The nature of multi-word vocabulary among children with English as a first or additional language and its relationship with reading comprehension.

Sara Ashley Smith

Hertford College
Department of Education

University of Oxford

Abstract

Vocabulary is well acknowledged as playing a critical role in language and reading development for young children, particularly for children learning English as an Additional Language (EAL) in school (Scarborough, 2001; Stahl & Nagy, 2005). However, most previous research on vocabulary has focused on measuring individual words and failed to examine knowledge of multi-word phrases, despite corpus evidence that these items are common in the English lexicon (Erman & Warren, 2000). The nature of multi-word vocabulary knowledge and its possible contribution to literacy skills among children remains underexplored, possibly due to a lack of available suitable measures. The current thesis details the development and administration of an original multi-word phrase task containing transparent, semi-transparent and non-transparent verb + object phrases to 108 British monolingual English speakers and learners with EAL in school years 3, 4 and 5. Results showed a strong effect of item transparency, even greater than frequency. Year 4 monolingual English speakers had significantly higher scores than year 3 monolingual learners on non-transparent items, while among learners with EAL year 3 and 4 performances were similar and year 5 learners' scores were significantly higher. The second phase of the study explored the contribution of multi-word phrase knowledge to reading among 40 year 4 monolingual English speaking children and Bengali speakers with EAL. Multiple regression analysis showed that multi-word task performance accounted for a significant amount of variance in reading scores, when controlling for non-verbal intelligence, receptive and expressive single word vocabulary and language background. These findings are of import for increasing our understanding of vocabulary development among young learners and provide insight into the particular needs of learners with EAL.

TABLE OF CONTENTS	
Chapter 1: INTRODUCTION	p.1
Chapter 2: LITERATURE REVIEW	p.12
2.1 Formulaic Language: An introduction	p.12
2.2 Definitions and frameworks	p.13
2.2.1 Terminology	p.14
2.2.2 Conceptual Perspectives	p.15
2.2.3 Collocations versus idioms	p.18
2.2.4 Opacity, transparency and formulaic language	p.19
2.2.5 Formulaic language and theory: Novelty versus formula	p.19
2.3 Multi-word phrase and first language acquisition	p.20
2.3.1 Learning English	p.21
2.3.2 Encountering novel figurative language later in the learning process	p.23
2.3.2.1 The Global Elaboration Model	p.24
2.4 Multi-word phrases in the mental lexicon	p.26
2.4.1 Theoretical models	p.26
2.4.2 Processing multi-word phrases	p.28
2.5 Theoretical models of reading	p.28
2.6 Reading Fluency	p.32
2.7 The role of vocabulary in reading comprehension	p.33
2.8 The Lexical Restructuring Model	p.34
2.9 Vocabulary breadth versus depth	p.37
2.10 Multi-word phrases and reading comprehension	p.38
2.10.1 Reading comprehension and L2 learners	p.38
2.10.2 Reading comprehension and multi-word units for children	p.39
2.11 Learners with EAL	p.44
2.11.1 Learners with EAL and language development	p.44
2.11.2 EAL reading comprehension	p.45
2.11.3 Vocabulary, reading comprehension and learners with EAL	p.47
2.11.4 Cross-linguistic similarity	p.49
2.11.5 UK based learners with EAL	p.51
2.12 Existing tests of formulaic knowledge	p.56
2.13 Rationale for the current study and original contribution	p.62
Chapter 3: METHODOLOGY	p.64
3.1 Introduction	p.64
3.2 Phase 1: Creating and piloting a test of multi-word phrase knowledge	p.66
3.2.1 Test design and development	p.66
3.2.2 Material	p.75
3.2.3 Sampling procedures for Phase 1	p.76
3.2.4 Testing procedures for Phase 1	p.79
3.2.5 Pilot results for Phase 1	p.79
3.2.6 Format changes made after the pilot	p.81
3.3 Phase 2: Validation of the MPT	p.82
3.3.1 Materials	p.84
3.3.2 Sampling procedure for Phase 2	p.88
3.3.3 Testing procedures for Phase 2	p.90
3.4 Phase 3: Correlations between reading comprehension and multi-	p.90

word…	
3.4.1 Materials	p.91
3.4.2 Sampling procedure for Phase 3	p.92
3.4.3 Testing procedure for Phase 3	p.92
3.5 Ethical considerations	p.93
3.5.1 Central University Research Ethics Committee	p.93
3.5.2 Language considerations	p.93
3.5.3 Inclusion of select students	p.94
3.5.4 School day disruption	p.94
3.5.5 Training	p.94
Chapter 4: RESULTS	p.96
4.1 Introductions	p.96
4.2 Phase 2 results: Validating the MPT	p.96
4.2.1 Descriptive statistics	p.96
4.2.2 MPT internal consistency	p.100
4.2.3 Test-retest reliability	p.101
4.2.4 MPT concurrent validity	p.102
4.2.5 Item facility and Item discrimination	p.104
4.2.6 Test and group effects	p.107
4.2.7 Phase 2 conclusions	p.120
4.3 Phase 3 results: Relationship with reading	p.121
4.3.1 Descriptive statistics	p. 122
4.3.2 Correlations between measures	p.124
4.3.3 Regression models	p.126
4.3.3.1 MPT overall contribution to YARC score variance	p.128
4.3.3.2 MPT Transparent contribution to YARC	p.130
4.3.3.3 MPT Semi-transparent contribution to YARC	p.131
4.3.3.4 MPT Non-transparent contribution to YARC	p.132
4.3.4 Phase 3 conclusions	p.134
Chapter 5: DISCUSSION AND CONCLUSION	p.135
5.1 Introduction	p.135
5.2 The nature of multi-word phrase knowledge among monolingual …	p.135
5.3 The Multi-word Phrase Task	p.140
5.3.1 Strengths of the MPT	p.140
5.3.2 Limitations of the MPT	p.142
5.3.3 Directions for future research with the MPT	p.145
5.3.4 Testing multi-word phrases	p.146
5.4 The relationship between multi-word phrase knowledge and reading	p.146
5.5 Implications with regard to pedagogy	p.150
5.6 General study limitations	p.151
5.6.1 Sampling	p.152
5.6.2 Instruments	p.153
5.6.3 Absences of L1 measures	p.155
5.6.4 Working with young learners	p.155
5.7 Conclusions	p.156
REFERENCES	p.160
APPEDICES	p.i

Chapter 1: INTRODUCTION

Assessment of vocabulary in reading comprehension has tended to focus on measuring the number of words in a text the individual knows, or *vocabulary breadth* (Pearson, Hiebert & Kamil, 2007). However, in order to process words effectively the learner must also have knowledge regarding a number of other aspects of the word, such its grammatical functions, register, stylistic characteristics, appropriate contextual usage, and the other words it combines with to form formulaic phrases (Milton, 2009). One area of vocabulary of increasing interest is the role multi-word phrases, formulaic phrases such as idioms and collocations, play in language acquisition and processing. There are a variety of definitions and conceptualizations of formulaic sequences; one prevailing element is that formulaic phrases function similarly to individual lexical items, with multiple words expressing a singular meaning or functions (Martinez & Schmitt, 2012; Wray, 2002). This is particularly evident when dealing with opaque formulaic language. With the help of corpus linguistics, it is now well accepted that such items are ubiquitous in naturally occurring discourse and are an essential element of language (Erman & Warren, 2000; Sinclair, 1991). As a vehicle for meaning, phraseological language offers a capacity to express a concept in a concise and effective manner. Research evidence indicates that these phrases are often learned and stored holistically (Wray, 2002, 2008) and consequently, it has been posited that these items might be better conceptualized as 'giant lexical items' in and of themselves (Nippold, 1998, p.106).

While much previous research has focused on how children develop the ability to comprehend figurative language and how young learners approach novel idioms (Cain et al., 2005, 2008; Gibbs, 1987, 1991; Levorato & Cacciari, 1992, 1995), at present little is known about the nature of multi-word phrase knowledge among children at various

stages of language development. Few studies have attempted to measure the amount of phrase vocabulary children have at different ages, perhaps in part due to a lack of appropriate available measures. This area of inquiry may be of particular import in relation to children who speak a language other than English at home and learn English as an Additional Language (EAL) in school. Many studies have found that learners with EAL lag behind monolingual English speaking peers in vocabulary and reading comprehension (Bialystok, 2010; Garcia, 1991; Verhoeven, 1990) and it is therefore possible that multi-word phrases present a particular challenge for these learners, especially since these items have been found to be particularly difficult for adult English as a foreign language learners (Barfield & Gyllstad, 2009).

1.1 Rationale for the current study

Formulaic opaque multi-word phrases, such as idioms, are well acknowledged to present a challenge both to young learners (Cain, Towse & Knight, 2009) and adult second language (L2) learners (Barfield & Gyllstad, 2009) and are frequently occurring, with as many as 500 multi-word phrases ranking in the top 5,000 most commonly used word families (Martinez & Schmitt, 2012). There is evidence that formulaic language is of benefit for language processing speed (Pawley & Syder, 1983; Gibbs et al., 1997). The predictability of these fixed expressions is thought to facilitate reading and reading related skills such as encoding, decoding (Poulsen, 2005, p. 77), fluency (Wray, 2002) and increased reading speed (Siyanova-Chanturia et al., 2011; Conklin & Schmitt, 2006). However, despite their frequent and facilitative nature, non-transparent items are thought to pose a challenge to young learners because they offer two potential meanings and must therefore be deciphered (Cain et al., 2009).

There has been much research focusing on the strategies children use to decipher novel idioms, the language skills requisite for understanding figurative language, as well

as the age range during which these skill develop: generally age 7 to 11 years (Cain et al., 2009; Gibbs, 1987, 1991; Levorato & Cacciari, 1995, 1999; Nippold & Martin, 1989; Nippold, Moran & Schwatz, 2001; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995); though there is evidence that even at age 11 or 12 years, idiom comprehension skills are not fully comparable to those of adults (Cain et al., 2009). Little previous work, however, has attempted to and measure and compare any productive elements of figurative language knowledge among children at different points during this age range or explored how this knowledge compares to knowledge of transparent multi-word items. This may be due to a lack of available child-appropriate measures of multi-word phrase knowledge.

For first language (L1) English speaking children, the ability to comprehend non-literal language is hypothesized to develop between roughly between the ages of 7 and 11 years (Levorato & Cacciari, 1995). The Global Elaboration Model, or GEM, posits that children develop the capacity for understanding figurative language over the course of the primary school years; the same linguistic skills that underlie reading and listening comprehension also enable figurative language comprehension (Levorato & Cacciari, 1995). According to the GEM, children initially process unknown figurative language part by part, using the components to piece together meaning; however, as they age children acquire the ability to infer figurative meaning from the surrounding context and they will employ more holistic strategies. These skills, essential for processing and retaining figurative language, develop during this age range of 7 to 11 years (Levorato & Cacciari, 1992, 1995, 1999) and by the end of primary school, children have usually acquired a high enough level of figurative competence to abandon a piecemeal approach (Levorato & Cacciari, 1995). There is a healthy body of evidence that children do develop the ability to understand figurative language generally in this

manner (Cain et al., 2005; Gibbs, 1987, 1991; Levorato & Cacciari, 1995, 1999; Nippold & Martin, 1989; Nippold, Moran & Schwatz, 2001; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995), although at age 11 years idiom comprehension may not be completely adult-like (Cain et al., 2009). As lexical knowledge and general language ability increases in a typically developing child, so too does the ability to elaborate language, process incoming language and reconcile this new information with information that came before, the same skills needed for understanding both non-figurative and figurative language (Levorato & Cacciari, 1995). The general age range identified by the GEM as a developmental period for processing and retaining figurative language (7 years – 11 years) will be the focus age range for the present study (7 years – 10 years).

For monolingual English-speaking children, research suggests that vocabulary knowledge at ages 5 and 6 years can be strong predictors of later reading outcomes (Scarborough, 2001; Cunningham & Stanovich, 1997; Roth, Speece and Cooper, 2002) and the larger the child's vocabulary the faster they will learn subsequent words (Biemiller, 2005; Nation, 2001). However, previous assessment of vocabulary in children has often failed to include consider multi-word phrases as vocabulary items, despite argument that these word combinations are not processed or stored as separate words and should be included in measures of vocabulary size (Martinez & Schmitt, 2012, Nippold, 1998; Read, 2000; Wray, 2002).

While vocabulary is generally a strong predictor of reading and listening comprehension (Anderson & Freebody, 1981; Vermeer, 1992); reading, in turn, has been shown to increase vocabulary (Anderson & Freebody, 1985; Stahl & Nagy, 2005; Taguchi, 1997). Children who do not successfully acquire basic word reading skills in the early years may then 'decelerate in comprehension' as instruction transitions from

'learning to read' to 'reading to learn' (Chall et al., 1990), a pattern that has been observed in a number of studies investigating children with EAL (Lesaux, Crosson, Kieffer & Pierce, 2010). This phenomenon has been called 'The fourth grade slump' by Chall, Jacobs and Baldwin (1990) who pointed out that by late elementary reading demands increase and content becomes increasingly abstract and complex (Chall et al., 1990; Kieffer & Lesaux, 2007).

This reciprocal relationship between vocabulary and reading can create a vulnerable situation for students with fewer opportunities to acquire background knowledge of vocabulary or have less exposure to English print, such as for example, some learners with EAL (approximately 17.5% of pupils in the UK (NALDIC, 2012)). Attainment studies have shown that achievement of learners with the EAL in the UK can be much lower than of monolingual English speaking peers (Burgoyne et al., 2009); monolingual English Speakers tend to have higher levels of educational achievement overall throughout schooling (DCSF, 2006/2007) and there is a significant attainment gap in A level scores between EAL and monolingual English speaking pupils when other variables such as socio-economic status (SES) are controlled for (NALDIC, 2006). Research points to vocabulary as a particular variable that constrains language comprehension for learners with EAL (Garcia, 1991; Verhoeven, 1990). Comparisons of vocabulary size between monolingual English speakers and children with EAL have shown that the EAL children tend to have smaller English receptive vocabularies (Bialystok, 2010). While a number of factors may contribute to this achievement gap, studies have consistently shown that vocabulary has a significant impact on school outcomes (Hutchinson et al., 2003). EAL children who begin school with little or no English must rapidly acquire the language in order to follow the same curriculum as L1 classmates who enter school already having 5 or 6 years of language learning. Despite

having equal intelligence and some increased ability related to their bilingualism (Bialystok, 2011), much research in both the UK and North America has shown that EAL school children often lag behind their monolingual peers in reading comprehension (Hutchinson et al., 2003; Lesaux et al., 2007).

Two UK based studies (Burgoyne et al., 2009; Hutchinson et al., 2003) exploring reading comprehension among children with EAL reading comprehension have shown significant differences between monolingual children and those with EAL in performance on reading comprehension; Hutchinson and colleagues conducted a longitudinal study collecting data at three time points and found that children with EAL lagged behind monolingual English speaking children by an average of one year in reading performance, as measured by the Neale Analysis of Reading Ability (NARA), in year 2 through till year 4, and without closing vocabulary gaps between the two language groups (Hutchinson et al., 2003). It is worth noting, however, that Hutchinson et al. compared approximately 43 children in each group to each other, 86 in total, and all learners with EAL were only from Indo-Iranian language family backgrounds. More importantly, all learners with EAL included in the study were learning to read Arabic outside of school – a potential confounding factor if this decoding training influenced other reading related skills (Hutchinson et al., 2003). This is especially worth considering in light of Burgoyne et al.'s (2009) study with 92 year 3 monolingual English speakers and learners with EAL – the study found learners with EAL focus on decoding when reading text, which in turn leads to comprehension difficulties; the authors suggest that learners with EAL struggle to read for meaning and are constrained by significantly weaker vocabulary knowledge. However, as with Hutchinson et al. (2003), all participants learning EAL were from Indo-Iranian language backgrounds and most (87%) attended Mosque 5 or more days a week, therefore were also likely learning

to read in Arabic outside of school, although Burgoyne et al. do not specifically report whether or not they were having language training (Burgoyne et al., 2009).

There is evidence that non-transparent language poses comprehension problems for adult English L2 learners (Barfield & Gyllstad, 2009; Cooper, 1999; Liontas, 2002; Martinez & Murphy, 2011). Studies have found the more opaque an item is, the more an L2 learner will struggle with comprehension (Cooper, 1999; Liontas, 2002). Martinez and Murphy (2011) presented adult L2 learners of English with paired texts containing the same lexical items, one text with items arranged into multi-word phrases of varying opacity and the other with only literal usage. Texts with non-transparent multi-word phrases were significantly less well comprehended than those with only literal language (Martinez & Murphy, 2011).

Interestingly, L2 learners do not seem to always be aware of their own phraseological knowledge gaps (Bishop, 2004; Martinez & Murphy, 2011). Bishop (2004) found that readers looked up the meanings of unknown words significantly more often than they looked up unknown formulaic phrases, perhaps indicating that the learner is unaware phrase components have taken on a holistic meaning. Martinez and Murphy (2011) found learners significantly overestimated their comprehension of the non-literal phrase texts, indicating the expressions were either unnoticed or learners incorrectly assumed they understood them.

It is therefore of interest to explore non-transparent multi-word phrase knowledge among both in monolingual English-speaking children and children with EAL. The current study will compare transparent and non-transparent multi-word phrase knowledge between children of different ages as well as compare children from monolingual English speaking backgrounds to those with EAL. Vocabulary knowledge is one of many important factors for understanding reading comprehension and reading

achievement outcomes; despite the clear link between single word vocabulary and reading, little research thus far has explored multi-word language development in monolingual English speaking children and none with children with EAL, despite the fact that these phrases are ubiquitous in daily speech and regularly appear in primary school curricula (Nippold, 1991). These items may therefore have an impact on, and a reciprocal relationship with, reading comprehension.

Additionally, there are some gaps in the literature regarding UK reading research with learners with EAL. Hutchinson and colleagues' 2003 study of UK learners with EAL included only learners with EAL who were learning Arabic reading outside of schooling, an additional decoding

1.2 Structure of the thesis

The aim of the current research is to explore the nature of verb + noun multi-word phrase knowledge among monolingual English speakers and learners with EAL in school years 3, 4 and 5 (corresponding to ages 7/8 years, 8/9 years and 9/10 years, respectively). The current study contains three phases; Phase 1 entails the creation of a verb + object multi-word phrase task (MPT) for use with children between 7 and 10 years old. Phase 2 consists of administering the task to 108 monolingual and EAL learners along with a batter of vocabulary assessments. Phase 3 of the study consists of administering the MPT to 40 learners in year 4 along with a battery of vocabulary and reading measures in order to explore the relationship between multi-word phrase vocabulary and reading comprehension.

Most tasks previously used with this age group have been multiple choice or idiom explanation tasks – few have asked for any productive elements of phrase knowledge. The current study attempts to ameliorate this paucity of research by measuring young learners' knowledge of multi-word phrases at three different time

points (school year 3, school year 4 and school year 5). The current study compares knowledge of transparent verb + object phrases with knowledge of figurative verb + object phrases of comparable overall frequency and composed of comparably frequently occurring components. The goal of Phase 2 of the present study is to measure children's knowledge of these multi-word phrases at these three different developmental time points to shed light on how transparency affects performance on the verb + object phrase task and if the effect differs between the ages of 7 and 10 - the age span during which some researchers hypothesize that the ability to analyze and understand idiomatic vocabulary develops (Cain & Towse, 2008; Levorato & Cacciari, 1992, 1995; Wray & Perkins, 2000), and to explore if this knowledge differs between monolingual English speakers and learners with EAL.

Phase 3 of the study seeks to address the gap in our understanding regarding the role of multi-word knowledge as an element of vocabulary that may contribute to variance in performance on four areas of reading ability (single word reading, reading accuracy, reading rate and reading comprehension) among young learners. 20 monolingual English speaking children and 20 Bengali speaking children with EAL in year 4 were given a battery of language assessments, including the MPT, expressive and receptive single word vocabulary measures, reading assessments and a non-verbal intelligence measure to explore how MPT performance might predict variance in reading measure outcomes.

The following chapter of this thesis, Chapter 2, will provide a review of existing literature related to the present study. It will present background information regarding multi-word vocabulary including a review of how it is distinct from single word vocabulary, how it is learned and stored, how it is measured and categorized, and its place within the field of linguistics including both historical and theoretical perspectives.

The section will also present an overview of the terminological debates regarding formulaic language. Chapter 2 will next examine multi-word vocabulary in first language acquisition and how young English speakers approach and process opaque multi-word phrases. The chapter will then go on to review reading: theoretical models, the components of reading comprehension, the contribution of vocabulary in general to reading and the contribution of multi-word units specifically. Next, the chapter will outline the current research regarding learners with EAL, specifically their language and reading development, and elaborate on the role multi-word phrases might play in vocabulary acquisition for these children. Finally, the second chapter will review existing tests of formulaic knowledge to evaluate appropriateness for the current study.

The third chapter of this thesis, Chapter 3, will detail the methodology of the present research project. The chapter first presents information regarding test creation and pilot, including test design and development, sampling procedure and piloting. The chapter will subsequently examine informed changes made after the pilot. Next, Chapter 3 describes Phase 2 of the study, including validation and administration of the MPT. This section also describes the sampling and testing methods of Phase 2. The chapter will then present the final phase of the study, Phase 3, including sampling procedures, methodology and materials. Lastly, the chapter expands upon the ethical concerns relevant to the current study and how the researcher addressed these issues where possible.

Chapter 4 presents the results from Phases 2 and 3 of the current study. This chapter contains detailed descriptive findings from the MPT administration in Phase 2, followed by statistical evaluations of internal consistency, Item facility and Item discrimination and an exploration of correlations between the task and other language measures. Effects of Transparency, Language Background, Frequency and Year Group

will all be presented and discussed in detail. Next, Chapter 4 will describe the findings from Phase 3, including descriptive statistics and the multiple regression models evaluating the contribution of multi-word vocabulary to reading outcomes and explore differences between groups.

The fifth and final chapter of the thesis elaborates on the research findings presented in Chapter 4 and discusses the limitations of the project. In addition, Chapter 5 proposes directions for future research and possible consideration for further exploration of this topic. The chapter will first expound upon the findings regarding the MPT, its strengths and weaknesses, as well as make proposals for potential adjustments and improvements to the measure. This section will also seek to characterize insights gained from study that could be beneficial for future research. Chapter 5 then discusses the nature of multi-word vocabulary and places the findings within the context of the literature. The differences between the two language groups included in the present study, monolingual English speakers and learners with EAL, will be explored. Limitations of the current study will also be discussed. Lastly, the chapter will review of findings of the study overall and draw final conclusions.

Chapter 2: LITERATURE REVIEW

2.1 *Formulaic Language: An introduction*

Linguists have long observed humans' capacity to store entire phrases (proper names, clauses, poems, song lyrics, even passages of prose) as holistic chunks, sometimes without considering or even understanding the meaning of the individual components (Wray, 2002). Evidence has shown that English speakers can process these multiple words together as a bundle (Wray, 2002). *Processing* in this context has two main meanings: a developmental definition referring to the acquisition of representations of linguistic knowledge or *developmental processing,* and the act of then using these representations of knowledge in real time, or *knowledge processing* (Sharwood Smith, 1991, p.10). Sometimes breaking down frequently used set phrases can even destroy the meaning – the whole phrase used in the customary manner can produce an entirely different meaning than the individual parts. Idioms are a prime example – the meaning can be entirely illogical or unfeasible when the parts are analyzed individually but as a whole it can be understood and the speaker's intended message is easily interpreted (e.g., It's raining cats and dogs). Many researchers contend that these word combinations are not processed, stored or recalled as separate words but can instead be conceptualized as one 'giant lexical unit' and should be included in measures of vocabulary size (Martinez & Murphy, 2011; Read, 2000; Wray, 2002). The predictability of these commonly used fixed expressions is thought to facilitate encoding, decoding (Poulsen, 2005, p. 77), processing speed (Gibbs et al., 1997) and fluency (Wray, 2002).

Formulaic phrases account for a sizable percentage of commonly used language (Nattinger & DeCarrico, 1992); high estimates are 58.6% for spoken and 52.3% for written language (Erman & Warren, 2000) to more modest estimates of 30% in conversational English and 21% in academic writing (Biber et al., 1999). As much as 70% of the language used by adult English speakers has been posited to be formulaic in some way (Altenberg, 1990). Corpus studies have shown that multi-word phrases are extremely common in English, with nearly 500 collocations ranking within the top 3,000 most commonly used word families (Martinez & Murphy, 2011). It has been suggested that if the formulaic language that English speakers know (counting names, titles, songs, poems, proverbs and onward) were counted up as lexical items, the number would match single word vocabulary

counts, doubling the number of items kept in the passive mental lexicon from about 60,000 to 120,000 (Jackendoff, 1995; Miller, 1991). The high frequency of these expressions suggests that they are an essential component of English language vocabulary knowledge and the mental lexicon (Wray, 2002).

Vocabulary knowledge plays a key role in reading and consequently, formulaic language may also have an influence on reading ability – underscoring the importance of exploring this relatively unexamined aspect of vocabulary. Phrasal knowledge has also been shown to be particularly difficult for L2 English learners (Barfield & Gyllstad, 2009) and so might similarly present challenges for the population of young EAL students in the United Kingdom (UK) – a population that research has shown to lag behind monolingual English speaking counterparts in vocabulary and reading comprehension (Burgoyne, Kelly, Whiteley & Spooner, 2009).

This chapter will first present the current literature on multi-word vocabulary in English, with a discussion of the varied theoretical and terminological perspectives, followed by a review of the presence of multi-word vocabulary phrases in early child language. The chapter will then provide an overview of reading with theoretical perspectives. Next, learners with EAL and their language development will be discussed, integrating the earlier material regarding reading and vocabulary. Finally, the presently existing measures of multi-word knowledge will be considered, keeping in mind the theoretical perspectives on multi-word vocabulary, the mental lexicon and the needs of young learners.

2.2 *Definitions and frameworks*

Setting typological classifications for formulaic language presents a complex task for linguists, made more difficult by the multitude of theoretical frameworks and differing methods of identification and operationalization. For example, frequency-based perspectives posit that words form a unit by nature of how frequently they occur together, relying heavily on corpus evidence and probability of word co-occurrence to determine phrasal cohesion (Barfield & Gyllstad, 2009; Sinclair, 1991). More phraseological perspectives are less focused on frequency as a way of determining multi-word units and instead emphasize the semantic and syntactic features of the words (Barfield & Gyllstad, 2009; Nesselhauf, 2005). These strings of words have many different names and

definitions, the permutations of which will be discussed in further depth later in this chapter. For the purposes of the present study, the items used in the measurement task will be referred to as *multi-word phrases*, words that co-occur together more than could be predicted by chance, appear to be stored and processed holistically and, in the case of the semi-transparent and non-transparent items, express a holistic meaning. The current study does not seek to resolve the debates regarding classification of formulaic language and the author acknowledges that the items used in the task fall on the continuum of proposed categories. *Formulaic language*, as used in the current study, will be a coverall term to describe this type of language, encompassing fixed idioms and routines, as well as the less fixed collocations and phrases. Some theoretical perspectives might consider some or all of the target items used in the current study to be either idioms or collocations based on their high co-occurrence in corpus evidence, while others would only classify them as formulaic phrases. The transparent items, for example, do not have a holistic definition and their meaning can be derived from the parts. This research will focus only on measuring the phrases themselves rather than exploring all the various parameters that have been used to define them.

2.2.1 *Terminology*

Formulaic language has long been recognized as ubiquitous in English (Wray, 2002). However, despite their prevalence, it has proven difficult for linguists to reconcile the many different definitions, operationalisations, and methods for identifying multi-word units. A multitude of terms and categories have been presented in the literature, some with overlapping conceptualizations. This terminological medley is further complicated by the many differing theoretical perspectives on the phenomenon. Terms that have been used to describe these types of strings include: *amalgams*, *chunks*, *collocations*, *complex lexemes*, *composites*, *fixed expressions*, *formulaic language*, *fossilized forms*, *frozen phrases*, *gestalt processes*, *idioms*, *multi-word chunks*, *non-compositional phrases*, *phrasal units*, *prefabricated routines*, *rote utterances*, *set phrases*, and *underanalyzed chunks of speech*. This list is by no means exhaustive and should simply serve to underscore the diversity of opinion and perspective within the field. Much of this complexity is rooted in the differences in historical perspective and theoretical framework that have been used to discuss formulaic language (Barfield & Gyllstad, 2009).

2.2.2 *Conceptual perspectives*

Past research has generally approached multi-word phrases from two different perspectives: the *frequency-based*, which conceptualizes these phrases as co-occurring words that form a unit by nature of how frequently they co-occur and the *phraseological perspective*, which instead emphasizes the semantic and syntactic features of the phrases (Barfield & Gyllstad, 2009). The frequency based perspective, first explored by Firth (1952; 1957), Halliday (1961; 1966) and later Sinclair (1987; 1991), focuses heavily on analysis of text and probability of word co-occurrence (Barfield & Gyllstad, 2009). The phraseological view, championed by Cowie (1981; 1988; 1991; 1998), Howarth (1996; 1998) and Nesselhauf (2005), treats multi-word phrases as word combinations and instead emphasizes the semantic and syntactic features as most salient (Barfield & Gyllstad, 2009). This tradition focuses less on frequencies and statistical significance and more on the transparency of the phrases themselves (Barfield & Gyllstad, 2009). These perspectives should be considered overlapping rather than dichotomous and often yield similar results for 'qualifying' phrases as formulaic chunks. The current study will combine elements from both traditions, the idea that a multi-word phrase has holistic meaning and that its components co-occur together statistically significantly more often than could be expected by chance alone.

Phraseological perspective

The phraseological perspective developed out of second language (L2) English pedagogy, as linguists attempted to provide L2 learners with learning materials based on formulaic language. This perspective focuses on a multi-word unit as word combinations with varying amounts of flexibility or fixedness, with a heavy emphasis on creating typological classifications. The phraseological perspective does not incorporate ideas of frequency or statistical likelihood of co-occurrence, hallmarks of the frequency perspective. Instead, the phraseological perspective emphasises phrase opacity and combination typology (Barfield & Gyllstad, 2009). Cowie (1981; 1998) used the work of 4 key linguists from the 1940's; Russian linguists Vingrov and Amosova who focused on ideas of 'collocation restriction' and Harold Palmer and A. S. Hornby, reformers and innovators of English language teaching in Japan, to develop classifications and categories for English phrases. Cowie (1991) defines *collocations* as two or more associated words that co-occur

in a particular variety of grammatical settings and *restricted collocation* or idioms as phrases in which one lexeme – often the verb – originally 'a technical sense or a long-established figurative sense which has lost most of its analogical force' (Cowie, 1991, p.102). Critics of the phraseological perspective argue that this definition does not consider the psychological reality of how phrases are processed (Erman, 2009). Many multiword phrases are processed as whole units but are not metaphorical or restricted enough to meet the standard of the phraseological framework, but are still very much non-compositional and processed holistically (Erman, 2009).

Frequency-based perspective

The primary proponents of the frequency-based perspective on formulaic phrases are Firth, in the 1950's, Halliday in the sixties and more recently, Sinclair. Firth posited that there were 5 overlapping levels of meaning transmission in language – collocation, phonology, phonetics, situational context and syntax (Firth, 1957). In a groundbreaking move, Firth separated the idea of words that commonly occur together in phrases and formulaic sequences (known as 'collocability') from the syntactic elements of a word (1957). As such, he posits that a listener might be better able to comprehend the meaning of a particular lexical item through other words it combines with in commonly used phrases – the meaning of the word 'tea' could be better established with the words 'strong', 'weak' or 'afternoon' (Firth, 1957). Firth (1952) also proposed that words themselves do not have intrinsic meaning alone, and instead have a 'functional value' based on usage (Firth, 1952, p.23). Although he did not specifically provide definitions for types of formulaic language, instead focusing on strings of concurring words, he provided the famous quote so often used in discussions of formulaic language: ' You shall know a word by the company it keeps' (Firth, 1957, p.179).

Halliday elaborated upon Firth's pioneering work and was the first to use the word *collocate* to refer to a commonly co-occurring item (Halliday, 1966). He theorised that these phrases could be conceptualised as words with syntactical relationships that could be distinguished and organised based on their probability of occurring within a particular textual proximity (Halliday, 1961). Halliday proposes two categorical axes for defining words, the paradigmatic and the syntagmatic, for these proximal co-occurring word relationships. These two axes would allow for analysis of highly likely relationships

between words and also enable the prediction of co-occurring words. This work was pivotal in its conceptual framework that introduced the idea of a multi-word phrase having a core 'node' and associated collocates that would occur in a set space, known as the *span* (Halliday, 1966).

Sinclair's pivotal perspective on multiword phrases has been potentially the most influential. He proposed two key principles for understanding how meaning is conveyed – the open-choice principle and the idiom principle (Sinclair, 1987; 1991). The open-choice principle posits that for every choice of word in a given text there are many potential individual lexical options, that can be plugged into these open meaning 'slots' (Sinclair, 1991). However, this 'slot and filler' model is constrained by the 'idiom principle' – that proffers that there are also a variety of formulaic sequences that could be slotted in as wholes, making up one slot insertion (Sinclair, 1991). In Sinclair's opinion, the idiom principle is the primary of the two and has dominance over the open-choice principle; therefore written language is first shaped by the relationships in multiword units (Sinclair, 1991).

Sinclair also built upon Halliday's concept of collocational span, showing that 95% of collocates occur within 4 spaces of the node, as in within 4 words before and after (Jones & Sinclair, 1974). These findings came from the pivotal COBUILD project, a computational lexicography research study in which many of Firth's original theories were examined using a corpus language database. Sinclair was also first to categorise the patterns of a different lexical items and their corresponding collocates – a practice that spawned the more recent technique of 'lexical bundle analysis' in which researchers can search a corpus for strings of words with formulaic characteristics (Barfield & Gyllstad, 2009).

Generative perspective

There is such a broad spectrum of opinion on how to define this unique facet of language, from definitions that disregard elements of syntax entirely and focus only on probability and node-collocate relationships (corpus linguistics) to phraseological definitions that crucially stipulate that one component of the phrase must be both fixed and restricted (Howarth, 1998). Another perspective is the generative perspective, which holds that these chunk units are not fully compositional (in that they are not created in a systematic matter from separate parts) hence not 'generated' and consequently not

compositional. This framework defines a collocation as having a component that is used in a 'specialised manner' – most often meaning a figurative manner (Howarth, 1998). Phrases may frequently co-occur and be considered a unit or a collocation by a corpus definition alone (such as 'cut a hole' or 'drink a beer') but lack this specialised restriction that something like 'catch a cold' has (Nesselhauf, 2003; Howarth, 1998). It is this figurative characteristic that presents learners with so much difficulty, working as a feedback loop to reinforce predictable combinations that co-occur due to fixedness, consequently becoming more 'collocationally linked' (Erman, 2009).

2.2.3 *Collocations versus idioms*

Formulaic language encompasses fixed *idioms* and *routines*, as well as less fixed *collocations*. Definitions vary – some define idioms as a type of extremely fixed, opaque collocation (Barfield & Gyllstad, 2009), others separate the two, arguing that a collocation by definition allows variation of the parts depending on context and is more syntactically free than an idiom (Erman, 2009). For example, a collocation might be "burst into tears", which can be adjusted to many other forms, including "almost burst into tears" "bursting into tears" "bursting into giant tears" "burst into happy tears and laughter" or "do nothing", "does next to nothing" "did more than nothing" "don't do nothing!", while an idiom would be more fixed and more opaque, for example: "at the drop of a hat", or "burn a hole in one's pocket" or "have one's chickens come home to roost". Idioms have been conceived of as "complex bits of frozen syntax, whose meanings cannot be derived from the meaning of their constituents, that is, whose meanings are more than simply the sum of their individual parts." (Nattinger & DeCarrico, 1992, p. 33). Cowie (1998) segregates idioms from other types of formulaic phrases (collocations, chunks, fixed speech, etc.) by defining idioms as 'semantically specialized', while the latter are simply 'pragmatically specialized' (Cowie, 1998, p.132-133). Cain and colleagues (2005, pg. 65) define an idiom as 'a figurative expression that usually can be interpreted literally but that takes a non-literal meaning when used in a specific context.' Idioms depart from their piece-by-piece meanings and are metaphorical in some manner, making them incomprehensible if analyzed individually or without etymological understanding (Moon, 1992). Cain and Towse (2008) add the definition parameter that an idiom can also have a literal meaning, causing additional difficulty for young learners; an idiom like 'break the ice' can in fact

mean to break ice or to start a conversation in a socially awkward setting. Cain and Towse (2008) propose that idioms can be analysed by examining alternate meanings of the word components involved and the phrase's non-literal meaning. Transparent idioms are defined as those that are *semantically analysable* – there is a clear connection between the figurative and the literal meaning, while opaque idioms cannot be guessed at from the parts (Cain & Towse, 2008). Erman (2009) emphasizes the lack of fixedness and lack of syntactic restriction in a collocation, resulting in many more potential structures and collocate nodes that are possibly with an idiom. However, despite increased flexibility, Erman (2009) asserts that collocations are stored and processed holistically, similar to idioms.

2.2.4 *Opacity, transparency and formulaic language*

Another point of contention within the discussion of formulaic language is the role of transparency; whether or not a phrase must have a completely holistic, un-analysable meaning to be considered formulaic, or if frequency of co-occurrence is qualification enough. Cowie (1998) argues that there is no clear distinction between transparent and non-transparent items and instead posits that there should be a continuum of analyzability for formulaic language. Wood (1986) also argues for a sliding scale, with idioms falling somewhere between totally transparent and non-transparent, but closer to the latter. Creating clear delineations on this continuum naturally presents a challenge. Wray (2002) points out that formulaic sequences are fluid and can be componential when used in one setting and holistic in another, making it difficult to set a dividing line after which phrases are totally non-transparent. Nesselhauf (2005) builds on this idea by presenting a transparency continuum for multi-word phrases (idioms and collocations), categorizing word combinations from transparent to totally opaque and fixed. According to the scale, a transparent item is one in which the constituent parts are all used with their first, literal dictionary meaning, a semi-transparent item is one in which one of the parts (for example the verb in a verb + object phrase) is used in a figurative manner and no longer retains the first dictionary meaning, and a completely non-transparent item is one in which none of the parts carry their original, literal definition and must be viewed as a whole (Nesselhauf, 2005). This scale will serve as a basis for the measurement instrument developed for this study, and will be discussed further later in this thesis.

2.2.5 *Formulaic language and theory: Novelty versus formulaticity*

For many years the emphasis in linguistics was on the potential to create novel language (Chomsky, 1965). Humans clearly have the ability to process and analyze grammar, as well as produce new and creative utterances. It has been argued, however, that this is not actually what our brains 'do', most of the time, nor do our brains 'want' to do this. Instead, we would much prefer to rely on our wide and varied arsenal of prefabricated chunks that we keep in storage for fast, unanalyzed use (Wray, 2002). Corpus linguistics had given us new, post-Chomskian insights into the company different words like to keep and the patterns that emerge in this usage. These patterns have been described by Sinclair as the 'unrandomness' of language (Sinclair, 1991). Corpus evidence has shown that certain words clearly co-occur with each other more than could be predicted by chance. The prevalence of these co-occurrences in our language must then be reconciled with the theories of grammar that heavily emphasize novelty.

Some models do incorporate ideas of formulaic language well, or at the least are not in conflict with it (Tomasello & Brooks, 1999). However, other theoretical reference points are less accommodating. Theories of a single processing system for grammar are particularly complex when accounting for formulaic language. If, as these theories posit, the grammar system is entirely analytical and language structure is built around rules, which are relied on for the creation of utterance, then holistic formulaic processing presents an almost wholly opposite position (Wray, 2002). While it is true that we can produce entirely new language through analysis and calculation, doing so would be exhausting and sometimes result in unnatural sounding expressions (Pawley & Syder, 1983). There are a multitude of perfectly grammatical sentences that can be formed according to the rules and patterns of English, only a smaller set of these are commonly used (Pawley & Syder, 1983). We can comprehend and create novel utterances but we do not utilize this ability as often as we use ready-formed expressions. Wray attests that the significance of this 'capacity for novelty' is given an undeserved amount of attention in modern linguistic theory, given that most of our daily language is made up of prefabricated chunks (Wray, 2002, p. 11). This also indicates that opaque phrases must, at the very least, be an exception and stored holistically in the mental lexicon, as they are not generated at the point when need (Wray, 2002).

2.3 *Multi-word phrases and first language acquisition*

2.3.1 *Learning English*

There has long been evidence that young language learners can treat several words as a unit. Children are exposed to long streams of speech containing multiword chunks, some of which occur with great frequency. From these language streams children fish out strings of language and store them, even before necessarily understanding the pieces that make them up (Wray, 2002). The idea of the 'word' as a distinct unit is an adult construct and a child can happily use a holistic chunk without being able to use any of the unique parts in different combinations (Locke, 1997). Brown noticed this phenomenon in 1973 and concluded that such strings, when used by a very young learners still grappling with learning language, must in fact be 'generated by some simpler mechanism' – and noted that it seemed unlikely the child was constructing them from individual parts already known. Cruttenden (1981) proposes that chunk production is an essential phase for language acquisition, allowing the child to first learn items and then the system that organises them. Children keep these chunks as unanalysed wholes. These particular formulaic chunks that children acquire as they are first learning English follow a developmental trajectory; first, the child will use under-analysed strings that they take from the input around them, and second, the child will create set phrases and chunks to keep for later use (Wray, 2002). The role of these constructions is assumed to be to ease the processing burden (Wray, 2002). Tomasello and Brooks posit that children are probably learning language from both entire chunks and individual words, are able to decipher meaning both 'part to whole' and 'whole to part', and use holistic and analytical strategies to acquire language (Tomasello & Brooks, 1999, p. 166).

Under-analysed strings are sets of multiple words or morphemes with complex grammar or lexical knowledge beyond the child's understanding. The child pulls these chunks of language from the linguistic input around them and uses them as a whole. Children will imitate these strings with varying degrees of accuracy. Children are also taught a multitude of songs and poems, often relying on music or speech patterns to aid memorization. One commonly occurring example of this are the phrases repeated to children in songs or stories that the child may retain without actually comprehending – for example 'Goodnight bowl full of mush' from the rhythmic bedtime poem Goodnight Moon or ' How I wonder what you are' from the song *Twinkle Twinkle Little Star*.

Children often learn entire songs and poems totally unanalysed and rarely is it expected that they would actually 'know' all the lexical contents of the songs and poems they have learned (Wray, 2002). They can remember and repeat these phrases without necessarily knowing or considering where or what the word components in them are (Gleitman, Gleitman, Landau & Wanner, 1989). Peters (1982, p. 15) notes that 'there is no evidence that what the linguist recognises as unanalysed routines are recognised and stored by the child any differently from items that we would call single words.' It is particularly evident that a child is keeping an entire phrase as a unit of acquisition when the children produces, or indeed mis-produces the whole, failing to separate sounds at the appropriate points and running words together to approximate the whole.[1] In the case of song or chants, a reliance on rhythm or cadence can lead to some words ending up 'mis-learned', as weaker sounds can be left off or run together (Wray, 2002).

Probably the most commonly used unanalysed strings in the lives of English speaking children are ritualised routines – particularly those associated with politeness and social norms (Wray, 2002). These routines are emphasized by adults who talk about them, teach them explicitly and elicit them specifically. Wray (2002) points out that by demanding exact production of these 'magic' strings even serves to reinforce to children that some phrases cannot and should not be broken down.

Children acquire individual words as well as phrases; the size of a 'unit of acquisition' depends on the individual child and on the developmental stage (Peters, 1982). Rather than a 'one-word stage', it has been posited that a more accurate description would be a 'one-unit stage', in which a single long string can be produced as a unit (Peters, 1982). Peters (1982) notes that children begin learning language with both holistic and analytic strategies, only to later rely more on analytic techniques, isolating language down to minimal units like words and morphemes. Brown, (1973, p. 390) notes that the child must eventually shift to segmenting speech into morphemes because these are the 'ultimate units of grammatical rules'.

---

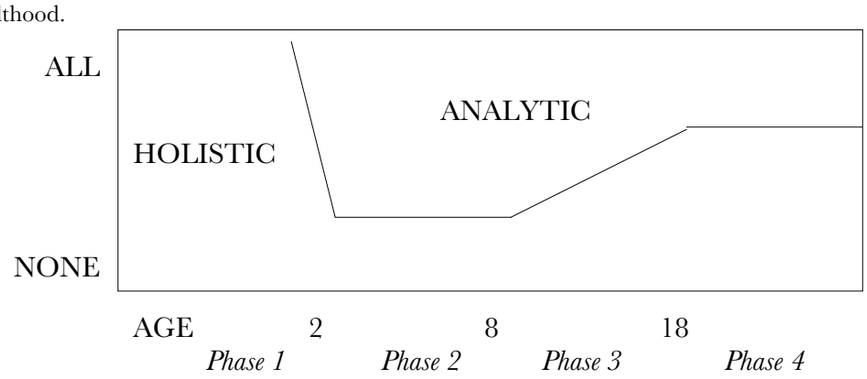
[1] In some particularly cute examples, the meaning of the string has not been fully understood by the child but is still used effectively. Wray (2002) gives the example of a girl less than two who picked up on the fact that when her mother said 'Time for a cup of coffee' the biscuits would also come out and the child would get one. She learned that approximating the phrase, saying ['taipə'kɒpə'kɒpɪ] produced the same result as asking directly for a biscuit – perhaps was even more successful! She had not separated out the individual sounds or words in the phrase. Children will also attempt to fit an unanalysed string into another language frame, again producing charming results. For example, the following from Wray (2002, p. 107); Mum: 'We're all very mucky.' Child: 'I all very mucky, too.'

Formulaic language is present in early child language in multiple ways – in some cases children will attempt to repeat a heard phrase before having the grammar or individual vocabulary words needed to generate such an utterance from scratch. In other cases, children will specifically learn a whole phrase for a particular purpose (songs, politeness routines). Multi-word phrases are hypothesized to play a role in language acquisition, learning the wholes and the pieces, as well as for achieving social and interactional ends.

2.3.2 *Encountering novel figurative language later in the learning process*

Young learners can retain chunks of language as wholes, without analysing them or breaking them apart, and there is evidence that adults too process chunks of language as wholes (Barfield & Gyllstad, 2009; Wray, 2002), see section 2.4. This then begs the question of what occurs when a young English speaker encounters a novel opaque phrase, after the 'adult construct' of the individual word has been taught to them, and how the meaning is understood. Nippold's (1998) *meta-semantic hypothesis* of figurative language comprehension proposes that semantic analysis plays a key role in understanding non-opaque idioms, as the literal meanings provide clues regarding the figurative meaning. When encountering idioms, Nippold proposes that adult learners can connect the components of a novel opaque phrase to the idiomatic meaning by performing some form of semantic analysis. At what age this analysis begins to occur remains a point of debate, although most research points to it being an early skill that develops through schooling years (Levorato & Cacciari, 1995; Wray & Perkins, 2000). See Figure 1 for a representation how the approach to novel idioms is hypothesized to change over time. The literature generally posits that children initially process formulaic language holistically, followed by an analytical period during which formulaic language use diminishes, until the teen years when the learner will return to more a more, but not entirely, holistic approach (Gibbs, 1991; Wray & Perkins, 2000).

Figure 1: Relative proportions of holistic and analytic involvement in language processing from birth to adulthood.



(Adapted from Wray & Perkins, 2000)

2.3.2.1 *The Global Elaboration Model*

The Global Elaboration Model, or GEM, developed by Levorato and Cacciari, (1999), posits that the same skills that make reading and listening comprehension possible also enable figurative language comprehension. Comprehending figurative language is accounted for by lexical development and general language abilities, which the (typically developing) child develops to elaborate language, process it and fit it in with the language information that came before, just as is needed for non-figurative language (Levorato & Cacciari, 1995). According to the GEM, children begin by analysing incoming language part by part, word by word, focusing on piecing together meaning within a sentence or a chunk of language, however, as their language development progresses, they acquire the skills needed to infer less obvious meanings of figurative language using surrounding context, rather than an literal word-by-word interpretation. According to the authors, these skills, essential for processing and retaining figurative language, develop between the ages of 7 and 11 (Levorato & Cacciari, 1992, 1995,1999). By the end of primary school, children have usually acquired a level of figurative competence that no longer requires an individual strategic approach to deal with figurative language (Levorato & Cacciari, 1995).

The GEM posits that the idiom itself will be more easily understood the better it can be connected to a literal interpretation of its meaning – or *semantic analysability*. Semantic analysability refers to the connection between the literal and figurative meanings of an idiomatic expression – for example, 'to get away with murder' is easily connected to meaning of avoiding punishment, while 'kick the bucket' is less obviously understood as meaning to die (Cain et al., 2005). There are also 'quasi-compositional' idioms that have some identifiable meanings and produce a figurative whole that might partially map on to the component parts, for example 'spill the beans' (Levorato & Cacciari, 1995). It is worth acknowledging that semantic analysability may be subjective; what appears to be an obvious connection to one adult might not be interpreted in the same way by another adult or seem entirely different to a child.

An unfamiliar idiom can be approached in 2 ways; semantically analysed or guessed at from context, and these 2 strategies are tied to different stages in language development (Levorato & Cacciari, 1995). The traditional view of idiomatic expressions is that these

chunks are processed as wholes – giant lexical units, and therefore when a young child attempts to analyse an idiom literally, this is read as the child failing to pull the whole from his or her mental lexicon and consequently breaking the string apart (Nippold & Martin, 1989). A more advanced learner will abandon this compositional analysis strategy and learn the expressions as wholes (Levorato & Cacciari, 1992). The GEM posits that for semantically analysable expressions a child initially uses the meaning of the idioms' components to guess at a figurative meaning and progresses from piecemeal elaboration to looking for a larger meaning for the whole phrase(Levorato & Cacciari, 1995). As learners acquire more language and develop inferential abilities they will try to make the meaning of the phrase consistent with the whole text and discard mismatched literal interpretations (Levorato & Cacciari, 1995).

There is evidence that children do approach new idioms in the manner that the GEM describes, particularly evidence that idioms are better understood in context (Cain et al., 2005; Gibbs, 1987, 1991; Levorato & Cacciari, 1995; 1999; Nippold & Martin, 1989; Nippold, Moran & Schwatz, 2001; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995). Levorato and Cacciari (1999) administered a multiple choice idiom comprehension task to Italian primary school children. They found that 9-year-old learners could guess at the meaning of a transparent idiom even without a surrounding context but a 7 year old could not (Levorato & Cacciari, 1999). The 9 year old participants had much higher comprehension for transparent idioms that were semantically analysable than the 7 year old learners, while younger learners showed increased comprehension of semantically analysable idioms within a surrounding short-story context than opaque idioms (within a context). The context was read to the participant and presented in text form. However, the task used might have also disadvantaged younger learners who would have weaker expressive language abilities and might therefore struggle to explain the meanings of the idioms, not because they couldn't guess at the meanings but simply because they didn't have the language integration and disambiguation skills to do so. Levorato and Cacciari (1999) posit that while skills at semantic analysis are early developing, these skills become stronger and more important as children develop. They argue that the role of context may be a secondary check, that learners will first try to analyse the phrase and then compare and narrow this meaning to fit the context within which it appears (Levorato & Cacciari, 1999).

Research has also been done with children younger than age 7. Gibbs (1987) found that 5 year olds could much better explain the meaning of transparent idiomatic phrases than opaque phrases, but again, this less accurate performance by the 5 year old participants could be related to the children's expressive language abilities. Gibbs (1991) also found that learners between ages 8 and 9 could much better understand transparent idioms than learners between 6 and 7. One element that seems especially important for younger learners when compared to older learners is the presence of context (Gibbs, 1987, 1991; Levorato & Cacciari, 1992). Young learners find it much easer to understand the meaning of an idiom presented in a context, from learners as young as 5 years old to teenagers. Gibbs (1991) asked participants between the ages of 5 and 9 to explain the meanings of transparent and opaque idioms, with and without context and found that although all participants' results improved with a surrounding story context around the target phrase, children ages 8 and 9 used semantic analysis when attempting to analyse phrases without a context, with more success the older they were (8 year olds had correct analysis 37% of the time, 9 year olds 42%).

Overall, it does appear from the literature that the ability to guess at the meaning of a figurative phrase from surrounding context arises first around age 7 and by age 9 children are also able to guess based on semantic analysis, even without a context. One potential critique of the GEM is that there is so much evidence that frequently used formulaic language is processed like giant lexical units, so while the theory may explain how unknown idioms are tackled by young learners, it does not address how already stored multi-word units are acquired or their role in language (Arnon 2009; Wray, 2002). Additionally it assumes that the connections between literal and figurative meaning are standard and equally clear to all children.

2.4 *Multi-word phrases in the mental lexicon*

2.4.1 *Theoretical models*

In order to consider the role of multi-word phrases as vocabulary items in the mental lexicon, we need a theoretical basis for understanding how they are stored, organized and processed in the brain. Early researchers such as Swinney and Cutler (1979) and Bobrow and Bell (1973) proposed that each idiom has its own lexical entry in the mental lexicon that is in turn connected to related lemma. The individual word

components of the phrase are not considered to have an impact on the representation of the unit (Swinney & Cutler, 1979). The theory was supported by evidence that participants could much more quickly decide that an idiom was a valid English phrase than a control literal expression (Swinney & Cutler, 1979). However, subsequent findings contradicted certain aspects of the non-componential conceptualization. Idiom phrases have been found to be susceptible to semantic priming for the words contained in the idiom (Peterson, Burgess, Dell & Eberhard, 2001). For example, presenting the words 'kick the...' was a successful primer for both a literal word item ('ball') as well as for an idiomatic ending ('bucket') (Peterson et al., 2001). However, Peterson and colleagues also found that what they refer to as 'concreteness', or transparency, of an item had no effect on processing speed among presented idiomatic sentences, indicating that the participants had not processed the literal meanings of these figurative phrases (Peterson et al., 2001).

More recently, hybrid models combining compositional and non-compositional elements have also been introduced (Titone & Connine, 1999). One such hybrid model has been proposed by Cutting and Bock (1997) who suggest that idioms are separate lexical entities on one level of the mental lexicon and do have their own 'lexical concept node'. This node can activate other related lexical concepts, for example other multi-word phrases with related meanings. However, activation is not limited to these conceptual connections. Activation also spreads along the lemma networks connected to the individual lexical items in the phrase (Cutting & Bock, 1997). A phrase such as 'kick the bucket' would not only activate concepts regarding death and other set phrases relating to this concept ('meet your maker') but also individual words related to 'kick' and 'bucket', such as 'ball' or 'pail' (Cutting & Bock, 1997). Research has provided support for this representation of multi-word phrases in the mental lexicon. Sprenger, Levelt and Kempen (2005) found that idioms could not only be primed on the conceptual level, but also with phonological word primes – providing support for a single, lexical entry. They also found that the final word of an idiom served as a primer for single phonological and semantically related words, indicating that the literal meaning of the component word is activated in the mental lexicon – supporting a more hybrid model of multi-word phrase representation in the mental lexicon (Sprenger et al., 2005). It seems that the dilemma of how multi-word phrases are stored in the mental lexicon is more complex than either of the original two proposed

models can alone account for and a hybrid must be considered, at least until future research can further unravel the issue.

2.4.2 *Processing multi-word phrases*

Research evidence from neuroscience has shown that figurative language and idiomatic language has a distinct place in the mental lexicon, very literally! Figurative language is one of the few language related processes that evidence indicates is processed in the right cerebral hemisphere, rather than the left, which is more generally associated with most aspects of language (the right also processes irony) (Mashel, Faust, Hendler & Jung-Beeman, 2008). Studies have found that patients with only right hemisphere damage struggle to match idioms with corresponding pictures representing meaning, especially in comparison to patients with equally severe left hemisphere damage (Kempler et al., 1999). Functional Magnetic Resonance Imaging (fMRI) research has shown that typical participants being asked to consider ambiguous idioms (that is idioms that could be interpreted literally or non-literally) will process the literal meanings faster than the idiomatic meanings (Mashal et al., 2008) but that these meanings will activate different regions of the brain: a literal interpretation will activate the left hemisphere, and a figurative interpretation will activate the right and left. The same participants would have exclusively left hemisphere activation when asked to process unambiguous, literal sentences (Mashal et al., 2008). This provides supporting evidence that there is something unique about figurative language and that models of lexical processing for single words or literal language should not be applied directly to formulaic language, particularly not figurative formulaic language, which the brain may address in a different manner.

Single word vocabulary is well documented as playing a significant role in reading and the impact of the former on the latter has motivated a substantial amount of reading research. Multi-word vocabulary is recognized as extremely prevalent and different from single word vocabulary from both a grammatical and a neurological perspective, as discussed in the previous section. This evidence of multi-word vocabulary as distinct consequently calls for a distinct examination of the nature of this aspect of vocabulary knowledge and its relationship to reading – findings from investigations of single word vocabulary cannot be assumed to apply to multi-word items.

2.5 *Theoretical models of reading*

The simple view of reading

Potentially the most well known (and well-researched) theoretical model of reading is the 'simple view of reading' posited by Gough and Tunmer (1986). This theory explains reading as a combination of two skills: language comprehension, as measured by listening comprehension, and word decoding ability. This relationship is often presented as RC=D x L. This model has been supported by numerous empirical studies (Hoover & Gough, 1990; Joshi & Aaron, 2000; Savage, 2001; Tunmer & Hoover, 1992). Within this work, researchers will often use listening comprehension as a proxy to estimate language comprehension. Findings from these studies have shown that listening and decoding together do account for a great proportion of variance in reading comprehension outcomes (Gough, Hoover & Peterson, 1996; Hoover & Gough, 1990; Johnston & Kirby, 2006; Savage, 2006).

While these skills clearly play some role in reading comprehension, there is still an open debate as to precisely how decoding and listening comprehension combine to result in reading comprehension. In a sense, the debate is over what kind of equation should the model be: addition, multiplication or a combination of the two. Gough and Tunmer (1986) initially posited a multiplication equation – reading comprehension was the product of the two components. This would mean that if a reader were a good decoder with weak language comprehension then reading comprehension would be quite low. It would also be this way in the case of the inverse, high language comprehension with low decoding skills.

This multiplication equation model is supported by Hoover and Gough's (1990) study with approximately 250 Spanish-English bilingual children who were tracked longitudinally from grade 1 through 4. Children were given an annual assessment battery measuring language, cognition and reading ability. Over the five-year duration of the study participants were enrolled in bilingual schooling that eventually transitioned into only English and received English and Spanish reading instruction. Data from the annual assessments was analyzed using hierarchical regression modelling and results from this analysis showed that decoding skill and linguistic comprehension did each account for a significant amount of variance in reading scores. In addition, the product of these two scores was even better able to predict scores in reading that either variable alone. Findings

showed that indeed, reading comprehension was a product of the two skills (Hoover & Gough, 1990).

However, contrasting research has found this model to fall short in other settings. Savage (2006) found that a sum equation of the two components is a much better fit than the product model for teenagers with weak reading skills. Chen and Vellutino (1997) found that the product of the two skills did not explain any unique variance in reading comprehension scores when controlling for using the sum. D + LC was just as good at predicting reading outcomes as D + LC + (D x LC). This additive model differed from the original in that it would mean that strength in one skill component could be compensatory and mitigate weakness in the other – in the end resulting in reading comprehension success. This model is also well supported by work with high functioning adults with developmental dyslexia who have overcome very weak decoding skills and have good reading comprehension as a result of strong linguistic comprehension (Wilson & Lesaux, 2001). Hatcher and Hulme (1999) investigated language ability among L1 English children in a remedial reading instruction program and found a correlation between vocabulary and reading comprehension, but no connection between reading comprehension and word reading accuracy. Similarly, in a two year study with 90 L1 English children, Muter and colleagues found that reading comprehension relied on semantic and syntactic abilities while word recognition was more dependent on phonological awareness (Muter, Hulme, Snowling & Stevenson, 2004).

Understanding reading becomes more complex when viewed longitudinally. Some evidence shows the simple model might need to adjusted as readers grow older. While it does appear that young readers initially rely heavily on word decoding skills, research with older readers has shown that reading comprehension scores are better explained by listening comprehension and the contribution of single word reading skills is minimized (Adlof, Catts & Little, 2006; Gough, Hoover & Peterson, 1996). Chen and Vellutino (1997) reported that the skills needed for reading and the relationships between these skills change over time. Young learners rely heavily on decoding for reading comprehension, whereas for older learners language comprehension is a stronger determinant of reading ability (Chen & Vellutino, 1997). Vellutino and colleagues posit that higher level skills,

such as vocabulary knowledge and syntax, become more significant as reading ability develops (Vellutino, Tunmer, Jaccard, & Chen, 2007).

Despite widespread supporting evidence and acceptance, the simple view of reading is not without criticism. One point of contention is how to define and operationalise the construct of word decoding. Hoover and Gough (1990) used phonological decoding (the ability to translate printed graphemes onto their phoneme counterpart). Ouellette and Beers (2010) instead argue in favour of word recognition (the ability to recognize the entire word) for a more complete representation. In the early stages of learning to read, straightforward phonological decoding is a good indicator of this skill, however, words that deviate from the orthographic-phonological code (yacht, stomach-ache, biscuit*)* must be learned and recognized as sight words (Harlaar, Cutting, Deaser-Deckard, DeThirne, Justice, Schatschneider, Thomson & Petrill 2010). As learners become stronger word decoders, their mental lexicon representations for these words expand to include phonological and semantic details and become sight words (Harm & Seidenberg, 2004; Perfetti & Hart, 2002). This means that word decoding can happen through a variety of means: phonological decoding or whole word reading. Ouellette and Beers (2010) argue that using word recognition instead of phonological decoding will improve the predictive value of the simple view of reading for reading comprehension outcomes. However, word recognition can involve other skills, for example orthographic and semantic knowledge, the latter of which could actually be considered an aspect of listening or language comprehension (Kirby & Savage, 2008).

There is evidence to support that many skills are involved in reading; Ouellette and Beers (2010) found phonological decoding and word recognition both made unique contributions to reading comprehension among 67 students in year 1 and 56 in year 6. All of the participants except for one were monolingual English speakers. Participants were given assessments of phonological awareness, decoding, word recognition, listening comprehension, oral vocabulary and reading comprehension. Ouellette and Beers (2010) posited that as children become stronger word readers (therefore among the year 6 learners) decoding would become a less important factor for reading comprehension and would account for less score variance. At this point, the contribution of vocabulary would increase. They also hypothesized that oral vocabulary knowledge would be a stronger

contributor to reading than 'linguistic comprehension' as measured by listening comprehension. Data was analyzed using hierarchical regression analysis and found that the contribution of oral vocabulary to reading outcomes was greater in year 6 than in year 1 and the contribution of decoding was less in year 6 than in year 1. In year 1 oral vocabulary did not predict reading comprehension beyond the other testing measures, but in year 6 it predicted a unique amount of variance in reading comprehension scores even when accounting for all other measures (Ouellette & Beers, 2010). The contribution of both oral vocabulary and listening comprehension were significant for reading outcomes, though the effect of listening comprehension was much smaller than that of oral vocabulary. These findings support Ouellette and Beers' hypotheses and provide evidence that the skills that contribute to reading and the amount that they contribute changes depending on the age and developmental stage of the reader. Considering this, it may be that multi-word vocabulary contributes more to reading comprehension at a later stage, for example, at a point when learners are reading texts containing them. This might also explain why they have been overlooked by research focusing on the skills needed for learning to read, as opposed to those important after decoding has been mastered - much as oral vocabulary has been shown to overtake decoding in import during later schooling years.

2.6 *Reading fluency*

When children first learn to read an orthographic language like English they are usually taught how to map orthography onto spoken phonemes, the goal being that eventually, after having read a word enough times, the child can read the same word automatically – rapidly, accurately and fluently – with correct expression and understanding (Shaywitz & Shaywitz, 2008). It has been pointed out by many researchers that teaching children to map letters onto sounds does result in accurate reading but not always fluent reading in English (Shaywitz, 2003). The topic of reading fluency and its importance is yet another area of contention in the complex discussion of the skill of reading.

Reading fluency has been defined as the process that occurs when individuals read accurately and rapidly while comprehending the material read (Lyon et al., 2003). Newer conceptualisations of reading fluency have emphasised the multi-componential view – the view that reading fluency is the result of the co-ordination of multiple-reading related

elements such as orthographic processing, phonological processing, syntactic processing and semantic awareness (Katzir et al., 2006; Wolf & Katzir-Cohen, 2001). Reading fluency can be thought of as the product of competency in these cognitive reading sub-skills (Benjamin & Gaab, 2011). This componential view of reading is relatively new, earlier frameworks considered fluency to be oral reading, speed and accuracy (Torgesen, Alexander, Wagner, Rashotte, Voeller & Conway, 2001). This newer perspective accurately allows reading assessment to consider lower-order skills essential for word recognition, for example phonological processing (Benjamin & Gaab, 2011). Proficiency in these lower order skills allows the reader to potentially automatize certain reading skills and read with greater speed and accuracy (Wolf & Katzir-Cohen, 2001).

The multi-componential view is well supported by research evidence. Reading disorders with fluency deficits respond only minimally to interventions focusing on lower-order skills like phonemic awareness (Lyon et al., 2003) and fluency in the lower level skills is not predictive of overall text fluency (Katzir et al., 2006). Instead, evidence points to multiple reading sub skills that predict word reading fluency (orthographic processing, phonological processing and rapid naming), while other skills are better predictors of overall text reading (reading rate, accuracy and comprehension) (Katzir et al., 2006). This is further supported by findings that young learners with phonological processing deficits and rapid naming deficits exhibit differing degrees of reading fluency deficit (Katzir et al., 2006).

2.7 *The role of vocabulary in reading comprehension*

One essential element for reading comprehension is vocabulary knowledge. Native speakers have grasped roughly 90% of the phonology, syntax and morphology they will ever acquire by the age of four, but, by contrast, a 6 year old only knows about a sixth of the number of words s/he will know by the end of secondary school at age 16 (Anglin, 1993). Between birth and the age of 6, children learn approximately 14,000 words (Dolloghan, 1994), although this number does not include multi-word phrases. Vocabulary learning is relevant at every age of schooling and different age groups have different patterns of word learning. Research with English L1 children suggests that vocabulary knowledge at age 6 is a strong predictor of later reading comprehension and vocabulary size in first grade can account for more than 30% of the variance in reading comprehension

by eleventh grade (Scarborough, 2001; Cunningham & Stanovich, 1997). Roth, Speece and Cooper (2002) found strong correlations between results from an oral vocabulary task with 5 to 6 year olds and reading comprehension measures taken both one and two years later. Additionally, the larger the child's vocabulary is the faster they will learn subsequent words (Nation, 2001). Biemiller (2005) reported that children in the lowest quartile for vocabulary knowledge added an average of 570 new root words in an academic year, while children with an average sized vocabulary added approximately 840 or more. As previously mentioned in this chapter, it has been argued that there are many different skills that contribute to reading and these skills are weighted differently as development progresses (Vellutino et al., 1994). Vocabulary knowledge is one of many important factors for understanding reading comprehension and reading achievement outcomes.

Vocabulary is generally a strong predictor of reading and listening comprehension (Anderson & Freebody, 1981; Vermeer, 1992) and reading in turn has been shown to increase vocabulary (Anderson & Freebody, 1985; Stahl & Nagy, 2005; Taguchi, 1997). Not accessing this 'snowballing' of more reading and resulting greater ability can also produce an opposite phenomenon for poor readers, known as *Matthews effects*, wherein weak readers have less practice, less exposure and consequently fall even further behind (Torgensen, 2000). As a result, it becomes increasingly difficult for struggling readers to catch up or close the gap between themselves and better performers.

This deficit accumulates over the years and by middle schooling, some weak readers will have such an impoverished sight word vocabulary that they will struggle to access grade-appropriate texts (Torgensen, 2000). This is an even greater risk for students with fewer opportunities to acquire background knowledge of vocabulary and less exposure to English print, for example, learners with EAL. This phenomenon has been called 'The fourth grade slump' by Chall, Jacobs and Baldwin (1990) who pointed out that by late elementary school, reading demands increase and content becomes progressively abstract and complex (Chall et al., 1990; Kieffer & Lesaux, 2007). Children who do not successfully acquire basic word reading skills in the early years typically then 'decelerate in comprehension' as instruction transitions from 'learning to read' to 'reading to learn' (Chall et al., 1990).

2.8 *The Lexical Restructuring Model*

Vocabulary knowledge has a strong influence on many important pre-reading skills: phonological awareness and orthographic knowledge and word identification (Chiappe, Chiappe & Gottardo, 2004; Nation & Snowling, 2004). One theoretical perspective on this relationship is the Lexical Restructuring Model (LRM) (Metsala & Walley, 1998). This model posits that as the lexical vocabulary of a child grows, the child becomes able to separate out smaller and smaller phonemic units. According to the LRM, young children start out learning whole words that they can eventually break apart to the syllabic and then phonemic level (Metsalla & Walley, 1998). This breaking up into sounds happens to each word according to frequency, age of acquisition and neighbourhood density (Walley, Metsala & Garlock, 2003). In order to learn to read, children must first understand the phoneme components of a word in their mental lexicon to then map it onto a grapheme representation. In addition to contributing to pre-reading skills, vocabulary knowledge is thought to benefit word identification skills (Levelt, Roelofs & Meyer, 1999). The mental lexicon has both a semantic and phonological entry for each vocabulary item – young readers with a weak phonological entry for a particular word, or many words, might struggle to map that phonological code onto an orthographic pattern (Levelt et al., 1999).

When discussing the contribution of vocabulary to reading it is important to disambiguate *receptive* and *expressive* vocabulary. Receptive vocabulary, also known as passive vocabulary, refers to the words an individual can recognize and expressive vocabulary, or active vocabulary, are words that the individual can produce. These two vocabulary categories differ in the skills and processes they involve and therefore may relate to reading differently as well. For example, expressive vocabulary requires accessing more semantic knowledge than receptive vocabulary does and would therefore be expected have a greater impact on word recognition (Wise, Sevcik, Morris, Lovett &Wolf, 2007). In a study with 13 American English speaking poor readers (below 26th percentile in word identification) and 49 good readers (above 29th percentile) found that the good and the bad readers differed significantly in expressive vocabulary but not in receptive (Chiappe, Chiappe & Gottardo, 2004). They also found expressive vocabulary to have a stronger correlation with a number of reading skills (phoneme deletion and blending, word identification and non-word identification). The authors interpret this to mean that expressive vocabulary relates more strongly to these reading and pre-reading skills and is a

more phonologically representative task than receptive vocabulary. These findings could clearly be confined to this age group (years 1, 2 and 3) and might also have been impacted by the authors' decision to combine this spread of ages into one participant group. This choice to combine the three years could potentially confound findings, given what we know regarding the differential contributions of various linguistic skills as reading develops over time. Readers in year 1 might be very different from those in year 3. McBride-Chang and colleagues (1997) found that among 142 American kindergarteners, kindergarteners' expressive vocabulary predicted performance on a phoneme deletion task, as well as the way performance on the task developed, again lending support to the LSM theory that vocabulary influences pre-reading skills.

Findings are not entirely homogeneous. Nation and Snowling (2004) tested 72 students, first at age 8, and again at age 13. Oral language skills (a combination of expressive vocabulary, listening comprehension and semantic skills) at the first testing session accounted for significant variance in reading comprehension as measured at both time 1 and time 2, more than just listening comprehension or expressive vocabulary. Findings showed that expressive vocabulary and listening comprehension predicted a significant amount of variance in reading comprehension scores, specifically word recognition and exception word reading; listening comprehension was the stronger predictor for word recognition. The authors interpret this to mean that a richer vocabulary facilitates reading words that do not match up along standard grapheme-morpheme patterns. Children with weaker language skills also had lower word recognition than predicted from their non-word reading skills and non-word reading and phonological skills did contribute to variance in word recognition at age 8 and 13. There are a few potential issues with this study. Firstly, exception word reading was only measured at age 13, not at age 8; therefore the study cannot control for where learners started in exception word reading and or compare how this skill at 8 might contribute to the skill at 13. It also becomes difficult to be sure that vocabulary at 8 makes in independent contribution to word reading at 13 without being able to control for word reading at 8. Additionally, the study measured only exception words and does show a clear relationship between exception word reading and it may be of interest to know if this relationship also exists for non-exception words. It should also be noted that oral language proficiency was not

measured by a standardized assessment, but through a combination of expressive vocabulary, listening comprehension and a semantic assessment,

Wise and colleagues (2007) conducted structural equation modelling analysis using data from 279 learners between ages 7 and 8 with a diagnosed reading disability and found that both receptive and expressive vocabulary related to general reading skills, (receptive more strongly) and that only expressive vocabulary was significantly related to word identification. These findings differ from those of Chippe et al., but this could be due to differences in sample. Chippe and colleagues had a balanced sample of strong readers and readers. Differences in findings could also be related to the specific measures used by each study. The variation in specific study findings makes clear that 1) vocabulary does play a significant role in reading and 2) that the exact nature of that role has not been fully revealed in the current body of literature. Current explorations of vocabulary may overlook some essential components of that relationship.

2.9 *Vocabulary breadth versus depth*

Previous research involving reading comprehension and vocabulary development has primarily focused on counting the number of words the individual knows, or *vocabulary breadth* (Pearson et al., 2007). However, there are other things the learner must know about these lexical items in order to use them effectively; grammatical functions, register, stylistic characteristics, appropriate contextual usage, idioms and collocations. This aspect of vocabulary knowledge is called as *vocabulary depth*, and refers to the quality of the knowledge (Milton, 2009).

There are fewer studies researching vocabulary depth as independent from breadth, but what has been done points to a strong involvement of vocabulary depth in reading. In a study with 60 year 4 Canadian monolingual and bilingual children, Ouellette (2006) showed that vocabulary depth, as measured by the Test of Word Knowledge (TOWK) word definitions and synonyms subtest, predicted 8% of the variance in reading comprehension scores. Additionally, researchers have hypothesized that more detailed semantic representations in the lexicon (or greater vocabulary depth) in turn leads to faster encoding and quicker phonological code retrieval from the mental lexicon (Ouellette, 2006). This would point to greater word identification with greater vocabulary depth. Vocabulary depth in EAL children remains relatively unexplored, possibly because there

are few available measures for use with children. Some previous studies have relied on measurements of breadth, ignoring depth or assuming that it expands in parallel with breadth, a potential oversight as studies have shown that in second language learners it sometimes does not (Bahns & Eldaw, 1993; Crutchley, 2007; Milton, 2009; Wray, 2002).

2.10 *Multi-word phrases and reading comprehension*

2.10.1 *Reading comprehension and L2 learners*

There is evidence that non-transparent language poses comprehension problems for L2 readers (Cooper, 1999; Liontas, 2002; Martinez & Murphy, 2011). Both Cooper (1999) and Liontas (2002) explored the learning comprehension for opaque idioms and found that the more opaque the item was the more an adult L2 learner struggled to understand it. However, these totally opaque idioms like those used in this task are actually quite uncommon. Martinez and Murphy (2011) presented adult L2 English learners with paired texts containing the same lexical items, one text with the items arranged into multi-word phrases of varying transparencies and the other without. Participant reading comprehension decreased significantly when the texts contained multiword expressions, even though the individual components were known (Martinez & Murphy, 2011).

Interestingly, L2 learners do not seem to always be aware of their own phraseological knowledge gaps (Bishop, 2004; Martinez & Murphy, 2011). Bishop (2004) found that readers looked up the meanings of unknown words significantly more often than they looked up unknown multi-word expressions, leading him to conclude that these phrases go unnoticed by the learner. Martinez and Murphy (2011) asked learners to report their estimated comprehension at the end of each presented text. Learners significantly overestimated their understanding of the texts containing the set phrases, indicating that the expression were either unnoticed or learners incorrectly thought that they had understood them (Martinez & Murphy, 2011).

There are a few potential criticisms of the Martinez and Murphy (2011) study. Firstly, the presented texts were generated for the test and as a result, read as highly inauthentic. The peculiar, potentially unfamiliar writing style of these texts might have influenced learner comprehension or learner perceptions of comprehension. Secondly, the idiomatic phrases included in the texts were manipulated for frequency, but not for

transparency. Since transparency has been found to have an effect on learner understanding (Cain et al., 2009; Cain et al., 2005), presenting texts that contain idioms ranging in transparency without manipulating or considering this variable could present a possible confounding element, an issue similarly faced by Nippold and colleagues (2001) when exploring idiom comprehension in context among children. Even so, this work provides a valuable insight into an under-explored area of vocabulary.

2.10.2 *Reading comprehension and multi-word units for children*

Little research thus far has explored this aspect of language development in native English speaking children and no research has explored it in EAL children, despite the fact we know that these phrases are ubiquitous in daily speech. Not only that, they also occur in primary school curriculum; Nippold (1991) found that between 6 and 10% of sentences in books in American reading programs for 8 to 12 year olds contain idiomatic language. These items are could have an impact on reading comprehension.

Children with learning disabilities have been shown to have delays in pragmatic language skills and the skills needed to untangle figurative meanings from literal meanings. Instead, these learners have been shown to approach language in a piece-meal manner and are less likely to integrate information from a context and use this information to check for overall text cohesion (Cain et al., 2005). Context has been shown to play a key role in figuring out the meaning of a new idiom (Levorato & Cacciari, 1999; Gibbs, 1991), at the very least by clueing the reader in that a figurative interpretation is needed. Consequently, weak readers may have difficulty understanding figurative vocabulary in a written context. Evidence from inconsistency detection tasks (tasks which require the test taker to identify when one proposition in a text contradicts or conflicts with another proposition in the same text) tasks show that strong comprehenders and older children are better able to monitor their understanding of a text and check for overall coherence in a text than young learners or weak comprehenders (Cain et al., 2005).

Levorato, Nesi and Cacciari (2004) conducted a study in which 7 to 9 Italian speaking year olds were presented with idioms in a written context. A standardized battery was used to split learners into thirds (poor, medium and good comprehenders). Error analysis revealed that poor comprehenders were more likely to select an incorrect literal answer choice than the better comprehenders – the authors interpret this to be indicative of

less attention to overall context and reading ability being a key point of idiom comprehension. Poor comprehenders were retested 8 months later for reading comprehension and idiom comprehension – the learners who had improved in reading also improved in idiom comprehension (Levorato et al., 2004). Strong performers were nearly at ceiling on the task, potentially indicating that the task might be too easy. Additionally, learners may have performed well because they were already familiar with the target items. It is unclear if learners were figuring out idiom meanings from context, as the authors intended, or if they were already familiar with the items.

However, there are a few points that the authors fail to address – the learners who were retested 8 months later and found to also have improved in reading comprehension and idiom comprehension might well have encountered more idioms as they progressed with reading skills. They might have also improved in other areas of reading related skills, such as reading speed, and this might have been a facilitating factor in improving their idiom task performance. In addition, it is not clear if reading the context was what facilitated idiom comprehension for the strong comprehenders. The study might have been improved by also presenting the same participants with idioms without a context, or including idioms that could be guessed at only with context (completely unanalysable) and some semantically analysable idioms to better tease out the true contribution of context.

Cain and colleagues (2005) studied 14 good comprehenders and 14 poor comprehenders between the ages of 9 and 10 years of age in the United Kingdom to see how they would approach completely novel and familiar (novel idioms were again, direct translations of idioms in other languages that had no English equivalent) idioms and semantically analyzable and unanalysable idioms. Learners were matched on vocabulary knowledge and word reading so that variation in performance on the idiom task could be more clearly attributed to text processing abilities, rather than other skills. Idioms were presented in and out of context. Target idiom recognisability and semantic analyzability was evaluated by 12 undergraduate students and the task itself was evaluated by another 12 undergraduate students.

Findings showed that children were much better at understanding idiom meaning in context. While poor and good comprehenders performed comparably when figuring out semantically analyzable items in context, poor comprehenders were significantly less able to

explain the meanings of opaque idioms in context than good comprehenders. There was a main effect for reading skill with idiom recognition – good comprehenders were far more likely to recognize 'real' idioms than weak comprehenders (effect of .142), and transparent idioms were recognized more than opaque (effect of .122). As expected, good comprehenders had higher scores than weak comprehenders.

Children were only slightly more able to recognize real semantically analyzable idioms than real opaque idioms without a supporting context, and real idioms were more easily understood than the novel idioms generated for the study presented without a context. When a context was presented, real idioms were again better understood than novel ones, suggesting that previous exposures to the idioms could play a role in establishing meaning, especially considering that the children could not provide explanations for these items without context. Perhaps the participants had been exposed to the idioms frequently enough to have stored a partial definition in the mental lexicon that could be activated or completed with the help of contextual support. Semantically analyzable items and opaque items were found to be much better interpreted in context, however strong comprehenders significantly outperformed poor comprehenders in this scenario. This falls in line with previous findings that poor comprehenders struggle to make inferences from context (Cain, Oakhill, Barnes & Bryant, 2001; Cain & Oakhill, 1999). Without context, both poor and strong comprehenders were able to generate meanings for some novel, semantically analyzable items, findings that coincide with those of Levorato and Cacciari's (1999) which indicate that children in this age range are able to analyse new idioms.

It is worth noting that the children included in this study as 'poor comprehenders', were classified as such by the study authors based on having performed at least 12 months below age-appropriate level on a reading accuracy measure (the NARA). The poor and strong comprehenders did differ significantly in their performance on the NARA reading comprehension measure ($t[26]=10.72$, $p<.001$). However, just because these children are poor comprehenders does not necessarily mean that they would be expected to perform similarly to children with diagnosed learning disabilities and apply literal interpretations. Abrahamsen and Burke-Williams (2004) found that learning disabled children chose more literal explanations in a forced choice idiom task, however, when presented an idiom

explanation task, the same children were much more likely to come up with meanings that fit the overall story context, indicating task type might play a role in whether or not children consider figurative meaning. In the context of Cain and colleagues' (2005) study, both groups of children were aware that a figurative meaning was needed and sought to come up with one rather than relying on a literal interpretation, even when the correct answer was not achieved. Poor comprehenders were much more likely to select ' I don't know responses', perhaps indicative of having fewer strategies to rely on when trying to figure out unknown idioms; previous research has suggested that poor comprehenders are less aware of comprehension strategies (for example, skimming the text for key information) (Cain, 1999).

All the participants performed most poorly on the novel-opaque items – perhaps these items were difficult because the participants did not even have 'partial' definitions in their mental lexicons from a few previous exposures and were only able to rely on supporting textual context for meaning clues. However, it could also indicate that the items themselves were flawed in some manner that made them particularly incomprehensible. Additionally, since there is evidence that some children can perform differently on idiom explanation tasks than forced choice idiom tasks (Abrahamsen & Burke-Williams, 2004), perhaps the idiom explanation task placed a greater burden on metalinguistic skills and performance was consequently affected. Lastly, it should be noted that the number of participants in each group was small – only 14. It might be of benefit to repeat the study in the future with more participants. 'Poor comprehenders' as a group have been shown to be varied, some have weak semantic skills, some struggle with syntax, (Cain & Towse, 2008). A group of 14 poor comprehenders might be very heterogeneous and in turn affect results.

Duthie and colleagues (2007) performed a study similar to that of Cain and Towse (2008) reporting that 12 year olds with better reading and listening comprehension outperformed 12 year old poor comprehenders on a task selecting the meanings of idioms from the context of a short story, though the shared variance was not especially high, R^2=.36 for reading comprehension and R^2=30 for listening comprehension. The authors attribute this low shared variance to differences in the transparency of target idioms presented in the task – a factor not manipulated by the authors. This provides additional incentive to

explore the connection between transparency and reading comprehension. A higher connection between reading comprehension and opaque idioms would be expected since deciphering the meaning of these idioms is only possible with the help of surrounding written context while semantically analyzable idioms can be guessed with semantic analysis. Additionally, it's possible that good comprehenders were simply more familiar with the target idioms due to greater exposure and consequently did not need to use context to figure out meanings.

Cain and Towse (2008) conducted two sets of experiments comparing 7-8 year olds' and 9-10 year olds' understanding of idioms of varying degrees of transparency and in and out of context. The goal of this project was to explore the theoretical frameworks behind idiomatic speech by presenting learners with entirely novel idioms to assess how these items were processed. Cain presented learners with idioms that were either transparent or opaque, familiar or created entirely for the measure, and in or out of a background context. The novel items were direct translations of idioms found in other languages that hold no equivalent in English. The study included 40 participants, 20 between the ages of 7 and 8 years old and 20 between 9 and 10 years old. Children were presented with 24 target idioms, half 'real' idioms and half novel. Within the two sets of real and novel idioms, 6 were transparent and 6 were opaque, with transparent items being those the authors felt could be figured out with semantic analysis and opaque being those that did not have a meaning that the authors felt could be derived. Participants of the task presented in multiple-choice form and were told to choose the 'right one'. Of the 4 answer options; the target answer, a non-literal interpretation that fit with the context, a non-literal interpretation that did not fit with the context and a literal interpretation.

Results showed that the two groups performed comparably in their understanding of the transparent items and their semantic analysis abilities. Both groups were able to work out the meanings of transparent items presented without context significantly more often than would be predicted by chance. There was a small effect of real over novel items (effect size of .16) but the small strength leads the authors to conclude that the real idioms were not familiar to the participants and had to be 'figured out', just as the novel idioms were. The greatest effect size was for context (.77). Similar to previous findings by Cain et al. (2005), the authors conclude that poor comprehenders struggled to make use of

surrounding context to decipher the meanings of opaque idioms, and that this weakness cannot be accounted for by weak semantic processing ability alone (Cain & Towse, 2008).

2.11 *Learners with EAL*

2.11.1 *Learners with EAL and language development*

EAL learners, or English language learners (ELL) as they are known in North American contexts, are children who go to school in English but speak another language at home with their parents and family. These children are bilinguals or emergent bilinguals. This is a sizable group in the UK and growing quickly. In 2012, there were a reported 995, 230 EAL pupils in state-funded schools in England, or roughly 17.5% of pupils (NALDIC, 2012). There has been concern that the British school system may not be adequately addressing the needs of these pupils as attainment studies show that achievement of learners with the EAL in the UK can be much lower than monolingual English speaking peers (Burgoyne et al., 2009) who also tend to have higher levels of educational achievement overall throughout schooling (NALDIC, 2006/2007) potentially indicating that learners with EAL are not receiving what is needed for literacy and academic achievement. This underachievement carries all the way through to the last stages of schooling. The UK Department for Education and Skills (now the Department for Education) reported a significant attainment gap in A level scores between EAL and L1 pupils when other variables such as socio-economic status (SES) were controlled for (NALDIC, 2006). A number of factors may contribute to this gap, however research consistently has shown that vocabulary, a common area of weakness for EAL children, has a significant impact on school outcomes (Hutchinson et al., 2003). EAL children who begin school with little or no English must rapidly acquire the language in order to follow the same curriculum as L1 classmates who enter school already having 5 or 6 years of language learning. It is troubling but unsurprising then that research continually reports that EAL learners are weak in reading comprehension when compared to the L1 speakers with whom they are expected to perform (Hutchinson et al., 2003).

Learners with EAL do not lag behind in all language based-skills. Studies from Canada, the UK and the US have found that when socio-economic status, age, gender and non-verbal intelligence are controlled for, learners with EAL and L1 learners perform equally well on word reading ability and word decoding (Hutchinson et al., 2003; Lesaux,

Orly, & Siegel, 2006; Lesaux, Rupp & Siegel, 2007; Lesaux, Geva, Koda, Siegel, & Shanahan, 2008; Nakamoto, Lindsey & Manis, 2007). The access that young EAL learners have to two languages can have numerous benefits; it can accelerate the development of both verbal and non-verbal abilities and has been positively associated with cognitive flexibility as well as earlier development in attention and inhibition (Bialystock, 2011). Young bilinguals have the ability to think about the system behind the language and this can even be advantageous in early literacy (Lesaux & Geva, 2006; Genesee & Geva, 2006).

Despite generally being assumed to have equal intelligence and some increased abilities related to their bilingualism, such as better working memory and increased phonological sensitivity (Bialystok, 2011), much research in both the UK and North America has shown that EAL school children lag behind their monolingual peers in reading comprehension (Hutchinson et al., 2003; Lesaux et al., 2007). While children with EAL are extremely gifted in some aspects of language learning and may appear conversationally fluent, or read aloud competently, their academic vocabularies can be far more limited. This usually becomes noticeable during or after Year 4 when the reading curriculum moves from the concrete to the abstract and from 'learning to read' to 'reading to learn' (Bialystok, 2007; Chall, 1983).

2.11.2 *EAL reading comprehension*

Literacy is arguably the greatest legacy of schooling, opening countless doors for the educated individual. While adults take this skill for granted, becoming literate can be a daunting task and many children struggle during the early years of schooling. Learning to read poses a particular hurdle for the rapidly growing population of children with EAL enrolled in UK schools. Numerous studies examining academic outcomes for children with EAL have shown that although these children possess strong phonological decoding skills, many children with EAL perform more weakly than their native English speaking (L1) peers in measures of reading comprehension (Hutchinson et al., 2003; Lesaux et al., 2007). Literacy has a profound effect on academic outcomes for these children, and research focusing on language and literacy development among learners with EAL is essential for improving the lives of the individuals and informing national curriculum and policy (Bialystok, 1999).

This explanation is especially salient in light of evidence that young children with EAL can sometimes mask weak language comprehension by compensating with strong decoding skills, only to later hit a 'slump' in reading comprehension around years 4 and 5, when texts become more complex (Hutchinson et al., 2003; Leach, Scarborough & Rescorla, 2003). Spoken interaction is very different from text, since there is much more information to help with context, such as gestures, tone, facial expression (Mancilla-Martinez & Lesaux, 2010). Additionally, written texts tend to contain more complex vocabulary and language structure than daily conversation (Mancilla-Martinez & Lesaux, 2010).

Cummins (1984) addresses these difficulty differences with his model of EAL language acquisition. The model situates language proficiency requirement along two bisecting difficulty continua, one based on the degree of cognitive involvement needed and the other on the amount of context. Cognitive involvement ranges from easy *Basic Interpersonal Communication Skills* (BICS) to more challenging *Cognitive Academic Language Proficiency* (CALP) (Cummins, 1984). BICS are the types of daily interactions that an English learner might encounter such as talking to a peer about concrete subjects while CALP situations are more demanding, such as a class lecture, and may contain abstract concepts or require skills such as evaluation or analysis (Cummins, 1984). Research indicates that BICS take usually only roughly 2 years to develop while CALPS can take 5 to 7 years (Cummins, 1984). The second difficulty continuum addresses the importance of context. The more context in a given situation, referred to in the model as *context-embedded*, the more clues the learner has to decipher meaning. Spoken language provides a communicative situation in which there can be extra hints regarding meaning, such as facial cues, gestures, tone or listener-speaker interaction (Cummins, 1984). On the other end of the continuum are *context-reduced* situations. These scenarios provide less contextual help to language learners, for example reading a textbook might be a context-reduced scenario (Cummins, 1984). This theoretical model allows further insight into why some learners with EAL are able to perform well in cognitively undemanding BICS situations or with the help of context-embedded scenarios, but may struggle when encountering more difficult CALP context-reduced environments, such as reading assigned lesson material.

The Simple View of Reading and the learner with EAL's profile

The task of reading, which has been previously discussed in this thesis, is the process of learning to recognize orthographic representations of words and connect them to words already in the mental lexicon. This is already a difficult task for some monolingual English speaking children and some do struggle, and possibly even more challenging for children who do not speak English at home. These children must learn English in school, acquiring structure and vocabulary of English in early schooling, then learn to decode it. The challenge of learning to read in a language one is not yet proficient in is a sizable one. Texts presented to students at the primary level typically contain simple vocabulary that the readers will already be highly familiar with. At this point, word reading skills are most important and most predictive of reading comprehension outcomes (Mancilla-Martinez & Lesaux, 2010). Little by little children improve at word reading and the importance of language comprehension will increase. As the children progress in school, the vocabulary in texts that they are expected to understand will become more complex and language comprehension will become a strong predictor for reading comprehension.

Studies with American learners with EAL show that after several years of English schooling, learners with EAL do not always achieve age appropriate language comprehension and so consequently do not have age appropriate reading comprehension (Lesaux & Geva, 2006). The aforementioned simple view of reading does provide some insight into these findings – a number of studies have found that the model can fit an EAL sample, with some adjustments (Hoover & Gough, 1990; Proctor, Carlo, August & Snow, 2005; Nakamoto et al., 2008). Proctor and colleagues (2005) and Hoover and Gough (1990) found that by later elementary schooling, language comprehension was a much stronger predictor than other skills for learners with EAL's reading abilities. Nakamoto and colleagues (2008) however, found that both language comprehension and word reading to be significant predictors of reading outcomes for 6th grade Spanish and English speakers, with word reading as a stronger predictor.

2.11.3 *Vocabulary, reading comprehension and learners with EAL*

Research points to one particular limitation that constrains language comprehension for learners with EAL: vocabulary (Garcia, 1991; Verhoeven, 1990).

Comparisons between the vocabulary size of monolingual (L1) and children with EAL have revealed that children with EAL have smaller English receptive vocabularies (Bialystok, 2010). If children with EAL start schooling at a comparative disadvantage with significantly smaller English vocabularies than their monolingual peers, then it is even harder to catch up (Gregory, 1996; Hutchinson et al., 2003; Bialystok, 2010). Listening comprehension, which is often used as a proxy measure of language comprehension, has also been found to be more highly correlated with vocabulary knowledge among learners with EAL than among monolingual English speakers (Droop & Verhoeven, 2003). Research among learners with EAL has found L2 oral fluency to be limited in accounting for L2 decoding but that it plays a significant role in reading and listening comprehension (Durgunolu, Nagy & Hancin Bhatt, 1993; Geva & Siegel, 2000).

Nakamoto, Lindsey and Manis (2007) tested year 1 Spanish-speaking learners with EAL and found them to have adequate English word reading, adequate reading comprehension and low vocabulary knowledge in comparison to monolingual peer norms. When these same learners progressed to year 3, their word reading remained average, their vocabulary remained low and their reading comprehension declined in comparison to monolingual peer norms. Mancilla-Martinez and Lesaux (2010) conducted a longitudinal study following Spanish speaking learners with EAL from before they began schooling until age 11 and found that English language comprehension accounted for all the unique variance in reading comprehension outcomes over this time. However, even at age 11, word reading was a better predictor of reading outcomes then vocabulary. In both Mancilla-Martinez and Lesaux (2010) and Nakamoto and colleagues (2007), the sampled children were approximately 1 standard deviation below the national average, leading Mancilla-Martinez and Lesaux to posit that the developmental shift from word reading skills having the greatest impact on reading comprehension to language comprehension (often measured with listening comprehension) being most important only applies to age-appropriate comprehenders. For the participants in Mancilla-Martinez and Lesaux's study, the learners at year 5 were reading at a year 2 grade level. Though they had average word reading skills, their vocabulary development had plateaued and remained at the approximate level of a 'normal' 9-year-old monolingual English speaker. This finding is similar to those reported by Jean and Geva (2009) with Canadian learners with EAL from a

diverse group of language backgrounds who also showed good word reading skills and weak vocabulary at approximately the same age.

This combination of abilities and lack of ability (competent word reading and low lexical knowledge) is know as a *specific comprehension deficit* in monolingual English speakers and is somewhat uncommon (5%-15% of the population) (Mancilla-Martinez & Lesaux, 2010). However, Lesaux (2006) posits that this profile is prevalent among Spanish speaking learners with EAL in the US. It seems possible that this could also be true of learners with EAL in the UK. The developmental profile of the learner with EAL is distinct from that of the monolingual English speaker - raising the possibility that a curriculum developed for a different ability profile (monolingual English speakers) may not be as appropriate or successful for learners with EAL. The benefit of instruction tailored to a student's individual needs and abilities is well documented (Connor, Morrison & Katch, 2004). Learners with EAL might benefit from increased or specialized vocabulary instruction, especially when taking into account that, for some learners with EAL, schooling is their main exposure to English and little English vocabulary will be encountered outside of class (Mancilla-Martinez & Lesaux, 2010).

2.11.4 *Cross-linguistic similarity*

In the UK, learners with EAL come from a wide and diverse range of linguistic backgrounds. The greatest percentage of UK students with EAL are speaker of Indo-Iranian languages, with Punjabi accounting for 1.7%, Urdu accounting for 1.6% and Bengali accounting for 1.3% of the UK primary school population (NALDIC, 2012). This still leaves many other language backgrounds; census data has identified over 240 languages spoken by learners with EAL in UK schools (NALDIC, 2012). This diversity in language background presents a wholly different context than the US, where the overwhelming majority of the learners with EAL are Spanish speakers (Mancilla-Martinez & Lesaux, 2010). A more homogeneous language population makes biliteracy programmes more feasible, as it is easier to allocate resources to teaching the home language when many of the learners with EAL speak the same one. This also raises questions regarding the generalisability of findings from US based studies with learners with EAL to the UK context, especially considering the closeness of Spanish and English when compared to a language such as Bengali and the current body of research on cross-linguistic similarity.

Learning a new language is generally considered less difficult when the new linguistic system is similar to the already known L1 (Kaushanskaya & Marian, 2008). This is mainly explained by the fact that the learner can rely on long-term memory and stored knowledge of the L1. For example, if the two languages share much of their phonological sounds, the learner will already have these sounds in their phonological knowledge. The learning benefits of phonological similarity between the L1 and the target language have been well established for adults (Ellis & Beaton, 1993a) and children (Gathecole, Willis, Emslie & Baddeley, 1991). Ellis and Beaton (1993a) found a high correlation between how well participants learned a novel word and how well that word conformed to their L1's phonological patterns. Learning to read in a new language requires coordination of both new phonological information and new orthographic patterns; therefore, it is hypothesized that cross linguistic similarity should have an impact on this process and facilitate mapping auditory information on to the written (Kaushanskaya & Marian, 2008). Orthographic similarity between the L1 and L2 has been found to have a positive effect on foreign language vocabulary learning (Ellis & Beacon, 1993b). Cross-linguistic similarity is a strong enough variable that it has been shown to moderate the effects of age of acquisition (DeKeyser, 2000). Kaushanskaya and Marian (2008) found that aurally presented artificial vocabulary items that matched English phonological patterns were better retained than those that did not by learners who were asked to recognize the words, both aurally and from written presentation, in an immediate and delayed test. The authors' goal was to mimic the process of recognizing written words that were aurally learned. Kaushanskaya and Marian (2008) controlled for potential memory differences among the adult groups with non-word repetition and digit span memory measures. There has also been evidence that a larger vocabulary in the L1 leads to improved performance in L2 word learning among adults, when the phonological systems of the two languages are close (Kaushanskaya & Marian, 2008). This is hypothesized to relate to support provided by the L1 phonological network in the mental lexicon and research showing that the strength of L1 lexical-network phonological representations correlates with L1 vocabulary knowledge (Kaushanskaya & Marian, 2008).

Lindsey and colleagues (2003) found that among Spanish speaking learners with EAL taking part in a bilingual curriculum, expressive vocabulary measured at the beginning

of kindergarten accounted for a significant amount of variance in word-identification scores at the end of year 1. Lindsey also found that receptive vocabulary measured at the beginning of kindergarten accounted for a significant amount of unique variance in reading comprehension, Spanish for English and English for Spanish (Lindsey, Manis & Bailey, 2003). This prediction also held true when looking at either Spanish vocabulary to English reading or English vocabulary to Spanish reading suggesting a common influence across language perhaps related to cross-linguistic similarity.

However, there have also been studies with contrasting findings. Just as some researchers have hypothesized that Spanish speaking learners with EAL in the US benefit from some cross-linguistic similarity between Spanish-English, and perhaps rely on vocabulary cognates, other studies have found that Spanish vocabulary ability may not contribute any unique variance to English reading outcomes, indicating knowing more Spanish words might not necessarily help with English reading (Manis, Lindsey & Bailey, 2004; Nakamoto et al., 2008; Mancilla-Martinez & Lesaux, 2010). Proctor and colleagues (2005) found Spanish vocabulary to account for only a very small amount of variance in English reading comprehension among biliterate Spanish speaking learners with EAL in elementary school receiving formal Spanish instruction in 4th grade.

Taken together, this evidence still creates a strong argument that research findings from other EAL contexts, particularly studies among Spanish speaking learners with EAL, cannot necessarily be generalized or applied to learners who come from language backgrounds that share less phonological similarity with English than Spanish (a relatively similar system) and less orthographic similarity (possibly not even an alphabetic system, in the case of Chinese speakers) in the case of biliteracy. This reinforces the need for UK based research to inform UK policy and the development measures intended for use with UK learners with EAL.

2.11.5 *UK based learners with EAL*

At present, only a small number of studies have focused on UK based learners with EAL in comparison to the sizable body of American and Canadian research. Attainment studies show that achievement of pupils with the EAL in the UK is much lower than of monolingual English speaking peers (Burgoyne et al., 2009) and that monolingual English speakers have overall higher levels of educational achievement throughout schooling

(NALDIC, 2006/2007). This potentially indicates that something about the current system might be resulting in learners with EAL not receiving the necessary education for literacy and academic achievement (NALDIC, 2012). This has been the focus of much of the research with young learners with EAL in the UK, and emphasis has been on what might be missing from the UK curriculum for learners with EAL. Evidence suggests that these learners struggle not with decoding but with understanding the text – often they are referred to as being poor comprehenders, able to read individual words but struggling with the content (Burgoyne et al., 2009; Hutchinson et al., 2003; Yuill & Oakhill, 1991).

Out-of-school literacy instruction has also been discussed as a potential contributing factor for this problem – a unique factor in UK settings. Some children with EAL in the UK attend additional classes at their Mosques or community centres, where they are taught to decode Arabic texts and recite from memory, but not taught to comprehend content (Rosowsky, 2001). It has been posited that this might have a detrimental effect on English language learning if learners try to translate this model of reading in Arabic onto reading in English, as it could result in strong decoding skills with minimal or no understanding of meaning (Rosowsky, 2001). Overall, this still remains an area of speculation.

Successful reading requires both word reading and language comprehension ability; the UK EAL profile of strong decoding and weaker language comprehension is initially viable but in the long run presents a problematic combination of skills and weaknesses (Oakhill, Cain & Bryant, 2003). This chapter has already emphasized that reading is a complex and multifaceted skill that requires a number of different, multi-componential processes and cerebral resources, weakness in even just some of these areas can let down the whole. Nation (2004) and Nation and Angell (2006) identified background knowledge, vocabulary knowledge, inference skills, working memory and comprehension monitoring as essential aspects of reading ability. As previously mentioned, research with UK based learners with EAL has shown them to have lower vocabulary knowledge in comparison to monolingual English speaking peers which in turn has a negative impact on reading (the ability to understand words in a text is essential for comprehending the text overall) (Hutchinson et al., 2003). Learners with EAL can be at a disadvantage, as they begin school with essentially a 5 year deficit when it comes to learning English words. As previously discussed, according to Cummins (1984), children need approximately two years

to acquire basic interpersonal skills and five to seven years to develop cognitive academic language proficiency. This is problematic when it comes to learning to read and the key role of word knowledge. For young children much early vocabulary learning comes from parent interaction, for example having a parent read a story (Nagy, Herman & Anderson, 1985). Learners who do not have English in the home are missing these experiences in English.

Frederickson and Firth (1998) compared phonological awareness (as measured by the PhAB, Phonological Assessment Battery) among 89 children identified as having Specific Learning Difficulties (SpLD) and 50 randomly selected UK learners with EAL who spoke Sylheti at home and had been in English schooling for 5 years. Participants were matched based on age and IQ. Both participant groups (SpLD and EAL) performed more weakly on reading accuracy measures than monolingual peers, learners with EAL performed much more strongly on single word reading measures than text reading, the opposite being true for the SpLD learners. This lends further support to the theory that SpLD learners compensate for weak decoding skills with semantic knowledge while learners with EAL are often competent decoders who compensate for a lack of semantic knowledge. However, this study only provides insight into this sample and this group of learners and findings are not generalisable.

The achievement gap between learners with EAL and their monolingual peers is mainly attributed to weaker English, but the nuances of this discrepancy, the precise areas of weakness and potential strength are still not fully understood. Hutchinson and colleagues (2003) conducted a 3 year longitudinal study comparing literacy development among 43 monolingual English speakers and 43 learners with EAL from Guajarati, Urdu, Punjabi, Bengali and Pashto language backgrounds. The participants with EAL were also all learning to read Arabic outside of school. Learners were matched based on gender, age and ability (as measured by the Raven's colour matrices) and given an annual testing battery for 3 years. Administered measures included: the Neale analysis of reading ability (Neale, 1997), a set of passages graded for reading accuracy and comprehension; the Raven Coloured Matrices (Raven, Raven & Court, 1990) involving the completion of coloured visual-spatial patterns intended to assess non-verbal reasoning ability; the Test of Reception of Grammar (TROG), a multiple choice test that measures English grammatical

knowledge; and the Test of Word Knowledge (TOWK), discussed in more detail in the following chapter. Results showed that learners with EAL had lower levels of vocabulary knowledge and reading comprehension when compared to monolingual learners, but performed similarly on reading accuracy. This falls in line with previous evidence that learners with EAL do not struggle with reading accuracy, and indeed when viewed longitudinally the learners with EAL displayed similar developmental patterns with reading accuracy as monolingual peers. Findings did show learners with EAL to lag behind monolingual peers in receptive vocabulary by approximately 2 years (Hutchinson et al., 2003). This gap was evident at the beginning of the study and did not appear to be closing at any point during or at the end. The authors attribute the lower levels of reading and listening comprehension among the learners with EAL participant group to this vocabulary deficit; monolingual English speaking participants consistently outperformed learners with EAL in these two comprehension areas. Learners with EAL also lagged behind monolingual peers by one year in reading comprehension. Both groups showed significant year-to-year improvement on all comprehension measures, but monolingual English speakers made significantly more progress annually, particularly during the first year of the study.

Relationships between the measures were comparable among the two groups. Expressive vocabulary in Year 2 predicted unique variance in reading and listening comprehension for participants in Year 4 for learners with EAL, and the authors point out in this study receptive vocabulary did not have the same predictive power for reading or listening comprehension that expressive did. This leads the authors to argue that receptive vocabulary measured in Year 2 did not necessarily indicate whether the learner had the required language skills to understand presented material (written or aural) and that assessment of receptive vocabulary will not necessarily be able catch language deficits in Year 2; expressive vocabulary is a better measure for recognizing deficits.

Of significance when considering the performance of the learners with EAL is that they were all learning Arabic outside the classroom. This decoding practice might contribute to some reading-related skills, such as reading accuracy, for which the study reported very high scores, although perhaps not all. Hutchinson reminds us why these comparatively lower reading comprehension scores among learners with EAL are

meaningful, saying 'within the classroom, children learning EAL have the same opportunities to learn as their monolingual peers, but vocabulary learning, either from a verbal or written experience, is enhanced by comprehension, so that children learning EAL are at a disadvantage." (Hutchinson et al., 2003, pg.30).

Burgoyne and colleagues (2009) examined reading comprehension skills among learners with EAL and monolingual English speakers in year 3 living in the UK. Participants were given a battery of measures for: reading accuracy (the NARA-R, a 6 passages reading comprehension measure), a single world reading measure (the WRAT 3, which requires learners to read out progressively more difficult words until a threshold is reached), expressive and receptive vocabulary measures (one word picture vocabulary) and two measures of listening comprehension (the NARA-R and LTCS, both audio recording followed by multiple choice questions). Burgoyne and colleagues (2009) compared findings between the two groups and found that children with EAL performed significantly more weakly on expressive and receptive vocabulary measures ($p<.001$, main effect size of .219 and .152 respectively) than monolingual English speakers. Vocabulary was found to be the greatest predictor of reading comprehension for both monolingual pupils and learners with EAL. Expressive vocabulary had a unique contribution to listening contribution to listening comprehension for both groups, although stronger for monolingual English speakers. Though learners with EAL had significantly higher scores for single word reading and test accuracy ($p=.02$ and p.022, respectively) than monolingual English speaking peers, they had significantly weaker reading comprehension (effect size of language background on reading comprehension $p<.001$, effect size of .154). For learners with EAL, receptive vocabulary did not make a significant unique contribution to reading comprehension, though expressive did. By contrast, for monolingual English speakers expressive vocabulary made no unique contribution to predicting reading comprehension, but receptive vocabulary was a significant predictor for reading comprehension outcomes (Burgoyne et al., 2009). This leads the authors to conclude that the factor differentiating the two groups was vocabulary. Learners with EAL outperformed monolingual English speakers on single word reading and reading accuracy, and after this was controlled for children with EAL underperformed in reading comprehension (Burgoyne et al., 2009). Again, perhaps this heightened word reading accuracy without text comprehension could

be attributed to the training in Arabic decoding these learners had. Burgoyne and colleagues (2009) also concluded that learners with EAL with Arabic language training may mask comprehension difficulties by focusing on reading accuracy.

Evidence from these studies are aligned with some results found by researchers in other contexts; children with EAL seem to have more limited vocabularies than monolingual English speaking children and this in turn constrains their reading comprehension (Hutchinson et al., 2003). Research indicates that this lower level of vocabulary knowledge persists, implying greater emphasis is needed earlier to establish a vocabulary base that can support later development (Burgoyne et al., 2009). In the UK, the added dimensions of dissimilarity between the L1 and English and potential complications rising from out-of-school reading instruction create a unique context that necessitates UK based research.

2.12 *Existing tests of formulaic language knowledge*

Having outlined the potential import of multi-word vocabulary for young learners and the absences of existing relevant literature, the next task is to measure multi-word vocabulary knowledge. Although multi-word phrases are a well-established area of learner difficulty, relatively few experimental studies have attempted to measure phrasal knowledge, possibly due to the difficulty in operationalising and measuring the construct. Many of the existing tests are unsuitable for the current study or for use with young learners due one or more of the following characteristics; unsystematic selection of test items (e.g. Biskup 1992; Bahns & Eldaw, 1993; Bonk, 2001; Farghal & Obiedat, 1995; Keshavarz & Samimi, 2007; Jaen, 2007; Rinaldi, 2000), too few test items (e.g. Bahns & Eldaw, 1993; Farghal & Obiedat, 1995), minimal or no measures of reliability (e.g. Biskup 1992; Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Rinaldi, 2000), lack of attention to the frequency of the multi-word phrase as whole, the component words or both (e.g. Biskup 1992; Bahns & Eldaw, 1993; Bonk, 2001; Farghal & Obiedat, 1995; Jaen, 2007; Mochizuki, 2002; Rinaldi, 2000), possible lack of measurement validity due to internal assumptions (Bonk, 2001; Farghal & Obiedat, 1995) or a reading comprehension heavy format (Martinez & Murphy, 2011).

Many of the existing test formats present a slightly different conceptualization of the construct of multi-word phrase knowledge. If multi-word phrases are conceptualized as a

complete lexical unit, processed and stored as a whole, then a productive test of this knowledge should test the production of the whole phrase, not a particular component (Revier, 2009). However, a number of the productive tests rely on presenting parts of a whole phrase to elicit the remaining component, providing very little information about knowledge of the phrase as a whole (Marton, 1977; Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Bonk, 2000).

Early measurement tools focused on using translation tasks to have students produce English multi-word phrases from phrases in their mother tongue. One such investigation by Biskup (1992) presented Polish and German university students with L1 phrases containing *verb + noun* collocations and *adjective + noun* collocations to translate into English. Data analysis showed the two groups produced the same mean number of correct responses, although Polish learners demonstrated different strategies. However, the study provides no information about how the target test items were selected and no measures of reliability.

Bahns and Eldaw (1993) administered a translation task and a cloze sentence completion task to German university students studying English. The test format provided a complete sentence in German followed by an English translation containing a target *verb + noun* collocation missing the verb component. The researchers assumed that if the test takers knew the phrase they would be able to provide the verb from the noun node, and if they did not they would paraphrase from the German equivalent. Results from the multi-word phrase test were compared to single word vocabulary knowledge, as measured by the same test, showing a substantial lag in mutli-word phrase knowledge. From this the authors conclude that phrase knowledge does not develop in parallel with single word knowledge in adult learners.

These results are fascinating; however, there are a few potential problems with the testing instrument. First, the test itself only consisted of 15 target test items, possibly too few to justify the robust conclusions drawn. Longer tests are generally considered more reliable, as more test items help to cancel the effect of measurement errors (Brown, 1983, cited in Gyllstad, 2005). Target test items were selected from learning materials and evaluated by native speakers but not evaluated for frequency or strength of co-occurrence. Additionally, the target test items chosen were those with a strong German equivalent. The

most interesting finding of the study is the lag between mutli-word phrase knowledge and single word vocabulary; however, the single word vocabulary measure was taken from the same phrase measure developed by the researchers – without any reliability measures – potentially confounding results.

Farghal and Obiedat (1995) presented L1 Arabic university students with a translation and cloze task, similar to that of Bahns and Eldaw (1993). The study created 2 different test formats with 11 target multi-word phrases in each – one format presented Arabic sentences to be translated into English, the second presented collocate pairs (for example, strong tea/weak tea) in a cloze task with one missing component. The same criticisms levelled at the previous testing instruments also apply here; the small number of test items, unsystematically selected target multi-word phrases and lack of reliability measures. The test format also relies on the assumption that the relationships between the dichotomous pairs will be obvious to the test takers. For example, in the following test question: 'Some people like salty soup, but others like _____ soup' (Farghal & Obiedat, 1995), the correct answer is 'bland' soup, but test takers might be able to provide other feasible responses.

Bonk (2001) created a test battery containing three subtests: a *verb + object* multi-word phrase test eliciting the verb, a *verb + preposition* multi-word phrase cloze test eliciting the preposition and a multiple choice task presenting four usages of a figurative verb requiring the test taker to identify the incorrect one. No information is given about how target test items were selected. The test was validated by 10 native speakers and analysis for internal consistency found the test to be reliable, with the exception of the *verb + preposition* subtest. Results from 98 university students showed a correlation between multi-word phrase knowledge and a measure of general English language ability. This study is relatively robust but the format requires a large amount of reading, and as a result might inadvertently measure reading comprehension. Another disadvantage to the measurement tool is the decision to elicit components of the multi-word phrases rather than the whole.

Mochizuki (2002) administered tests of multi-word phrase knowledge, paradigmatic knowledge and general vocabulary knowledge to Japanese university students to explore how their language ability progressed over the academic year with 75 hours of language

lessons. The paradigmatic language test and multi-word phrase knowledge test required testees to find links between single lexical items in a multiple-choice format. The tests were administered twice, at the beginning and end of the year. Results showed a significant increase in multi-word phrase knowledge but not in paradigmatic word knowledge or vocabulary. The test itself contained a high number of test items (72) and internal validity is reported. The test format presents single, de-contextualised items, which might be a questionable method for testing multi-word phrases.

Barfield (2003) created a test to measure recognition of *verb + noun* multi-word phrases and compared this skill to single noun and verb recognition. The instrument presented 120 test questions, 20 pseudo-phrases and 100 genuine multi-word phrases using the same 40 verbs and a multiple choice scale to accept or reject the multi-word phrases. The test was administered to 93 Japanese university students from 4 different academic disciplines. Analysis showed that the test had high internal reliability (Cronbach's alpha of .97) and distinguished well between participants from different disciplines. The testing instrument is systematic, although some of the 'non-phrases' used in the test would be considered acceptable in certain contexts.

Interestingly, participants showed higher recognition of nouns than verbs and a higher recognition of verbs than *noun + verb* multi-word phrases. The 20 most correctly identified phrases were those that contained nouns and verbs in the top 3,000 most common English words, as reported by the BNC. Barfield interprets this to mean that the frequency of the components had a strong influence on overall phrase recognition. This conclusion is also supported by Arnon and Snider's research with phrase component frequency and voice onset time among L1 English speakers (Arnon, 2009; Arnon & Snider, 2010). The fact that participants had greater recognition of individual words than multi-word phrases could also lend support to the idea that there is a lag between single word knowledge and phrasal knowledge for L2 learners.

Gyllstad (2009) created two original formats for measuring receptive knowledge of multi-word phrases, which he administered to upper-secondary and university L1 Swedish students. The formats, COLLEX and COLLMATCH, tested *verb + object* phrases containing high frequency words (85% of the verbs and 80% of the nouns were among the 3000 most frequently used English words as measured by the BNC) in a de-contextualised

item testing format. For the COLLEX, test takers were given 50 test questions and required to select proper *verb + noun* phrases from a list containing pseudo – phrases of the same noun and a different verb. If the test taker guessed the answer, he/she indicated this by ticking a box next to the test item. A bilingual version of the test was also presented featuring a translation of target phrase in the L1 but this had little effect on test outcomes. Gyllstad notes that this might have even distracted learners and encouraged negative transfer. The COLLMATCH format presents test takers with a grid showing 3 verbs vertically along the left most column and 6 nouns along the top row. The task of the learner is to match the verbs with the nouns in order to create feasible combinations. Gyllstad describes this format as receptive but requiring slightly more knowledge.

The two test formats had an acceptable measure of internal consistency (Cronbach's Alpha of .83 overall for the COLLEX and .80 for the COLLMATCH, although the number was much lower for native speakers for whom there was very little variance). The test was administered twice and the results for the two formats correlated highly with each other. Generally results showed that test scores increased with increased exposure to English. Second and third term university students didn't perform significantly differently from each other, and third term students did not perform significantly differently from native speakers. Vocabulary size was seen to make a significant contribution to phrasal knowledge; learners with smaller vocabularies were less able to recognize phrases. The only possible area of criticism for this test is that the format, while reliable, may be more focused on rejecting non-phrases than identifying or knowing the meaning of real multi-word phrases. This may explain why, unlike previous studies testing productive knowledge (Bahns & Eldaw, 1993; Farghal & Obiedat, 1995), Gyllstad (2005) found no lag in phrasal knowledge.

Crutchley (2007) explored comprehension of idiomatic *verb + particle* constructions among native English speaking 6-11 year olds using the only test reviewed here specifically created for young learners, the Assessment of Comprehension and Expression (ACE 6-11) subtest for non-literal language. The test contains 15 particle verbs used in a non-literal fashion. While this is not a measure of multi-word phrase knowledge, it is included in this review because it is a measure of non-literal language. The test taker is presented with a full sentence containing a particle verb and must identify the meaning

from 4 written or pictorial answer options. The test was standardized in the UK and Irish Republic with approximately 800 pupils and has a high internal consistency.

In Crutchley's study, a ceiling effect was found for the test; an increasing number of children answered each test item correctly as age increased. The test included 'distracter' incorrect answers presenting various literal interpretations of the verb, the particle or the whole string. Interestingly, no changes were found as age increased, indicating that from ages 6 to 11 strategies for interpreting unknown test items did not change. Analysis of the children's strategies showed that the participants treated the verbs as a unit, and did not analyze the components alone. This leads Crutchley to conclude that Wray's hypothesis is correct - children process language as chunks, only to then be broken apart when needed (Crutchley, 2007; Wray, 2002). Crutchley asserts that when a child does not know the meaning of a presented string the child will chose a meaning for the string based on context.

The ACE 6-11 is a well-validated test and will be used in the present study as a measure of concurrent validity to correlate with the original tests, as well as to provide an additional measure of non-literal language understanding. The test alone is quite short, only 15 items, and receptive. Particle verbs are only a small element of vocabulary depth, and this measure alone would not be sufficient to make any robust conclusions regarding multi-word phrase knowledge.

The final testing instrument reviewed here is the CONTRIX format developed by Revier (2009), which is the primary model for the productive testing instrument used in this study. The CONTRIX format presents a sentence with a gap at the end for an entire multi-word phrase. The test taker must form the phrase from a 3 by 3 or 3 by 2 matrix of constituent parts by selecting one word from each column. The target test items were *verb + noun* multi-word phrases, divided into categories based on how literally components are used – transparent, semi-transparent and non-transparent.

Revier administered this test to L1 Danish secondary school and university students. The test had a high internal consistency (Cronbach's alpha =.89), and discriminated well between age levels and scores increased with greater exposure to English. Data analysis also found that the three test sections significantly differed from

each other. This test presents a clear, reliable format for testing somewhat productive knowledge of multi-word phrases without requiring a high amount of reading.

2.13 *Original contribution of the current study*

English multi-word phrases present particular difficulty for even very advanced foreign language learners. It has been argued that learners will simply 'pick them up' with enough exposure, as native speakers do. The importance of these vocabulary items is not totally understood, particularly with respect to young learners, but phrasal knowledge has been shown to have an impact on reading comprehension for adult L2 speakers (Martinez & Murphy, 2011). At present there is only a sparse amount of literature examining this aspect of vocabulary acquisition in monolingual English speaking children, and no empirical studies have ever examined these aspects of children with EAL's vocabulary development, though they are developmentally unique from both native speaker children and foreign language learners. Therefore, an important first step is to identify the nature of monolingual English speaker children and children with EAL's multi-word vocabulary knowledge at different points in their development. The limited number of existing reliable tests of phrasal knowledge presents an additional challenge. Currently, there are no other measures of noun + verb multi-word phrases suitable for young learners. The first phase of this study will develop and pilot a multi-word phrase measurement task. The second phase of this study will see its administration to young learners in order to better explore the nature of multi-word phrase vocabulary among primary school learners and examine the role phrase transparency plays in phrase knowledge. The third and final phase of the study will explore the contribution of multi-word vocabulary to variance in reading scores among year 4 learners with English as their only language or an additional language, in hopes of providing direction for future investigation into the role phrasal vocabulary plays in reading for young learners. No previous studies have examined this relationship among this age group, possibly due to the lack of measurement tools available. The multi-word phrase task developed for this study will be suitable for future use child language research. Findings from the current study could provide insight into language acquisition, help to inform EAL curriculum and policy, and pave the way for future research using the instruments. A robust body of research has shown vocabulary to be an essential component of reading comprehension, yet single word vocabulary measures fail to

fully explain discrepancies in reading comprehension outcomes between monolingual English speakers and learners with EAL. The next essential step for reading research is to explore the role of multi-word vocabulary in reading comprehension and language development over the primary school years.

Chapter 3: METHODOLOGY

3.1 *Introduction*

The current study consists of three phases. Phase 1 of the study involves the creation and pilot of a task designed to measure multi-word phrase knowledge, referred to as the Multi-word Phrase Task (MPT). Phase 2 explores the MPT as a measure (including internal reliability, evaluation of individual items, test-retest, and an investigation of concurrent validity) and attempts to measure multi-word phrase knowledge among samples of children in years 3, 4 and 5. The third phase entails the administration of the MPT to year 4 students to explore a possible relationship with reading comprehension. Ethical considerations and overall limitations of the project will be discussed at the end of this chapter. The three research phases are outlined in Table 3.1.

Research Aims

One aim of the current research project was to develop a new measure of British English multi-word phrase knowledge appropriate for use with monolingual English speaking children and learners with EAL in years 3, 4, and 5 in order to address the following research questions:

1. Can verb + object multi-word phrase knowledge among monolingual English speaking children and in learners with EAL between the ages of 7 and 10 be measured by a discrete test? If so, what is the nature of verb + object multi-word phrase among these learners and how do they compare across year and language group?

2. What effect does transparency have on verb + object multi-word phrase knowledge for monolingual English speakers and learners with EAL between the ages of 7 and 10?

3. What is the relationship between verb + object multi-word phrase knowledge and reading comprehension among monolingual English speaking children and learners with EAL in Year 4?

Table 3.1: Research outline

Research phases	Participants	Instruments	Sessions
Phase 1 – Creating and piloting the MPT	35 participants total -16 year four students, 8 learners with EAL and 8 monolingual English speakers -9 year three students, 3 learners with EAL and 6 monolingual English speakers - 10 year five students, 3 learners with EAL and 7 monolingual English speakers	- MPT - Vocabulary measure - Language background questionnaire	- Two per student. Session 1: MPT and interview Session 2: Vocabulary measure and language background questionnaire
Phase 2 – Validating the MPT	105 participants in total -32 year three students, 12 learners with EAL and 20 native English speakers - 47 year four students, 16 learners with EAL and 31 monolingual English speakers - 26 year five students, 9 learners with EAL and 17 monolingual English speakers	- MPT - Test of Word Knowledge (TOWK) Expressive and Receptive vocabulary - Wechsler Abbreviated Scale of Intelligence (WASI) non-verbal IQ measure - Assessment of Comprehension and Expression (ACE) 6-11 - British Picture Vocabulary Scale (BPVS) - Language background questionnaire	-Three sessions per student, sessions 1 and 3 occur 4 weeks apart. Session 1: MPT, ACE 6-11 and language background questionnaire, WASI, Session 2: TOWK and BPVS. Session 3: MPT re-test 4 weeks later.
Phase 3 – MPT correlations with reading comprehension	- 40 participants total, year 4. 20 learners with EAL and 20 monolingual English speakers	- MPT - TOWK. Expressive and Receptive - WASI -York Assessment of Reading Comprehension (YARC)	-One session per student. Session 1: MPT, TOWK, WASI and YARC.

3.2 *Phase 1: Creating and piloting a test of multi-word phrase knowledge*

Little research has previously examined multi-word phrase knowledge among young monolingual English speakers or young learners with EAL; therefore, one goal of the present study was to explore the nature of multi-word vocabulary at different points in development. In order to do so, a new measure of multi-word phrase knowledge appropriate for use with this population needed to be developed.

3.2.1 *Test design and development*

Defining and operationalising the construct

For the purpose of the current study, a 'multi-word phrase' was defined as a composite unit of words that express meaning as a whole and co-occur together more often than would be expected by chance. Multi-word phrase knowledge is conceptualized as an independent construct; therefore, knowing a multi-word phrase means knowing the entire phrase, the meaning as a whole, the lexical components and the necessary grammar (Revier, 2009). If a multi-word phrase is stored as 'one giant lexical unit', as has been hypothesized (Wray, 2002), then a test of multi-word phrase knowledge should aim to test the entire phrase. This eliminates using existing test formats that provide a part of the phrase and elicit the missing component.

Selecting target phrases

Verb + object multi-word phrases were chosen as the target phrases for testing because 1) they are numerous in English, occurring in both high and low frequencies; 2) they range from entirely literal to completely non-literal; and 3) in English they tend to form the core of an utterance and often convey essential information because they contain the content words (the verb and object) and affect the structure of the sentence (Gyllstad, 2005; Nesselhauf, 2005). This allows for the development of test items with completely literal *verb + object* phrases, such as 'catch mice'; and with more opaque

phrases using the same verb, such as 'catch fire' and 'catch a cold'. Comparisons can then be made between non-literal phrases and literal phrases consisting of the same or similar lexical components.

Table 3.2: Target phrases by verb component and transparency

VERBS	Transparent	Semi-transparent	Non-transparent
break	break a bone	break the silence	break the ice
carry	carry an umbrella	carry a risk	carry the day
catch	catch mice	catch fire	catch a cold
change	change direction	change trains	change hands
cut	cut a hole	cut jobs	cut corners
hold	hold hands	hold a conversation	hold your tongue
get	get a message	get a taxi	get the sack
pay	pay a bill	pay attention pay a visit	pay the price
run	run a race	run a business run tests	run the show
take	take the money	take your pick take the lead	take sides

Test item transparency

Test items were divided into three categories based on opacity: t*ransparent, semi-transparent* and *non-transparent.* This classification system is adapted from Nesselhauf (2005) and Revier (2009), who adapted Nesselhauf's framework and used it successfully in his validated test of collocational knowledge for Danish speaking adult L2 learners of English. Classification is based on the semantic properties of the word items in the phrase. If the verb and object are both used with their literal, first dictionary entry meaning, the phrase is 'transparent'; if the noun is literal but the verb takes on a non-literal meaning, the phrase is 'semi-transparent'; and if neither is used in a literal sense and the phrase forms a whole meaning that could not be derived from the individual literal parts, then the phrase is 'non-transparent' (Revier, 2009). See Table 3.2 for target phrases presented by transparency. This range of item transparency enables the research to isolate the effect of transparency on multi-word phrase knowledge. As discussed in the previous chapter, there are many different

perspectives on how to define and categorize formulaic language. Regarding the transparent test items, these are frequently co-occurring items that would meet a frequency based definition for a collocation. However, they can easily be broken apart, are not fixed, can be analysed word for word and do not express a holistic meaning. The non-transparent items, by contrast, can be conceptualized as whole chunks, fixed idioms that cannot be analysed as parts and cannot have other pieces substituted in. Therefore, by some definitions the target items on the MPT are all collocations, by other theoretical frameworks the transparent items are collocations and the non-transparent are idioms, but by all definitions using frequency as a determiner, the target items do qualify as multi-word units, with components that co-occur more than can be explained by chance.

Test item frequency

Test item frequency was measured using the British National Corpus (BNC) and the Child Language Data Exchange System (CHILDES). Frequencies for target phrases ranged from 4 occurrences in the BNC to 384. Target phrases were formed using 10 high frequency verbs (in the top 1,000 most frequently used words in English

Table 3.3: Target multi-word phrases and number of occurrences in the BNC

Transparent	Semi-transparent	Non-transparent
carry an umbrella – 3	run tests – 8	hold my tongue – 20
catch mice – 4	cut jobs – 23	carry the day – 28
cut a hole – 16	change trains – 26	catch a cold – 33
break a bone – 16	hold a conversation – 33	run the show – 36
run a race – 27	carry a risk – 45	break the ice – 44
get a message – 61	get a taxi – 57	cut corners – 57
change direction – 112	take your pick – 76	take sides – 66
hold hands – 141	break the silence – 81	get the sack – 75
pay the bill – 145	pay a visit – 132	pay the price – 191
take the money – 151	run a business – 180	change hands – 276
	catch fire – 211	
	take the lead – 375	
	pay attention – 384	

according to the BNC). For each verb a minimum of three target phrases were developed: one transparent, one semi-transparent and one non-transparent. See Table 3.3 for a list of final test items appearing in the MPT and frequency of entire phrase in the BNC. These target items were selected because: 1) there existed a set of at least three multi-word phrases that fit the framework for transparency (there had to be an existing verb + object multi-word phrase that was transparent, semi-transparent and non-transparent), 2) they fit the criteria for component frequency (the nouns had to be in the 5,000 most frequently used English nouns) and 3) they could fit into the aforementioned framework. One potential limitation of the current study is the fact that the frequency classifications used were based on a corpus of adult language, the BNC, rather than child-directed or child-produced speech. Unfortunately, CHILDES does not accommodate searching for entire phrases, and at present no suitable corpus of child speech is available and so the BNC was used. However, the BNC is not a measure of child-directed language or language that learners with EAL might come into contact with. To help ensure that the words contained in the MPT target items were familiar to participants, a vocabulary test containing all individual lexical items contained in the target test items was administered during the pilot phase of the research. See section 3.2.2 for more detailed information regarding the vocabulary measure and section 3.2.5 for vocabulary measure results.

It has been hypothesized that component frequency might contribute to overall multi-word phrase recognition (Arnon, 2009; Barfield, 2003). Keeping this possibility in mind, all of the individual lexical items used in the test were checked for frequency in two different corpora, the BNC and CHILDES. CHILDES does not allow for whole phrase searching. Instead, CHILDES was primarily used to explore the frequency of the test item lexical components in child directed speech to ensure that these individual

components occurred frequently enough in the corpora to likely be familiar to young learners. See Appendix A for frequency details for individual lexical items. All the individual lexical items in the target test items occur in the top 5,000 most frequently used English nouns or the top 5,000 most frequently used English verbs, as measured by the BNC. It is therefore likely that participants are familiar with the individual items on the MPT.

Frequency comparability across transparency categories

As little is presently known about this topic, restricting the item sample to only one frequency band would limit testing and might inhibit possible insights that could come from presenting a range of items. Items were categorized into frequency bands, presented in Table 3.4, high mid and low. Two analyses were done to ensure a comparable frequencies across the transparency categories. A one-way ANOVA was used to compare frequency, as measured by number of BNC occurrences, among the three transparency categories, and showed no significant differences, [$F(2,30)=.275$, $p=.761$]. In addition, a Kruskal-Wallis test was also used to compare binned frequency (low, mid and high) among the three transparency categories and found no significant difference in distribution between transparencies, $H=2.119$, 1 d.f., $p=.347$. The three transparency categories have a comparable distribution of item frequency when grouped into bins.

In addition to the BNC, test item frequency was evaluated by 21 adult British speakers of English who self-identified as 'native speakers' of British English, ranging in age from 25 to 63. Adult participants were recruited via University departmental emails and personal connection to the researcher. These adults were asked to rate item familiarity using the Titone and Connine (1999) scale, ranking items from 1 (never seen or heard) to 7 (frequently seen or heard). A copy of the survey can be found in

Table 3.4: Target multi-word phrases by frequency		
0-40 occurrences in the BNC	40-100 occurrences in the BNC	100+ occurrences in the BNC
carry an umbrella – 3 catch mice – 4 cut a hole – 16 break a bone – 16 run a race – 27 hold my tongue – 20 carry the day – 28 catch a cold – 33 run the show – 36 hold a conversation – 33 run tests – 8 cut jobs – 23 change trains – 26 13 items	get a message – 61 get a taxi – 57 take your pick – 76 break the silence – 81 cut corners – 57 take sides – 66 get the sack – 75 carry a risk – 45 break the ice – 44 9 items	pay a visit – 132 run a business – 180 catch fire – 211 take the lead – 375 pay attention – 384 pay the price – 191 change hands – 276 change direction – 112 hold hands – 141 pay the bill – 145 take the money – 151 11 items

Appendix B. Survey responses found that the overall categories transparent, semi-transparent and non-transparent items did not significantly differ in familiarity. Mean frequency scores were 5.26 for transparent items, 5.38 for semi-transparent items and 5.57 for non-transparent items. A one-way ANOVA showed that transparent, semi-transparent and non-transparent items did not differ significantly in familiarity overall [$F(2,570)=2.075$, $p=.126$].

Mutual Information

Target multi-word phrases were further confirmed as phrases using the Mutual Information (MI) index, a measure of the strength of word co-occurrence (Gyllstad, 2009). MI evaluates how frequently selected words occur together in a corpus and if this occurs more frequently than would be predicted by chance, as calculated by the frequencies of the individual lexical items in the same corpus (Church & Hanks, 1990). An MI of 2.5 is considered the threshold for indicating significantly more co-occurrence than would be predicted by chance and an MI of 3 or greater denotes a 'significant collocational link' (Church & Hanks, 1990). The MIs of all test items were

evaluated using COLLOCATE phrasal extraction software on the BNC, and were found to co-occur significantly more than could be explained by chance. See Appendix C for a table of test item MI scores. MI was not intended as an experimental variable in the current study. While some research has indicated that MI could play a role in multi-word phrase acquisition (Ellis, N., Simpson-Vlach, & Maynard, 2008), the purpose of calculating MI in the current study was to ensure that all of the test items had a high enough MI to indicate they were strong multi-word phrases and to better isolate transparency as a variable. A one-way ANOVA was used to compare MI across the three transparency categories to ensure that "collocational strength" of items in the three transparency categories was comparable; a non-paramatric test, a Kruskal-Wallis test, was conducted to compare the distribution of MI scores between the transparency categories showed that there was no significant difference between the three categories $H=1.772$, 2 d.f., $p=.412$. A one-way ANOVA also showed no significant differences between means [$F(2,30)=.429$, $p=.655$]. Manipulating MI might be an interesting direction for future research.

Test format

The MPT was developed for young learners specifically. Pre-existing tests of multi-word phrase knowledge would not have been appropriate for the current study as they have all been developed for either adult or young adult language learners. Young language learners differ from older learners in at least three key areas: growth, literacy and vulnerability (McKay, 2006). Understanding these differences is essential for effective assessment. Children vary as individuals, like adult learners, but they also develop at different rates. Bialystok (2001) refers to the adult mind as the 'stable state', whereas the child's mind is in a constant state of growth. It is therefore difficult to say at what age exactly a particular developmental process will occur for an individual child.

However, for a reliable test, the cognitive demand of the task must be commensurate with age-related abilities (McKay, 2006). Regardless of language ability, young children may not be able to respond to tasks that extend beyond their experience of the world. Additionally, young learners have limited attention spans and may abandon a task they find difficult or lengthy (McKay, 2006). Special attention was paid to avoid making the MPT too reliant on reading comprehension skills, potentially confounding results. The sentence completion prompts were kept lexically and structurally simple and contained only literal language. The researcher is a speaker of American English, therefore 7 adult British English speakers evaluated all sentence prompts before the pilot phase began. Sentence content required minimal background knowledge and contained only topics hopefully familiar to the age group such as school, friendships and food. The MPT was designed to be short while still containing enough test items to be valid, so as not to tire young learners.

The test format was created based on a pre-existing test format called the CONTRIX, a format for testing multi-word phrase knowledge in adults validated by Revier (2009). In the CONTRIX format, a prompt sentence is followed by a matrix of words. The test taker chooses one word from each column to form a phrase that completes the sentence. In the original format three different phrases could be formed using the grid, only one of which made sense in the context of the sentence prompt. This format relies heavily on reading comprehension since the test taker has to have thoroughly comprehended the sentence to determine which combination of words makes the most sense in the sentence context. As discussed in Chapter 2, EAL children have been shown to lag behind age-matched monolinguals on measures of reading comprehension. As a consequence, this format could potentially disproportionately disadvantage these EAL test takers. If a child were to spot a valid

phrase but not understand a key lexical item in the prompt, s/he might view the item as a feasible response, as it would be a grammatically viable, real English phrase. In order to place emphasis on the ability to produce multi-word phrases in the MPT, another type of response matrix was developed, in which only one viable English phrase could be produced. All other lexical items in the matrix matched the context but could not combine to form a phrase.

Example 1: Original CONTRIX

The quickest way to win a friend's trust is to show that you are able to _____.

tell	a/an	joke
take	the	secret
keep	_	truth

(Revier, 2009, p. 127)

In the original CONTRIX format above, if a test taker doesn't fully understand the prompt, both 'keep a secret' and 'tell the truth' are feasible answers. However, in a MPT version of the same question, the child could still answer the question, if s/he knew that 'keep a secret' was a correct phrase and saw all the other options were incorrect.

Example 2: Adjusted CONTRIX: MPT format

The quickest way to win a friend's trust is to show that you are able to _____.

be	on	loyalty
pay	the	secret
keep	a	game

Phase 1 of the project piloted both formats and results are reported in section 3.2.5 of this chapter.

Scoring of the MPT is dichotomous – test takers receive 1 point for a correct response and 0 points for an incorrect response. No partial credit is given for an incorrect response, as the goal of the test is to elicit whole multi-word phrases. Scores are divided into three categories: Transparent (10 items), Semi-Transparent (13 items) and Non-Transparent (10 items). A score for each category is calculated along with a total score.

Administration

The MPT is administered individually to each test taker. The test administrator first reads the instructions to the test taker and then an example is presented and reviewed. The test instructions can be found in Appendix D. To avoid potential confounds, such as reading ability, the test is read aloud in its entirety to the test taker, including prompts and matrix box words. After the answer choices have been selected, the entire sentence with the new ending is read aloud to the test taker. The test was paper and pencil; read an unlimited number of times to reduce potential strain on working memory.

3.2.2 *Materials*

Language Background Questionnaire

An adjusted version of the Beech and Keys (1997) Language Preference Questionnaire was administered to each child in the pilot study. Additional questions focusing on family, sibling relationships, and out of school language study were added (see Appendix E for language background questionnaire). The primary purpose of the questionnaire was to screen for possible confounding variables, for example, students who were selected by teachers as learners with EAL but lived with one monolingual English speaking caregiver or parent and one non-English speaker, or students who were selected as monolingual English speakers but spoke another language at home

with one caregiver or parent. Also excluded were students who were listed as monolingual English speakers but had attended full-time primary school in another language outside of the UK for any amount of time.

Vocabulary Measure

A vocabulary measure was developed for this study to explore students' receptive knowledge of the literal, first dictionary meaning of the individual words that make up the target multi-word phrases on the MPT. As previously discussed in this chapter, all of the individual nouns and verbs in the target multi-word phrases occur on the BNC's lists of the 5,000 most frequently used nouns and verbs, respectively. It is therefore highly likely that children between the ages of 7 and 10 would be familiar with these words. Corpus frequency counts, like those provided by the BNC, have been previously used by other vocabulary measures as a basis for assuming that young learners would be familiar with target test items, for example the ACE 6-11 (Adams, Cooke, Crutchley, Hesketh & Reeves, 2001). However, the BNC is not intended to reflect child language specifically or child directed language and is not a corpus of language that the participants would necessarily be exposed to, whether monolingual English speakers or learners with EAL. This presents a potential limitation. For extra assurance, the researcher presented 8 learners, 4 with EAL and 4 monolingual English speakers, with the individual word vocabulary measure. See Appendix F for the vocabulary measure. The test presented each individual word followed by three sentences, one with a correct usage of the word. Sentences were read aloud to the students who then had to select which sentence had the correct word usage.

3.2.3 *Sampling procedures for Phase 1*

A pilot study was carried out during October and November of 2010 to trial the overall feasibility of the proposed study and to examine the appropriateness of the

MPT for use with children in years 3, 4 and 5. Both formats of the MPT measure, the original CONTRIX and the MPT version were piloted to explore student preferences and measure internal consistency. These two formats were combined into one test in order to better evaluate any differences in test taker performance between the two formats. The pilot also provided an opportunity to evaluate the assessment timetable and the demands of the schedule on the participants.

Participants

Thirty-five students participated in the pilot; emphasis was placed on recruiting Year 4 students, as they should represent the middle of the primary school population. See Table 3.5 for participant information. Participants were recruited from primary schools in Oxfordshire and Swindon, as these were the nearest areas with students who fit the recruitment criteria for learners with EAL. The researcher identified approximately 45 schools in commuting distance from the Oxford and Swindon area with a population of 30% or higher learners with EAL that were not already participating in a research project to the researcher's knowledge. The researcher then contacted the head teachers by post, email and phone to invite the schools to participate (see Appendix G). Two of the contacted schools agreed to take part in the study, one school in Oxford and one school in Swindon (see Appendix H for additional descriptive information on the schools that participated in the current study). Individual student participation was on an opt-in basis. The researcher provided the schools with consent forms and parental information pamphlets to distribute to potential participants that the school felt met the recruitment criteria. Parents or caregivers provided written consent for their child to participate (see Appendix I for consent form and parental information pamphlets). Classroom teachers distributed consent forms to eligible students and then presented the signed forms to the

researcher. The researcher did not know the names of students who did not meet the recruitment criteria or declined to participate. For the purpose of this study, a learner with EAL was defined as a child who speaks any language other than English at home with all of his/her parents or primary caregivers, as established by teacher selection, school records and research administered language background questionnaires. This definition was based on the National Association for Language Development in the Curriculum assessment of pupil achievement among minority ethnic groups and lower achieving groups (NALDIC, 2004). Students who spoke English with one parent or caregiver and another language with another were not included in the study. Participants were included only if they had been in English speaking school in the United Kingdom since reception (age 5 or 6) and had minimum school attendance of 75% from the previous or present year, depending on time of recruitment. Children with a diagnosed special educational need (SEN) were also excluded from the study to avoid potential confounding factors. For the pilot and validation phase of this study the learners with EAL were not limited by first language 1) as no comparisons were made between groups and 2) to ensure that the test would be suitable for different types of learners. Participants with EAL in phase 1 had the following first languages: Bengali (4), Spanish (1), Russian (1), Polish (1), Turkish (2), Tetum (1) and Urdu (4). See Table 3.5 for participant information.

Table 3.5: Participant information for Phase 1 pilot study

Year 3	N	Male	Female	Mean age
Monolingual English speakers	7 (Oxford)	4	3	7.6
Learners with EAL	3 (Oxford, Swindon)	2	1	7.35
Year 4	N	Male	Female	Mean age
Monolingual English speakers	8 (Oxford)	4	4	8.26
Learners with EAL	7 (Oxford, Swindon)	5	2	8.64
Year 5	N	Male	Female	Mean age

Monolingual English speakers	7 (Oxford)	2	5	9.47
Learners with EAL	3 (Oxford, Swindon)	2	1	9.1

Participants for the vocabulary measure were selected based on teacher willingness to allow for longer sessions, with one exception. The lowest performing participant was intentionally selected by the researcher to take part in the vocabulary measure, as it was suspected that this student might be less likely to know the words than higher performing peers. 8 students took the additional vocabulary measure; details are presented below in Table 3.5.1.

Table 3.5.1: Participant information for Phase 1 vocabulary measure

	Year 3	Year 4	Year 5
Monolingual English Speaker	0	2 (Oxford)	2 (Oxford)
Learners with EAL	1 (Oxford)	3 (Oxford, Swindon)	0

The learners with EAL had the following first languages: Tetum (1), Turkish (2), and Urdu (1). 5 of the participants were male and 3 were female.

3.2.4 *Testing procedure for Phase 1*

Data was collected in individual assessments in quiet settings outside of the classroom. All students who took the additional vocabulary measure took it after the MPT. Language background questionnaires were administered after the MPT and, when applicable, after the vocabulary measure.

3.2.5 *Pilot results for Phase 1*

Table 3.6: Performance means from Phase 1

	Mean (out of 40)	Standard Deviation
Year 3	24.44	5.68
Year 4	23.5	6.92
Year 5	29.2	7.18
Monolingual English speakers	27.52	6.46
Learners with EAL	22.14	6.63

Phase 1 of the data collection revealed that monolingual English speaking children and learners with EAL in years 3, 4 and 5 can successfully complete the MPT. The MPT took a maximum of 36 minutes to administer and a minimum of 8 minutes, with an average time of 16 minutes. Table 3.6 presents performance means from the MPT pilot administration. Table 3.7 presents internal reliability (as measured by Cronbach's alpha) from the pilot administration. The pilot version of the MPT had 40 items, later versions of the test have 33.

Table 3.7: Cronbach's alpha from Phase 1	
MPT overall	.853
MPT new format	.805
CONTRIX original format	.891
Removing test items no participant answered correctly	.925
Transparent items	.788
Semi-transparent items	.886
Non-transparent items	.753

The MPT showed an acceptable level of internal reliability (.853) with this small sample. Results from the pilot administration showed that there was a difference in the way young learners performed on the two different test formats: the original CONTRIX format, in which other multi-word phrases that do not fit in the sentence context and the MPT format, in which no other multi-word phrase can be formed. When presented with a test item they did not know, participants spent much longer trying to find a correct answer in the newer format, as no other 'real' combinations could be formed. After spending extra time working on the item, they would sometimes produce the target phrase, or settle for a nonsense combination. By contrast, when presented an unknown test item in the original format, children were much quicker to simply select another real multi-word phrase from the matrix that they recognized, even when it didn't fit the prompt sentence content. A few participants even articulated this when interviewed about which test items were most difficult. As a

result, it was decided to use the newer test format containing only one real multi-word phrase per matrix for the MPT, given the similar internal reliability scores for both formats and considering the small number of participants (Cronbach's alpha of .805 for the newer format and Cronbach's alpha of .891 for the original). The slightly lower internal reliability score for the MPT format could also be explained by the fact that a greater number of the test items that no participant answered correctly, simply due to chance, happened to be in the MPT format. See Appendix J for the final version of the MPT used in Phases 2 and 3.

All students performed well on the vocabulary measure. Scores ranged from 36 to 48 out of 48 total points, indicating that the vocabulary words in the test items were familiar to students. Only one participant scored below 46 on the test; this was the participant who performed the lowest on the MPT. See Table 3.8 for details. Results from this measure, along with the frequency counts detailed in section 3.2.1, support the assumption that most students in years 3, 4 and 5 will be familiar with the lexical items that make up the target multi-word phrases.

Table 3.8: Mean scores from Phase 1 Vocabulary Measure			
	N	Mean (out of 48)	Standard Deviation
Participants	8	45.63	3.96

3.2.6 *Format changes made after the pilot*

After test administration, the researcher asked participants to review the test and discuss which questions were most difficult or easy and why. From these results, it was decided that some sentence prompts should be altered (See Appendix K for the pilot version of the MPT). The researcher elected to change the proper names used in the sentence prompts to be more orthographically shallow. It was also decided that the transparent target multi-word phrase for the verb *carry* should be changed from 'carry a

gun' to 'carry an umbrella'. The researcher also reviewed texts in the pilot schools for appropriate themes and styles.

The pilot MPT contained target multi-word phrases formed from 13 different verbs; however after the pilot, 3 of these verbs (draw, make, play) and their corresponding test items were removed. The pilot showed that the test was slightly too long for convenient testing and the researcher shortened the test. In addition, discussions with participants revealed that particular multi-word phrases were problematic for learners or were difficult to fit into clear sentence prompts. These phrases were removed and replaced with another phrase from the same transparency category that fit test format criteria. If no replacement phrase could be found, the entire verb and all three phrases were removed from the test in order to maintain the framework of each verb forming at least one multi-word phrase in each of the three transparency categories.

Findings from the pilot also led to the removal of the 'I don't know'/ 'Pass' option that allowed test-takers to avoid answering a test question. Many children never chose to pass, while a few others over-relied on the option in order to move more quickly through the test. When the researcher asked students to go back to these test questions, the students were sometimes able to come up with the correct multi-word phrase, if prevented from choosing 'I don't know'. The 'I don't know' option was helpful during the pilot phase as it allowed insight into which questions children found more difficult answer. For grading purposes 'I don't know' was considered an incorrect answer on the pilot tests. In subsequent test administrations children were not allowed to 'pass' on a test question but were encouraged to guess when unsure.

3.3 *Phase 2: Validation of the MPT*

The goal of the second research phase was to establish test reliability and validity. Demonstrating test validity is the process of accumulating evidence to show an instrument measures what it intends to. Reliability refers to the ability of the measure to yield the same results repeatedly (Fulcher & Davidson, 2007). There are a number of different ways to compile evidence of validity and reliability. This study sought to demonstrate overall validity and reliability through the following: concurrent validity (how highly test scores correlate with scores on measures of related skill), internal consistency (for example, seeing that low scorers get the same difficult questions wrong and the same easier questions correct), construct validity (how well the test is able to distinguish between learner groups, in this case age, and between the different transparency categories), and test-retest reliability (making sure that test takers achieve similar scores when given the test multiple times) (Henning, 1987).

Test-retest coefficient is a statistical measure of temporal stability, indicating how much the individual's normative score has changed during the elapsed period of time between test administrations (Guttman, 1945). Test-retest reliability is obtained by administering the same test with a set amount of time elapsed between test times and then correlating the scores from the two sessions (Guttman, 1945). For a reliable test, any change in test scores should be small and would be a reflection of either a degree of progression in the ability being measured, improvement caused by the previous test administration, or simply random. It is important when conducting test-retest reliability to consider the rate at which the measured skill develops. Vocabulary is still growing between ages 7 and 10, so the elapsed window for test-retest needed to be relatively small to avoid a large change in performance related to language development. Pearson language tests recommend a test-retest interval of 4 weeks for the vocabulary measurements aimed at this age group (Pearson Speech and Language,

2011). It is unknown how multi-word phrase knowledge develops among these children, so the recommendations for testing general vocabulary development in this age group were followed.

3.3.1 *Materials*

To establish concurrent validity, Phase 2 included a number of other measures of vocabulary that were administered to investigate correlations between the MPT test scores and scores on these related measures. These instruments are discussed in detail below. Along with the MPT, the second research phase included a range of standardised measures of related skill, presented here in detail. All of these instruments helped to build a strong body of evidence for concurrent validity. As little is known about multi-word phrase knowledge in this age group, it is difficult to say which language based skills are most related. Consequently, a variety of vocabulary-based measures were given to participants to establish concurrent validity.

Scores for the additional administered assessments will be presented as raw scores, not standardized, for two primary reasons. Firstly, some of the measures have been standardized separately with EAL learners (details in following sections) and some have not. Standardized score results would therefore show a range in scores that would only sometimes reflect a scoring adjustment for language difference. A standardized score reported for an EAL learner on a measure with separate standardization will show how far the score deviates from the norms for EAL learners at that age, for example, 85 – one standard deviation from the mean. However, a very different raw score for a monolingual speaker might also translate to 85 – one standard deviation from the mean for monolingual learners. These two learners might therefore have very different accuracy on the assessment itself but have comparable standardized scores on this assessment. The same two learners might then show very different scores on another

assessment, one without separate standardized scores for learners with EAL. Secondly, using standardized would make differences between learners of different ages less immediately apparent, as scores would be adjusted by age; two learners, one in year 3 and one in year 4, might both have a standardized score of 85, for their age, but differ from one another. Raw scores will therefore be reported for maximum insight into differences between ages and language backgrounds.

Non-Literal Comprehension subtest of the Assessment of Comprehension and Expression (ACE 6–11).

The Non-Literal Comprehension subtest is one of seven subtests that make up the Assessment of Comprehension and Expression (ACE 6–11) developed by Adams, Cooke, Crutchley, Hesketh & Reeves (2001). The ACE 6-11 is a measure of receptive and expressive syntactic, semantic and discourse-level language abilities for use with children between ages 6 to 11. The subtest consists of 15 particle verb phrases. The test taker selects the answer from 4 choices that best fits with the meaning of the spoken particle verb phrase. The test consists of a set of 'pictureable' particle verbs, with four picture choices as potential answers and a set of 'non-pictureable' items with four written answer choices. All test items are read aloud to the test taker. Many particle verbs phrases can be analysed literally or semantically and the answer options contain three 'distracters' that reflect literal interpretations of the verb or draw on semantic information.

The instrument was standardized with a stratified sample of 790 learners from 50 different schools in the United Kingdom. The test was not separately standardized or validated with learners with EAL; however, they were included in the final standardization sample proportionate to the population. The Non-Literal Comprehension subtest had an acceptable, though not high, level of internal reliability

with an average Cronbach's alpha of 0.61 and an overall test-retest reliability of 0.8 (Crutchley, 2007). The makers of the ACE 6-11 consider this an acceptable level of internal reliability and the test has been published. However, other researchers in testing and assessment consider .7 to be the minimum acceptable Cronbach's alpha score (Fulcher & Davidson, 2007) and so multiple measures of related skill are used in the current study, rather than relying on the ACE alone as a measure of vocabulary knowledge. The ACE 6-11 was able to differentiate between learners of different ages (Crutchley, 2007). Test items were chosen from a corpus of child language, on the assumption that the frequency of a particle verb in child language would reflect the frequency of the verb in child-directed speech (Adams et al., 2001).

Test of Word Knowledge; Receptive, Expressive (TOWK)

The Test of Word Knowledge (TOWK) is a standardised test of vocabulary knowledge intended for use with children aged 5 to 17 (Wiig & Secord, 1992). The TOWK has been thoroughly vetted and previously used in a number of literacy studies and with learners with EAL (Hutchinson et al., 2003; Nation & Snowling, 1998). The current study administered the Receptive Vocabulary and Expressive Vocabulary subtests. The Receptive Vocabulary subtest aims to assess the child's receptive vocabulary knowledge by having the child match pictures with spoken words (noun or verb). Test items become increasingly difficult and testing ends when the child has chosen five consecutive wrong answers or has finished all 42 items. The Expressive Vocabulary subtest functions as a reverse of the Receptive Vocabulary: the tester presents the child with pictures to name. There are 32 items and, as with the Receptive Vocabulary subtest, testing ends when the child has chosen five consecutive wrong answers or reached the end of the test (Wiig & Secord, 1992, p. 28).

The TOWK was standardized with a sample of 2,243 students in the United States. Approximately 10% of this standardization sample was made up of learners with EAL enrolled in mainstream English language only education (Wiig & Secord, 1992, p. 45). Further validity and reliability was explored with a mixed ability sample of 140 children: 48 8 year olds and 92 11 year olds (Wiig & Secord, 1992, p. 50). Results found the TOWK to be valid and reliable in that the test differentiated well between the groups and correlated highly with other measures of language ability and subtest inter-correlations were all acceptable (above .5). Internal reliability was calculated for all subtests and sample population groups and was found to be quite high (Cronbach's alphas from .85 to .92). Test-retest reliability for all subcomponents was also acceptable, ranging from .46 and .96 (Wiig & Secord, 1992).

British Picture Vocabulary Scale – II (BPVS)

The BPVS is a receptive listening vocabulary measure that has been validated in the UK with monolingual English speakers and learners with EAL (Dunn, Dunn, Whetton, & Pintilie, 1982). The researcher reads a word aloud to the child, who selects a picture that best represents the meaning from 4 options (Dunn et al., 1982). The test is discontinued when the test taker answers 8 or more items incorrectly in a 12-item set.

The BPVS – II is an updated version of the original edition. The original BPVS was standardized with a general UK sample and a sample of learners with EAL from the UK in 1982 (Dunn et al., 1997). The second edition was standardized again in the UK including learners with EAL in 1995 with a sample of 2571 children (Dunn et al., 1997). Internal reliability was high with a Cronbach's alpha of 0.93 (Dunn et al., 1997). Validity was evaluated separately for the sample of learners with EAL. In all

samples, the test discriminated well between groups and was found to be valid (Dunn et al., 1997).

Wechsler Abbreviated Scale of Intelligence for Children (WASI) non-verbal IQ measure

The WASI is a test that measures nonverbal intelligence among children ages 6 to 16. This standardised test has been designed for use with non-proficient speakers of English and requires a minimal amount of language. The test consists of four subtests; however, the current study used only one subtest, Matrix Reasoning. Test takers are presented with a series of patterned pictures and must choose from 5 answer choices the picture that best completes the pattern. Testing is discontinued when the test taker incorrectly answers 4 items in a row or 4 out of 5 consecutive items.

The WASI was standardised in the United States with a sample of 2,200 children aged 6 and 16 years 11 months, and in the UK with 780 children (Wechsler, 1992). Both standardization samples included a variety of participants, including gifted children, children with mental disabilities and children unable to speak (Wechsler, 1992). The test has been used in previous studies with children with learning disabilities and language specific disorders as well as average and above average students, and should be suitable for use with learners with EAL (Lobello, 1991). The WASI was validated with other measures of achievement and memory (Wechsler, 1992). The primary purpose of including the WASI was to ensure that the monolingual English speakers and children with EAL were comparable in general cognitive skill; however, correlations between MPT and WASI scores were also explored. See Chapter 4: Results for further details.

3.3.2 *Sampling procedure for Phase 2*

Phase 2 of the study was carried out between January and December of 2011. Primary schools in the Cornwall, Coventry, Oxfordshire, Slough and Swindon area were invited to participate. Sampling was done using convenience sampling method based on schools it was possible for the researcher to travel to. More than 90 schools were contacted. During this phase, schools with smaller populations of learners with EAL or no learners with EAL were contacted as well. All primary schools in Oxford not currently participating in another research project (to the researcher's knowledge) were contacted. Schools in Swindon and Slough with larger populations of learners with EAL were also specifically contacted. Additionally, the researcher contacted one school in Coventry, one school in Cornwall and one school in Oxfordshire where the researcher had personal connections. Three schools in Oxfordshire agreed to take part in the study, one school in Swindon, one school in Slough, one school in Coventry and one school in Cornwall.

Participants

The same recruitment criteria were applied as detailed in Phase 1 – recruitment was on an opt-in basis and only children with no diagnosed SEN who had been in English language schooling since reception were included. A minimum of 20 children were recruited for each year group, a minimum of 10 EAL students and 10 monolingual English speakers. The participants with EAL had the following first languages: Bengali (14), Chinese (1), Egyptian Arabic (1), German (1), Hindi (1), Italian (1), Kiswahili (1), Konkani (4), Malayalam (2), Portuguese (1), Somali (1), Tagalog (2), Tamil (2), Turkish (1) and Urdu (6). As in the pilot phase, participating learners with EAL came from a variety of first language backgrounds to ensure that the test would be suitable for different types of learners and so the sample would be more representative of the general UK population. See Table 3.9 below for participant information.

Table 3.9: Participant information for Phase 2				
Year 3	N	Male	Female	Mean age
Monolingual English speakers	20 (Oxfordshire, Cornwall)	8	12	7.43
Learners with EAL	12 (Oxford, Swindon)	4	8	7.71
Year 4	N	Male	Female	Mean age
Monolingual English speakers	31 (Oxfordshire)	13	18	8.22
Learners with EAL	16 (Oxford, Swindon)	5	11	8.6
Year 5	N	Male	Female	Mean age
Monolingual English speakers	17 (Oxford, Cornwall)	7	10	9.88
Learners with EAL	12 (Oxford, Swindon)	3	9	9.5

3.3.3 *Testing procedure for Phase 2*

The same testing procedure was applied as in Phase 1. Tests were administered in the same order to each participant in a quiet location outside of the classroom. The researcher met with each participant at least twice for approximately 30 minutes. 87 Phase 2 participants were selected to participate in test-retest to evaluate the MPT's reliability. The researcher met with these participants for a third session 4 weeks after the first and re-administered the MPT. These students were chosen by convenience and school availability. See Table 3.10 for test-retest participant information. See Chapter 4: Results for Phase 2 findings.

Table 3.10: Participant information for Phase 2 test-retest			
	Year 3	Year 4	Year 5
Monolingual English speakers	9 (Oxfordshire)	31 (Oxfordshire)	10 (Oxford, Cornwall)
Learners with EAL	12 (Oxford, Swindon)	16 (Oxford, Swindon)	9 (Oxford, Swindon)

3.4 *Phase 3: Relationships between reading comprehension and multi-word phrasal knowledge.*

The final phase of the study explored relationships between reading comprehension and multi-word phrase knowledge among monolingual English learners

and EAL learners in year 4. For this phase, only year 4 students were selected, as they represent the middle of the age group the MPT was developed for and it was not within the scope of the present research to extend Phase 3 across three year groups.

3.4.1 *Materials*

The third research phase involved 3 standardised measures in addition to the MPT, the WASI non-verbal IQ measure, the TOWK Expressive and Receptive and the York Assessment of Reading Comprehension (YARC). Both the WASI and the TOWK were used in Phase 2 and were previously discussed.

York Assessment of Reading Comprehension (YARC)

The YARC is a standardised measure of reading comprehension intended for use with learners between ages 5 and 11. The test measures three components of reading ability: comprehension, decoding and fluency (as measured by reading rate) (Snowling, Stothard, Clarke, Bowyer-Crane, Harrington, Truelove, Nation & Hulme, 2009). Four different scores from the YARC will be used: a single word reading score, a reading accuracy score, a reading comprehension score and a reading rate score. The test is administered individually to students by the researcher. Learners are presented first with a Single Word Reading Test to determine appropriate starting level for passage reading, and then are given progressively more complex passages to read aloud. The Single Word Reading Test will also provide a single word reading score. After each passage the test taker is asked a series of comprehension questions. Scores are given for the total time taken to read the passage (Rate), the number of errors (Accuracy) and the responses to the comprehension questions (Comprehension). Testing discontinues when the test taker makes a predetermined number of errors (15 in early stage passages, 20 in later passages) (Snowling et al., 2009).

The YARC was standardised in the UK with a stratified sample of 1376 children. Learners with EAL comprised 14.02% of the standardisation sample (Snowling et al., 2009). Each section of the test was found to be reliable using Cronbach's alpha (results from .55 to .77) and Rasch analysis (results from .66 to .81) (Snowling et al., 2009). Test scores correlated highly with results from parallel measures of reading comprehension, indicating concurrent validity (Snowling et al., 2009).

3.4.2 *Sampling procedure for Phase 3*

Participants

Phase 3 of the study was carried out between November of 2011 and January of 2012. The same primary schools in Cornwall, Coventry, Oxfordshire, Slough and Swindon discussed in Phase 2 were all invited to participate. The same schools in Coventry and Cornwall agreed to take part. The same recruitment criteria were used as in the previous phases – recruitment was opt-in and only children with no diagnosed SEN who had been in English language schooling since reception with acceptable attendance records were included. 40 year 4 students were recruited, 20 students with EAL and 20 monolingual English speakers. Phase 3 of the study involves comparisons between monolingual English speakers and learners with EAL, so to minimize potential confounds only learners with EAL from the same language background (Bengali speakers) were included in the study. See Table 3.11 below for participant information.

Table 3.11: Participant information for Phase 3

	N	Male	Female	Mean age
Monolingual English speakers	20 (Cornwall)	10	10	8.31
Learners with EAL	20 (Coventry)	13	7	8.38

3.4.3 *Testing procedure for Phase 3*

The same testing procedure was applied as in the first two phases of the study. The four testing instruments (TOWK Expressive and Receptive, WASI, MPT and YARC) were administered in the same order to each participant, to ensure a break halfway through the session would not interrupt a longer assessment and there would be a mix of assessment types, in a quiet location outside of the classroom. The researcher met with each participant once for approximately 40 minutes. See Chapter 4: Results for Phase 3 findings.

3.5 *Ethical considerations*

There are particular ethical considerations raised by this study due to the nature of the research design.

3.5.1 *Central University Research Ethics Committee*

Any work conducted with young people requires fastidious attention to research ethics. No part of the current study was undertaken without ethical approval from the Central University Research Ethics Committee. The IDREC protocol MSD/IDREC/2005/P2.1 for studies using non-invasive methods with typically developing children in schools was followed precisely. All aspects of the study adhered strictly to the protocol recommendations, including researcher training, participant recruitment, confidentiality, duty of care and data protection. The researcher observed the recommendations made by the British Psychological Society (BPS) and The British Educational Research Association (BERA), with particular attention to the *Children, Vulnerable People and Vulnerable Adults* guidelines.

3.5.2 *Language considerations*

There are additional language considerations for the present study, because the sample contains participants with EAL. It is possible that the parents or guardians who were asked to opt-in to the study on behalf of their children were not fully competent in

English. The researcher discussed these concerns with appropriate school representatives and adjusted procedures as needed. At one school the EAL co-ordinator suspected certain parents might have limited English and the school provided translated copies of the participant information to the parents. Additionally, the researcher attempted to take cultural practices and norms into account when interacting with participants and their families by dressing appropriately for the religious environment in certain schools and ensuring the parental information pamphlets showed a range of children from diverse ethnic backgrounds. The researcher took cultural background into account when creating the sentence prompts and names on the MPT, as well as when conversing with the children.

3.5.3 *Inclusion of select students*

The current research only involved some of the children in any particular classroom. To minimize the negative impact this selection might cause, the researcher spoke with classroom teacher about interacting with the class as a whole where appropriate. In three cases, the researcher volunteered in the classroom while the project was taking place. The researcher was not aware why students were excluded from the study so as to protect student privacy.

3.5.4 *School day disruption*

The current study required some accommodation from the schools and teachers. The researcher attempted to minimize this demand as much as possible. The researcher consulted the classroom teachers regarding convenient times for participating students to leave the classroom. Scheduling was discussed in detail to ensure minimal disruption and inconvenience.

3.5.5 *Training*

The researcher has been trained by the Educational Studies Research Council approved University of Oxford Education Department research methodology course.

Chapter 4: RESULTS

4.1 *Introduction*

This chapter presents the findings from Phases 2 and 3 of the current study. Section 4.2 will detail the findings from the validation phase (Phase 2). This section will first present descriptive information regarding participant performance on the MPT and the additional administered measures, followed by general analyses of instrument reliability and validity for the MPT such as: internal consistency scores for the MPT, test-retest reliability findings, correlations between the MPT and other administered measures, Item facility, and Item discrimination scores. Section 4.2 will also present analyses of the effect of transparency, frequency and group (year group and language background) on MPT performance accuracy. Section 4.3 will detail findings from the third research phase (Phase 3), in which the MPT was administered along with the YARC reading measure. This section (4.3) will present participant performance means, correlations between the MPT and additional administered measures, as well as regression models exploring the unique contribution of the MPT overall and each separate MPT section to YARC performance scores.

4.2 *Phase 2 results: Validating the MPT*

The goals of the second phase were to: 1) demonstrate that the MPT measures what it intends to, 2) demonstrate that the MPT is a reliable measurement tool, 3) show that it is suitable for the intended population and 4) explore the nature of *verb + object* multi-word phrase knowledge among learners in years 3, 4 and 5. There are a number of different ways to compile evidence of validity and reliability, as detailed in the previous chapter. This study sought to provide evidence of validity and reliability by showing the MPT to have concurrent validity (how well MPT test scores correlate with scores on measures of related skill, in this case measures of single word vocabulary, the

ACE 6-11, the BPVS and the TOWK), construct validity (a large test effect showing differences between the test transparency categories and evidence that the test differentiates between age groups), internal consistency (for example, seeing that low scorers get the same difficult questions wrong and the same easier questions correct, as measured with a Cronbach's Alpha), and test-retest reliability (evaluating how closely participant scores correlate after multiple test administrations, in the present case, after a 4 week interval) (Henning, 1987).

4.2.1 *Descriptive statistics*

Normality of the data

In order to run the parametric inferential statistical tests needed for the intended analyses outlined in the previous section (4.2), the data set for the current study had to first meet certain statistical assumptions; namely, it needed to be normally distributed and there had to be homogeneity of variance. A normal distribution refers to the fit of the data along a so-called "normal curve"; a distribution of data that is symmetrical and bell-shaped with data tending toward a central value, and roughly 99% of all data falling within three standard deviations of this central point (Field, 2009). Data normality is generally evaluated by measuring *skew* and *kurtosis*, both terms refer to the degree by which data deviates from this normal curve. The distribution of a data set can violate normality by spreading too much or too little horizontally (skew) or by being too high or too low vertically (kurtosis). Homogeneity of variance is the assumption that variance within the different sample groups (in this case, language background groups and year groups) is similar and that overall the variance is the same throughout the data.

To evaluate normality of the data for the current study, the skew and kurtosis of the data were computed for each measure (ACE 6-11, BPVS, MPT, TOWK

Expressive, TOWK Receptive, WASI) and each sub-section of the MPT (MPT transparent, MPT semi-transparent, MPT non-transparent). For a modestly sized sample such as the one in the present study to be considered normal, the skew and kurtosis should be below the threshold of 3.29; higher scores indicate a statistically significant deviation from normality (Field, 2009). Kolmogorov-Smirnov tests are often used to evaluate normality, however, this test is only appropriate for measures with many different potential result scores. Some of the measures used in the current study have as few as 10 items (MPT transparent); therefore, a Kolmogrov-Smirnov test may not reliably distinguish normality from abnormality (Field, 2009). Measures with limited potential outcomes, like the MPT transparent score with only 10 different possible scores, are more likely to deviate from a normal curve when measured by a Kolmogrov-Smirnov test. Consequently, a lack of normality on the Kolmogrov-Smirnov test alone is not a barrier to using parametric tests, as the data may still be normally distributed (Field, 2009).

Homogeneity of variance is the assumption that variances within different sample groups (in this case, language background groups and year groups) are similar and that overall the data variance is the same throughout the data set. This is usually measured using Levene's test, a statistical measure that tests the hypothesis that the variance in outcomes for a variable (in this case, each of the administered measures) are the same for more than one sample group. In the current study, more than one group was sampled (3 different age points, 2 different language background groups) therefore it is possible that the variances in this samples differs.

See Table 4.1 below for the results of a Kolmogrov-Smirnov test for normality and Levene's test for homogeneity of variance. Most, but not all of the data was found to be normally distributed and met standards for homogeneity of variance.

Table 4.1: Tests of normality (skew and kurtosis) and homogeneity of variance for Phase 2 data

	Normality		Homogeneity of Variance	
	Skew	Kurtosis	Levene's Test	Significance
ACE	1.79	1.22	14.339	.000**
MPT	1.97	.52	2.406	.124
BPVS	1.61	1.93	.212	.646
TOWK Expressive	.79	5.01**	10.544	.002*
TOWK Receptive	2.16	1.5	12.967	.000*
WASI	.99	1.38	2.707	.103
MPT transparent	4.29**	6.7**	7.921	.006*
MPT semi-transparent	1.7	.91	2.59	.11
MPT non-transparent	2.0	1.37	9.4	.003*

* indicates significant deviation from normality at p<.05, **p<.001

It is generally considered acceptable to proceed with most types of statistical analysis if the data meets one of the two assumptions, although normality is considered more important than homogeneity of variance for conducting parametric tests (Field, 2009). In the current study, two of the measures used violated both assumptions: the TOWK Expressive scores and MPT transparent scores. However, because there were so few items on the MPT transparent (10) and most learners scored well, though not at ceiling, it would have been difficult for this data to be 'distributed normally'. Additionally, the MPT transparent was not a distinct measure but a subcomponent of the overall MPT, which did meet both assumptions. While TOWK Expressive also did not meet either assumption, it was only one measure among all of the measures used in the current analysis and so should not cause results to be inaccurate. In addition, the type of analysis used, ANOVA, is particularly robust and functions well even with data that violates these normality assumptions, so long as groups do not differ too greatly in size, in this case a minimum of 12 and maximum of 31 (Moore, McCabe & Craig, 2010).

However, for extra assurance, non-parametric tests will also be presented in addition to parametric tests, to confirm parametric findings.

Participant means

Having established that the data is suitable for parametric statistical evaluations, with accompanying non-parametric tests to confirm significant findings where possible, the following section will present descriptive information regarding MPT participant performance, presented by year group and language background. Tables 4.2 and 4.3 present details of participant performance on the MPT; Table 4.2 presents mean MPT scores (out of a possible 33 points) and standard deviations, Table 4.3 shows performance scores for each transparency section, presented as percentage of items answered correctly in that section. Transparency section scores are presented in percentages, rather than raw scores, since the transparency categories contained differing numbers of questions (10 and 13) and a raw score would therefore be potentially misleading.

Table 4.2: Raw scores (out of 33 test questions) on multi-word phrase task presented by year and language group

	Overall Group		Monolingual		Learners with EAL	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
All years	*20.34*	*5.74*	*22.44*	*4.99*	*16.78*	*5.176*
Year 3	17.97	5.31	19.30	5.96	15.75	3.08
Year 4	20.06	5.71	22.77	3.73	14.81	5.26
Year 5	23.41	4.96	25.53	3.64	20.42	5.16

Table 4.3: MPT mean scores (in %) and standard deviations presented by transparency category

	Overall Group		Monolingual		Learners with EAL	
	Mean	SD	Mean	SD	Mean	SD
Transparent						
All years	85.09	16.60	90.59	12.44	75.75	18.66
Year 3	79.69	17.13	82.50	18.03	75.00	15.08

Year 4	85.53	17.67	92.26	7.17	72.50	24.08
Year 5	90.34	12.39	97.06	5.88	80.83	13.11
Semi-transparent						
All years	61.92	22.38	69.38	19.55	49.23	21.32
Year 3	53.06	20.79	57.65	23.70	45.42	12.06
Year 4	61.23	22.85	72.52	15.47	39.38	18.70
Year 5	72.79	19.08	77.47	14.98	66.17	22.78
Non-transparent						
All years	37.78	20.06	44.26	19.87	26.75	15.09
Year 3	30.94	18.38	35.50	20.89	23.33	9.85
Year 4	34.68	17.55	41.61	16.75	21.25	9.57
Year 5	50.34	20.61	59.41	16.00	37.50	20.06

Overall, on a descriptive level, learners did well on the MPT – learners did not appear to be at floor or ceiling and overall older learners outperformed younger learners; scores appear to be higher among older learners. As can be seen in Table 4.3, transparent items appear to be more accurately answered than the semi-transparent and non-transparent items. These comparisons will be explored more fully in section 4.2.6.

Before comparing the two language groups' MPT performances, it is of interest to consider their performances on the measures of single word vocabulary. An independent samples t-test comparing the two language groups' collapsed performance on the WASI revealed no significant differences between the language background groups [$t(106)=.8$, $p=.183$]; all participants scored within the normal range and the two groups can be considered to be of comparable cognitive skill overall. However, the two language groups did differ significantly on all the additional language measures, other than the WASI and the TOWK-Receptive at year 3. See Table 4.4 for three one-way ANOVA comparisons of the two language groups on all additional measures, including WASI, separated by year group.

Table 4.4: 3 one-way ANOVAs comparing monolingual English speakers and learners with EAL performance on additional measures

Year 3			
	WASI	F(1,30)=1.79, p<.191	
	ACE 6-11	F(1,30)=24.39, p<.000	***
	BPVS	F(1,30)=7.22, p<.012	*
	TOWK-Expressive	F(1,30)=20.13, p<.000	***
	TOWK-Receptive	F(1,30)=4.09, p<.052	
Year 4			
	WASI	F(1,45)=1.23, p<.293	
	ACE 6-11	F(1,45)=16.04, p<.000	***
	BPVS	F(1,45)=24.55, p<.000	***
	TOWK-Expressive	F(1,45)=23.69, p<.000	***
	TOWK-Receptive	F(1,45)=27.08, p<.000	***
Year 5			
	WASI	F(1,27)=.591, p<.45	
	ACE 6-11	F(1,27)=14.04, p<.001	***
	BPVS	F(1,27)=11.91 p<.002	**
	TOWK-Expressive	F(1,27)=8.29, p<.008	**
	TOWK-Receptive	F(1,27)=7.69, p<.010	*

* indicates significance at p<.05, ** p<.005, ***p<.001

The two groups differed significantly from each other at all three year points. In year 3, they differed only on the ACE 6-11, BPVS and TOWK-Expressive. In years 4 and 5, the two groups differed on all additional measures, ACE 6-11, BPVS and TOWK-Expressive and TOWK Receptive. This is in line with previous research that has found learners with EAL to lag behind monolingual peers in vocabulary (Bialystok, 2010; Garcia, 1991; Verhoeven, 1990). Since these learners differed on measures of single word vocabulary in years 3, 4 and 5 (excluding TOWK Receptive in year 3), one might also predict that they would differ on a measure of a different aspect of vocabulary, namely multi-word vocabulary. An analysis of how the groups compared in MPT performance is presented in 4.2.6, after an examination of MPT validity and reliability, so it can be assured that the MPT performance presents insight into multi-word phrase knowledge.

4.2.2 *MPT internal consistency*

Cronbach's alpha is a commonly used statistical measure of internal consistency that offers a scale measurement of test reliability by indicating if low scorers can answer only easier questions correctly and high scorers can answer both easy and difficult test items (Field, 2009). Typically values of above .7 are considered acceptable for ability measures and above .8 are considered acceptable for psychometric assessments (Fulcher & Davidson, 2007). The MPT has an acceptable internal reliability (.85), overall (See Table 4.5). In the interest of being thorough, results are presented for all year group and language background combinations, but the overall sample score for all year groups is the most significant, as some of the groups when separated by language and year group have as few as 12 participants. Therefore, the Cronbach's alpha score for these smaller sample groups should be viewed as purely exploratory.

Table 4.5: Cronbach's alpha results for the MPT for overall sample and language groups

	Year 3	Year 4	Year 5	All
Overall sample	.825	.864	.832	.850
Monolingual English speakers	.864	.89	.832	.825
Learners with EAL	.511	.627	.765	.826

4.2.3 *Test-retest reliability*

A test-retest coefficient is a statistical measure of temporal stability, indicating how much the individual's score has changed during the elapsed period of time between test administrations. Test-retest reliability is obtained by administering the test twice with a set amount of elapsed time in between. Scores from the two administrations are then correlated (Guttman, 1945). For a reliable test, any change in score should be small and reflect either a progression in the ability being measured, improvement caused by the previous test administration or a random change in performance. A smaller sample of 87 participants was selected to take the MPT a

second time, 4 weeks after the first administration, to evaluate temporal stability. Selection was based on convenience and school willingness to participate. See Table 4.6 for descriptive statistics from the first and second MPT administration.

Table 4.6: MPT test-retest descriptive statistics

	N	Mean	SD
First Administration	87	20.8	5.62
Second Administration	87	21.26	4.80

A Pearson's correlation between the first and second test scores of the test-retest group found that the MPT does have acceptable temporal stability (.869**, significant at the $p < 0.001$ level).

4.2.4 *MPT concurrent validity*

Concurrent validity is a method of evaluating how well an instrument measures the target construct by correlating scores with other validated measures of the same or similar skills. As no other tests of *verb + object* phrase knowledge exist for this age group, the previously mentioned vocabulary tests (the ACE 6-11, the BPVS, the TOWK Expressive and Receptive) were administered as measures of similar language skills. The WASI was used primarily as a screening measure; however correlations with the MPT were also explored, since little is presently known about multi-word phrase knowledge among this population of learners. Pearson's correlations between MPT scores and other measures are presented in Table 4.7.

Table 4.7: Correlations between MPT scores and other additional measures for overall sample and language groups

	ACE 6-11	BPVS	TOWK Expressive	TOWK Receptive	WASI
Overall sample	.718**	.594**	.753**	.667**	.499**
Monolingual English speakers	.638**	.690**	.538**	.623**	.390**
Learners with EAL	.614**	.656**	.582**	.651**	.479**
Year 3	.586**	.732**	.494**	.45**	.472**
Monolingual English	.59**	.711**	.4	.62**	.5*

speakers					
Learners with EAL	.45	.6*	.54	-.26	.07
Year 4	.74**	.81**	.68**	.78**	.5**
Monolingual English speakers	.58**	.71**	.53**	.42*	-.08
Learners with EAL	.66**	.69**	.45	.84**	.64**
Year 5	.69**	.6**	.69**	.67**	.27
Monolingual English speakers	.54*	.51*	.47*	.58**	.28
Learners with EAL	.55*	.4	.71**	.54	.19

* Correlation significant at the .05 level ** Correlation is significant at the 0.01 level (2-tailed).

As is apparent from the table, the MPT correlated with the other measures of vocabulary, providing evidence of concurrent validity. Interestingly, there was an unexpected correlation between the MPT and the WASI. This could potentially indicate that the MPT measured more than simple vocabulary knowledge and that multi-word phrase knowledge might have a relationship with other skills shown to be linked to non-verbal IQ measures, for example reading comprehension (Berninger, Abbott, Vermeulen & Fulton, 2006; Tiu, Thompson, & Lewis, 2003). It could also be possible that the novel format of the MPT (the sentence prompt with puzzle box) required some non-verbal reasoning skills to complete. This will be considered in more detail in the Chapter 5: Discussion. Correlations between the three MPT transparency sections and the other measures of vocabulary are listed below in Tables 4.8, 4.9 and 4.10.

Table 4.8: Correlations between MPT transparent scores and vocabulary measures for overall sample and language groups

	ACE 6-11	BPVS	TOWK Expressive	TOWK Receptive	WASI
Overall sample	.552**	.594**	.513**	.583**	.496**
Monolingual English	.380**	.413**	.449**	.414**	.456**

speakers					
Learners with EAL	.431**	.569**	.337*	.507**	.488**

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Table 4.9: Correlations between MPT semi-transparent scores and vocabulary measures for overall sample and language groups

	ACE 6-11	BPVS	TOWK Expressive	TOWK Receptive	WASI
Overall sample	.702**	.670**	.655**	.707**	.450**
Monolingual English speakers	.605**	.553**	.506**	.621**	.326**
Learners with EAL	.642**	.646**	.635**	.658**	.444**

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Table 4.10: Correlations between MPT non-transparent scores and vocabulary measures for overall sample and language groups

	ACE 6-11	BPVS	TOWK Expressive	TOWK Receptive	WASI
Overall sample	.576**	.719**	.550**	.544**	.393**
Monolingual English speakers	.578**	.743**	.446**	.501**	.341**
Learners with EAL	.288	.408**	.400*	.353*	.250

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

The greatest differences in correlations between monolingual English speakers and learners with EAL occurred among the MPT non-transparent correlations. The ACE 6-11 correlated highly with MPT non-transparent scores for monolingual speakers but not at all for learners with EAL. This may point to differences between the two language background groups in non-transparent multi-word phrase knowledge. While for monolingual learners, knowledge of non-literal phrasal verbs correlated with non-transparent multi-word phrase knowledge, there was seemingly no such relationship for learners with EAL.

4.2.5 *Item facility and Item discrimination*

Classical test theory has relied on two statistical measures to determine the appropriateness of test items: Item facility and Item discrimination. Item facility measures the difficulty of a test item, given the participant sample (Bachman, 1990). Item discrimination, also known as Item-total correlation, measures how well a given item discriminates between individual test takers, those who score well on the test and score poorly on the test; in a sense, Item discrimination measures how indicative performance on a given item is of overall test performance (Henning, 1987). The sample of test takers are divided evenly into performance groups and a Pearson Correlation is taken between response to a test item and overall test scores (Henning, 1987). Item facility refers to the percentage of students answering that item correctly; a score approaching 1 for an item would indicate that the item is too easy, while a score approaching 0 would indicate the item is too difficult. Standards for Item discrimination usually dictate that items with scores below .3 or .25, depending on the standard, do not discriminate well between test takers and could potentially be revised (Fulcher & Davidson, 2007). Table 4.11 presents Item facility and Item discrimination results for all MPT test items.

Table 4.11: Item facility and Item discrimination for MPT test items

Item	Item facility	Item discrimination	Item	Item facility	Item discrimination
catch a cold*	0.97	0.26	takes the lead	0.50	0.39
cut a hole	0.86	0.36	cut jobs	0.51	0.36
paid the bill*	0.93	0.22	run a race	0.71	0.37
run a business	0.70	0.49	get a taxi*	0.97	0.15
take sides	0.48	0.48	break the silence	0.47	0.41
take your pick	0.66	0.37	pay the price	0.53	0.44
pay attention	0.70	0.46	carry an umbrella*	0.96	0.30
carry a risk	0.30	0.49	change direction*	0.93	0.29

get the sack	0.42	0.66	break the ice	0.11	0.47
change trains	0.77	0.44	took the money*	0.94	0.15
catch fire	0.65	0.21	carried the day*	0.07	0
cut corners	0.25	0.50	holding hands	0.89	0.41
pay a visit	0.71	0.38	changed hands	0.41	0.33
hold my tongue	0.38	0.44	break a bone	0.86	0.34
run tests	0.44	0.49	get a message*	0.86	0.05
run the show*	0.17	0.09	catch mice	0.83	0.30
hold a conversation	0.40	0.35			

* indicates an item outside Item facility or item discrimination boundaries.

Item facility and Item discrimination indicated that seven items on the MPT were too easy (catch a cold, paid the bill, get a taxi, carry and umbrella, change direction, took the money, get a message) and two items on the test were too difficult (run the show, carried the day). These items have been marked on the above table with an asterisk. According to item response theory, all items in a test affect each other and therefore a test should be viewed holistically (Fulcher & Davidson, 2007). Performance on one test question is affected by the preceding items and in turn affects the subsequent items. This means that the removal of these items would require a re-validation of the entire instrument, something outside the scope of the current project.

In addition, there are compelling reasons for keeping these items in the MPT, particularly the 'too easy' items. Removing these items might have a detrimental effect on the overall test in two ways. First, having easier items, particularly at the beginning of the test, allows the researcher to quickly assess if the child has grasped the concept of the MPT matrix, a novel format. Secondly, young learners can easily become disheartened when taking a test on which they feel they are performing poorly (McKay, 2006). This could be particularly acute with a test like the MPT on which the learner

knows they 'can't find' the answers. Including easier questions might keep learners from giving up or becoming demoralised during the test and consequently putting in less effort. This phenomenon was evident to the researcher frequently throughout test administrations. Learners would manifest increased energy and pace after quickly finding an easy answer.

Regarding the two 'too difficult' items, future research could remove these items and revalidate the instrument or attempt to re-write the sentence prompts to improve them. Additionally, the test might be trailed again, this time with higher performers or Year 6 students. The top three highest scoring students answered these two items correctly, so it is possible that with a larger sample or a sample that includes more high performers these items might discriminate better. Even with these too difficult and too easy items the MPT demonstrated reliability and validity, indicating that the inclusion, while possibly redundant, did not prohibitively affect the test.

4.2.6 *Significant effects: transparency and group (year and language)*

Based on the analyses presented in the previous section, the MPT can be considered a reliable and valid instrument; one that measures what it intends to measure and is suitable for use with this population. The following section, 4.2.6, will explore the differences between the three transparency sections of the MPT (the effect of transparency) and differences between participant groups (language background and year group) using ANOVAs.

Assumption of sphericity

Before comparing MPT performance scores across transparency sections or sample groups using an ANOVA analysis, the data must be found to meet the assumption of sphericity. Mauchly's test indicated that the assumption of sphericity had been met for the main effect of transparency [$\chi^2(2)=3.41$, $p<.181$] but had been violated

for the effect of frequency [$\chi^2(2)=11.12$, $p<.004$] and the interaction between transparency and frequency [$\chi^2(9)=41.69$, $p<.001$]. The degrees of freedom for these effects and interaction effects will be reported using the Greenhouse-Geisser estimate of sphericity, a correction for data that does not meet the assumption of sphericity (Field, 2009). The Greenhouse-Geisser corrected measures for frequency and the interaction between frequency and transparency were quite close to one ($\varepsilon=.906$ and $\varepsilon=.843$, respectively), indicating that the data was close to being spherical, and therefore, that an ANOVA would still be a suitable statistical test. As previously mentioned, non-parametric tests were also used to confirm significance.

Effect of transparency

The first analyses run were to explore the effect of transparency on score accuracy. Scores for transparent, semi-transparent and non-transparent items were compared to determine if increased opacity (less transparency) resulted in items being less accurately answered and, conversely, if increased transparency resulted in more accurately answered items; the independent variable in the analysis was transparency category (transparent, semi-transparent, non-transparent) and the dependent variable was performance accuracy. These means can be found in Table 4.3. For the sample overall, there was a strong effect of transparency [$F(2, 101)=256.38$, $p<.001$, $\eta2= .811$]. Weighted means contrasts showed that transparent items were answered accurately significantly more than semi-transparent items [$F(1,102)=438.32$, $p<.001$, $\eta2=.811$] and semi-transparent items were answered accurately significantly more than non-transparent items [$F(1,102)=110.78$, $p<.001$, $\eta2=.521$]. This provides further evidence that there is indeed a substantive difference between transparent, semi-transparent and non-transparent test items. Transparency did not interact with year group [$F(4, 204)=2.14$, $p=.078$, $\eta2= .04$] or language background [$F(2, 101)=.53$, $p=.59$, $\eta2= .01$];

across all three year groups and for both language groups, transparent items remained more accurately answered than semi-transparent items, which were in turn more accurately answered than non-transparent items.

Differences between language backgrounds

The two language groups' performances on each section of the MPT were compared at each year point using one-way ANOVAs, Bonferroni post-hoc tests and non-parametric Mann-Whitney U tests to confirm significance. See Table 4.12 for differences between language background groups at each of the three year points.

Table 4.12: One-way ANOVAS and Non-Parametric tests comparing MPT performance between monolingual English speakers and learners with EAL across year groups

Year 3					
	ANOVAS				
	Total		F(1,30)=3.643, p<.066		
	Transparent		F(1,30)=1.457, p<.237		
	Semi-transparent		F(1,30)=2.766, p<.107		
	Non-transparent		F(1,30)=3.558, p<.069		
	Non-Parametric Tests	Mann-Whitney U	Z	Exact Sig.	
	Total	73.5	-1.816	.07	
	Transparent	84.00	-1.434	.170	
	Semi-transparent	82.00	-1.495	.146	
	Non-transparent	78.00	-1.680	.107	
Year 4					
	ANOVAS				
	Total		F(1,45)=36.188, p<.000		***
	Transparent		F(1,45)=18.101, p<.000		***
	Semi-transparent		F(1,45)=39.376, p<.000		***
	Non-transparent		F(1,45)=20.105, p<.000		***
	Non-Parametric Tests	Mann-Whitney U	Z	Exact Sig.	
	Total	40.00	-4.679	.000	***
	Transparent	117.00	-3.059	.002	**
	Semi-transparent	44.00	-4.610	.000	***
	Non-transparent	71.00	-4.054	.000	***
Year 5					
	ANOVAS				
	Total		F(1,27)=09.828, p<.004		**
	Transparent		F(1,27)=20.454, p<.000		***
	Semi-transparent		F(1,27)=01.895, p<.169		

Non-transparent		F(1,27)=10.704, p<.003		**
Non-Parametric Tests	Mann-Whitney U	Z	Exact Sig.	
Total	45.00	-2.543	.011	*
Transparent	28.50	-3.482	.001	***
Semi-transparent	77.00	-1.118	.283	
Non-transparent	40.50	-2.757	.005	**

* indicates significance at p<.05, ** p<.005, ***p<.001

As Table 4.12 shows, among year 3 learners, monolingual English speakers did not differ significantly from learners with EAL in any of the transparency categories, when analysed using an ANOVA [total score: F(1,30)=3.643, p<.066; transparent items: F(1,30)=1.457, p<.237; semi-transparent items: F(1,30)=2.766, p<.107; or non-transparent items: F(1,30)=3.558, p<.069] or using a Mann-Whitney U test (total score: p<.07, transparent: p<.170, semi-transparent: p<.146, non-transparent: p<.107). This is particularly interesting when considering that at year 3, monolingual English speakers and learners with EAL performed significantly differently on all the additional other language measures, excluding TOWK Receptive language.

However, in year 4 the two language groups differed in all areas: total score [F(1,45)=36.188, p<.000], transparent [F(1,45)=18.101, p<.000], semi-transparent [F(1,45)=39.376, p<.000] and non-transparent items [F(1,45)=20.105, p<.000]. Finally, in year 5 monolingual English speakers performed significantly more accurately than learners with EAL on the test overall [F(1,27)=09.828, p<.004] as well as transparent items [F(1,27)=20.454, p<.000], and non-transparent items [F(1,27)=10.704, p<.003], but did not differ in performance on semi-transparent items [F(1,27)=01.895, p<.169]. This is in accordance with the differences on other vocabulary measures seen between the language groups and these two year points (see Table 4.4 for reference).

Differences between year groups, separated by language background

Since the language background groups differed significantly in years 4 and 5, between year group differences were examined separately for each language background using one-way ANOVAs, Bonferroni post-hoc tests and non-parametric Kruskal-Wallis tests. It is important to note that while the overall sample was 108, when separated by year and language, some of the sample groups contain as few as 12 participants. Results are therefore purely exploratory and must be interpreted cautiously. See Table 4.13 for comparisons of monolingual English speakers at year 3, 4 and 5.

Table 4.13: One-way ANOVA, Post-Hocs and Non-Parametric tests comparing monolingual speakers' MPT performance across years 3,4 and 5 and across transparency categories.

ANOVA				
Total			F(2,65)=9.046, p<.000	***
Transparent			F(2,65)=8.279, p<.001	***
Semi-transparent			F(2,65)=5.708, p<.005	**
Non-transparent			F(2,65)=8.832, p<.000	***
Bonferroni Post-Hoc				
Total				
Year 3 to 4			p<.026	*
Year 4 to 5			p<.137	
Year 3 to 5			p<.000	***
Transparent				
Year 3 to 4			p<.011	*
Year 4 to 5			p<.490	
Year 3 to 5			p<.001	***
Semi-transparent				
Year 3 to 4			p<.018	*
Year 4 to 5			p<1.00	
Year 3 to 5			p<.010	*
Non-transparent				
Year 3 to 4			p<.714	
Year 4 to 5			p<.005	**
Year 3 to 5			p<.000	***
Non-Parametric Kruskal-Wallis Test	DF	Chi Squared	Sig.	
Total	2	11.512	.003	**
Transparent	2	11.334	.003	**
Semi-transparent	2	5.736	.057	*
Non-transparent	2	13.939	.001	***

* indicates significance at p<.05, ** p<.005, ***p<.001

These further analyses revealed that for monolingual English speakers there was a significant difference in task performance across all transparencies over the 3 different year points. This indicates that monolingual learners in years 3, 4 and 5 differed in total performance [F(2,65)=9.046, p<.000], transparent item performance [F(2,65)=8.279, p<.001], semi-transparent item performance [F(2,65)=5.708, p<.005], and non-transparent item performance [F(2,65)=8.832, p<.000]. This difference between the 3 year groups was further confirmed by non-parametric Kruskal Wallis tests.

To better understand between which year groups these differences occurred, Bonferroni post-hoc tests were conducted. Comparing year 3 monolinguals to year 4 monolinguals, year 4 learners had significantly more accurate performance on the MPT overall (p<.026), transparent items (p<.011) and semi-transparent items (p<.018), however, not on non-transparent items (p<.714). Comparing year 4 and 5, year 5 monolingual learners showed significantly more accurate performance on non-transparent items (p<.005), but not on semi-transparent items (p<1.00) or transparent items (p<.490) or overall MPT score (p<.137).

Learners with EAL presented a differing profile when compared across year groups. See Table 4.14 for one-way ANOVA, post-hoc comparisons and non-parametric tests comparing learners with EAL's MPT performance across year groups 3,4 and 5.

Table 4.14: One-way ANOVA, Post-Hocs and Non-Parametric tests comparing learners with EAL's MPT performance across years 3,4 and 5

ANOVA		
Total	F(2,37)=5.320, p<.009	**
Transparent	F(2,37)=0.687, p<.510	
Semi-transparent	F(2,37)=7.078, p<.002	**
Non-transparent	F(2,37)=5.418, p<.009	**
Bonferroni Post-Hoc		
Total		

Year 3 to 4			p<1.00	
Year 4 to 5			p<.010	*
Year 3 to 5			p<.049	*
Transparent				
Year 3 to 4			p<1.00	
Year 4 to 5			p<.760	
Year 3 to 5			p<1.00	
Semi-transparent				
Year 3 to 4			p<1.00	
Year 4 to 5			p<.002	**
Year 3 to 5			p<.032	*
Non-transparent				
Year 3 to 4			p<1.00	
Year 4 to 5			p<0.10	
Year 3 to 5			p<.045	*
Non-Parametric Kruskal-Wallis Test	DF	Chi Squared	Sig.	
Total	2	7.148	.028	*
Transparent	2	1.078	.583	
Semi-transparent	2	9.681	.008	**
Non-transparent	2	6.662	.036	**

* indicates significance at p<.05, ** p<.005, ***p<.001

The ANOVA found that there was a significant difference between learners with EAL in years 3, 4 and 5 on the MPT overall [F(2,37)=5.320, p<.009], semi-transparent items [F(2,37)=7.078, p<.002] and non-transparent items [F(2,37)=5.418, p<.009] but no difference between transparent items [F(2,37)=0.687, p<.510]. Bonferroni post-hoc tests showed that among learners with EAL, there were no significant performance differences between learners with EAL in year groups 3 and 4, no difference in total scores (p>1.00), transparent item performance (p>1.00), semi-transparent item performance (p>1.00) or non-transparent item perfomance (p>1.00). However, year 5 learners with EAL performed significantly more accurately than year 4 learners on semi-transparent items (p<.002) and year 5 learners with EAL performed significantly more accurately than year 3 learners with EAL on non-transparent items (p<.045). There were no significant differences between learners with EAL in years 3,

4 and 5 on transparent item scores, even through learners were not at ceiling (mean percentage scores out of a hundred of 75, 72.5, 80.83, respectively).

Interaction effects

A mixed repeated measures ANOVA with two between-subjects variables (language background tested at two levels: monolingual English speakers and learners with EAL, and year group tested at three levels: years 3, 4 and 5) and two within-subjects variables (transparency tested at 3 levels: transparent, semi-transparent and non-transparent, and frequency tested at 3 levels: high, mid and low) was conducted to explore possible interaction effects. All findings should be considered exploratory; effects are significant at $p<.05$.

Frequency descriptive statistics

Though frequency was not intended as a variable and was not intentionally manipulated, it was considered in the analysis to ensure that a possible frequency effect was not overlooked. Means and standard deviations for MPT performance by frequency category are presented again for reference in Tables 4.15, 4.16 and Graphs 4.17, 4.18.

Table 4.15: MPT score means presented by frequency category and across year group

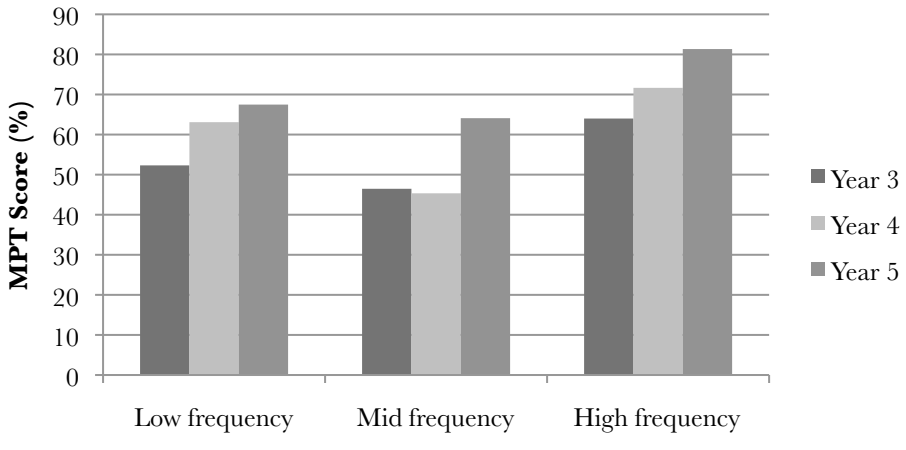
Overall Sample	N	Mean	Standard Deviation
Low frequency	108	59.10	17.49
Mid frequency	108	59.19	22.88
High frequency	108	68.54	20.54
Year 3	N	Mean	Standard Deviation
Low frequency	32	52.31	17.26
Mid frequency	32	46.47	20.65
High frequency	32	64.0	19.49
Year 4	N	Mean	Standard Deviation

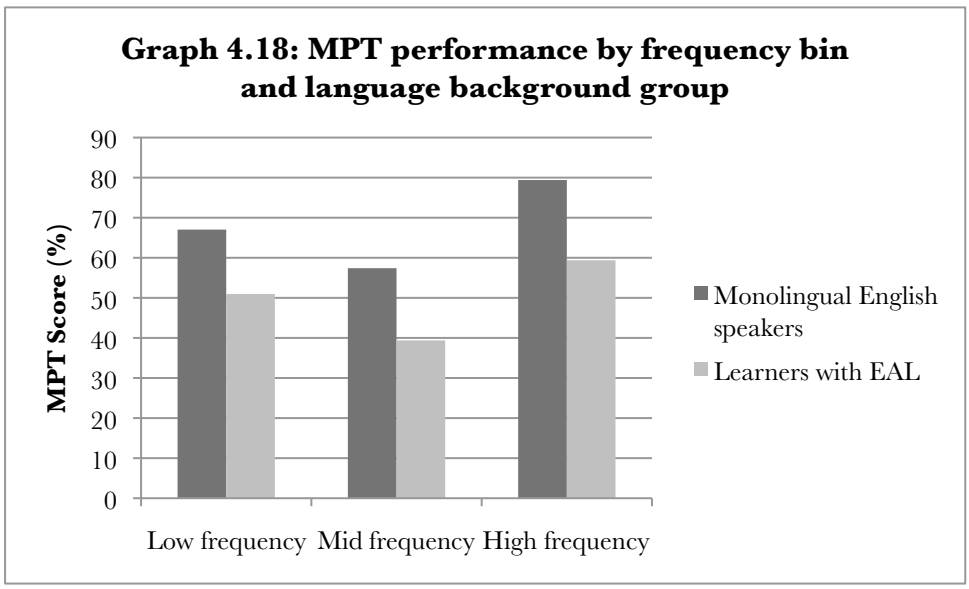
Low frequency	47	63.09	17.59
Mid frequency	47	45.34	20.5
High frequency	47	71.66	21.82
Year 5	N	Mean	Standard Deviation
Low frequency	29	67.48	13.93
Mid frequency	29	64.09	24.17
High frequency	29	81.34	15.68

Table 4.16: MPT score means presented by frequency category and by language background

Monolingual English speaker	N	Mean	Standard Deviation
Low frequency	68	67.03	15.03
Mid frequency	68	57.4	22.27
High frequency	68	79.4	16.43
Learners with EAL	N	Mean	Standard Deviation
Low frequency	40	50.95	16.87
Mid frequency	40	39.38	19.32
High frequency	40	59.38	20.85

Graph 4.17: Overall sample MPT performance by frequency bin and year group





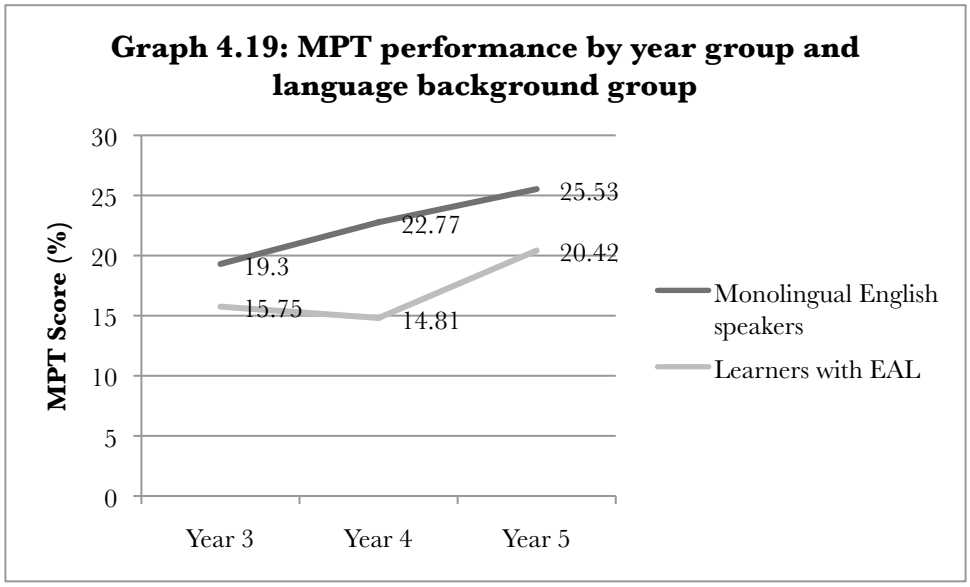
Mixed repeated measures ANOVA

A significant interaction effect occurred between transparency and year group [F(4, 204)=2.59, p<.038, η2=.048] but not between transparency and language background [F(2, 204)=.457, p<.634], indicating that language background did not have an impact on the transparency effect, and that non-transparent items remained the most difficult, followed by semi-transparent and transparent items. There was no interaction between transparency, year group and language background [F(4,204)= 1.025, p<.395].

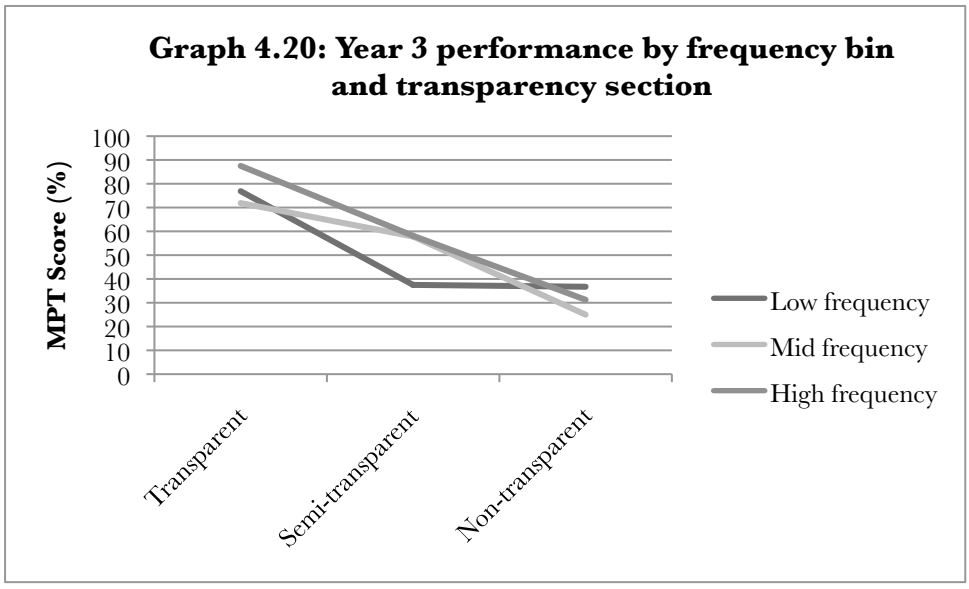
There was also a significant effect for frequency [F(1.81, 184.74)=21.23, p<.001, η2=.172], though much weaker than the effect of transparency. Contrasts showed that high frequency items are more accurately answered than mid frequency items [F(1,102)=46.151, p<.001, η2=.312] and mid frequency items are more accurately answered than low frequency items [F(1,102)=25.31, p<.001, η2=.199]. This effect validates the construct of frequency as a binned category and indicates that there is a significant difference between the three frequency categories used in the current study.

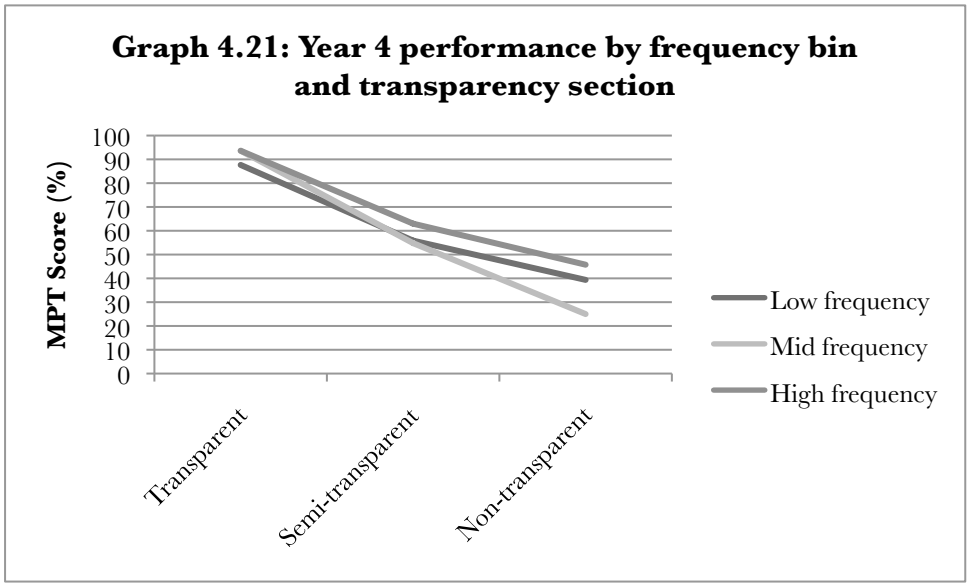
There was a significant effect for year group [F (2,102)=10.2, p<.001, η2=.167], but no effect of language background [F(1,102)=1.79, p<.184]. There was, however, a significant interaction between year group and language background [F(2,102)=12.17,

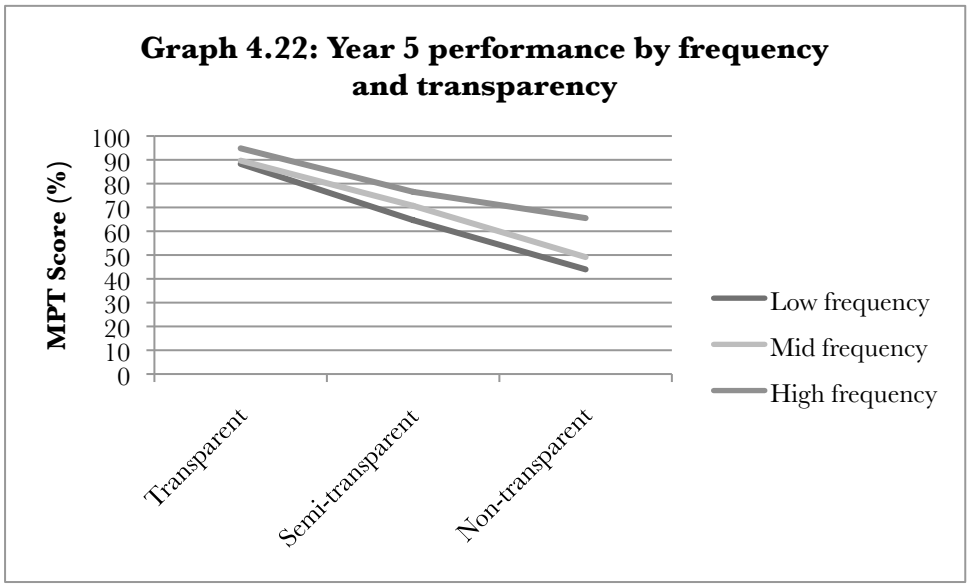
p<.001, η2=.193]. Graph 4.19 presents MPT scores by language background and year group.



There was a significant interaction between transparency and frequency [F(3.372, 343.910)=3.49, p<.012, η2=.03] as well as between transparency, frequency and year group [F(6.743,343.91)=3.635, p<.001, η2=.067] and a smaller four way interaction between transparency, frequency, year group and language background [F(6.743,343.91)=2.62, p<.013, η2=.049]. Graphs 4.20, 4.21 and 4.22 present MPT scores by transparency, frequency and year group.







No significant interaction was found between frequency and year group [F(3.622, 404.66)=.907, p<.454], nor between frequency and language background [F(1.811, 184.74)=2.029, p<.139] indicating that low frequency items remained the least accurately answered and high frequency the most accurately answered, regardless of year group or language background. There were no three way interactions between frequency, year group and language background [F(3.622,266.13)=.596, p<.649] nor between transparency, frequency and language background [F(3.372,343.91)=2.10, p<.092].

The ANOVA analyses revealed three significant main effects (transparency, frequency and year group) and five significant interactions (transparency x frequency, year group x language background, transparency x year group, transparency x frequency x year group, transparency x frequency x year group x language background). The strongest effect by far was that of transparency [$\eta 2 = .715$], providing evidence of MPT construct validity. There was a clear difference in performance accuracy level between the transparent, semi-transparent and non-transparent items. The effect of frequency was less pronounced; findings showed that the items that occur more frequently in the BNC are easier for learners to produce on the MPT, although transparency plays a more important role in item difficulty. The MPT also discriminated between age groups, as evidenced by the effect of year group. There was no significant effect of language background on MPT performance, indicating that monolingual English speakers and learners with EAL answered the same items more or less accurately, respectively, and did not differ significantly in MPT performance overall, although the performance means for learners with EAL were lower than those of the monolingual learners.

Transparency, frequency and year

To further explore the interaction between transparency, frequency and year, a one-way ANOVA with post-hoc Bonferroni t-test comparisons was conducted to explore how participant performance varied by year and by transparency. Results of the Bonferroni comparisons are presented in the Table 4.23.

Table 4.23: Bonferroni post-hoc mean comparisons between year groups for overall sample

		Years 3 and 4	Years 4 and 5	Years 3 and 5
Transparent	Low	.079	1.00	.106
	Mid	.018*	1.00	.127
	High	.256	1.00	.18
Semi-transparent	Low	.017*	.59	.001*
	Mid	1.00	.024*	.139

	High	1.00	.14	.041*
Non-transparent	Low	1.00	.673	.238
	Mid	1.00	.002*	.005*
	High	.237	.062	.001**

* p is significant at .05, ** p is significant at .001

Years 3 and 4 performed significantly differently from each other only on transparent mid frequency items and semi-transparent low frequency items (p=.018 and p=.017, respectively). Years 4 and 5 differed significantly from each other on semi-transparent mid frequency items (p=.024) and non-transparent mid frequency items (p=.002). The difference in performance on non-transparent high frequency items between year 4 and year 5 was approaching significance (p=.062). When comparing years 3 and 5, there was a significant difference in performance on semi-transparent low frequency items (p=.001), semi-transparent high frequency items (p=.041), non-transparent mid frequency items (p=.005) and non-transparent high-frequency items (p=.001).

In addition, a one-sample t-test was conducted to explore performances differences between the mean scores from the three transparency categories and frequency bins, without considering year group. A Bonferonni correction set p<.006 as the threshold for significance. All 9 categories of transparency by frequency differed significantly from one another. Means and t-tests from this comparison are presented in Table 4.24.

Graph 4.24: T-test comparisons for overall sample performance separated by transparency and frequency

Transparency and Frequency		Mean	Standard deviation	t(107)=
Transparent	Low	84.63	21.29	41.32**
	Mid	86.11	34.74	25.76**
	High	92.13	15.53	61.65**
Semi-transparent	Low	52.78	30.25	18.13**
	Mid	59.95	25.56	24.37**
	High	65.19	29.24	23.17**
Non-transparent	Low	39.81	16.05	25.79**
	Mid	31.48	30.64	10.68**

High	46.76	37.61	12.9**

** indicates significance at the p<.001 level

Overall, transparent items were most accurately answered, followed by semi-transparent and non-transparent items; within each transparency category, high frequency items were most accurately answered, followed by mid frequency and, finally, low frequency items. Only the non-transparent items deviated from the pattern of the highest score mean for high frequency items, followed by mid frequency items and finally, low frequency items with the lowest mean. Low frequency non-transparent items showed a higher mean score (among this sample) than mid frequency non-transparent items. One potential explanation for this deviation could be related to which multi-word items were classified as non-transparent low frequency. There are only between 1 and 4 multi-word items in each of the 9 categories, so one multi-word item could impact the overall score. In this case, one of the multi-word items found to be 'too easy' according to the Item facility analysis, *catch a cold*, was classified as a low frequency multi-word item, based on number of occurrences in the BNC. This may be masking the true nature of the multi-word phrase's frequency for these participants, as the item is clearly well known to primary school learners and might be higher frequency in child directed speech, for example. Unfortunately, CHILDES does not allow a frequency search for whole multi-word items. If this item were removed completely, the mean score for low frequency non-transparent items would have been 20.68 and non-transparent items would also follow the high, mid, low frequency difficulty pattern. This finding points to a need for future multi-word vocabulary research with children that would take into account potential differences in item frequency between adult speech and child directed speech and further supports calls for more extensive corpora of child language and child directed language.

4.2.7 *Phase 2 conclusions*

The goal of Phase 2 of the current study was to address research questions 1 and 2. Firstly, can multi-word *verb + object* phrase knowledge be measured among monolingual English speaking children and in learners with EAL between the ages of 7 and 10? Secondly, if it indeed can be measured, what is the nature of this knowledge among these learners? Findings from the present study indicated that *verb + object* phrase knowledge can indeed be measured among this learner group. Results showed that the MPT is reliable and valid, based on analyses of test-retest reliability, internal reliability, concurrent validity and construct validity. The MPT yielded acceptable outcomes for both test-retest reliability and internal reliability; it also significantly correlated with many of the additional measures of single word vocabulary. Analysis showed that the MPT distinguished adequately between year groups – ANOVAs revealed that among both monolingual learners and learners with EAL there was a significant difference in overall MPT score between year groups 3, 4 and 5. There was a strong effect of transparency, indicating that there is a substantial difference in the challenge posed by transparent, semi-transparent and non-transparent multi-word phrases. This finding provides further evidence that opaque phrases present a disproportionate challenge to young learners when compared to more transparent phrases, even when accounting for phrase frequency.

A further goal of the study was to identify whether MPT findings would vary across year group (i.e., is there a developmental change across 3 years of primary school corresponding to ages 7 years though 10 years) and whether monolingual children and children with EAL would differ on MPT performance at the three data collection time points. When comparing results from monolingual English speakers to those of learners with EAL, findings showed that learners in year 3 did not differ significantly on any of the transparency sub-sections, although the difference in overall

task scores was approaching significance. Among year 4 children, the two language groups differed in all areas; in year 5, monolingual English speakers performed significantly more accurately in all transparency categories except for semi-transparent items. These significant differences in multi-word phrase knowledge between the two language background groups are in accordance with the significant differences in performance between the two groups in years 4 and 5 on the additional language measures administered, as would be expected. However, this is not case with year 3 learners. At year 3, the two language background groups do not differ significantly in multi-word phrase knowledge but do differ on all other language measures. This difference is discussed in greater depth in Chapter 5: Discussion.

In addition, there were significant differences between the two language groups at years 4 and 5; a significant difference in transparent and semi-transparent item knowledge for monolingual English speakers between years 3 and 4, and a difference in non-transparent item knowledge between years 4 and 5, while transparent and semi-transparent were not different. It would appear from these findings that between years 3 and 4, monolingual English speakers have greater knowledge of transparent and semi-transparent multi-word phrases, without a corresponding increase in their knowledge of non-transparent phrases.

By contrast, learners with EAL in year 4 did not perform significantly more accurately than learners in year 3 on any item category. Learners with EAL in year 5 did perform significantly more accurately than learners in year 4 on semi-transparent and non-transparent test items, with no significantly higher scores on transparent items between any of the three different year points. The differences between the two language background groups and the lack of difference in transparent item accuracy

between the three year points among learners with EAL will be addressed in more detail in Chapter 5: Discussion.

4.3 *Phase 3 results: Relationships with reading*

For phase 3 of the current study, the MPT was administered to 40 year 4 learners, 20 monolingual English speakers and 20 Bengali speaking learners with EAL, along with two vocabulary measures (the TOWK Expressive and TOWK Receptive) and a reading comprehension measure (the YARC) containing 4 reading element subtests (Reading accuracy, Reading comprehension, Reading rate and Single word reading). This battery of assessments was administered with the goal of exploring a possible relationship between reading and multi-word phrase knowledge.

4.3.1 *Descriptive statistics*

Normality of the data

The importance of normality and homogeneity of variance has been previously discussed in section 4.2.1. None of the measures used in Phase 3 violate both the assumption of normality and the assumption of homogeneity of variance, therefore it is acceptable to continue to analyse the data using parametric tests. See Table 4.25 for tests of normality and homogeneity of variance among Phase 3 measures.

Table 4.25: Tests of normality (skew and kurtosis) and homogeneity of variance for Phase 3 data

	Normality		Homogeneity of Variance	
	Skew	Kurtosis	Levene's Statistic	Significance
YARC Single word reading	-1	.045	1.34	.208
YARC Accuracy	1.07	1.94	.001	.970
YARC Reading Rate	4.62**	5.87**	1.60	.288
YARC Reading Comprehension	2.59	1.55	2.69	.109
TOWK Expressive	1.02	.17	2.31	.14
TOWK	.002	1.01	1.98	.17

Receptive				
WASI	.80	1.16	.917	.344
MPT	1.26	.61	.094	.76
MPT Transparent	3.82**	3.48**	.39	.54
MPT Semi-transparent	.44	.56	.88	.35
MPT Non-transparent	.04	.61	1.24	.72

** indicates significant deviation at p<.001

Participant Means

Tables 4.26 and 4.27 below detail participant performance on the administered measures and the results of a one-way ANOVA comparing scores on all administered measures between the two language background groups. As Table 4.26 shows, there were differences between monolingual English speakers' and learners with EAL's performances. Monolinguals had a mean overall performance of 20.2 (out of 33) and learners with EAL had a mean overall performance of 14.6, significantly different means [F(1,39)=11.12, p<.002]. The two language background groups also had significantly different performances in all three transparency categories, transparent [F(1,39)=8.58, p<.006], semi-transparent [F(1,39)=11.09, p<.002] and non-transparent items [F(1,39)=5.48, p<.03]. In addition, the two language groups differed significantly in performance on the two administered single word vocabulary assessments, the TOWK Expressive [F(1,39)=28.86, p<.000] and Receptive [F(1,39)=5.13, p<.03]. The two language groups did not differ significantly on the WASI [F(1,39)=.47 p<.5], indicating they are of comparable overall cognitive ability. Interestingly, the two language groups also did not differ in performance on three of the four reading assessments, the YARC Single word reading [F(1,39)=.044, p<.835], YARC Accuracy [F(1,39)=1.26, p<.268] and YARC Reading rate [F(1,39)=3.91, p<.055], although the difference between the groups in YARC Reading rate is approaching signficance. The ANOVA did show significant differences between the two language background groups

in one element of the YARC reading assessment, however, the YARC Reading comprehension [F(1,39)=9.42 p<.004].

Table 4.26: Phase 3 MPT mean scores (overall score out of 33 and transparency scores in percentage) and standard deviations by language background

	Overall Group		Monolingual English speakers		EAL	
	Mean	SD	Mean	SD	Mean	SD
MPT	17.40	5.96	20.2	4.95	14.6	5.65
MPT Transparent	75.00	21.96	85.00	16.70	65.00	22.36
MPT Semi-transparent	51.1	22.05	60.75	20.71	41.45	19.32
MPT Non-transparent	33.25	16.39	38.00	14.36	28.5	17.25
TOWK Expressive	15.23	3.2	17.3	1.98	13.15	2.83
TOWK Receptive	20.28	4.33	21.75	3.27	18.8	4.82
YARC Reading accuracy	43.95	9.32	45.6	9.73	42.3	8.83
YARC Reading comp	53.95	10.23	58.45	7.07	49.45	11.05
YARC Reading rate	59.83	18.15	65.3	12.59	54.35	21.32
YARC Single word	35.05	10.40	35.4	8.76	34.7	12.04
WASI	14.2	5.96	14.85	5.43	13.55	6.52

Table 4.27: One-way ANOVAs comparing monolingual English speakers and learners with EAL performance on all measures

WASI	F(1,39)=.47 p<.5	
TOWK-Expressive	F(1,39)=28.86, p<.000	* * *
TOWK-Receptive	F(1,39)=5.13, p<.03	*
MPT	F(1,39)=11.12, p<.002	* *
Transparent items	F(1,39)=8.58, p<.006	* *
Semi-transparent items	F(1,39)=11.09, p<.002	* *
Non-transparent items	F(1,39)=5.48, p<.03	*
YARC Single word reading	F(1,39)=.044, p<.835	
YARC Accuracy	F(1,39)=1.26, p<.268	
YARC Reading Rate	F(1,39)=3.91, p<.055	

YARC Reading comprehension	F(1,39)=9.42 p<.004	**

* indicates significance at p<.05, ** p<.005, ***p<.001

4.3.2 *Correlations between measures*

To better understand the relationships between the various measures used in Phase 3, Pearson's correlations were run between all the administered measures. None of the predictor variables correlated with each other at a value of greater than .8, indicating there was not an issue with multi-collinearity and all variables could be viewed as distinct (Field, 2009). However, the MPT overall and MPT subcomponent scores do obviously present a potential issue of multi-collinearity, as the MPT overall score is a composite of these subcomponent scores. This issue is addressed in the following section. Tables 4.28, 4.29 present these correlations between the administered measures.

Table 4.28: Correlations between administered measures for overall learner sample

	TOWK Exp.	TOWK Rec.	YARC Acc.	YARC Reading comp	YARC Reading rate	YARC Single word	WASI
MPT	.634**	.288	.396*	.706**	.629**	.620**	.416**
MPT Transparent	.593**	.217	.512**	.586**	.471**	.381*	.143
MPT Semi-transparent	.532**	.406**	.365*	.649**	.491**	.489**	.266
MPT Non-transparent	.313*	.334*	.392*	.659**	.711**	.600**	.308
TOWK Expressive		.275	.396*	.513**	.298	.256	.218
TOWK Receptive	.275		.030	.158	.240	.104	.170
YARC Accuracy	.396*	.030		.624**	.582**	.769**	.255
YARC Read. comp.	.513**	.158	.624**		.791**	.711**	.042

YARC Reading rate	.298	.240	.582**	.791**		.744**	.191
YARC Single Word	.256	.104	.769**	.711**	.744**		.290
WASI	.218	.170	.255	.042	.191	290	

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed)

Many of the measures correlated highly with one another as would be expected, particularly the different components of the YARC. The WASI correlated only with the MPT, something also found and discussed in the previous section.

Table 4.29: Correlations between administered measures presented by language background group

Monoling. / Eng. / EAL	TOWK Exp.	TOWK Rec.	YARC Acc.	YARC Reading comp	YARC Reading rate	YARC Single word	WASI
MPT	.731** / .339	.153 / .150	.305 / .681**	.381 / .702**	-.070 / .714**	.517 / .799**	.442 / .398
MPT Transparent	.590** / .353	.351 / -.073	.042 / .565**	.381 / .527*	-.070 / .598**	-.054 / .665**	-.020 / .182
MPT Semi-transparent	.522* / .247	.211 / .380	.183 / .557*	.545* / .606**	.074 / .662**	.356 / .683**	.176 / .307
MPT Non-transparent	.560* / -.060	.369 / .199	.333 / .466*	.554* / .658**	.402 / .681**	.538* / .667**	.117 / .410
TOWK Expressive		.378 / -.079	.294 / .454*	.510* / .245	.297 / .071	.315 / .309	.308 / .129
TOWK Receptive	.378 / -.079		.135 / -.163	.583** / -.245	.477 / .020	.332 / -.017	-.109 / .284
YARC Accuracy	.294 / .454*	.135 / -.163		.500* / .738**	.691** / .522*	.728** / .844**	.173 / .307
YARC Read. Comp.	.510* / .245	.583** / -.245	.500* / .738**		.690 / .798**	.602** / .862**	.133 / .175
YARC Reading rate	.297 / .071	.477* / .020	.691** / .522*	.690 / .798**		.851** / .742**	.217 / -.094
YARC Single Word	.315 / .309	.332 / -.017	.728** / .844**	.602** / .862**	.851** / .742**		.302 / .281
WASI	.308 / .129	-.109 / .284	.173 / .307	.133 / .175	.217 / -.094	.302 / .281	

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed)

There were some differences in correlations between the two language background groups. Among monolingual speakers, there was a very high correlation between both the MPT overall, and its transparency components, and the TOWK

Expressive. By contrast, among the learners with EAL there was no significant correlation between the TOWK Expressive and the MPT or any of its sub-sections. This will be considered in more depth in Chapter 5: Discussion. Interestingly, the YARC measures correlated highly with the MPT and the transparency subsections for learners with EAL. For monolingual English speakers, the YARC accuracy and the YARC reading rate did not correlate with the MPT or any of its subsections; the YARC comprehension test correlated only with the MPT semi-transparent and MPT non-transparent, and the YARC single word reading correlated with the MPT non-transparent.

4.3.3 *Regression models*

The goal of the final research phase (Phase 3) was to explore the relationship between MPT performance and YARC outcomes. To do so, the current study used multiple regression analysis. Multiple regression analysis is a statistical method that uses a set of independent variables to predict a dependent outcome variable. There are three commonly used types of regression models: simultaneous, stepwise and hierarchical – the method used in the current study. In a simultaneous regression, all predictors are used together in the analysis; for a stepwise analysis a statistical model chooses which predictors to include in which order and for a hierarchical regression, the researcher can control how predictors are included, usually based on theoretical motivation (Cohen, 2001). Hierarchical multiple regression analysis allows the researcher discretion over the order variables are entered; the relative import of a particular variable can then be measured based on how much it contributes to outcome prediction after the earlier predictors are accounted for (Cohen, 2001). This allows the researcher to see the amount of prediction accounted for by a late entered variable above and beyond those previously entered. The first regression shows the researcher

the variance accounted for by one set of independent variables, the next will include the original and the new, showing the contribution of the second over the first.

Hierarchical regression is appropriate for situations like the current study in which the independent variables are correlated with each other, so that the effect of one variable (for example, non-transparent multi-word knowledge) can be seen while controlling for others. This is accomplished by accounting for the incremental change in the adjusted R squared at the inclusion point for each variable. Hierarchical regression is preferable over stepwise regression that has been the target of much criticism (Cohen, 2001). Statistical software programs like the one used in the current study, SPSS, can misrepresent the number of statistical tests run for the regression model and report the degrees of freedom too low, consequently potentially over-reporting statistical significance (Cohen, 2001). Additionally, there is potential for misreporting the importance of a predictor variable that is statistically dependent on another variable already in the analysis. For these reasons, hierarchical multiple regression analysis is used in the present analysis.

Hierarchical regression models were run to examine the amount of unique variance in YARC outcomes that could be predicted by the MPT overall and by each of the MPT transparency sub-sections. Hierarchical models were used to find the unique contribution of the predictor variables to variance in YARC outcome scores. Separate models were run for each of the 4 YARC outcome measures (YARC Single word reading, YARC Accuracy, YARC Reading rate, YARC Reading comprehension). The analysis provides a standardised Beta co-efficient, which identifies the relationship between each predictor variable and the outcome measure when the influence of the other variables has been accounted for. The higher the Beta coefficient, the stronger the relationship between the predictor and the outcome. The regressions presented

also report how much outcome variance is explained by a predictor after the effect of all other predictors has been controlled for. The additional variance that each predictor accounts for is reported as the R squared change for that variable; R-squared is the explained variation over the total variation. Because of the relatively small size of the sample, 40 participants, adjusted r squared will be used rather than r squared throughout these analyses. The overall MPT score could not be used in the regression models along with the sub-section scores, as this would present an issue with multi-collinearity. The MPT overall score is made up of the three section scores so this would be, in a sense, counting the same thing twice (Field, 2009). Instead, separate models were run for the MPT overall and for the individual section scores, allowing the researcher to separate out transparency and explore the unique contribution of each specific section. Previous analyses presented above found that participants perform significantly differently on the subsections and so this analysis might provide more specific information about which aspects of the MPT have a relationship with reading comprehension. Each model shows the predictor variable in the last step of the hierarchy, after the additional measures that correlate with it (for example, WASI, TOWK, language background) have been controlled for. Language background is coded as a binary measure; due to the way language background was coded, the negative beta co-efficient for Language background indicates that learners with EAL performed more poorly on the outcome variable, hence the negative variance.

4.3.3.1 *MPT overall contribution to YARC*

The following regression models show the unique contribution of the MPT overall score as a predictor variable for YARC score outcomes. Language background, WASI score and TOWK Expressive scores were entered as separate steps in the

regression models before the predictor measures. See Table 4.30 for the regression analysis of the MPT overall for YARC subtest outcomes.

Table 4.30: **Regression analysis for MPT overall contribution to variance in YARC subtest outcomes**

Variables entered at each step	Standardised β	Adjusted R-Squared	F change	Sig. F change	
YARC Single word					
Step 1: Language Background	.034	-.025	.044	.835	
Step 2: WASI	.29	.035	3.352	.075	
Step 3: TOWK Expressive	.351	.081	2.856	.10	
Step 4: MPT Overall	.79	.41	21.29	.000	***
YARC Accuracy					
Step 1: Language Background	.18	.007	1.26	.268	
Step 2: WASI	.24	.04	2.27	.140	
Step 3: TOWK Expressive	.45	.13	4.86	.034	
Step 4: MPT Overall	.434	.211	4.73	.037	*
YARC Reading rate					
Step 1: Language Background	.31	.07	3.91	.055	
Step 2: WASI	.008	.09	.003	.958	
Step 3: TOWK Expressive	.176	.11	.69	.412	
Step 4: MPT Overall	.851	.48	24.76	.000	***
YARC Reading comprehension					
Step 1: Language Background	.45	.18	9.42	.004	**
Step 2: WASI	.143	.18	.96	.333	
Step 3: TOWK Expressive	.37	.23	3.69	.063	
Step 4: MPT Overall	.11	.47	17.13	.000	***

* indicates significance at p<.05, ** p<.005, ***p<.001

The total adjusted r-squared showed that the models overall accounted for 41% of the variance in YARC Single word reading outcomes; 21% of the variance in YARC Accuracy outcomes; 48% of the variance in YARC Reading rate outcomes and 47% of

the variance in YARC Reading comprehension. The MPT overall score accounts for a significant contribution to all of the YARC outcomes. After controlling for other variables, the MPT uniquely accounted for (change in r-squared) 33% of variance in Single word reading, 8% of variance in Accuracy, 37% of variance in Reading rate and 24% of variance in Reading comprehension.

4.3.3.2 *MPT Transparent contribution to YARC*

The following regression models show the unique contribution of the MPT transparent section score to variance in YARC score outcomes. Language background and TOWK Expressive scores were entered as separate steps in the regression models. See Table 4.31 for a regression models showing MPT transparent score contribution to variance in YARC outcomes.

Table 4.31: **Regression analysis for MPT transparent score contribution to variance in YARC subtest outcomes**

Variables entered at each step	Standardised β	Adjusted R-squared	F change	Sig. F change	
YARC Single word					
Step 1: Language Background	.7	-.025	.044	.835	
Step 2: TOWK Expressive	.34	.049	3.95	.054	
Step 3: MPT Transparent	.304	.302	14.4	.001	***
YARC Accuracy					
Step 1: Language Background	.18	.007	1.26	.27	
Step 2: TOWK Expressive	.49	.123	6.06	.02	
Step 3: MPT Transparent	.35	.184	3.74	.061	
YARC Reading rate					
Step 1: Language Background	.31	.07	3.91	.055	
Step 2: TOWK Expressive	.17	.06	.7	.409	
Step 3: MPT Transparent	.63	.31	14.17	.001	** *
YARC Reading comprehension					
Step 1: Language	.45	.18	9.42	.004	**

Background					
Step 2: TOWK Expressive	.39	.25	4.41	.043	*
Step 3: MPT Transparent	.43	.35	6.96	.012	*

* indicates significance at p<.05, ** p<.005, ***p<.001

The total adjusted r-squared showed that the model accounted for 30% of the variance in YARC Single word reading outcomes; 18% of the variance in YARC Accuracy outcomes; 31% of the variance in YARC Reading rate outcomes and 35% of the variance in YARC Reading comprehension. The MPT transparent score made a significant contribution to variance in YARC Single word reading, Reading rate, Reading comprehension. After controlling for other variables, 25% of the variance in Single word reading, 25% of the variance in Reading rate and 10% of the variance Reading comprehension can be uniquely explained by MPT transparent scores.

4.3.3.3 *MPT Semi-transparent contribution to YARC*

The following regression models show the unique contribution of the MPT semi-transparent score as a predictor variable for YARC score outcomes. Language background, TOWK Expressive and TOWK Receptive scores were entered as separate steps in the regression models before the predictor measures. See Table 4.32 for regression models showing MPT semi-transparent score contribution to variance in YARC subtest outcomes.

Table 4.32: **Regression analysis for MPT semi-transparent score contribution to variance in YARC subtest outcomes**

Variables entered at each step	Standardised β	Adjusted R-squared	F change	Sig. F change	
YARC Single word					
Step 1: Language Background	.03	-.025	.04	.84	
Step 2: TOWK Expressive	.41	.05	3.95	.05	*
Step 3: TOWK Receptive	.08	.03	.24	.63	

Step 4: MPT					
Semi-transparent	.67	.29	14.36	.001	***
YARC Accuracy					
Step 1: Language					
Background	.18	.007	1.26	.268	
Step 2: TOWK					
Expressive	.49	.123	6.06	.02	**
Step 3: TOWK					
Receptive	-.063	.103	.151	.7	
Step 4: MPT					
Semi-transparent	.443	.206	5.67	.023	*
YARC Reading rate					
Step 1: Language					
Background	.306	.07	3.91	.055	
Step 2: TOWK					
Expressive	.193	.06	.7	.409	
Step 3: TOWK					
Receptive	.172	.06	.75	.39	
Step 4: MPT					
Semi-transparent	.144	.242	9.85	.003	**
YARC Reading comprehension					
Step 1: Language					
Background	.45	.18	9.42	.004	**
Step 2: TOWK					
Expressive	.39	.25	4.41	.04	*
Step 3: TOWK					
Receptive	-.05	.23	.01	.92	
Step 4: MPT					
Semi-transparent	.58	.4	.2	.001	***

* indicates significance at p<.05, ** p<.005, ***p<.001

The total adjusted r squared showed the model accounts for 39% of the variance in YARC Single word reading outcomes; 21% of the variance in YARC Accuracy outcomes; 24% of the variance in YARC Reading rate outcomes and 40% of the variance in YARC Reading comprehension. The MPT semi-transparent score made a significant contribution to the models for YARC Single word reading, Accuracy, Reading rate, Reading comprehension. After controlling for other variables, 26% of the variance in Single word reading outcomes can be uniquely explained by MPT semi-transparent, as well as 11% of Accuracy scores, 18% of Reading rate and 17% of Reading comprehension.

4.3.3.4 *MPT Non-transparent contribution to YARC*

The following regression models show the unique contribution of the MPT non-transparent score as a predictor variable for YARC score outcomes. Language background, TOWK Expressive and TOWK Receptive scores were entered as separate steps in the regression models before the predictor measures. See Table 4.33 for regression analysis of MPT non-transparent score contribution to variance in YARC subtest outcomes.

Table 4.33: **Regression analysis for MPT non-transparent score contribution to variance in YARC subtest outcomes**

Variables entered at each step	Standardised β	Adjusted R-squared	F change	Sig. F change	
YARC Single word					
Step 1: Language Background	.034	-.03	.044	.84	
Step 2: TOWK Expressive	.412	.05	3.95	.05	*
Step 3: TOWK Receptive	.082	.03	.24	.628	
Step 4: MPT Transparent	.622	.327	16.92	.000	***
YARC Accuracy					
Step 1: Language Background	.18	.007	1.26	.27	
Step 2: TOWK Expressive	.49	.123	6.06	.02	*
Step 3: TOWK Receptive	-.063	.103	.151	.7	
Step 4: MPT Transparent	.49	.281	9.95	.003	**
YARC Reading rate					
Step 1: Language Background	.31	.07	3.91	.055	
Step 2: TOWK Expressive	.17	.11	.7	.409	
Step 3: TOWK Receptive	.14	.13	.75	.392	
Step 4: MPT Transparent	.49	.31	9.32	.004	**
YARC Reading comprehension					
Step 1: Language Background	.45	.18	.2	.004	**
Step 2: TOWK Expressive	.39	.25	.09	.04	*
Step 3: TOWK	-.016	.23	.00	.92	

Receptive					
Step 4: MPT					
Transparent	.5	.41	.19	.001	***

* indicates significance at p<.05, ** p<.005, ***p<.001

The total adjusted r-squared showed this model accounts for 33% of the variance in YARC Single word reading outcomes; 28% of the variance in YARC Accuracy outcomes; 31% of the variance in YARC Reading rate outcomes and 41% of the variance in YARC Reading comprehension. The MPT non-transparent score made a significant contribution to the model; when controlling for all other variables, the MPT non-transparent score contributed 30% of the variance in Single word reading, 18% of the variance in Accuracy, 18% of Reading rate and 18% of Reading comprehension.

Other measures

Regression models were also run for the other variables (language background, WASI, TOWK Expressive and TOWK Receptive). These models are presented in Appendix L. Language background did not account for a significant amount of variance in any YARC subtest scores. The TOWK Expressive and TOWK Receptive also did not account for a significant amount of variance on any of the YARC outcomes. The WASI, however, accounted for a significant amount of variance in YARC Reading rate score, when all other variables were controlled for, contributing 7% of the variance.

4.3.4 *Phase 3 conclusions*

The final phase of the current research project explored a possible relationship between reading, as measured by YARC performance, and *verb + noun* multi-word phrase knowledge, as measured by the MPT, among year 4 monolingual English speakers and Bengali speaking learners with EAL. Regression analyses showed

that the MPT as a whole accounted for a significant amount of variation in reading outcomes for all 4 YARC measures. Relationships between reading and the different transparency subcomponents of the MPT were also explored. MPT transparent scores were found to account for a significant amount of variance in YARC Single word reading, YARC Reading rate, and YARC Reading comprehension. MPT semi-transparent scores accounted for a significant amount of variance in all 4 YARC scores and MPT non-transparent scores also accounted for a significant amount of variance in all four YARC scores (Single word reading, Accuracy, Reading rate and Reading comprehension).

Chapter 5: DISCUSSION AND CONCLUSIONS

5.1 *Introduction*

The primary goal of this research project was to begin to fill the considerable gap in our existing understanding of children's multi-word vocabulary knowledge. First, the current project developed a measure to assess knowledge of multi-word phrases. Then, with this measure, the study investigated whether children would be able to manifest knowledge of multi-word verb + object phrasal units of differing transparency. A further goal of the study was to identify whether the findings vary across year group (i.e., is there a developmental change across 3 years of primary school corresponding to ages 7 though 10 years) and whether monolingual children and children with EAL would differ on this measure at the three time points. The study then sought to explore the relationship between multi-word phrase vocabulary and reading comprehension among a smaller subset of year 4 learners with EAL and monolingual English speakers.

5.2 *The nature of multi-word phrase knowledge among monolingual and bilingual learners*

The findings revealed a strong effect of opacity on item difficulty, as expected. Transparent items were most accurately answered, followed by semi-transparent items and non-transparent items. This finding provides further evidence that these types of opaque phrases present a disproportionate challenge to young learners, especially when considering how commonly they are used.

Results provide further support for the GEM and are in line with previous findings regarding this age group; monolinguals in year 4 had higher scores than those in year 3 on transparent and semi-transparent items, but only between years 4 and 5 (or ages 8 and 9), was there a difference in non-transparent performance. During this age window, it is posited that learners develop figurative competence and sensitivity to the

semantic characteristics of figurative language, increasing their idiom comprehension and acquisition (Cain et al., 2009; Levorato & Cacciari, 1995, 1999). There is evidence that learners will continue to develop these skills beyond the age range included in the present study (Cain et al., 2009) and future research might explore administering a measure of multi-word phrase knowledge to older children.

When comparing monolingual English speakers to learners with EAL, year 3 learners in the two language background groups did not differ significantly on any of the transparency sub-sections, although the difference in overall task scores was approaching significance. Interestingly, the two language groups did differ significantly on all of the additional language measures at this age. Among year 4 children, the two language groups differed in all areas; in year 5, monolingual English speakers performed significantly more accurately in all categories except for semi-transparent items. These significant differences in multi-word phrase knowledge are in accordance with the significant differences between the two language groups on the additional language measures administered, as would be expected, except in year 3.

Interestingly, at year 3, the two language background groups do not differ significantly in multi-word phrase knowledge but do differ on all other language measures. This lack of difference may be rooted in a broader lag in English language skills and a related lag in reading development in combination with figurative language skills being slightly later developing; a smaller vocabulary might hinder reading, resulting in less exposure to multi-word items and slower acquisition of new items. If the greater number of multi-word vocabulary items that year 4 and 5 monolingual English speakers have in comparison to year 3 monolinguals is related to increased exposure and retention of idiomatic language through exposure to written material, then perhaps participants with EAL do not experience a sudden burst in idiomatic

vocabulary items because they have not acquired a sufficient amount of single vocabulary words to access written material. The language skills that underlie theses developmental advances in monolingual speakers between ages 7 and 10 years and result in increased acquisition and comprehension of opaque multi-word items might not yet have been acquired by learners with EAL in this same age window. This could explain the lack significant differences between the two language groups in year 3, at the beginning, followed by significant differences in years 4 and 5 as monolingual learners advance.

The significant differences between the two language groups at years 4 and 5 are of interest when considering how performance at each age point differed within the two language groups. There was a significant difference in transparent and semi-transparent item knowledge for monolingual English speakers between years 3 and 4, and a difference in non-transparent item knowledge between years 4 and 5, while transparent and semi-transparent remained comparable. It would appear from these findings that between years 3 and 4, monolingual English speakers have greater knowledge of transparent and semi-transparent multi-word phrases, without a corresponding increase in their knowledge of non-transparent phrases. In year 5, when the children are 9 and 10 years old, there is an increase in non-transparent multi-word phrase vocabulary, perhaps as a result of increased reading at this time. This is in line with theories that children between 7 and 11 are acquiring an improved ability to understand increasingly opaque and figurative language in context during these years (Levorato & Cacciari 1992, 1995, 1999). If children are acquiring an improved ability to understand figurative language in context during these years, as posited by Levorato and Cacciari (1992, 1995, 1999), perhaps they are also retaining more of these non-transparent items when they do recognize them in written material instead of overlooking them or analyzing

them as parts, as L2 learners have been found to do (Bishop, 2004; Cooper, 1999; Liontas, 2002). It may be that this is evidence of a figurative language vocabulary 'burst', as Wray (2002) has hypothesized, a result of suddenly 'seeing' idioms or having much more exposure to idiomatic language through reading, or both.

By contrast, learners with EAL in year 4 did not perform significantly more accurately than in year 3 in any item category. Learners with EAL in year 5 performed significantly more accurately than learners in year 4 on semi-transparent and non-transparent test items, with no significant differences in transparent item scores between any of the three year points. These findings could be a reflection of the sample; the 16 learners with EAL in year 4 could have been particularly low performing participants. Indeed, the standard deviation in mean scores among this group was relatively high (5.26). However, it is not the highest for any sample sub-group (among year 3 monolingual English speakers the standard deviation was 5.96) and comparable to the overall sample standard deviation (5.74). As previously stated, learners were matched on WASI performance and so should be of comparable cognitive ability.

While non-transparency does appear to make an item less accurately answered, there was a significant difference between performance on non-transparent item between the youngest at the oldest participants in both language groups. By contrast, transparent item performance did not differ significantly between any of the three year points for learners with EAL; year 5 learners with EAL did not perform significantly more accurately than those in years 3 or 4. More accurate performance on these items among the older participants would be expected, and was seen among monolingual participants, and in year 3 monolingual English speakers and learners with EAL were comparable. One would hypothesise that young children would be able to derive the transparent *verb + object* phrases using knowledge of the component parts, even though

they may not be able to 'recognise' non-literal phrases not previously encountered and therefore be unable to produce them. The lack of score differences on the transparent items among older learners with EAL then might reflect a weakness in another area, perhaps a lack of ability to dismiss non-real answer choices or a lack of single word vocabulary knowledge or less general exposure to English. It is unlikely that weaker reading comprehension skills played a role in this, as the test required no reading. Overall, the lack of increase in transparent item scores for learners with EAL across the year groups is worrying and could be an indication of a differential rate or pattern of growth in their transparent multi-word knowledge relative to monolingual English speaking children.

The findings from the MPT revealed a strong effect of transparency on item difficulty, as predicted. Transparency seems to be a powerful variable in determining how children respond to these multi-word phrases. This finding provides further evidence that these types of opaque phrases present a disproportionate challenge to young learners, especially when considering how commonly they are used. These findings underscore the importance of transparency and support generative theories, like those proposed by Nesselhauf (2005) and Erman (2009). These theories posit that the opacity of these items makes them distinct from items that simply co-occur frequently; as opposed to the frequency definition, which stipulates that two items with comparable MI scores and comparable frequency should be equally challenging, in this case they were not. It also provides support for Cain et al.'s (2011) assertion that the 2 possible meanings (one literal and one figurative) are what make these items challenging and lends support to Cowie's (1991) assertion that there is a distinct difference between items that co-occur (collocations) and more fixed idiomatic items (restrictive collocations).

As predicted, there was a strong relationship between the other measures of vocabulary administered to participants and the multi-word phrase task. Of greater interest are the differing relationships between performance on the multi-word phrase task and the other administered vocabulary measures between the two language groups. There were fewer significant correlations between the task overall and the additional measures among learners with EAL, particularly in year 3. Results also showed no significant correlation between performance on the multi-word phrase task non-transparent test items and performance on the ACE 6-11 among learners with EAL in year 3, though these two measures correlated for all other ages (and for year 3 monolinguals). This is interesting considering the fact that phrasal verbs, the target items on the ACE 6-11, could be considered another type of multi-word or another type of multi-word unit. Phrasal verbs have also been identified as an area of particular difficulty for second language learners, but the ACE 6-11 has been validated with learners with EAL and should not have been too difficult for this age group. Previous studies have found differing relationships or differing strengths of relationships between various language measures for monolingual and bilingual children. For example, listening comprehension, often used as a proxy measure of language comprehension, has been found to be more highly correlated with vocabulary knowledge among learners with EAL than among monolingual English speakers (Droop & Verhoeven, 2003). However, one would hypothesize that there would be a relationship between these two measures of non-literal language, and so it is of particular note that there is none. This might be an interesting direction for future research, as the correlation between these two is strong for monolingual English speakers and non-significant for learners with EAL.

5.3 *The Multi-word Phrase Task*

5.3.1 *Strengths of the MPT*

The first necessary step for addressing the research questions presented in the current study was to develop a measure of multi-word phrases suitable for use with the target sample. No pre-existing assessments were appropriate, due either to issues stemming from operationalisation of the construct (for example, using one component of a phrase to elicit the other or relying heavily on reading comprehension) or to being ill-suited to young learners. A theoretical framework emphasizing the phrases as holistic units - rather than, for example, formats that use one component to elicit another component - led to a format that tested the items as wholes. The adapted CONTRIX format used in the present study proved very well-suited to young learners. The semi-productive framework allowed the learner to demonstrate knowledge of the phrase and the sentence gap structure provided a guiding context so it could be clear that the child intended to select a non-literal interpretation of the verb + object phrase. Additionally, the task was designed not to be overly tiring. Children did not have to explain the meaning of the target choice or provide elaborations as is required by some idiom explanation tasks, nor did they have to read to themselves. The intent was to minimize the cognitive load placed on test takers. As a result, the assessment could present more target items. By requiring less expressive language ability - for example not presenting a task that requires a young child to explain the meaning of a presented idiom - it is also less likely that a child's inability to sufficiently describe or express their ideas might be misrepresented as a lack of knowledge of the target phrase. Finally, the 'puzzle box' format proved to be accessible to young learners. They were able to complete the test items and even the lowest performers got the most accurately answered items correct, indicating that they grasped the format.

Statistical evaluations of internal reliability, item facility and item discrimination all showed that the MPT was a valid and reliable test. Correlations with other measures of vocabulary provided evidence of concurrent validity, although another measure of this exact aspect of vocabulary was not available for comparison. The measure is clearly suitable for use with this population and age range, British learners in years 3, 4 and 5 with English as a first or additional language. Performance on the test was not at ceiling when administered to year 5 learners and so future research might explore whether the measure may be suitable for older learners as well. Learners with EAL from a variety of language backgrounds were included in Phase 1 of the study; therefore, the measure should be appropriate for learners with EAL from the L1s included.

5.3.2 *Limitations of the MPT*

There are limitations to the MPT. First and foremost, the measure had not been standardized and thus far has only been used with the 180 participants that took part in the present study. Before any further implementation of generalizable conclusions the measure should be trialed with a much larger and more diverse sample. Additionally, the measure may be unsuitable for English language learners who are learning English as a foreign language in a different setting or who are not attending school exclusively in English. The age range that the MPT was used with was very specific and it may be too easy for older child learners or adult L2 English learners, but could easily be trialed to examine suitability for advanced L2 adult learners. The format and puzzle box framework may be far too challenging for children below year 3 in UK schooling, or for children from different schooling traditions.

At present the MPT is only appropriate for British learners. Learners from different schooling cultures might not be familiar with testing generally or with this type

of sentence completion task. All of the language and contexts used in the MPT were highly 'British' in nature and all the frequency counts are based on British English currently in use in England. Therefore, this test may not be suitable for learners of other Englishes, even monolingual English speakers from countries other than England. However, with some adjustment the MPT could be made appropriate for speakers of other English backgrounds.

One potential limitation of the current study is the fact that the frequency classifications used were based on a corpus of adult language, rather than child-directed speech. For example, one item on the task in the non-transparent mid frequency category that was frequently answered correctly by learners was the phrase 'catch a cold', a phrase that could arguably be of much higher frequency in child-directed speech than in adult usage. Future studies seeking to explore or manipulate frequency as a variable may wish to ensure that the frequency measures are more appropriate for young children to more carefully investigate the role that frequency plays in processing multiword phrases. Overall, high frequency items were more accurately answered and mid-frequency, which were more accurately answered than low frequency. An exception to the high frequency, mid frequency, low frequency pattern was found among the non-transparent items. For these items, non-transparent high frequency items were the most accurately answered, followed by non-transparent low frequency items. Lastly and least accurately answered were non-transparent mid frequency items. This pattern may be a reflection of the fact that the frequency classifications used in the present study were based on a corpus of adult language, rather than child-directed speech. For example, one item on the task in the non-transparent mid frequency category that was frequently answered accurately by learners was the phrase 'catch a cold', a phrase that could be of much higher frequency in child-directed speech than in

adult usage. Future studies seeking to explore or manipulate frequency as a variable may wish to ensure that the frequency measures are more appropriate for young children to more carefully investigate the role that frequency plays in processing multi-word phrases.

The multi-word phrase task was somewhat sensitive to age group differences, but could have been more sensitive. The ANOVA run for monolingual English speakers at all three age points showed a significant difference in MPT scores across the three year groups ($p<.000$) and at least one significant difference between each of the three year points – between years 3 and 4, a difference in total score, transparent score and semi-transparent score; between years 4 and 5, a difference in non-transparent scores. The ANOVA run for EAL learners comparing all three age points also showed a significant difference in scores across the three year groups ($p<.009$). However, among EAL learners, there were no significant differences between year points 3 and 4 and no significant differences in transparent scores across all three year points ($p=.51$). There was a significant difference between year points 4 and 5 (on semi-transparent performance) and between year 3 and 5 (on semi-transparent and non-transparent). The MPT was there, perhaps, more sensitive to age group differences among monolingual speakers than learners with EAL, particularly the transparent items. This might be a worthwhile area of further exploration; at present we know very little about the nature of this type of vocabulary knowledge, perhaps there are no significant bursts or increases in multi-word knowledge between years 3 and 4 among learners with EAL. However, this lack of difference might also indicate that the measure could be improved or altered to be better suited for learners with EAL. This might be done through a qualitative exploration with EAL learners and asking them to review the items

and their interpretation of the questions or by changing the target items themselves to better reflect the language these learners come into contact with.

5.3.3 *Directions for future research with the MPT*

Beyond the limitations regarding appropriate test takers, the MPT itself could potentially be improved. Administering the test can be tiring; it requires substantial reading aloud and is generally on the long end of the time spectrum for vocabulary assessments aimed at this age group. There are a few potential ways to ameliorate this. One possible change could be to remove some of the test questions – perhaps fewer test items could still yield equally good or better validity and reliability results. Another potential alteration to the test could be to change how it is administered, such as eliminating the tester and the pen and paper aspect of the testing. The MPT puzzle format could easily be presented via computer, one test question at a time. Learners could listen to items read aloud an unlimited number of times and then click on the desired puzzle box selections, or touch if the test were presented on a touch-screen. These words would then appear at the end of the sentence in the blank area, represented as an open underline in the MPT, and the whole answer would be read aloud to the test taker. The test taker could then alter his or her selection or select an 'accept' option presented in a child friendly format, such as a smiley face button or a 'thumbs up' image to continue on the next question. This format alteration would minimize tester fatigue and allow the test taker to be more 'in control' of the testing experience. It would also allow for extremely easy computer scoring and error analysis. For example, it could generate data regarding which test items takers spend the longest trying to answer or whether or not a test taker changes a response after hearing the entire final phrase read aloud. A computer-presented format would also eliminate any potential tester effect. Now that it is clear that the MPT puzzle-box format and the test

items are appropriate for young learners, it may be of value to consider further development to make improvements, such as these suggested here, should the test be standardized for larger scale use.

5.3.4 *Testing multi-word phrases*

Further research into developing measures of multi-word phrases has many potential areas to focus on. While the MPT in its present form is suitable for future use in some settings, it would also be of benefit to standardize it with a much larger sample and possibly adjust it to be appropriate for speakers of English other than British English. Another productive research direction could be to address frequency in child-directed speech of child produced speech. The current measure uses frequency counts for British English, but a future measure could compile a corpus of the multi-word phrases that occur in child directed speech and examine these frequencies. Another method might be to pull target test items from curriculum and reading material aimed at the intended age group and use phrase prevalence in these materials to assess the frequency of an item.

When moving forward with creating measurement tools for this age group and this language construct, an important element seems to be ensuring that the measure is eliciting holistic knowledge of the phrases without imposing too much on the child. If this skill of examining and understanding non-literal language is still developing, as hypotheses like those of Levorato and Cacciari (1992) suggest, then an overly complex assessment could place a huge demand on what is already a confusing area for the learner. This should be kept at the forefront of testing goals and research with the MPT has shown this to be possible.

5.4 *The relationship between multi-word phrase knowledge and reading*

The present research did more than create an absolute measure of vocabulary for future use. The current study also presented, for the first time, evidence that multi-word phrase knowledge can be a significant predictor of reading comprehension. Results from the third research phase showed that multi-word phrase knowledge clearly has an influence on certain aspects of reading. Performance on the multi-word phrase overall task accounted for a significant amount of unique variance in single word reading (33% of variance), reading accuracy (8%), reading rate (37% of variance) and reading comprehension (24% of variance), when controlling for the other variables that correlated. The separate transparency sections also accounted for significant amounts of score variance; transparent item scores made a significant contribution to the single word reading model (25%), reading rate (25%), reading comprehension (10%). Semi-transparent items explained 26% of variance in single word outcomes, 11% in accuracy, 18% in reading rate and 17% in reading comprehension. Non-transparent item scores made a significant contribution to the single word reading model (30%), the accuracy model (18%), the reading rate model (18%) and the reading comprehension model (18%), controlling for other variables.

While single word vocabulary is already well acknowledged as significant for reading as discussed in the Literature Review. The current study contributes new findings that multi-word vocabulary knowledge also has a relationship with reading, though it is distinct from single word vocabulary from a neurological and grammatical perspective. It has been argued that this distinctness necessitates separate investigation of multi-word items; assumptions regarding knowledge of individual lexical item cannot be identically applied multi-word phrase knowledge (Wray, 2002). The results from this research support this assertion: MPT scores accounted for more variability in reading outcomes than single word vocabulary scores. These findings provide further

evidence that multi-word phrases are an essential part of language knowledge and have a strong influence on reading ability. As few previous studies have specifically addressed multi-word phrase knowledge in connection with reading among this population, it is helpful to have an idea of the comparative importance of multi-word phrase knowledge for reading in relation to other types of vocabulary knowledge. The findings indicate that, at least at this age, multi-word knowledge plays a significant role in accounting for variance in certain reading outcomes – far more than expressive or receptive single word vocabulary as measured by the TOWK. It is possible that stronger readers or learners who read more often will encounter more language and consequently more multi-word phrases, and therefore would perform better when given a measure of multi-word knowledge.

It is of interest that multi-word phrase knowledge made a significant contribution to the prediction for YARC reading rate. Previous findings have shown greater multi-word phrase knowledge to correlate with faster language processing speed in adult monolingual English speakers and L2 English learners (Arnon, 2009, Wray, 2002). The presence of known formulaic language speeds up reading; formulaic sequences have been shown to be read more quickly than non-formulaic equivalents (Siyanova-Chanturia et al., 2011; Underwood et al., 2004) and the presence of formulaic language is thought to facilitate processing by easing the burden on memory (Martinez & Schmitt, 2012). This supports similar findings that greater multi-word phrase knowledge correlates with language processing speed in adult monolingual English speakers and L2 English learners (Arnon, 2009; Wray, 2002, 2009). However, it is not yet clear if knowing more formulaic language speeds up reading for young learners, or if both faster reading and knowledge of phrasal items like those on the

MPT might both be the result of another element – such as amount of time spent in the practice of reading.

The relationship between multi-word knowledge and reading could reflect a number of different things; general vocabulary knowledge and general exposure to written text might increase both reading skill and multi-word phrase knowledge. Ideally, there should be no task overlap between the MPT and the YARC. The MPT does not require any reading but it is possible that the tasks are more easily completed for children who are stronger readers and able to read the items in the puzzle box over while formulating a response. The regression analyses only controlled for vocabulary and language background, there are many other aspects of language and many other skills that still might contribute to both reading outcomes and MPT outcomes – semantic knowledge, higher order text integration skills, the ability to incorporate new language information and reconcile it with previous information. A number of elements in this relationship remain unknown and therefore results should be interpreted cautiously.

Among the correlations between the multi-word task and other measures, the WASI only correlated with the multi-word task, not the other two vocabulary measures. This could potentially be attributed to the format of the task; although it did not require any reading comprehension, it might have possibly placed a demand on other problem solving skills. It may also be that learners with greater reading abilities at this age also have higher IQ scores and in turn encounter more phrasal language. Additionally, there is a likely reciprocal relationship between reading skill and multi-word vocabulary knowledge. It should be noted that the lowest performing participants still reliably answered the easiest test items correctly, indicating that the test format was understood.

The differences in correlations between the MPT and the other measures (WASI and TOWK) between the two language groups provide additional evidence that multi-word phrase knowledge as measured by the MPT is distinct from single word knowledge. Of particular interest is the lack of correlation between TOWK expressive scores and MPT performance among the learners with EAL. Among monolingual English speakers there was a strong correlation between these two skills, $p<.01$. This further reinforces the need to examine the language development of learners with EAL rather than assuming they develop and perform as monolinguals. There are some differences in correlations between the two language groups. Among monolingual speakers there is a very high correlation between the MPT overall and its components and the TOWK Expressive. By contrast, among the learners with EAL there is no significant correlation between the TOWK Expressive and the MPT or any of its sub-sections. The Literature Review discussed differences in language development and reading between monolingual English speaking children and their peers with EAL. Previous studies with this population of UK learners with EAL have found that TOWK Expressive performance differs greatly from monolingual English speakers even when matched for other language measures (Burgoyne et al., 2009).

These findings provide a springboard for future research. Identifying the predictive role of multiword phrase knowledge in reading comprehension suggests that both researchers and teachers should focus more attention on this aspect of vocabulary if we wish to help promote reading comprehension in both EAL and L1 children. Research emphasis should be placed on better understanding how this knowledge develops, identifying more precisely its impact on reading outcomes, and the potential for including multi-word phrases as a focus of vocabulary instruction and evaluation.

5.5 Implications with regard to pedagogy

In L2 English classes idioms and formulaic phrases are sometimes explicitly taught. Given evidence of a relationship with reading comprehension – though not causal – it may be worth considering also explicitly teaching multi-word phrases, particularly opaque phrases, as vocabulary items. Additionally, a review of the curriculum to evaluate the presence of idiomatic or non-transparent multi-word items in materials intended for young learners and reconsider appropriateness and difficulty level in reference to the presence of opaque multi-word items.

It is likely there is a reciprocal relationship between reading and multi-word knowledge; more reading likely exposes learners to greater numbers of multi-word phrases and in turn greater knowledge of the phrases improves reading. Future research may consider examining this relationship further. Perhaps longitudinal research projects might follow groups of learners receiving supplemental reading exposure and track possible gains in multi-word phrase knowledge compared to a group not exposed to extra reading. Another option could be to provide a group of learners with an intervention consisting of explicit discussion and instruction regarding multi-word vocabulary along with increased reading, then compare multi-word vocabulary gain among this group to learners in an increased reading only condition. Learners might also be exposed to two types of reading materials, one group would receive texts with a significantly higher density of figurative language and at the end of the intervention, reading and vocabulary gains could be evaluated. Most authentic, though perhaps most difficult, would be to allow readers to progress without interference and monitor their reading exposure for amount of multi-word items, then correlate this with final reading outcomes.

5.6 *General study limitations*

There are limitations to the current study. Some limitations are a result of the research being a doctoral thesis and therefore constrained by the available resources (e.g., time and finances). These limitations should be kept in mind when considering the findings of the current research.

5.6.1 *Sampling*

By limiting the study to children who speak a small group of home languages, findings regarding these learners with EAL cannot be applied to the broader EAL population. Additionally, the sampling procedures (for example, not contacting schools currently in 'Special Measures' as a result of a low Office of Standards in Education report) and the limited physical size of the range of sampling locations mean that the findings should not be generalised to smaller EAL classrooms or learners in other areas. To avoid classroom or teacher effect, no participant group – such as all year 3 learners with EAL or all year 4 monolingual English speakers – was entirely made up of learners from a single classroom and when possible contained a mix of learners from different schools as well. The intention of the research was to describe, not to explain, and accordingly findings are limited in this regard.

The number of participants who took part in Phase 2 of the study in the monolingual English speaking and learners with EAL group were unbalanced (68 and 40, respectively). This does have the potential to impact results; however, both groups were larger than the minimum required to perform the statistical analyses used, which should serve to somewhat mollify this concern. Additionally, ANOVAs rather than independent samples t-test were used when appropriate and the statistical tests used should be robust enough to successfully compare groups of differing sizes.

An additional limitation is presented by the participant sample in phase 3. These two learner groups came from multiple classrooms in two different schools; one

in an urban setting (EAL learners) and one rural (monolingual English speakers). The schools are comparable in some aspects, but not perfectly matched. Both schools had comparable overall OFSTED ratings ("Good") but did differ in percentage of pupils receiving free or reduced priced lunches (urban school: 40%, rural school: 19%) and percentage of pupils with a special education need (urban school: 11%, rural school: 15%). The schools had somewhat similar, but not identical attainment results in percentage of pupils reaching key stage goals in reading (urban school: 75%, rural school: 89%), writing (urban school: 71%, rural school: 86%) and mathematics (urban school: 89%, rural school: 84%). In phase 3, the two groups of learners were not compared to each other in any analysis and instead were combined to form a larger participant group, 50% EAL and 50% monolingual. The participants themselves were not matched, though did not differ significantly in non-verbal ability, YARC single word reading, YARC accuracy or YARC reading rate. They did differ in performance on measures of vocabulary (TOWK and MPT) as well as YARC reading comprehension (see Results Chapter for further details). Findings from phase 3 are intended to be exploratory; the intention was to examine if MPT scores could explain variance in reading outcomes among this participant group (a combination of learners with EAL and monolingual speakers) to provide a stepping stone for further research. While the findings are not generalizable to any other population other than these two schools, results do indicate that a relationship between the two variables does exist and warrants future research attention.

5.6.2 *Instruments*

Another limitation of the current study is the use of the many other standardized measures. While these measures have been previously evaluated and are commonly used with the sample population, they are potentially imperfect, as all

assessments are, and some details may be lost or misrepresented when relying so heavily on so many of these instruments. As little is known about multi-word phrase knowledge in young learners, and even less about multi-word phrase knowledge among young learners with EAL, the instruments used in the study may not paint a complete picture of all aspects of multi-word phrase knowledge. Given the complex nature of lexical knowledge and multi-word phrases, it is difficult to gain a complete understanding of the phenomenon from testing only some related components. However, the hope is that the information gained can be illuminating even without encompassing all aspects, as so little is presently known about multi-word vocabulary in children. There is still a lack of understanding as to all the skills and cross-linguistics factors that contribute to reading, particularly reading for learners with EAL; as such is it difficult to test for all aspects of reading and lexical knowledge within the constraints of one study. The intention of the present research was to utilize enough different types of measures to investigate language and literacy development without exhausting the participants. Bialystok and Herman (1999) note that while particular areas have been identified as key in learning to read for monolinguals, research has been contradictory in relating these skills to bilingual learning. L1 learning has been shown to be rooted in oral competence with literary language, conceptual understanding of print and symbols and meta-linguistic awareness such as phonological awareness. All of these components have been showed to be affected by bilingualism (Bialystok, 2007). It is possible that using the measures developed to explore reading in monolingual children, particularly a single measure of reading comprehension, might not provide complete information about all specific aspects of EAL literacy. However, a more elaborate and intensive preliminary investigation of the nature of EAL literacy before

developing a multi-word phrase test (phases 1 and 2) and using it (phase 3) was not within the scope of the present study.

5.6.3 *Absence of L1 Measures*

Certain skills from the first language (L1) have been shown to have a strong correlation with L2 literacy for learners with EAL (Gabriele, Troseth, Martohardjono & Otheguy, 2009). Research with young L1 Spanish learners with EAL has shown that L1 oral skills have a relationship with L1 phonological awareness and research has demonstrated a stronger relationship between L1 literacy and L2 literacy than between L2 oral skills and L2 literacy (Durgunoglu, Nagy, & Hancin-Bhatt, 1993; Uccelli & Paez, 2007). Reese, Gernier and Gallimore (2000) found that acquisition of L2 reading can proceed even with limited L2 oral competence, as long as oral fluency in the L1 is quite strong. While it would have been ideal to include L1 measures in the present study, the lack of available reliable L1 measures combined with the heterogeneity of the participant group made doing so beyond the scope of the present study. There is a broader need for future bilingualism research that expands beyond only L2 measures in order to gain a more complete understanding of the L1 and L2 component skills that make up the process and how they connect together differently for bilingual learners. It is hoped that the language background questionnaire used in the current study ameliorated the issue to some degree by providing information about the participants' L1 experiences to help ensure the learner with EAL participants came from comparable backgrounds.

5.6.4 *Working with young learners*

There are a number of aspects of testing that are particularly problematic when assessing young learners. Young learners behave very differently depending on the setting and an artificial testing situation can have an influence on the data collected

there (McKay, 2006). Adults understand the importance of testing but young learners may not grasp the test's importance and consequently not give it the attention needed to full demonstrate their ability. It is hoped that by making the testing experience as comfortable as possible for the participants and by explaining the tests fully, these issues were at least minimized in this study. It is also possible that the tester, who was also the researcher, may have impacted test outcomes through her presence or her manner of test administration. The tester was aware of this potential influence and attempted to be of minimal influence. The same tester administered all of the tests to all of the participants, so this impact, if any, was consistent across the entire sample.

5.7 Conclusions

This study set out to measure *verb + object* multi-word phrase knowledge among monolingual English speaking children and children with EAL enrolled in UK schooling years 3, 4 and 5, to compare any differences in this knowledge between year and language groups and to explore possible relationships with other vocabulary measures. Findings revealed that multi-word phrase knowledge does increase between years 3, 4 and 5. There was a strong effect of transparency. Transparent phrases test items were most accurately answered by the learners, followed by semi-transparent items and finally, non-transparent items were the least accurately answered. Year 5 learners had significantly more accurate performance on non-transparent items than year 4 learners for both language background groups, although there was no significant difference between year 3 and year 4 learners. Monolingual English speaking learners in year 4 performed significantly more accurately than year 3 learners on transparent and semi-transparent items, while among learners with EAL there were no significant differences between year groups on transparent items, but year 4 learners did perform significantly more accurately on semi-transparent items than year 3 learners. This more

accurate performance on non-transparent items in year 5 might be indicative of a burst in idiomatic or non-literal vocabulary, or may reflect increased exposure through reading development. Correlations with other administered vocabulary measures showed a strong relationship between performance on the MPT and performance on measures of other types of vocabulary knowledge, although these relationships differed between the two language background groups.

This study set out to measure verb + object multi-word phrase knowledge among monolingual English speaking children and children with EAL enrolled in UK schools in years 3, 4 and 5, age 7-10 years, to compare any difference in this knowledge between the year and language groups. Findings revealed that non-transparent multi-word phrase knowledge does differ at points between the ages of 7 and 10 years for both language groups; year 5 learners had significantly better performance on non-transparent items than year 3 learners. There was a strong effect of transparency; transparent phrase items were most accurately answered by learners, followed by semi-transparent and finally, non-transparent. Noticeably, the two language background groups (monolingual English speakers and learners with EAL) did not differ in year 3 but in years 4 and 5, monolingual English speakers were significantly more accurate in many item transparency categories. Additionally, there were no significant year group differences on transparent item performance found among learners with EAL. Year 5 learners with EAL did, however, perform significantly more accurately on semi-transparent and non-transparent items than year 3 learners.

Figurative multi-word phrases are a key element of the English lexicon and their role in language development for young learners remains under-explored, despite evidence of their common and challenging presence in English usage and the well-acknowledged role of vocabulary in language and reading. The current study was

intended to be exploratory in nature and findings should not be generalised to the wider population. However, results do provide some insights on the nature of multi-word vocabulary among this age group and minimally demonstrate the need for more research in the area. Multi-word phrase knowledge may be revealed as an area of vocabulary of significant import, especially when considering a possible relationship with reading comprehension, and should not be ignored. Future research may consider attempting to measure other types of formulaic or phrasal knowledge among children, particularly those with EAL, as well as exploring how this knowledge may contribute to language development and reading related skills at different ages.

The final component of the current research explored the relationships between YARC performance and verb + noun multi-word phrase knowledge as among year 4 monolingual English speakers and Bengali speaking learners with EAL. Regression models showed that the MPT as a whole accounted for a significant amount of variation in reading outcomes for all 4 YARC measures. MPT transparent scores accounted for a significant amount of variance in YARC Single word reading and YARC Reading rate, and were approaching significance for YARC Reading rate. MPT task semi-transparent scores alone did not account for a significant amount of variance in any YARC score. MPT task non-transparent scores accounted for a significant amount of variance in YARC Single word reading and YARC Reading rate scores. Correlations among the various other measures administered alongside the MPT differed between the two language groups.

Multi-word phrases are a key element of the English lexicon and their role in language development for young learners remains underexplored, despite the well-acknowledged role of vocabulary in language and reading. The current study was intended to be exploratory in nature; these results do provide some insights on the

nature of multi-word vocabulary within this age group. This is an area of vocabulary learning that is of potential import, especially when considering a possible relationship with reading comprehension, and should not be neglected simply because measuring it presents a challenge. Future research may consider attempting to measure other types of formulaic or phrasal knowledge among children, as well as exploring how this knowledge may contribute to reading related skills at different stages of development.

The amount of unique variance in reading outcomes accounted for by the multi-word phase task provides further evidence of the import of this type of vocabulary and underscores the need for further investigation. Even with the limitations of the current study, these findings provide valuable insight into multi-word vocabulary among young learners. This current study supports theories that multi-word phrases are distinct from single word vocabulary and provides evidence that this aspect of vocabulary also has a distinct relationship with reading. This study is one of few attempts to measure this aspect of vocabulary knowledge in children, both monolingual English speakers and children with EAL, and can serve as a stepping-stone for further investigation into this area. These findings emphasize a need for greater attention on multi-word phrase vocabulary; hopefully the MPT and the findings from Phase 3 of the current study will provide a foundation with a tool and initial evidence that future researchers can use and build upon.

REFERENCES

Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of comprehension and expression 6–11* (ACE 6–11). Windsor: NFERNelson.

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component?. *Reading and Writing,* 19(9), 933-958.

Altenberg, B. 1(990). 'Speech as linear composition'. In: Caie, G. Haastrup, K., Jakobsen, A.L., Nielsen, J.E., Sevaldsen, J., Specht, H. and Zettersten, A. (Eds.), *Proceedings from the Fourth Nordic Conference for English Studies, Vol. 1. Department of English, University of Copenhagen*, 133-143.

Anderson, R. C., & Freebody, P. (1985). Vocabulary knowledge. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 343–371). Newark, DE: International Reading Association.

Anglin, J.M. (1993) Vocabulary development: A morphological analysis. *Monographs of the Society of Research in Child Development, 58(10).*

Arnon, I. (2009). *Starting Big: the role of multi-word phrases in language learning and use.* Unpublished dissertation, Stanford University.

Arnon, I. & Snider, N. (2010). More than words: frequency effects for multi-word phrases. *Journal of Memory and Language,* 62, 67-82.

Bahns, J. & Eldaw, M., (1993). Should we teach EFL students collocations? *System,* 21(1).

Barfield, A. (2003). *Collocation Recognition and Production: Research Insights.* Chuo University, Japan.

Barfield, A. & Gyllstad, H. (2009) *Researching Collocations in Another Language.* Basingstoke: Palgrave Macmillan.

Beech, J.R. & Keys, A. (1997). Reading, vocabulary and language preference in 7- to 8-year old bilingual Asian children. *British Journal of Educational Psychology*, 67, 405-414.

Benjamin, C. F., & Gaab, N. (2011). What's the story? The tale of reading fluency told at speed. *Human Brain Mapping.*

Berninger, V. W., Abbott, R. D., Vermeulen, K., & Fulton, C. M. (2006). Paths to reading comprehension in at-risk second-grade readers. *Journal of Learning Disabilities*, 39(4), 334-351.

Bialystok, E. (1999). Cognitive Complexity and Attentional Control in the Bilingual Mind. *Child Development,* 70(3), 636-644.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition.* New York: Cambridge University Press.

Bialystok, E. (2007). Acquisition of literacy in bilingual children: A framework for research. *Language learning*, 57, 45.

Bialystok, E. (2010). Global–Local and Trail-Making Tasks by Monolingual and Bilingual Children: Beyond Inhibition. Developmental Psychology. 46 (1), 93–105

Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(4), 229.

Bialystok, E. & Herman, J. (1999) Does bilingualism matter for early literacy? *Bilingualism: Language and Cognition 2,* Cambridge University Press: Cambridge, 35-44.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade instruction. In A. Hiebert & M. Kamil (Eds.) *Teaching and Learning Vocabulary: Bringing Research to Practice*, Mahwah, N.J.: Erlbaum, 223-242.

Bishop, H. (2004). Noticing formulaic sequences: a problem of measuring the subjective. *LSO Working Papers in Linguistics*, 4, 15-19.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations. A Polish/German empirical study. In P.J.L. Arnaud and H. Béjoint, (eds), *Vocabulary and Applied Linguistics*, 85-93, London: Macmillan.

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, 1(3), 343-346.

Bonk, W.J. (2001). Testing ESL Learners' Knowledge of Collocations. In T. Hudson and J.D. Brown, (Eds), *A Focus on Language Test Development: Expanding the Language Proficiency Construct Across a Variety of Tests. (Technical Report #21)*, pp. 113-142, Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Brown, R. (1973). *A first language.* London: Allen & Unwin.

Burgoyne, K., Kelly, J., Whiteley, H., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology*, 79, 735-74.

Cain, K. (1999). Ways of reading: How knowledge and use of strategies are related to reading comprehension. *British Journal of Developmental Psychology*, 17(2), 293-309.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and writing*, 11(5), 489-503.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition,* 29(6), 850-859.

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,* 90, 65-87.

Cain, K. & Towse, A. (2008). To get hold of the wrong end of the stick: Reasons for poor idiom understanding in children with reading comprehension difficulties. *Journal of Speech, Language and Hearing Research*, 51, 1538-1549.

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.

Chall, J.S. (1983) *Stages in Reading Development*, New York, McGraw-Hill.

Chall, J.S., Jacobs, V.A., & Baldwin, L.E. (1990). *The reading crisis: Why poor children fall behind.* Cambridge, MA.: Harvard University Press.

Chen, R.S. & Vellutino, F.R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research*, 29 (1), 1-24.

Chiappe, P., Chiappe, D. L., & Gottardo. A. (2004). Vocabulary, context and speech perception among good and poor readers. *Educational Psychology, 24,* 825-843.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

Church, K. & Hanks, P, (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics* 16 (1), 22-29.

Cooper, T.C. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly*, 33(2), 233-262.

Connor, C. M., Morrison, F. J., & Katch, L. E. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading*, 8(4), 305-336.

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223-235.

Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. *Vocabulary and language teaching*, 126-139.

Cowie, A. P. (1991). Multiword Units in Newspaper Language in Perspectives on the English Lexicon. A Tribute to Jacques Van Roey. *Cahiers de l'Institut de Linguistique de Louvain,* 17(1-3), 101-116.

Cowie, A.P. (1998). Phraseological Dictionaries: Some east-west comparisons. In A. P. Cowie (ed.) *Phraseology.* Oxford: Clarendon Press, 209-228.

Crutchley, A. (2007). Comprehension of idiomatic verb + particle constructions in 6- to 11-year-old children. *First Language.* 27 (3), 203-226.

Cruttenden, A. (1981). Item-learning and system-learning. *Journal of Psycholinguistic Research*, 10(1), 79-88.

Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. *Language proficiency and academic achievement*, 10, 2-19.

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

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. Memory & Cognition, 25(1), 57–71.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, 22(04), 499-533.

Dollaghan, C. (1994). Children's phonological neighborhoods: Half empty or half full? *Journal of Child Language.* 21. 257-271.

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

Duthie, J. K., Nippold, M. A., Billow, J. L., & Mansfield, T. C. (2008). Mental imagery of concrete proverbs: A developmental study of children, adolescents, and adults. *Applied Psycholinguistics*, 29(1), 151.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D.(1982). British Picture Vocabulary Scale. Windsor, UK:NFER-Nelson.

Durgunoglu, A., Nagy, W., & Hancin-Bhatt, B. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology,* 85, 453-465.

Ellis, N. C., & Beaton, A. (1993a). Psycholinguistic determinants of foreign language vocabulary learning. Language learning, 43(4), 559-617

Ellis, N., & Beaton, A. (1993b). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. The Quarterly Journal of Experimental Psychology, 46(3), 533-558.

Ellis, N., Simpson-Vlach, R. & Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, 42 (3), 375-396.

Erman, B. (2009). ' Formulaic language from a learner perspective: What the learner needs to know. 'In *Formulaic Language Volume 2: Acquistion, loss, psychological reality, and functional explanations.* Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen M. Wheatley (Eds). 27-51. John Benjamins Publishing: Philadelphia,
Pennsylvania.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29-62.

Farghal, M. & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Journal of Applied Linguistics* 28(4), 313-331.

Field, A. P. (2000). *Discovering statistics using SPSS for Windows: Advanced techniques for the beginner*. London: Sage.

Firth, J. R. (1952). Linguistic analysis as a study of meaning. *Selected papers of JR Firth, 1959,* 12-26.

Firth, J. R. (1957). Applications of general linguistics. *Transactions of the Philological Society*, *56*(1), 1-14.

Frederickson, N., & Frith, U. (1998). Identifying dyslexia in bilingual children: A phonological approach with inner London Sylheti speakers. *Dyslexia*, 4(3), 119-131.

Fulcher, G. & Davidson, F. (2007). Language testing and Assessment. An Advanced resource book. London: Routledge.

Gabriele, A., Troseth, E., Martohardjono, G. & Otheguy, R. (2009). Emergent literacy skills in bilingual children: evidence for the role of L1 syntactic comprehension. *International Journal of Bilingual Education and Bilingualism*, 12(5), 533-54.

García, G.E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly,* 26, 371–392.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12(3), 349-367.

Genesee, F. & Geva, E. (2006). Cross-Linguistic Relationships in Working Memory, Phonological Processes and Oral Language. In August, D. and Shanahan, T. (Eds) *Developing Literacy in Second-Language Learners.* London: Lawrence Erlbaum Associates, Publishers.

Geva, E., & Siegel, L. S. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing*, 12(1), 1-30.

Gibbs, R. (1987). Linguistic factors in children's understanding of idioms. *Journal of Child Language,* 14, 569-586.

Gibbs, R. (1991). Semantic analyzability in children's understanding of idioms. *Journal of Speech and Hearing Research,* 34, 613-620.

Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1989). Where learning begins: Initial representations for language learning. *Language: Psychological and biological aspects,* 150-193.

Gough, P. & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education,* 7, 6-10.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. *Reading comprehension difficulties: Processes and intervention,* 1-13.

Gregory, E. (1996). *Making Sense of a New World: Learning to Read in a Second Language.* London:Paul Chapman Publishing.

Guttman, L. (1945) A basis for analyzing test-retest reliability. *Psychometrika,* 10(4), 255-282.

Gyllstad, H. (2005). Words that go together well: Developing test formats for measuring learner knowledge of English collocations. In F. Heinat, and E. Klingvall, (Eds.). *The department of English in Lund: Working papers in linguistics. 5*, 1-31.

Gyllstad, H. (2009). Designing and Evaluating Tests of Receptive Collocation Knowledge: COLLEX and COLLMATCH. In Barfield, A. & Gyllstad, H. (Eds) *Researching Collocations in Another Language.* Basingstoke: Palgrave Macmillan, 125-138.

Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word*, 17(3), 241-292.

Halliday, M. A. (1966). Some notes on 'deep' grammar. *Journal of Linguistics*, 2(1), 57-67.

Harlaar, N., Cutting, L., Deater-Deckard, K., DeThorne, L. S., Justice, L. M., Schatschneider, C., Thomson, L. & Petrill, S. A. (2010). Predicting individual differences in reading comprehension: a twin study. *Annals of dyslexia,* 60(2), 265-288.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological review*, 111(3), 662.

Hatcher, P.J. & Hulme, C. (1999). Phonemes, rhymes, and intelligence as predictors of children's responsiveness to remedial reading instruction: Evidence from a longitudinal study. *Journal of Experimental Child Psychology*, 72, 130-153.

Henning, G. (1987*). A guide to language testing.* New York: Newbury.

Hoover, W.A. & Gough, P.B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127-160.

Howarth, P. A. (1996). *Phraseology in English academic writing*. Niemeyer.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, 19(1), 24-44.

Hutchinson, J., Whiteley, H., Smith, C., & Connors, L. (2003) The developmental progression of comprehension-related skills in children learning' *EAL Journal of Research in Reading.* 26(1), 19-32.

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E.-J. VanderLinden, & R. Schreuder (Eds.), Idioms. Structural and psychological perspectives (pp. 133–165). Hillsdale, NJ: Erlbaum.

Jaen, M. (2007). 'A Corpus-driven Design of a Test for Assessing the ESL Collocational Competence of University Students' *International Journal of English Studies.* Vol. 7 (2), 127-147

Jean, M., & Geva, E. (2009). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics*, 30(1), 153.

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing*, 19(4), 339-361.

Jones, S., & Sinclair, J. (1974). English lexical collocations. *Cahiers de lexicologie*, 24(1), 1974.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21(2), 85-97.

Katzir, T., Kim, Y., Wolf, M., O'Brien, B., Kennedy, B., Lovett, M., & Morris, R. (2006). Reading fluency: The whole is more than the parts. *Annals of Dyslexia*, 56(1), 51-82.

Kaushanskaya, M., & Marian, V. (2008). Mapping phonological information from auditory to written modality during foreign vocabulary learning. *Annals of the New York Academy of Sciences*, 1145(1), 56-70.

Keshavarz, M. H. & Salimi, H. (2007), Collocational competence and cloze test performance: a study of Iranian EFL learners. *International Journal of Applied Linguistics,* 17, 81-92.

Kieffer, M. & Lesaux, N. (2007). Breaking down words to build meaning: Morphology, vocabulary and reading comprehension in the urban classroom. The Reading Teacher, 61(2), 134-144.

Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading?. *Literacy*, 42(2), 75-82.

Leach, J.M., Scarborough, H.S. & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology,* 95, 211-224.

Lesaux, N., & Siegel, L. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology*, 39, 1005-1019.

Leseaux, N. & Geva, E. (2006). Synthesis: Development of Literacy in Language-Minority Students. In August, D. and Shanahan, T. (Eds) *Developing Literacy in Second-Language Learners.* London: Lawrence Erlbaum Associates, Publishers.

Lesaux, N., Orly, L. & Siegel, L. (2006) 'Investigating Cognitive and Linguistic Abilities that Influence the Reading Comprehension Skills of Children from Diverse Linguistic Backgrounds'. *Reading and Writing.* 19(1), 99-131.

Lesaux, N. K., Rupp, C. K., & Siegel, L S. (2007). Growth in Reading Skills of Children From Diverse Linguistic Backgrounds: Findings From a 5-Year Longitudinal Study'. *Journal of Educational Psychology.* 99 (4), 821-834.

Lesaux, N.K., Geva, E., Koda, K., Siegel, L.S. & Shanahan, T. (2008) Development of literacy in second-language learners. In D.August & T. Shanahan (Eds.) *Developing Reading and Writing in Second-Language Learners: Lessons from the Report of the National Literacy Panel on Language-Minority Children and Youth*, New York: Routledge, 27-59.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). Multiple perspectives on word production. *Behavioral and Brain Sciences*, 22(01), 61-69.

Levorato, M. & Cacciari, C. (1992). Children's comprehension and production of idioms: The role of context and familiarity. *Journal of Child Language*. 19 (2), 415.

Levorato, M. & 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.

Levorato, M. & Cacciari, C. (1999). Idiom comprehension in children: Are the effects of semantic analyzability and context separable? *The European Journal of Cognitive Psychology.* 11 (1), 51.

Levorato, M. C., Nesi, B., & Cacciari, C. (2004). Reading comprehension and understanding idioms: A developmental study. *Brain and Language*, 202, 4-16.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology*, 95(3), 482.

Liontas, J. (2002). Context and idiom understanding in second languages. *EUROSLA Yearbook*, 2, 155-185.

Lobello, S. (1991). A short form of the Wechsler preschool and primary scale of intelligence-revised. *Journal of School Psychology.* 29 (3), 229-236.

Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and language*, 58(2), 265-326.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of dyslexia*, 53(1), 1-14.

Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of educational psychology*, 102(3), 701.

Manis, F. R., Lindsey, K. A., & Bailey, C. E. (2004). Development of Reading in Grades K–2 in Spanish-Speaking English-Language Learners. *Learning Disabilities Research & Practice*, 19(4), 214-224.

Martinez, R., & Murphy, V. (2011). The effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, 45, pp. 267-290.

Marton, W. (1977). 'Foreign vocabulary learning as problem No. 1 of language teaching at the advanced level'. *Interlanguage Studies Bulletin* 2, 33-57.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2008). Hemispheric differences in processing the literal interpretation of idioms: Converging evidence from behavioral and fMRI studies. *Cortex*, 44(7), 848-860.)

McBride-Chang, C., Wagner, R. K., & Chang, L. (1997). Growth modeling of phonological awareness. *Journal of Educational Psychology*, 89(4), 621.

McKay, P. (2006). *Assessing Young Language Learners.* Cambridge: Cambridge University Press.

Metsala, J. L., & Walley, A. C. (1998). Spoken Vocabulary Growth and the Segmental Restructuring of Lexical Representations: Precursors to Phonemic Awareness and Early Reading Ability. *Word Recognition in Beginning Literacy*, 89.

Miller, G. (1991). The science of words. New York: Scientific American Library.

Milton, J. (2009) *Measuring Second Language Vocabulary Acquisition.* Multilingual matters.

Mochizuki, M. (2002). Exploration of two aspects of vocabulary knowledge: Paradigmatic and collocational. *Annual Review of English Language, Education in Japan* 13, 121-129.

Moon, R. (1992). Textual aspects of fixed expression in learners' dictionaries. In P.J.L. Arnaud & H. Bejoint (eds.) *Vocabulary & applied linguistics.* Basingstoke: Macmillan, 13-27.

Moore, M., McCabe, G., & Craig, B. (2010). *Introduction to the Practice of Statistics.* W.H. Freeman & Co Ltd.

Muter, V., Hulme, C., Snowling, M.J. & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40 (5), 665-681.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading research quarterly*, 233-253.

Nakamoto, J., Lindsey, K.A. & Manis, F.R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing*, 20, 691-719.

NALDIC (2006). Key Documents, Available online at: <http://www.naldic.org.uk/docs/resources/KeyDocs.cfm#S> (Accessed 05/12/2010)

NALDIC (2012). EAL Statistics, Available online at: <http://www.naldic.org.uk/research-and-information/eal-statistics/eal-pupils> (Accessed 10/12/2012)

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, K. & Snowling, M.J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language,* 39, 85-101.

Nation, K., & Angell, P. (2006). Learning to read and learning to comprehend. *London Review of Education*, 4(1), 77-87.

Nattinger, J.R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

Neale, M. (1997). *Neale Analysis of Reading Ability II.* Windsor. England: NFER-Nelson.

Nesselhauf, N. (2005). Collocations in a learner corpus. Amsterdam: Benjamins.

Nippold, M.A. (1991). Evaluating and enhancing idiom comprehension in language-disordered students. *Language, Speech and Hearing Services in Schools,* 22, 100-106.

Nippold, M. A. (1998). Later Language Development: The school-aged and adolescent years. Pro-Ed.

Nippold, M.A. & Martin, S. T. (1989). Idiom interpretation in isolation versus context: a developmental study with adolescents. *Journal of Speech and Hearing Research,* 32, 59-66.

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, 58-66.

Nippold, M.A. & Taylor, C. L. (1995). Judgements of idiom familiarity and transparency: A comparison of children and adolescents. *Journal of Speech and Hearing Research,* 45, 384-391.

Nippold, M.A., Moran, C. & Schwatz, I. E. (2001). Idiom understanding in preadolescents: Synergy in action. *American Journal of Speech and Language Pathology,* 10, 169-179.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and cognitive processes,* 18(4), 443-468.

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.

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing,* 23(2), 189-208.

Pawley, A. & Syder, F. (1983). ' Two puzzles for linguistic theory: Nativelike selection and nativelike fluency.' In *Language and communication,* J.C. Richards & R. W. Schmidt (Eds). 191-226. Longman: London.

Pearson Speech and Language. (<http://www.pearsonassessments.com>) (Accessed, 02.08.2010)

Pearson, D.P., Hiebert, E.H. & Kamil, M.L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42 (2), 282-296.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 67-86.

Peters, A. M. (1982). *Units of language acquisition.* Cambridge: Cambridge University Press.

Peterson, R. R., Burgess, C., Dell, G., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27, 1223–1237.

Poulsen, S. (2005). *Collocations as a language resource (A functional and cognitive study in English phraseology).* Ph. D. dissertation, University of Southern Denmark.

Proctor, P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-Speaking Children Reading in English: Toward a Model of Comprehension. *Journal of Educational Psychology*, 97(2), 246-256.

Raven, J. C., Raven, J., & Court, JH. (1990). *Standard progressive matrices.* Oxford, United Kingdom: Oxford Psychologists Press.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Reese, L., Gernier, H., & Gallimore, R. (2000). Longitudinal Analysis of the Antecedents of Emergent Spanish Literacy and Middle-School English Reading Achievement of Spanish-Speaking Students. *American Educational Research Journal.* 37(3), 633-662.

Revier, R. L. (2009). Evaluating a New Test of Whole English Collocations, In Barfield, A. & Gyllstad, H. (Eds) *Researching Collocations in Another Language.* Basingstoke: Palgrave Macmillan, 125-138.

Rinaldi, W. (2000) 'Pragmatic comprehension in secondary school-aged students with specific developmental language disorder' I*nternational Journal of Language & Communication Disorders,* 35(1), 1-29

Rosowsky, A. (2001) Decoding as a cultural practice and its effects on the reading process of bilingual pupils. *Language and Education*, 15 (1), 56-70.

Roth, F.P., Speece, D.L. & Cooper, D.H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research*, 95, 259-272.

Savage, R. (2001). The'simple view'of reading: Some evidence and possible implications. *Educational Psychology in Practice*, 17(1), 17-33.

Savage, R. (2006). Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading*, 10(2), 143-164.

Scarborough, H.S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S.B. Neuman & D.K. Dickinson (Eds.) *Handbook of Early Literacy Research: Volume 1*, New York, The Guilford Press, 97-110.

Sharwood Smith, M. (1991). *Plenary given at the Second Language Research Forum. University of California, Los Angeles.*

Shaywitz, S. E., & Shaywitz, B. A. (2003). Neurobiological indices of dyslexia. *Handbook of learning disabilities*, 514-531.

Shaywitz, S. E., & Shaywitz, B. A. (2008). Paying attention to reading: The neurobiology of reading and dyslexia. *Development and psychopathology*, 20(04), 1329-1349.

Sinclair, J. M. (Ed.) (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary.* Collins ELT.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Smith, M. S. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. Second Language Research, 7(2), 118-132.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K., & Hulme, C. (2009). *York Assessment of Reading for Comprehension*. GL Assessment: London.

Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. Journal of memory and language, 54(2), 161-184.

Stahl, S. & Nagy, W.(2005). *Teaching word meaning*. Mahwah, NJ: Lawrence Erlbaum.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. Journal of Verbal Learning and Verbal Behavior, 18, 523–534.

Taguchi, E. (1997). The effects of repeated readings on the development of lower identification skills of FL readers. *Reading in a Foreign Language*, 11, 97–119.

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. Journal of Pragmatics Special Issue: Literal and figurative language, 31(12), 1655–1674.

Tiu, R., Thompson, L., & Lewis, B. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities*, 36, 424-436.

Tomasello, M., & Brooks, P. J. (1999). Early syntactic development: A Construction Grammar approach. In M. Barrett (Ed.), *The development of language*. Hove, UK: Psychology Press.

Torgesen, J.K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, 15(1), 55–64.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K., & Conway, T. (2001). Intensive Remedial Instruction for Children with Severe Reading Disabilities Immediate and Long-term Outcomes From Two Instructional Approaches. *Journal of Learning Disabilities*, 34(1), 33-58.

Tunmer, W. E., & Hoover, W. A. (1992). Cognitive and Linguistic Factors in Learning to Read. *Reading acquisition*, 175.

Uccelli, P., & Paez, 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(3), 225-236.

Vellutino, F.R., Tunmer, W.E., Jaccard, J.J. & Chen, R. (2007). Components of reading ability: multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, 11 (1), 3-32.

Verhoeven, L. (1990). Acquisition of reading in a second language. *Reading Research Quarterly*, 25, 90-114.

Vermeer, A. (1992). Exploring the second language learner lexicon. In L. Verhoevenand & H.A.L. de Jong (Eds), *The Construct of Language Proficiency*, Amsterdam: Benjamins.

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, 16(1), 5-20.

Wechsler, D. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Psychological Corporation.

Wiig, E. H. & Secord, W. (1992) *Test of Word Knowledge*. USA: The Psychological Corporation.

Wilson, A. M., & Lesaux, N. K. (2001). Persistence of phonological processing deficits in college students with dyslexia who have age-appropriate reading skills. *Journal of learning disabilities*, 34(5), 394-400.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of speech, language and hearing research*, 50(4), 1093.

Wood, M.M. (1986). *A definition of idiom.* Bloomington, IN; Indiana University Linguistics Club.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, 5(3), 211-239.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2009). 'Identifying formulaic language: Persistent challenges and new opportunities.' In *Formulaic Language Volume 1: Distribution and historical change.* Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen M. Wheatley (Eds). 27-51. John Benjamins Publishing: Philadelphia, Pennsylvania.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. Language & Communication, 20(1), 1-28.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge: Cambridge University Press.