhero wore a cape.	The pirates always hid their treasure near the cape.
Unbalanced Homonym	capital	When in France, I always visit the capital.	We will expand the business as soon as we get the capital.
Unbalanced Homonym	card	To pay for the food the lady used a debit card.	To remove the shortest fibres prior to spinning she used a card.
Unbalanced Homonym	chop	I couldn't rest, I had too much wood to chop.	I enjoyed the pork chop.
Unbalanced Homonym	chord	The length of the arc was calculated through the chord.	The girl played a G minor chord.
Unbalanced Homonym	clock	In the classroom, the children couldn’t stop watching the clock.	The woman wore her stockings with embroidered clock.
Unbalanced Homonym	coach	They travelled to the concert by coach.	The rugby team gathered to listen to their coach.
Unbalanced Homonym	corn	The farmers worked hard harvesting the corn.	Being on his feet all day caused the man to develop a corn.
Unbalanced Homonym	count	The maths teacher helped the pupils learn to count.	He enjoyed the status he received from being a count.
Unbalanced Homonym	crash	Using your phone whilst driving will surely cause a crash.	His trousers were itchy and made of crash.

Unbalanced Homonym	dam	The torrential rain lasted so long that it burst the dam.	The calves depended on the milk from their dam.
Unbalanced Homonym	diet	Rodents are the primary constituent of an owl's diet.	I will have a Pepsi, but make sure it’s diet.
Unbalanced Homonym	drill	In her toolbox, she had a drill.	The school evacuated because of the fire drill.
Unbalanced Homonym	ear	Unfortunately, Robert has lost hearing in his right ear.	The seeds are found in the part of the plant called the ear.
Unbalanced Homonym	file	Once alphabetically organised, he put the documents in a file.	The carpenter smoothed the surface with a file.
Unbalanced Homonym	fleet	The battleships congregated in a fleet.	Although the woman was old, she was still very fleet.
Unbalanced Homonym	flight	The bird opened its wings ready for flight.	The burglars had been caught and needed to ascend into flight.
Unbalanced Homonym	flock	The hungry birds gathered in a flock.	The pillows were thin and lacking flock.
Unbalanced Homonym	foil	She wrapped her sandwich in tin foil.	In the sport of fencing, they use a foil.
Unbalanced Homonym	fry	She put sausages and eggs in the pan to fry.	When fishing, he caught a small fry.
Unbalanced Homonym	game	The children played a board game.	He protested against the hunting of game.
Unbalanced Homonym	gin	On her birthday Jen celebrated with her favourite gin.	The seeds were separated from the cotton using a gin.
Unbalanced Homonym	habit	She knew she needed to give up her bad habit.	The nun wore a long, black habit.
Unbalanced Homonym	hide	The boy didn't want to be seen so he decided to hide.	He skinned the animal then preserved its hide.
Unbalanced Homonym	hiding	Paula couldn't understand what her daughter was hiding.	The boy was naughty so his dad gave him a good hiding.
Unbalanced Homonym	hop	She could not put weight on her foot so she had to hop.	British beer is sometimes flavoured with Czech hop.

Unbalanced Homonym	horn	Impatiently, he honked the horn.	The bullfighter was injured by the animal’s horn.
Unbalanced Homonym	host	I stayed at my friend's house, she was a good host.	It was difficult to choose my favourite amongst such a large host.
Unbalanced Homonym	keen	The blade was very keen.	When her dog died, Olivia let out a terrifying keen.
Unbalanced Homonym	lap	The child sat on her lap.	The runners were on the final lap.
Unbalanced Homonym	launch	Everyone was scared when hearing about the missile launch.	The lifeboat service personnel left on the launch.
Unbalanced Homonym	lawn	It's been three months since my husband cut the lawn.	The shirt was white and made of lawn.
Unbalanced Homonym	lock	He secured his suitcase with a lock.	The mother had kept her baby daughter's golden lock.
Unbalanced Homonym	lumber	She sat at a makeshift table of unfinished lumber.	He walks with a heavy lumber.
Unbalanced Homonym	mail	On your birthday, it is nice to receive cards in the mail.	The knight was well protected wearing his mail.
Unbalanced Homonym	mat	Before entering a house, you should wipe your feet on the mat.	The artist gave his painting a border using a mat.
Unbalanced Homonym	mate	She had finally found her soul mate.	The game was over once the boy called check mate.
Unbalanced Homonym	meal	To celebrate the special occasion, the family went out for a meal.	She made the cake with meal.
Unbalanced Homonym	mint	She grew herbs in the garden, including mint.	The two coins were made at the same mint.
Unbalanced Homonym	mole	The holes in the ground had been made by a mole.	On her left arm, there was a small, raised mole
Unbalanced Homonym	mount	He used a walking pole to assist him in his mount.	The storm was so bad it began to hail.
Unbalanced Homonym	nap	The child had an afternoon nap.	She carefully sewed the fabric together following the direction of the nap.

Unbalanced Homonym	nip	The dog didn’t mean to hurt me, it was just a nip.	He had an expensive whisky and often took a small nip.
Unbalanced Homonym	novel	The young girl was developing characters for her novel.	The students found the whole lesson to be novel.
Unbalanced Homonym	pack	She couldn't fit all her clothes in to the one pack.	The policeman was crooked and easy to pack.
Unbalanced Homonym	pad	After injuring her knee, she protected it with a pad.	As she walked down the corridor you could hear her feet pad.
Unbalanced Homonym	palm	Before her speech, she had sweaty palms.	Coconuts grow on palms.
Unbalanced Homonym	park	The children played in the park.	For his driving test, he had to learn to bay park.
Unbalanced Homonym	peck	The bird gave the ground a peck.	She barely touched pudding, she just had a small peck.
Unbalanced Homonym	peel	On the kitchen table, someone had left an orange peel.	To get the pizza out of the oven, he used a wooden peel.
Unbalanced Homonym	pen	She wrote her essay using a pen.	The animals were kept in a small pen.
Unbalanced Homonym	perch	The bird used a branch as its perch.	While fishing, he caught a perch.
Unbalanced Homonym	pet	The dog, Lucky, was the family pet.	He had offended her and now she was in a pet.
Unbalanced Homonym	pit	They worked hard to dig a deep pit.	She made sure not to swallow the apricot pit.
Unbalanced Homonym	plump	The berries were sweet and plump.	She sat down with a plump.
Unbalanced Homonym	pool	It is possible to bathe in the natural sulphur pools.	Together, they formed a talent pool.
Unbalanced Homonym	port	The ships sailed towards the port.	At Christmas, they celebrated with a glass of port.
Unbalanced Homonym	pose	Models are famous for striking a pose.	It was a question she just would not pose.

Unbalanced Homonym	post	She put the birthday card in the post.	The notice was attached to the lamp post.
Unbalanced Homonym	pulse	After running the race, she had fast pulse.	Most people don't realise that peas are a pulse.
Unbalanced Homonym	pump	At the garden centre, Liz bought an antique water pump.	During the rehearsal, the girl lost her ballet pump.
Unbalanced Homonym	punch	The black eye was the result of a punch.	For the party, they made rum punch.
Unbalanced Homonym	race	He was tired after finishing the race.	She fought against discrimination based on race.
Unbalanced Homonym	racket	The tennis star played with her favourite racket.	They were annoyed as the neighbours were making a racket.
Unbalanced Homonym	rank	She was an army officer of high rank.	The milk was out of date and smelt rank.
Unbalanced Homonym	ray	She relaxed in the sun’s rays.	While snorkelling, he saw a school of rays.
Unbalanced Homonym	rear	When viewing the house, she noticed there was parking at the rear.	On the farm, the sheep took three years to rear.
Unbalanced Homonym	reef	She went snorkelling in the coral reef.	The wind was strong so she tied down the sails reef.
Unbalanced Homonym	relief	When he left, we all breathed a sigh of relief.	The column at the temple was covered with sculptured relief.
Unbalanced Homonym	rubber	He made a mistake in his homework but couldn't find a rubber.	The tyres were made from recycled rubber.
Unbalanced Homonym	sack	Santa came down the chimney with a present-filled sack.	The pirates hoped there was another city to sack.
Unbalanced Homonym	sage	He flavoured the meal with sage.	Having much experience on the matter, her advice was sage.
Unbalanced Homonym	scale	She weighed the flour using her new set of scales.	As he picked up the fish, he could feel it's scales.
Unbalanced Homonym	scrap	After eating the best of the chicken, they allowed the dog a scrap.	On the playground, the boys got in a scrap.

Unbalanced Homonym	settle	Before they went to court, she hoped the company would settle.	Next to the fire, she sat down on a settle.
Unbalanced Homonym	shed	He kept his toolbox in the shed.	The trees were bare, their leaves had been shed.
Unbalanced Homonym	sheer	Her new tights were sheer.	The wet road caused the car to sheer.
Unbalanced Homonym	sheet	After changing the duvet, she also changed the sheet.	Having written on both sides of the paper he needed a new sheet.
Unbalanced Homonym	shower	After the football, he needed a shower.	It would be sunny all day with a chance of a brief shower.
Unbalanced Homonym	soil	Plants need water, sun and plenty of soil.	He played in the mud causing his sleeves to soil.
Unbalanced Homonym	spell	Your name is often the first thing you learn to spell.	The witch cast an evil spell.
Unbalanced Homonym	spray	The plants needed watering so I gave them a spray.	The blossom tree grew a beautiful spray.
Unbalanced Homonym	squash	Fitting all the children in the car was a bit of a squash.	The soup was made from pumpkin and squash.
Unbalanced Homonym	staff	The shop was closed yesterday because they didn't have enough staff.	The wizard always carried his staff.
Unbalanced Homonym	stern	The new teacher was too stern.	The boat had a broken stern.
Unbalanced Homonym	stir	She gave her cup of tea a stir.	After getting arrested, he landed himself back in stir.
Unbalanced Homonym	strain	I don't think you've torn anything, maybe it's just a strain.	She was annoyed she didn't have a cheesecloth to strain.
Unbalanced Homonym	strand	The storm caused the boat to strand.	The cotton shirt had a lose strand.
Unbalanced Homonym	strike	The heavyweight boxer won the match after only one strike.	The workers were not happy with the conditions so they went on strike.
Unbalanced Homonym	strip	In the public changing room, she was afraid to strip.	He needed some paper so he tore off a small strip.

Unbalanced Homonym	swallow	In the garden, she could see the swallow.	Her sore throat made it hard to swallow.
Unbalanced Homonym	temple	Every Saturday, they went to worship at the temple.	I feel my migraine mostly in my left temple.
Unbalanced Homonym	tense	After the run, her hips were tense.	When speaking French she had trouble using the past tense.
Unbalanced Homonym	toast	When they were young they used to love cheese on toast.	When they got married, the groom made a toast.
Unbalanced Homonym	toll	For the evening service, the cathedral bells began to toll.	To drive across the bridge, she had to pay a toll.
Unbalanced Homonym	trace	The boat disappeared without a trace.	The horse was attached to the cart by a trace.
Metaphor	ape	The girl went to the zoo and saw an ape.	All Brian needs is one pint and he turns into an ape.
Metaphor	arm	The boy fell and broke his arm.	There was no room on the sofa, so he sat on the arm.
Metaphor	brain	In science, they learnt about the brain.	I wish I was as clever as Zoe, she is such a brain.
Metaphor	bun	For lunch, I bought chips in a bun.	For my ballet exam, I put my hair in a bun.
Metaphor	chair	He wanted to sit, but struggled to find a chair.	The dance society had chosen to appoint Katherine as chair.
Metaphor	chicken	I have always wanted a pet chicken.	Mary didn’t want to do the zip wire, she was a chicken.
Metaphor	cow	When I walked through the field, I saw a cow.	Since she started her diet, Mary had been such a moody cow.
Metaphor	dense	I could not see as the fog was very dense.	Paul never understands the homework, he is a little dense.
Metaphor	doll	The girls played with the baby doll.	Poppy is so lovely, she’s a doll.
Metaphor	dough	To make pizza, you first must knead the dough.	She bought a huge house, it cost a lot of dough.

Metaphor	dribble	The dog was cute, but he did dribble.	The basket ball player had to work on his dribble.
Metaphor	fold	He did the ironing and had only one sweater to fold.	Rumour has it that Cadbury's is about to fold.
Metaphor	fox	The animal in the garden was a fox.	She stole the sweets without anyone seeing, what a fox!
Metaphor	gem	The gorgeous ring had a huge, green, gem.	Nobody else had the same scarf as Sally, she found a gem.
Metaphor	grit	During the winter we spread icy roads with salt and grit.	I wanted to be a teacher but didn't have the grit.
Metaphor	hurdle	In P.E. the children had to jump over the hurdle.	Finding a new job is my next hurdle!
Metaphor	index	You can search for a specific word in the index.	The woman had a ring fitted for her finger, it was on her index.
Metaphor	key	James was locked outside, he had forgotten his key.	If she wanted the job, making a good impression was key.
Metaphor	lion	The roar came from the very hungry lion.	The soldier had the courage of a lion.
Metaphor	lip	The woman put some lipstick on her lips.	The liquid bubbled over the saucepan lip.
Metaphor	mouth	At the dentist, Luke was told to open his mouth.	A lot of litter had collected at the river mouth.
Metaphor	neck	The giraffe had a very long neck.	He held the bottle at it's neck.
Metaphor	nose	When I have a cold, I get a runny nose	To increase its speed, the aircraft had a pointed nose.
Metaphor	nucleus	A cells genetic material is held in its nucleus.	In the organisation, she is the nucleus.
Metaphor	nuts	I filled the bird feeder in the garden with some nuts.	The man wore a t-shirt in the snow, he was nuts!
Metaphor	pig	My favourite animal is a pig.	After eating the whole cake, Bruce felt like such a pig.

Metaphor	pillar	In the middle of the museum, there was a large pillar.	In the little town, the mayor was the pillar.
Metaphor	sheep	Sally went for a walk in the field and saw a sheep.	Tim was not a leader, he was more of a sheep.
Metaphor	shoulder	I wanted his attention, so I tapped him on the shoulder.	The tailor adjusted the blazer's shoulder.
Metaphor	snake	The insects were eaten by the snake.	The boy who lies to his friends is known as a snake.
Metaphor	spice	I like my curry to have a lot of spice.	I did the bungee jump to give my life some spice!
Metaphor	spill	She tripped, causing her drink to spill.	He lit the fire with a thin paper spill.
Metaphor	star	On a clear night you can see the northern star.	Jessica did so well in her solo, she was a star.
Metaphor	stick	The dog was tired from fetching the stick.	Lucy barely ate a thing – she was a stick!
Metaphor	tongue	I licked the ice cream with my tongue.	You tie your laces by the shoes tongue.
Metaphor	tooth	I went to the dentist with a wobbly tooth.	While combing my knotty hair, I snapped a tooth.
Metaphor	wing	The bird had broken its wing.	The meeting is in the east wing.
Metaphor	worm	In the garden I could see a bird eating a worm.	He got out of doing his chores, he's such a worm.
Metonym	alley	The construction workers were paving the alley.	We walked between the houses, down the narrow alley.
Metonym	arena	He went to watch a football match at the arena.	If there is a game on, it's too busy to park in the arena.
Metonym	attic	Under the roof, they discovered the attic.	He had a lot of boxes of toys which he kept upstairs in the attic
Metonym	bag	She waited at the hotel reception for the porter to take her bag.	We were cleaning the room and put all the rubbish in the bag.

Metonym	barrel	He injured his back trying to pick up the empty barrel.	Hannah loved pickles so much, she bought them by the barrel.
Metonym	basket	I had fun weaving my own wicker basket.	I don't sort my laundry, I just tip in the whole basket.
Metonym	bottle	They wanted to recycle the glass bottle.	The recipe called for a whole bottle.
Metonym	bucket	He couldn't make a sandcastle without his bucket.	The patio was dusty so he threw the whole soapy bucket.
Metonym	cage	John was busy painting the cage.	He took the hamster out of the cage.
Metonym	cellar	Beneath the house there was a dark cellar.	The thieves stole all the expensive wine from the cellar.
Metonym	chimney	Their new house had a tall brick chimney.	They told Jimmy that Father Christmas comes down the chimney.
Metonym	cup	He poured the tea into a cup.	When cutting down on coffee I limited myself to just one cup.
Metonym	fig	It was the first time Mary ate a fig.	I like fruit bars, my favourite is fig.
Metonym	glass	John fell and broke the glass.	I tripped with my water and spilt the whole glass.
Metonym	horse	He loved to ride his horse.	The restaurant sold stewed horse.
Metonym	lamb	The children went to the farm to pet a lamb.	On Sunday they would eat roast lamb.
Metonym	lemon	She went to the fruit bowl and picked up a lemon.	The cake tasted of lemon.
Metonym	maple	The tree in the garden was a maple.	The biscuit was pecan and maple.
Metonym	onion	When tending to the vegetable patch, Pete harvested the onion.	The sandwich she ate had cheese and onion.
Metonym	orange	At the market, she bought an orange.	His vitamins tasted of orange.

Metonym	pine	The forest was full of evergreen pine.	Her perfume smelled of pine.
Metonym	pipe	The house flooded because of a broken pipe.	The mouse ran up the pipe.
Metonym	plate	He went to the sink to wash up his plate.	He couldn't leave the table unless he had finished his plate.
Metonym	porch	In the front, the house had a beautiful porch.	In the evenings, they would sit out on the porch.
Metonym	potato	I went to the market and picked up a potato.	My favourite food is beans and mashed potato.
Metonym	rabbit	She gave a carrot to her rabbit.	The pie contained leek and rabbit.
Metonym	theatre	The little town had an old theatre.	My parents always enjoyed the theatre.
Metonym	tub	They sold the soup in a small tub.	When eating ice-cream, she would finish the whole tub.
Metonym	tube	I always recycle the toilet paper tube.	He used up all the toothpaste and left the empty tube.
Metonym	window	She couldn't play football in the house in case she broke a window.	She snuck out through her bedroom window.

Idioms

Item subtype class	Item	Sentence
Decompositional	to back the wrong horse	In all his years as a publisher, he had never backed the wrong horse
Decompositional	to be at death's door	Despite being discharged from hospital earlier in the week, Paul still looked as though he was on death's door.
Decompositional	to be in the dark	I don't like surprises as I can’t stand being in the dark
Decompositional	to be in the same boat	She kept complaining she was cold, but we were both in the same boat
Decompositional	to be kept on one's toes	The new teacher certainly kept all the students on their toes.
Decompositional	to be let off the hook	Luckily the party was cancelled so he was off the hook.

Decompositional	to be on the same page	Tim and Pete discussed the party arrangements to ensure they were on the same page.
Decompositional	to be on the wrong track	My boss said I was on the wrong track
Decompositional	to be saved by the bell	Thankfully, my colleague then came into the room and I was saved by the bell
Decompositional	to be the icing on the cake	After a terrible day, losing my keys was just the icing on the cake
Decompositional	to be the tip of the iceberg	He said she was boring and that was just the tip of the iceberg.
Decompositional	to call the shots	While his parents were away, his grandparents were calling the shots
Decompositional	to get off the ground	He knew that without a substantial investment, the business wouldn't get off the ground
Decompositional	to get out of hand	Her behaviour is really getting out of hand.
Decompositional	to go against the grain	All her family were doctors but, as she hated the sight of blood, she decided to go against the grain
Decompositional	to go around in circles	After arguing about where to eat for an hour, Abbie and Tom felt they were going around in circles.
Decompositional	to go down the drain	If the business fails, that's all of the investment money down the drain
Decompositional	to go downhill	After the first pub, the night went downhill.
Decompositional	to go on a wild goose chase	I feel like my manager purposely sent me on a wild goose chase.
Decompositional	to hang by a thread	After their argument, Richard and Kath's marriage was hanging by a thread.
Decompositional	to happen once in a blue moon	She only drank wine once in a blue moon.
Decompositional	to have a feel for it	I know I can master French as soon as I get a feel for it.
Decompositional	to have a mental block	Halfway through his presentation, he had a mental block.
Decompositional	to have pins and needles	She had been sitting down so long, she had pins and needles.
Decompositional	to hold all the cards	When it comes to pay cuts, management holds all the cards.
Decompositional	to keep an open mind	When trying food from another country, I always try to keep an open mind.
Decompositional	to knock one out cold	I watched the fight last night, the winner knocked his opponent out cold
Decompositional	to lose one's head	When yet another customer complained, the manager tried not to lose his head.
Decompositional	to make the grade	I wanted to be a dentist but couldn't make the grade.
Decompositional	to pass the buck	I hate doing my taxes, it's so tempting to pass the buck.
Decompositional	to pop the question	He got down on one knee and popped the question.
Decompositional	to preach to the choir	Teaching IT skills to millennials is preaching to the choir.
Decompositional	to pull the plug	The project was going nowhere and so they decided to pull the plug.

Decompositional	to put a sock in it	He couldn't listen to his sister complain any longer and told her to put a sock in it.
Decompositional	to run out of steam	Working two jobs meant he was sure to soon run out of steam.
Decompositional	to save one's skin	He lied and blamed the accident on his friend to save his skin
Decompositional	to steal the show	With her extravagant dress, she really stole the show.
Decompositional	to turn sour	After their row, their relationship turned sour.
Non-decompositional	to bark up the wrong tree	If you think I want to go to church with you, you're barking up the wrong tree.
Non-decompositional	to be a piece of cake	Suzie got an A+ on her maths test, she said it was a piece of cake.
Non-decompositional	to be as fit as a fiddle	Gran reassured us she was fit as a fiddle.
Non-decompositional	to be full of beans	Even after running around the park all day, the children were still full of beans.
Non-decompositional	to be given the cold shoulder	Since their argument, she had been giving him the cold shoulder.
Non-decompositional	to be head over heels	After spending the evening laughing at all his terrible jokes, she realised she was head over heels.
Non-decompositional	to be in a pickle	Do I wear the green or blue shoes? I'm in a pickle
Non-decompositional	to be in hot water	He came home late again to a note saying he was in hot water
Non-decompositional	to be in stitches	The comedy show was hilarious; the first guy had me in stitches
Non-decompositional	to be on the ball	When it came to her job, Jane was always on the ball.
Non-decompositional	to be on the fence	When asked of his political opinions, he always replied he was on the fence.
Non-decompositional	to be on the rocks	After fighting over a boy, Patsy and Dawn's friendship was on the rocks.
Non-decompositional	to be right as rain	After finishing his antibiotics, Grandpa said he felt right as rain.
Non-decompositional	to be the apple of one's eye	After having three boisterous sons, his delicate daughter was the apple of his eye.
Non-decompositional	to be the elephant in the room	They decided not to address the elephant in the room.
Non-decompositional	to be the last straw	When she discovered someone at work had used all her milk again, she decided it was the last straw
Non-decompositional	to be under one's thumb	His friends joked he was under the thumb.
Non-decompositional	to be under the weather	After catching a cold the week before, he still looked a little under the weather
Non-decompositional	to beat around the bush	She voices her opinions and doesn't beat around the bush.
Non-decompositional	to bite the bullet	If he wanted the raise, he decided he needed to bite the bullet.
Non-decompositional	to bite the dust	Judging from the noise the engine is making, the car may soon bite the dust.
Non-decompositional	to blow one's mind	The new Taylor Swift song really blew my mind.

Non-decompositional	to break a leg	Before Mary walked on to the stage, her sister told her to break a leg.
Non-decompositional	to burn the candle at both ends	Trying to find the work-life balance often resulted in Steve burning the candle at both ends.
Non-decompositional	to butter one up	If I need to ask someone a favour, I'll first try to butter them up.
Non-decompositional	to chew the fat	They made a pot of tea so they could chew the fat.
Non-decompositional	to clam up	As soon as he got on stage he clammed up.
Non-decompositional	to cook the books	I have a feeling the accountant is cooking the books.
Non-decompositional	to cut the mustard	I wanted to be a doctor but I couldn't cut the mustard.
Non-decompositional	to drive one up the wall	She loved her children but they could sometimes drive her up the wall.
Non-decompositional	to drop a line	I haven't spoken to Jan for ages. I must drop her a line
Non-decompositional	to face the music	After calling in sick for three days, he knew he'd eventually have to face the music.
Non-decompositional	to feel out of one's element	As a footballer on a rugby pitch, Harry felt out of his element.
Non-decompositional	to feel something in one's bones	It's going to rain tomorrow, I can feel it in my bones!
Non-decompositional	to float one's boat	Caffeine before bedtime really doesn't float my boat.
Non-decompositional	to get away by the skin of one's teeth	He escaped the bear by the skin of his teeth.
Non-decompositional	to get in a stew	Telling lies can get you in a stew.
Non-decompositional	to get into a lather	It's best not to talk about politics, it gets some people in a lather.
Non-decompositional	to get off on the wrong foot	After initially calling him the wrong name, Penny worried that her and her date had gotten off on the wrong foot.
Non-decompositional	to give a hoot	I'm bored of listening to her drama, I couldn't give a hoot!
Non-decompositional	to give the boot	She was so terrible at her job, her boss wanted to give her the boot.
Non-decompositional	to give the kiss of death	Rain at a barbecue is the kiss of death.
Non-decompositional	to give the slip	The police almost arrested the burglar but he gave them the slip.
Non-decompositional	to go off on a tangent	He started talking about cheese and then just went off on a tangent
Non-decompositional	to go off the deep end	After one drink, he went off the deep end.
Non-decompositional	to go out on a limb	If I didn't have the proof to justify this, I wouldn't go out on a limb.
Non-decompositional	to go overboard	She had wanted her birthday party to be special, but her mother worried she had gone overboard.
Non-decompositional	to go the whole nine yards	If we're going to throw a party, we've got to go the whole nine yards.
Non-decompositional	to have a chip on one's shoulder	He used to be the star player and now walks around with a chip on his shoulder.

Non-decompositional	to have an axe to grind	She said she didn't care but I knew she had an axe to grind.
Non-decompositional	to have ants in one's pants	He couldn't keep still and was moving like he had ants in his pants.
Non-decompositional	to have bigger fish to fry	I don't need the hassle, I've got bigger fish to fry.
Non-decompositional	to have egg on one's face	After being the reason the team lost the championship, Tim had egg on his face.
Non-decompositional	to have in the bag	The interview went well, I think I've got the job in the bag.
Non-decompositional	to have itchy feet	Although he only recently returned from Spain, Bill already had itchy feet.
Non-decompositional	to have on the cards	They had been happily married for years, but children were never on the cards.
Non-decompositional	to have plans up in the air	After Ben cancelled our date, my evening plans were up in the air.
Non-decompositional	to have the ball in one's court	I couldn't deal with his indecisiveness any longer and told him the ball was in his court.
Non-decompositional	to hit the hay	After a long day, he couldn't wait to hit the hay.
Non-decompositional	to hold one's horses	I was anxious to leave the house but my mum kept telling me to hold my horses.
Non-decompositional	to kick the bucket	While we were at school, the fish kicked the bucket.
Non-decompositional	to know the ropes	The manager promised the apprentice she would show him the ropes.
Non-decompositional	to let the cat out of the bag	Anne wanted to tell her friend about the surprise party, but had promised not to let the cat out of the bag
Non-decompositional	to make waves	Although I have lots of ideas for my new job, I'm afraid of making waves.
Non-decompositional	to mince one's words	Barry said I looked terrible today, he certainly doesn't mince his words.
Non-decompositional	to not have one's heart in it	The interview didn't go well, my heart wasn't in it.
Non-decompositional	to paint the town red	It's my birthday tonight and I can't wait to paint the town red.
Non-decompositional	to pull one's leg	Dad said the dog ran away but he was only pulling my leg.
Non-decompositional	to push the envelope	Her death-defying stunts always pushed the envelope.
Non-decompositional	to put all one's eggs in a single basket	The problem with dating is that people often don't want to put all their eggs in one basket.
Non-decompositional	to rain cats and dogs	There was a storm outside and it was raining cats and dogs.
Non-decompositional	to rub one the wrong way	I dislike Samantha's disposition, her attitude rubs me the wrong way.
Non-decompositional	to shake a leg	If he was going to make it into work on time, he would have to shake a leg.
Non-decompositional	to shoot the breeze	The team got together to shoot the breeze
Non-decompositional	to sit on a nest egg	Thanks to opening a savings account when she was young, Molly was now sitting on a nest egg.
Non-decompositional	to spill the beans	What happened at the party last night? Spill the beans!

Non-decompositional	to take the biscuit	She has always been spoilt but her latest outburst really takes the biscuit.
Non-decompositional	to take the bull by the horns	If she wanted to be successful, she decided she would have to take the bull by the horns.
Non-decompositional	to take the strain	The bus service can't take the strain.
Non-decompositional	to take to the cleaners	After the mechanics rude remarks, Mike wished he could take them to the cleaners.
Non-decompositional	to throw in the towel	He was getting too old and decided it was time to finally throw in the towel.
Non-decompositional	to tie the knot	After finally buying the dress and ordering the flowers, she couldn't wait to tie the knot.
Non-decompositional	to wear one's heart on one's sleeve	She fell in love too easily and was known to wear her heart on her sleeve.

Appendix C: Stimuli List for Chapter 4 (Homonyms)

Experimental Items

Type	Target	Correct_answer1	Correct_answer2	Incorrect_answer1	Incorrect_answer2
Balanced	**band**	A group that plays music	Something used to tie things together	Going to the movies	Wool used for knitting
Balanced	**beam**	A piece of wood holding the ceiling up	A ray of light	A small framed photo	Walking in the rain
Balanced	**blow**	When someone gets punched	Something you do to candles	Two people holding hands	When you get presents
Balanced	**bowl**	Something you eat from	Throwing a ball across the ground	Something you drink from	A thing you say when playing golf
Balanced	**cricket**	A noisy, green insect	A game played with wickets	A tiny, black insect	A game played with a hoop
Balanced	**fan**	Something you use to cool down	Someone who loves a certain person or sport	Something you put on to keep warm	Someone who sits and reads quietly
Balanced	**jam**	Tightly fitting something into a space	Something you put on toast	A feeling you get while packing	A big plate of pancakes
Balanced	**log**	A thick piece of tree trunk	An official daily record	A pile of sticks	A ships wheel
Balanced	**match**	Things that are the same	Something you use to light a fire	Things that are different	Something that you smoke
Balanced	**note**	A short letter	A musical tone	Reading a book	Playing the drums
Balanced	**organ**	A body part inside of a person	The musical instrument played in church	A way of painting your nails	A building with a steeple
Balanced	**pupil**	Someone who studies under a teacher	A body part that helps you see	Something you see when walking to school	A body part that helps you hear
Unbalanced	**ball**	Something you would throw at the beach	Somewhere you might dance	Somewhere you would build a sandcastle	Something you would do when camping
Unbalanced	**bay**	A cove with water	The nook by a window	A smooth peaceful river	A large stone archway
Unbalanced	**bolt**	A strong metal pin	When you run away quickly	The handle of a door	When you go on a dog walk
Unbalanced	**file**	Somewhere you store paper	A metal tool with a rough surface	A tool to staple paper together	A tool to hit nails with
Unbalanced	**lock**	A way of keeping people out	A curl of hair	Something you use to open doors	Something you brush your hair with
Unbalanced	**park**	A place to play with your friends	Putting a car in a particular place	Somewhere you can play tennis	When you drive fast
Unbalanced	**pen**	Something you write with	Something you keep animals in	Something you paint with	Something you feed horses
Unbalanced	**ray**	A thin strip of radiation	A type of fish	When there are clouds in the sky	A large boat
Unbalanced	**shed**	A wooden hut outside	A tree losing its leaves	A colourful potted plant	A green forest
Unbalanced	**strike**	To hit something hard	To refuse to work in protest	Somewhere you exercise	When people run in a race
Unbalanced	**swallow**	A small songbird	What you do when you eat	A large bird of prey	What you do when you cry
Unbalanced	**temple**	Somewhere people pray	A area on the side of your head	Somewhere you buy food	When you have a stomach ache

Control items

Type	Target	Correct_answer	Inorrect_answer1	Incorrect_answer2	Incorrect_answer3
Balanced	bacon	Meat you eat for breakfast	A cold breakfast food eaten with milk	A type of farm animal	A person who wears lots of pink
Balanced	banjo	A musical instrument like a guitar	A musical instrument you blow into	A type of cowboy hat	A person who wears cowboy boots
Balanced	barn	A large farm building	A type of house in the country	Something that farm animals eat	An animal that makes milk
Balanced	canoe	A small boat with pointed ends	A type of fast boat	The pole used to steer a boat	A type of reptile with rough skin
Balanced	elf	A magical creature with pointed ears	A magical creature that flies	A type of sledge drawn by horses	The horns of a deer
Balanced	grey	The colour between black and white	The colour between blue and red	The colour of the sun	The colour of the grass
Balanced	hula	A type of dance by Hawaiian woman	Someone who loves to dance	A type of tropical tree	A type of volcano
Balanced	igloo	A house built from blocks of snow	A type of penguin	Someone who is always cold	A small house built of bricks
Balanced	kilt	A tartan skirt worn by men	Shoes you wear to tap dance	A instrument played by Scottish people	A type of small dog
Balanced	kiwi	A green, furry fruit	A long, yellow fruit	A round, red fruit	A type of berry
Balanced	mango	An oval, tropical fruit	A large, hairy seed	A yellow, prickly fruit	A person who picks fruit
Balanced	muffin	A small, domed cake	A big cake with lots of cream	A ring shaped bread roll	An oval object laid by chickens
Unbalanced	newt	A small amphibian with a tail	A type of frog	A large, colourful fish	A type of flower that floats
Unbalanced	nun	A woman who has dedicated herself to religion	A man who has dedicated himself to religion	A lady about to get married	A way of lighting candles
Unbalanced	rice	Food eaten with chopsticks	A type of pasta	A black and white animal	A large vegetable
Unbalanced	rugby	A team game with an oval ball	A game played with rackets	A game played with wooden sticks	A game played on a checked board
Unbalanced	salad	A healthy meal made of vegetables	Hot food made from potatoes	Flavoured ice on a stick	What you use to brush your teeth
Unbalanced	seven	The number after six	A number with two digits	A letter in the alphabet	A type of colour
Unbalanced	snail	An animal with a shell on its back	A large reptile with flippers	A slimy animal with no shell	The lava of butterfly or moth
Unbalanced	squid	A sea animal with tentacles	A large mammal with a blowhole	A type of kangaroo	A type of animal that builds a nest
Unbalanced	tulip	A brightly coloured, cup shaped flower	A small white and yellow flower	A winged insect that makes honey	A tool used for digging
Unbalanced	vase	A glass container you put flowers in	A type of glass you drink from	Something you cook with	A utensil you eat with
Unbalanced	wasp	A winged insect with a sting	An insect that forms from a caterpillar	A small, red beetle with black spots	The home a bee lives in
Unbalanced	zebra	An black and white striped animal	An animal with a trunk and large ears	An African mammal with a long neck	A person who wears bright colours

Appendix D: Stimuli List for Chapter 5 (Metaphors)

Experimental items

Item	Correct Answer1	Correct Answer2	Incorrect Answer1	Incorrect Answer2
ape	A type of primate	A large, clumsy, person	A person who likes bananas	Someone who is elegant and graceful
gem	A type of jewel	Someone who is valuable	A type of hat	A person who you do not like
brain	An organ in the head	Someone who is smart	A bone in the foot	A person who tells good jokes
index	Something you might point with	An alphabetical list of something	An unorganised person	Something you use when planting flowers
lion	A large member of the cat family	A person who is brave	A type of small, brown cow	Someone who has a lot of hair
nut	A type of food contained in a hard shell	Someone who is a little crazy	A fruit that grows underground	Somewhere you go to read quietly
sheep	An animal that grows wool	A person who follows others	An animal that is extinct	Someone who is always cold
snake	A reptile with no legs	Someone who betrays their friends	An animal with eight legs	Someone who is always late
spice	Something you put in food to add flavour	Something that adds interest	Someone who likes to cook	Someone who is adventurous
star	Something that lights up in the night sky	Someone who is famous	A type of rocket	Someone who eats a lot
wing	The part of a bird that makes them fly	A section of a building	An aeroplanes' engine	The seat of a chair
worm	An invertebrate with a long body	A weak, horrible person	An animal you see at the zoo	Someone who is very tall

Control items

Item	Correct Answer	Incorrect Answer1	Incorrect Answer2	Incorrect Answer3
bike	A vehicle you ride with two wheels	A loud vehicle with four wheels	A train that runs on electricity	A type of rollercoaster
cactus	A prickly plant in the desert	A plant that grows on a river bank	A type of soft flower	An animal that lives in a hot country
coffee	A hot drink adults like	A cold drink made from strawberries	Something you eat on toast	A type of brown fruit
desk	A piece of furniture you work on	A piece of furniture you sit on	A person who fixes chairs	A person who wears glasses
glue	A thick liquid used to stick things together	A liquid that erases ink	A tool that staples paper together	A tool that hangs pictures on walls
lake	A body of water surrounded by land	A small pool of water by the sea	A piece of land used for farming	A person who likes to swim
mitten	Clothing that keeps your hands warm	Clothing that makes your feet cold	A type of woolly hat	A person who can sew
oval	A shape like an egg	A shape like the sun	A colour like the moon	The sound a chicken makes
pizza	A round food made from dough	A spicy food made from nuts	A red, fizzy drink	A person whose clothes are too big
sofa	Soft furniture you sit on	Cupboards in the kitchen	A type of sink	A person sits on the floor
tuba	A musical instrument you blow into	A musical instrument you stand on	A musical instrument you hit	Someone who marches in a band
tutu	A skirt worn by ballerinas	A shoe worn by dancers	A hat worn by builders	A coat worn by doctors

Appendix E: Stimuli List for Chapter 6 (Metonyms)

List A

Target	Target sentence	Sense	Correct Answer	Incorrect Answer
cup	He poured tea into his favourite cup	1	I accidentally broke the handle of my cup	Though I hate coffee, I sometimes have a cup
bag	Mum put all the rubbish in a bag	2	She remembered she left her money in the bag	She waited for the porter to take her bag
plate	At the sink, he washed up his plate	1	Dad gave me my dinner on a pink plate	He couldn't go upstairs until he finished his plate
alley	We walked between the houses, down the narrow alley	2	The burglar ran away, through the dark alley	The builders were paving the alley
glass	He fell and broke the glass	1	Mum asked me to clean my glass	I was so thirsty I drank the whole glass
cage	The hamster was so big, it filled the cage	2	Thick black smoke filled the cage	He was busy painting the cage
tube	I always recycle the toilet paper tube	1	I asked for Smarties and she passed the tube	He squeezed all the toothpaste from the tube
pipe	The mouse ran through the pipe	2	There was a blockage in the pipe	He decided to fix the broken pipe
tub	They sold the soup in a small green tub	1	For my art project, I needed an empty tub	When eating ice-cream, she would finish the whole tub
barrel	She loved eating pickles from the barrel	2	Playing with water, the children filled the barrel	He hurt his back picking up the wooden barrel
lemon	At the supermarket, she picked up a lemon	1	The orange was smaller than the lemon	The kitchen cleaner smelled of lemon
pine	Her perfume smelled of pine	2	The new air freshener was pine	The forest was full of evergreen pine

List B

Target	Target Sentence	Sense	Correct Answer	Incorrect Answer
cup	He was thirsty and drank the whole cup	2	Though I hate coffee, I sometimes have a cup	I accidentally broke the handle of my cup
bag	She was upset the airport had lost her bag	1	She waited for the porter to take her bag	She remembered she left her money in the bag
plate	She didn't like lasagne so left the whole plate	2	He couldn't go upstairs until he finished his plate	Dad gave me my dinner on a pink plate
alley	There were lots of bins lining the alley	1	The builders were paving the alley	The burglar ran away, through the dark alley
glass	When the cat ran past, I spilled my glass	2	I was so thirsty I drank the whole glass	Mum asked me to clean my glass
cage	He bought a brand new cage	1	He was busy painting the cage	Think black smoke filled the cage
tube	The lumpy yogurt was stuck in the tube	2	He squeezed all the toothpaste from the tube	I asked for Smarties and she passed the tube
pipe	The plumber removed the old pipe	1	He decided to fix the broken pipe	There was a blockage in the pipe
tub	She used up all the butter in the tub	2	When eating ice-cream, she would finish the whole tub	For my art project, I needed an empty tub
barrel	There was a hole in the old barrel	1	He hurt his back picking up the wooden barrel	Playing with water, the children filled the barrel
lemon	The cake tasted like lemon	2	The kitchen cleaner smelled of lemon	The orange was smaller than the lemon
pine	He told me the tree was pine	1	The forest was full of evergreen pine	The new air freshener was pine

CONTROL	Target sentence	Correct answer1	Correct answer2
acorn	On the tree, hung a very large acorn	She bent to pick up the polished acorn	I gave the baby squirrel an acorn
ankle	She fell and hurt her ankle	The mosquito bit him on the ankle	The man had a tattoo on his ankle
burger	At the restaurant, she ordered a burger	I always take the pickle out of my burger	I wanted a milkshake with my burger
cliff	Dad was tired after hiking up the steep cliff	We looked up at the enormous cliff	She carefully sketched a picture of the cliff
cookie	She ate the freshly baked cookie	My brother would not share his cookie	Mum would not give me a cookie
girl	She gave a flower to the pretty girl	He saved the best seat for the girl	I had a lovely conversation with the girl
hen	A female chicken is called a hen	I watched the farmer chase the hen	He quickly picked up the noisy hen
lamp	I tripped over whilst holding the lamp	I switched on the big red lamp	Mum wanted to buy a new lamp
owl	In the forest, I saw an scary owl	My brothers favourite animal is an owl	Perched on the branch was an owl
pond	My sister pushed me into the pond	There was moss on the top of the pond	In the garden we built a pond
tennis	He said his favourite sport was tennis	We spent the afternoon watching tennis	In the summer, I enjoy playing tennis
turtle	My friend has a pet turtle	A tortoise is similar to a turtle	I fed the lettuce to the turtle

Appendix F: Stimuli List for Chapter 7 (Idioms)

Target sentence		Correct Answer	Incorrect Answer 1	Incorrect Answer 2	Incorrect Answer 1
Figurative	After I said something mean, our friendship turned sour	Our friendship went downhill	Our friendship floated my boat	Our friendship blew my mind	Our friendship went overboard
Figurative	I think it might rain tomorrow; I have a feel for it	I feel it in my bones	I'll have to face the music	It’s got out of hand	I’m barking up the wrong tree
Figurative	After having a argument, their marriage was hanging by a thread	Their marriage was on the rocks	They would be saved by the bell	They're marriage was right as rain	They got away by the skin of their teeth
Figurative	I failed my assignment; the teacher said I was on the wrong track	I went off on a tangent	I was preaching to the choir	I went the whole nine yards	I had bigger fish to fry
Figurative	He was so nervous during his speech, he had a mental block	He clammed up	He wore his heart on his sleeve	He was on the ball	He was keeping an open mind
Figurative	We couldn’t get the new business off the ground	The business got off on the wrong foot	The business was up in the air	The business was baking the wrong horse	The business had a chip on it's shoulder
Figurative	The choice was hard so she had to 'take the bull by the horns'	She’d have to bite the bullet	She’d be the elephant in the room	She’d have to go against the grain	She would be given the boot
Figurative	Paul said I had a big nose; he doesn’t 'mince his words!'	He doesn’t beat around the bush	He tried to butter me up	He was the apple of my eye	He couldn't make the grade
Figurative	While I was at school, Dad said the rabbit 'kicked the bucket'	The rabbit bit the dust	The rabbit knew the ropes	The rabbit was calling the shots	The rabbit put all his eggs in one basket
Figurative	Granny was sick; she looked very 'under the weather'	She almost looked on death's door	It’s raining cats and dogs	She was as fit as a fiddle	She was under the thumb
Figurative	After I drank a fizzy drink, Mum said I had 'ants in my pants'	I was full of beans	I had pins and needles	I was taking the biscuit	I didn’t have my heart in it
Figurative	When I didn’t tell her, Lucy told me to 'let the cat out of the bag'	She said to spill the beans	She said I had it in the bag	She went on a wild goose chase	She said to paint the town red
Literal	After I said something mean, our friendship turned sour	We started to fall out	He became my best friend	He said something mean back	He told the teacher
Literal	I think it might rain tomorrow; I have a feel for it	I can’t explain why, I just know	I could feel the raindrops in my skin	I saw it on the news	I feel excited for it

Literal	After having a argument, their marriage was hanging by a thread	They might break up	They distracted themselves with some sewing	They said sorry and made up	Their marriage was better than ever
Literal	I failed my assignment; the teacher said I was on the wrong track	I didn’t follow the instructions of the task	I’d given it to the wrong person	I should have written about roads	I needed to do the homework again
Literal	He was so nervous during his speech, he had a mental block	He couldn’t remember his words	He went crazy	He started shaking	He had to have a snack
Literal	We couldn’t get the new business off the ground	The business failed before it even started	The business involved aeroplanes	The business was going well	We don’t enjoy our jobs
Literal	The choice was hard so she had to 'take the bull by the horns'	She had to make a decision quickly	She’d have to take some time to think	She had to ask her husband	She’d have to go to the farm
Literal	Paul said I had a big nose; he doesn’t 'mince his words!'	He didn’t try to say it nicely	He used complicated words	He doesn’t talk quietly	He has a big nose too
Literal	While I was at school, Dad said the rabbit 'kicked the bucket'	The rabbit died	He’d taken the rabbit out of it's cage	The rabbit had stopped eating	The rabbit had babies
Literal	Granny was sick; she looked very 'under the weather'	She still looked very ill	She thought it might rain	She looked like she was getting better	She looked upset
Literal	After I drank a fizzy drink, Mum said I had 'ants in my pants'	I was full of energy	I was covered in bugs	She could see my pants	I’d spilt my drink
Literal	When I didn’t tell her, Lucy told me to 'let the cat out of the bag'	To tell her my secret	I didn’t’ have to tell her	Her cat ran away	She wasn’t my friend anymore
Literal	They first thought Paul had stolen the money, but he was let off the hook.	He was found to be innocent	He was found to be guilty	The bank called him	Everyone was mean
Literal	When it comes to classroom awards, the teacher holds all the cards	The teacher has all the power	The teacher can't make up her mind	The teacher wants to play 'Snap'	The children got to choose
Literal	She couldn't sleep, her mind was 'going around in circles.'	She was thinking too much	She felt dizzy	She'd had a tiring day	The bed wasn't very comfy
Literal	Hannah worried so much, it was better to 'keep her in the dark.'	It was better to keep some things secret	It was better to tell her everything	It was better to turn the lights off	They said she was silly for worrying
Literal	With her posh dress, she 'stole the show.'	She was the centre of attention	She looked silly	She was a thief	She tripped over her dress
Literal	After running around after the children, Dad 'ran out of stream.'	Dad was exhausted	Dad felt wide awake	Dad couldn't breathe	Dad ran out of money

Literal	Before Mary walked onto the stage, her sister said 'break a leg.'	Her sister wished her good luck	Her sister told her not to fall	Her sister said she'd be great	Her sister was wearing a cast
Literal	When Dad said the dog ran away, he was 'pulling my leg.'	He was only joking	He hurt my leg	He was upset	The dog bit him
Literal	Suzie got an A+ on her test, she said it was 'a piece of cake.'	She thought it was easy	She found it really hard	It was a test on baking	She was hungry
Literal	His jokes were terrible but he had me 'in stitches!'	I laughed so much	I hurt badly myself	I don't find him funny	He made me cry
Literal	I accidentally double booked myself today, I'm 'in such a pickle.'	I don't know what to do	I'm so unorganised	I need a new diary	I need to buy lunch
Literal	He was struggling with his homework and 'threw in the towel'	He gave up	He asked his Mum for help	He had another try	He went for a swim
Control	He couldn't listen to her complain any longer and told her to 'stop talking.'	He told her to be quite	He walked away	He could find his socks	He comforted her
Control	He got down on one knee and 'asked her to marry him.'	He proposed to her	He asked her on a date	He hurt his knee	He asked where she wanted to eat
Control	She complained she was cold, but they were all 'in the same situation'	Everyone was cold	No one else was cold	They were all outside	No one minded her moaning
Control	While his parents were away, his grandparents were 'in charge.'	His grandparents made the rules	His grandparents weren't strict	He missed his parents	His grandparents slept a lot
Control	Dad's usually quite calm, but yesterday he 'lost control'	He got really angry	He crashed the car	He fell asleep	He couldn't stop laughing
Control	She tidied her room 'very rarely'	She didn't tidy her room a lot	She never tidied her room	Her room was already clean	She'd rather go to the movies
Control	They made holiday plans together to make sure everyone 'agreed'	To make sure everyone wanted the same things	To arrange two different holidays	To make sure everyone was awake	To see who didn't want to come
Control	The party wasn't going well so she decided to 'put an end to it.'	She decided to stop the party	She couldn't wait till it was over	She cried	She brought out more cake
Control	Having three-year-old twins will 'keep you busy'	It will make sure you stay active	It will make you tired	It will mean you sleep lots	It means you need two of everything
Control	By the end of term, the children were 'becoming difficult to control.'	They were becoming hard to manage	They were starting to calm down	They were all talking about their holidays	The teacher was excited for half-term

Control	When she got fired, she felt all her hard work had 'been wasted'	She felt it was all for nothing	She felt she had learnt a lot	She would miss her new friends	She was glad she was fired
Control	After breaking his Mums vase, Tim tried to 'pass the blame.'	He said it was somebody else's fault	He apologised	He tried to fix the vase	He said he would pay for a new vase
Control	After a long day, he couldn't wait to 'go to bed.'	He was really tired	He was wide awake	He was ready to go to the park	He hadn't made his bed this morning
Control	I was ready to leave the house but mum said to 'wait a moment.'	Mum wasn’t ready to leave	Mum was hurrying me out of the door	We were really late	I couldn’t find my boots
Control	I didn’t know what flavour ice-cream to get. I 'couldn't make up my mind.'	I was still considering my options	I wanted to buy strawberry	I don't like ice-cream	I decided to have cake instead
Control	A week after their first date, Pete knew he was 'in love.'	Pete really liked her	Pete wanted to break up	They might go to the cinema	They'd known each other a long time
Control	Sam kept eating Amy's' cereal, it 'really annoyed her.'	It made Amy angry	Amy didn’t mind	Amy didn't like cereal	Sam bought cereal
Control	He came home late again to a note saying 'she was angry.'	He got told off	He needed to take the children to school	He needed to buy milk	His wife was very happy
Control	We put on the kettle and 'had a chat'	We had a lovely conversation	We watched some television	We ate some biscuits	We had an argument
Control	I haven't heard from Auntie Sarah. I must 'give her a call'.	I should ring Auntie Sarah	Me and Auntie Sarah have fallen out	Auntie Sarah said she wall call me	My phone is broken
Control	After starting his new job, Ben tried not to 'cause trouble'	He tried to keep to himself	He made a lot of fuss	He told lots of jokes	He didn’t like his new job
Control	If he was going to be on time, he would have to 'hurry up.'	He would have to get a move on	He might cancel his plans	He was going very slowly	He woke up late
Control	After buying her dream dress, she couldn't wait 'to get married'	She was excited to be a wife	She was a guest at a wedding tomorrow	She didn’t like shopping	Her dress was beautiful
Control	She had a sum money saved for the future	She had some money in the bank	She spent all her money	She just bought a car	She needed to go to the bank