CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Understanding how time flies and what it means to be on cloud nine: English as an Additional Language (EAL) learners' metaphor comprehension

Annina K. HESSEL* and Victoria A. MURPHY

Department of Education, University of Oxford, UK
*Corresponding author: University of Göttingen - Educational Psychology, Waldweg 26, 37073 Göttingen, Germany. E-mail: annina.hessel@psych.uni-goettingen.de

Abstract

We explored the vocabulary and metaphor comprehension of learners of English as an additional language (EAL) in the first two years of UK primary school. EAL vocabulary knowledge is believed to be a crucial predictor of (reading) comprehension and educational attainment (Murphy, 2018). The vocabulary of five- to seven-year-old children with EAL was compared to that of English monolinguals (N = 80). Comprehension was assessed for both verbal (e.g., *time flies*) and nominal metaphors (*be on cloud nine*) of varying frequency. Results showed that children in year 2 (age six to seven years) had better comprehension than their younger (age five to six) peers, particularly for low-frequency metaphors. Children with EAL had weaker metaphor comprehension than their monolingual peers, particularly on a reasoning task. The results document how metaphor comprehension develops over the first critical years of schooling and indicates where learners with EAL differ from monolingual peers, thereby supporting targeted vocabulary teaching at primary schools.

Keywords: vocabulary; metaphor comprehension; English as an Additional Language

Introduction

Internationally, increasing numbers of children are growing up with more than one language. In many countries, they are learning an additional language to the majority language that is spoken in an otherwise fairly monolingual society such as Turkish in a German community or Maori in a New Zealand English community. Whatever the combination of their two (or more) languages, minority language learners show distinct learning patterns as a result of their learning context. While some suggest that bilingualism is associated with cognitive and social advantages (Bialystok, Craik, & Luk, 2012), some minority language learners develop their vocabulary and comprehension skills more slowly than their monolingual peers, which may then impede their reading abilities and overall academic success (Hutchinson, Whiteley, Smith, & Connors, 2003). The present study examines

CrossMark

figurative vocabulary comprehension, specifically metaphors, for learners of English as an additional language (EAL) in England, allowing us to expand on previous research on EAL vocabulary knowledge, and to further add to the evidence on metaphor processing.

English as an additional language learners

Partly as a consequence of increased global migration (from 156 million in 1990 to more than 210 million in 2013; Castles, 2013) more children find themselves in bilingual environments where they learn one language in the home, while being schooled in a different majority language (Murphy, 2014, 2018). In the UK in 2016, 20.1% of all primary school students were learners of EAL (Department of Education, 2013). Within this group, children with EAL learn many different first languages (over 360) (NALDIC, 2011) and vary in their educational attainment. While there is an international trend that children from ethnic minority families tend to underperform relative to majority language peers (OECD, 2006), a closer analysis of subgroups at higher risk of lower academic achievement revealed that multiple factors determine the academic outcomes of children with EAL in England, including ethnicity, time of arrival in the UK, and socioeconomic status (SES) (Strand, Malmberg, & Hall, 2015). Beyond those factors, Strand and Demie (2005) have shown that one of the key factors that determines the academic success of students with EAL is their fluency in English. Indeed, Whiteside, Gooch, and Norbury (2017) demonstrated that, when children with EAL are matched on English language fluency to non-EAL children, the children with EAL were either matched or outperformed the non-EAL children on a range of behavioural and cognitive measures. Of course, many children with EAL do not have the same English language skills as non-EAL children, but the success of learners of EAL with high English abilities only highlights the importance of supporting EAL children's language skills to help their overall attainment. More research is needed to fully understand which aspects of the language abilities of children with EAL may be associated with academic difficulties, and how they could be supported more effectively. Vocabulary is a promising place to start looking because children with EAL, as with other bilingual populations, have repeatedly been found to lag behind their monolingual peers in the number of words they know in the majority language (Bialystok, Luk, Peets, & Yang, 2009; Melby-Lervåg & Lervåg, 2014; Murphy, 2014; Pearson, Fernandez, & Oller, 1993; Spencer & Wagner, 2017). This lag in vocabulary development (in the majority language) is not surprising given that vocabulary seems to be particularly sensitive to availability of linguistic input. Since vocabulary acquisition requires learning numerous items, it is more likely to be affected by limited exposure. What is important is that vocabulary is a strong predictor of reading comprehension in general and in particular in children with EAL (Babayigit & Stainthorp, 2014). Furthermore, reading comprehension is an area in which learners of EAL have been demonstrated to lag behind monolingual peers (Spencer & Wagner, 2017). Taken together, a lag in vocabulary and reading comprehension may explain some of the difficulties of lower-achieving learners of EAL. At the same time, vocabulary in particular is a factor that teachers can directly support in the classroom. We thus focus on vocabulary development of learners of EAL in the hope of facilitating targeted language teaching at primary schools to support their needs.

Children with EAL tend to have smaller vocabulary sizes in one of their languages compared to their monolingual peers (Bialystok *et al.*, 2009; Pearson *et al.*, 1993). Vocabulary knowledge, however, is componential and can be measured in several ways. Vocabulary knowledge can also be operationalised as how well a word is known, a dimension which is sometimes referred to as VOCABULARY DEPTH. Furthermore, some words are regularly combined to make up collocations such as *pay attention* or idioms such as *to kick the bucket* (Nation, 2001). Several aspects of vocabulary depth have been shown to be particularly relevant to understanding the gap in vocabulary between (some) learners of EAL and their monolingual peers (Cameron & Besser, 2004; Kan, 2014; Martinez & Murphy, 2011; McKendry, 2014; Smith & Murphy, 2015). Importantly, it is vocabulary depth in particular that likely plays a crucial role in core skills that are educationally important, such as the ability to extract meaning from texts (Cain & Oakhill, 2014).

Vocabulary depth has been investigated in learners of EAL using standardised measures such as the Test of Word Knowledge (TOWK) *Multiple Contexts* subtest (e.g., McKendry, 2014). Experimental measures have also been developed to tap concrete aspects of multiword and figurative vocabulary that seem to be particularly relevant for the vocabulary depth of learners with EAL. For instance, monolingual primary school children and those learning EAL were compared on an experimental test probing their contextualised understanding of high-frequency collocations such as *by and large* through comprehension questions, followed by a later rating of their own level of understanding (Kan, 2014), a task that is difficult for foreign language learners of English (Martinez & Murphy, 2011). Learners of EAL had lower accuracy on the test compared to their monolingual peers, and their own high rating of their understanding suggested that they were not aware of their comprehension difficulties (Kan, 2014). Children with EAL, therefore, might experience difficulties understanding high-frequency collocations they do not notice themselves.

Other studies have investigated collocations such as *pay attention* (Smith & Murphy, 2015) and idioms such as *It's raining cats and dogs* (McKendry, 2014). A cross-sectional study with children aged seven to ten found that learners of EAL not only knew fewer single words across all years, but that older EAL children also did not produce many more collocations correctly than their younger peers, unlike monolinguals who experienced a 'burst of multi-word vocabulary' (Smith & Murphy, 2015). While differences between the age groups could be due to convenience sampling from the highly heterogeneous EAL population (Strand *et al.*, 2015), the findings suggest that monolinguals' growth in multiword vocabulary serves to increase the gap in vocabulary knowledge depth between monolingual learners and those with EAL in the later years of primary school. Regarding idioms (e.g., *around the clock*), McKendry (2014) compared the common variance of standardised tests of vocabulary depth and size (*Multiple Contexts* and *Expressive Vocabulary* of the TOWK; Wiig & Secord, 1992) with a customised test of idiom comprehension in a sample of UK ten- to eleven-year-old students. Monolinguals outperformed learners of EAL on the idiom comprehension task; however, all differences between the groups disappeared when vocabulary depth was partialled out. In contrast, when controlling for vocabulary size, differences across idiom types disappeared only for children with EAL, not for their monolingual peers (McKendry, 2014). This finding suggests that vocabulary depth captures a distinct variance between monolinguals and students with EAL, a variance that furthermore makes distinct contributions to reading comprehension (McKendry, 2014).

Taken together, these studies support the notion that, throughout their time at primary school, learners of EAL tend to lag behind their monolingual peers in their knowledge of multiword and figurative vocabulary. Furthermore, those difficulties likely extend into literacy. For example, children's performance on the collocational measure in Smith and Murphy (2015) made a significant and unique contribution to their reading comprehension scores. Measuring EAL comprehension of different types of vocabulary is particularly interesting as it is likely to help offer greater precision in our understanding of why some children with EAL struggle with (reading) comprehension, and how their processing patterns of different linguistic features contribute to their difficulties. With this endeavour in mind, the present study is pioneering in that it focuses on an under-researched type of figurative multiword vocabulary in the context of learners of EAL, namely metaphors.

Developing metaphor comprehension

What matters for successful metaphor comprehension is not just how well the individual words in a metaphor are known, but how well the child understands a metaphor as a whole. To understand the whole metaphor, children need both lexicalised knowledge about words and phrases and reasoning skills. Depending on the type of metaphor, the relevance of these different skills will vary. We thus chose to investigate different types of metaphors that are believed to require different skills. The first type of metaphors we chose to look at are metaphors that systematically combine words and ideas, such as *time flies*, and novel adaptations thereof (such as *minutes crawled by*). These metaphors are examples of what are often called 'conceptual metaphors'. Conceptual metaphors systematically map a concrete domain, such as space, onto an abstract domain, such as time (Lakoff & Johnson, 2003). For example, the prepositional metaphor *under control* links UP to POWERFUL, and the adjectival metaphor *raw facts* links IDEAS to FOOD (Lakoff & Johnson, 2003). For the current study, we will focus on conceptual metaphors of the type ABSTRACT NOUN + CONCRETE VERB, such as *time flies* or *thoughts escape*. Verbal metaphors have previously been investigated in developmental studies, and researching them allows us to consider the results of our study in the wider context of other research on conceptual metaphors (Özçalişkan, 2005a, 2007). Verbal metaphors have been found to show systematicity across languages (Turkish and English) (Özçaliskan, 2003). Table 1 illustrates the way that these abstract mappings can surface in various verbal metaphors that follow their logic. From a usage-based perspective, we can expect that, through the exposure to those systematic metaphors, children will gradually acquire their underlying patterns (Tomasello, 2008). Once children have acquired the patterns, understanding such metaphors would be expected to be easier, even for comparatively lower-frequency phrases such as *minutes crept by*. By testing verbal metaphors, we thus test children's knowledge of these types of vocabulary structures.

Metaphor comprehension, however, is more than a simple act of recalling lexicalised word meanings or word relations. Particularly for novel links between ideas, such as in *Paul is a pig* or *Tom felt like a rainbow*, understanding them has been argued to rely on understanding the implied similarity between two concepts (Camac & Glucksberg, 1984). Earlier developmental metaphor studies tended to focus on these NOMINAL METAPHORS that compare two entities, and found them be linked to domain-general reasoning abilities (Beaty & Silvia, 2013; Billow, 1975; Johnson, 1991; Kogan, Connor,

Table 1. Examples of Systematic Patterns in the Concepts Underlying Verbal Metaphors

	Abstract noun	Motion verb
Time	*time*	*flies*
	days	*pass*
	minutes	*drag*
Thought	*thoughts*	*cross (the mind)*
	ideas	*come*
	mind	*wanders*

Gross, & Fava, 1980). For example, Johnson (1991) found that nominal metaphors were interpreted more accurately by older, but not linguistically more proficient, learners (in their case, they compared monolinguals to Spanish–English learners of EAL). They concluded that the most important variable for children's comprehension of metaphors was not their language knowledge, but their reasoning skills (which were stronger in older children). The account of metaphors as recognising similarities and drawing analogies has fuelled computational (Gentner, Bowdle, Wolff, & Boronat, 2001) and early developmental research (Beaty & Silvia, 2013; Cicone, Gardner, & Winner, 1981; Waggoner, Messe, & Palermo, 1985). The degree to which metaphors are understood through either lexicalised knowledge or analogical reasoning is likely to vary according to the type of metaphor – nominal or verbal – that is being processed. We thus decided to consider both verbal and nominal metaphors when testing metaphor comprehension.

Another element in metaphor comprehension is their conventionality. Frisson and Pickering (2007) found that novel compared to conventional metonymies created additional processing effort in the reading process. They interpreted the additional effort as related to the need to come up with a plausible interpretation through conscious reasoning. This finding suggests that the conventionality of figurative phrases may influence the relative contribution of reasoning vs. language skills to comprehension. Furthermore, the conventionality of phrases matters particularly for learners of EAL. Kan (2014) and Martinez and Murphy (2011) demonstrated that even highly frequent multiword phrases can cause unnoticed comprehension difficulties in learners of EAL and L2 learners, respectively. Collectively, these studies outline a critical role of frequency or conventionality of a phrase for how it is processed. Conventionality was therefore operationalised as frequency in this study and acted as a continuous predictor of metaphor comprehension. One major difference between previous research and the present study is that we consider nominal and verbal metaphors of varying frequencies alongside each other, and link them to overarching questions on EAL vocabulary development.

A related and important methodological issue is how best to assess metaphor comprehension. Just as in traditional vocabulary tests that often combine subtests of receptive or productive knowledge, for example, different facets of metaphor comprehension can also be measured through different tasks. For instance, Özçalışkan (2007) used the following task combination: metaphors were presented orally to children embbeded in six to twelve short stories scaffolded by matching pictures. Metaphor comprehension was measured through two to three tasks, where

children first recalled the stories, answered a multiple-choice question posed by two puppets, and finally justified their responses. Each task measures metaphor comprehension at different levels: recall requires only the memory of the phrase and the ability to reproduce it. The multiple-choice questions and the reasoning tasks required the children to make sense of the metaphor in the story context. These two tasks thus involved also making linking inferences and comprehending metaphors at a higher level (Oakhill, Cain, & Nesi, 2016). The advantage of using several tasks is that metaphor comprehension can be measured at different difficulty levels. We thus decided to similarly measure different facets of metaphor comprehnsion by combining these same measures (recall, multiple-choice, and reasoning).

The age of the child is a further important element in metaphor comprehension. A number of studies have investigated developing metaphor understanding in primary school children to see at which age children begin to comprehend metaphors (Gardner, Kircher, Winner, & Perkins, 1975; Özçalışkan, 2007). Estimates range from four to five years, depending on the type of metaphor (Winner, Rosenstiel, & Gardner, 1976), the linguistic context, and tasks (Özçalışkan, 2007). There is converging evidence that metaphor comprehension remains challenging throughout the first years in primary school. In these early years, children with EAL are known to particularly lag behind their monolingual peers (Strand et al., 2015). We wanted to see how much children's metaphor comprehension differs in year 1 and 2, and identify whether there is a gap between children with EAL and English in this domain. We thus decided to test children with EAL and monolinguals in year 1 (ages five to six) and year 2 (ages six to seven), and to compare the metaphor comprehension in these important two years. We tested metaphor comprehension across a range of metaphor types that tap different degrees of phrase knowledge and reasoning skills (verbal vs. nominal of differing frequency), and through a variety of measures that tap different aspects of metaphor comprehension. We thus aimed to address the following research questions:

1. Do children in year 2 have a larger receptive vocabulary and better metaphor comprehension than children in year 1?
2. Do children with EAL differ relative to their monolingual peers in their receptive vocabulary and metaphor comprehension?

Children in year 2 were expected to score significantly higher on both receptive vocabulary and metaphor comprehension. Children with EAL were expected to score lower than monolinguals in receptive vocabulary and metaphor comprehension. Differences between language and year groups were expected to be larger on the areas of metaphor comprehension that most require efficient higher-level processing, and that rely most on analogical skills and efficient language processing. This would include understanding low-frequency and nominal metaphors, and performing on the reasoning measure.

Method

Participants

The participants were 43 monolinguals (20 girls) and 37 children with EAL (20 girls) in years 1 and 2. Year 1 and 2 refers to the first two educational levels in UK primary schools, which correspond to an age range of five to seven years of age (age range in current sample = 4;11–7;11). Language groups were matched on their chronological

age in both year 1 (M = 6;3, SD = 0;6 for children with EAL, and M = 6;2, SD = 0;3 for monolinguals) and year 2 (M = 7;1, SD = 0;6 for children with EAL, and M = 7;1, SD = 0;5 for monolinguals, all differences n.s.). The data were collected in three primary schools in Oxfordshire, UK. A comparison of area-related weekly household income levels for the schools' catchment areas suggests that the socioeconomic background of the children is fairly heterogeneous and represents different facets of society. Two children decided to discontinue the testing during the metaphor task because of fatigue, and some scores are missing for three other children due to technical problems or researcher error, resulting in 84 missing values (less than 3% of the data). All children were kept in the sample, and missing data was deleted case-wise in following analyses.

The children with EAL were defined as such by their school. To confirm this categorisation, a Language Background Questionnaire (LBQ) (a shortened version of the material used in McKendry & Murphy, 2011, see 'Appendix A') was administered to each child with EAL prior to testing. The questionnaire asks about the languages spoken with family and friends at home and at school. The children reported that their L1 was used to varying degrees at home, from some using it rarely ("Always/ mostly English"), to others reporting using and hearing it 50% of their time (range of scores 9–22, M = 13.6, SD = 3.8). The LBQ also asked children which language was spoken at home. The children in the EAL sample had a diverse set of home languages, including Arabic (n = 6), Urdu (n = 4), Spanish (n = 4), Russian (n = 2), French (n = 1), and Swahili (n = 2), as well as several languages spoken by one child only (Hebrew, Albanian, Pashto, Kikuyu, Malay, Chinese, Shona, Filipino, Polish, Lithuanian, Indonesian, and Japanese). Five children reported the use of a home language that they were unable to name.

To ensure that the EAL and monolingual groups were comparable on their non-verbal development, their performance was compared on the *Wechsler Abbreviated Scale of Intelligence* matrices subtest (WASI-II) (Wechsler, 2011). The WASI-II matrices test is commonly used in developmental research to compare target and control groups on their level of non-verbal reasoning. A *t*-test did not indicate any difference between the EAL (M = 12.0, SD = 4.6) and monolingual group (M = 11.8, SD = 4.2) in the WASI-II raw scores ($t(78) = -0.14$, $p > .05$). Furthermore, the two groups attended the same schools that were serving the same catchment areas, and can thus be assumed to come from similar socioeconomic backgrounds and receive comparable education. For further analyses, we will thus assume that the EAL and the monolingual group are reasonably comparable in their backgrounds and differ mainly on their language learning experiences.

Data-collection procedure

The children were tested individually in a quiet place in school while the rest of the class was participating in their normal in-class activities. Each testing session took approximately 30–40 minutes. Testing was initiated with a warm-up phase during which the researcher established a rapport with the child. After the warm-up, the researcher asked the children about their language exposure based on the LBQ, and administered the standardised measures and the metaphor comprehension test. Tests were administered in counterbalanced orders, apart from the questionnaire and the metaphor test that were the first and last measure, respectively.

Measures

RECEPTIVE VOCABULARY was tested with the *British Picture Vocabulary Scale* Second Edition (BPVS2) (Dunn, Dunn, Whetton, & Burley, 1997). Non-verbal reasoning was assessed with the matrices subtest of the *Wechsler Abbreviated Scale of Intelligence* (WASI-II) (Wechsler, 2011).

Given the lack of standardised measures, we developed a new test of METAPHOR COMPREHENSION. Our test assessed comprehension of both nominal (*to feel like a sinking shi*p) and verbal metaphors (*hours fly by*). Metaphors were sampled from previous studies (Özçalışkan, 2007; Waggoner *et al.*, 1985) as well as from an idiom dictionary (*Cambridge Idioms Dictionary*, 2006). Verbal metaphors were based on the domain mapping of 'time as movement in space', and 'ideas as moving objects', while nominal metaphors compared affective states and physical objects or situations. All metaphors were searched for in the *Oxford Children's Corpus* (OCC), a unique database compiling a sample of British school children's literature written for and by children (Banerji, Gupta, Kilgarriff, & Tugwell, 2013). To make sure that all metaphors were acceptable in the context of British English children's reading and writing, metaphors were only included in the study if they occurred at least once in the corpus. At the same time, we decided to include some extremely low-frequency items (with a count of 1 for example) as an acceptable but novel version of a metaphor. We wanted to include metaphors with few occurrences in the OCC since we expected them to be novel but interpretable, and since we were interested in how well children would understand metaphors that would require different degrees of recalling meanings or establishing new interpretations. To make sure that low-frequency metaphors were both novel and interpretable, we presented our metaphor stimuli to twenty-two adult speakers of English (24–56 years of age, mean age = 34) who took part in a questionnaire-based norming study. They were all students and professionals based at Oxford, had English as their native language, and half of them reported to also speak a second language. In the questionnaire, speakers rated our metaphor stimuli for familiarity ("How often have you heard or seen the following phrases before?") on a scale from 0 'Never' to 4 'Very often'. Next, they were asked to "tell us the meaning of each phrase in your own words". The meaning that participants reported was compared to the meaning that we expected the metaphors to have. Each entry was scored as either 1 if the reported and expected meanings matched, and 0 if they did not. Ten percent of the interpretability scores were double-checked by a second rater. We checked inter-rater reliability by calculating Cohen's kappa using the psych package (Revelle, 2018) in R (R Core Team, 2016). The inter-rater reliability was substantial (74% overlap, Cohen's kappa of .44), and all initial disagreements could be resolved when discussing the respective cases. The results of the norming showed that, as expected, the metaphors ranged from novel to familiar (M = 1.8, SD = 1.5, range 0–4), and familiarity was positively correlated with frequency in the OCC. Metaphors were highly interpretable to the adult speakers (M = 84%, SD = 36%), except for two with low scores (31% and 52%). To make sure that these stimuli did not bias the children's performance, we re-ran all our reported analyses without these two items and found that the results remained the same. It is possible that the metaphors were ambiguous only in isolation but that their meaning were clear in the story context we provided.

The twenty-four metaphors were presented in the context of 12 short stories (range = 67–107 words of length). Metaphors were presented in a story context since previous

research has found that, if figurative phrases are presented out of context, children under ten years of age are less able to demonstrate their understanding (Levorato & Cacciari, 1992). Half of the stories were adopted from previous studies (Özçalişkan, 2005b, 2007; Waggoner *et al.*, 1985), the other half were developed for this study but modelled after the stories used in previous research to ensure continuity. The appropriateness of the stories for young listeners was checked by examining the frequency of words used in stories through *Web Vocabprofile*. The results confirmed that all stories contained 80.96–89.90% from the 1000 most frequent words, a high proportion that indicates the appropriateness of the stories for a young population (Cobb, 2013). All metaphor items described a final action which was central to understanding the plot, as can be seen in the example in Figure 1 below: the story is open-ended until the metaphor is introduced. To understand the stories, children needed to perform various higher-level tasks with the metaphors. For example, they needed to understand the metaphor expressing the main character's emotional state and infer what had happened to them (for nominal metaphors), or the metaphor expressing the reason for the main character's behaviour and link them causally (for verbal metaphors). In both cases, children were required to make elaborative and/or cohesive inferences based on their understanding of the metaphors. The advantage of this task is that it allowed us to test children's text-level comprehension, while still being focused on the processing of metaphors. A potential drawback, however, is that the introduction of inferencing skills adds a potential source of noise, particularly since the nominal and verbal metaphors differed slightly in the inferences required. It is important to keep this complexity in mind in interpreting our findings.

Children heard two metaphors in each story (see above re metaphor pairs). We decided to include two metaphors per story to model our tasks as closely as possible on a series of preceding studies (Özçalişkan, 2005b, 2007) which had established reliability with these tasks, making reliability more likely in our context. Furthermore, the reuse of the same paradigms allows comparability across studies, which is particularly important in the context of researcher-designed measurement tools. Metaphors all appeared towards the end of stories, and the two metaphors appeared immediately or shortly after each other (range = 0–9 words distance between metaphors). We decided to accept this range since the main focus was on creating interesting and plausible story contexts for the chosen metaphors. Since all children heard the same stories, and since the spacing was not distinguishably different between metaphor types (M = 3.2, SD = 3.8 for verbal, and M = 1.3, SD = 0.8 for nominal metaphors, n.s.), any potential differences in spacing should affect both metaphor types and language groups equally, and should therefore not influence the comparisons drawn in our subsequent analyses. The two metaphors within one story relied on the same underlying domain mapping (e.g., *hours flash by* and *minutes hurry along*, both relying on mapping of '(quick) passing of time as (fast) movement in space'), and meant the same thing in the context of the story.

The metaphor test was administered in the following way: first, the three tasks were explained to each participant at the beginning of each session. The explanation was followed by a warm-up task that foreshadowed the multiple-choice question: puppets and children were shown squares of three different colours, and puppets were asked for their colour. The child had to choose the puppet who named the colour shown to the children. This procedure allowed the children to demonstrate that they had understood that they had to choose the puppet that answered the question correctly (Özçalişkan, 2007). After successful completion of the warm-up phase, the testing

began. Performance was audio-recorded for later analysis. All stories were presented orally: children listened to a recording of an adult native speaker of British English reading the stories. When listening, children looked at a picture story showing all actions in the story except for the event expressed through the metaphor (see 'Appendix B' for a full list of stimuli). As in previous studies with the same age group, pictures were included to help focus the children on the story when listening, and to support recall afterwards. We counterbalanced the order of stories and answer options, and the order of correct and incorrect answers, as well as which puppet gave which answer and in what order they spoke. The counterbalancing ensured that none of these factors would systematically influence the results.

We used three measures of metaphor comprehension – multiple-choice, recall, and reasoning. After listening to the story, children were asked "What did X do?" (where X stands for the character in the story) to elicit story recall. The researcher scaffolded the recall by encouraging children to continue ("And then?"), or to look at the pictures to help their memory. When participants appeared to have recounted all that they could remember, they were asked a forced-choice question which referred to the event expressed metaphorically (see Figure 1). Answers were given by two puppets (Sheep and Bear), and the child had to identify the puppet who provided the correct answer. After the child had made their choice, they were then asked "How did you know that Sheep/ Bear was right?" to elicit reasoning.

Questions and answers were modelled on preceding studies (Özçalişkan, 2005b, 2007). The two answer options followed the same sentence structure and were matched on syllable length for each question (range = 0–4 syllables difference). We decided to let the length of answers vary across stories to allow the necessary flexibility to create plausible answer options for each stimulus (range = 3–19 syllables). The answers were counterbalanced across puppets such that each puppet was associated with an equal number of correct and incorrect answers, and that between children, different puppets would be associated with the correct answers for different stimuli. It was then important to confirm that multiple-choice questions could only be answered when the information from the metaphor was included. This was achieved by piloting the stories and questions with a new group of adult native speakers (N = 32, aged 21–30 years), following a procedure adapted from a previous study (Özçalişkan, 2005b). During the pilot, participants read each story twice: first, they read all stories without metaphors and answered the comprehension questions, and then they read the stories again – this time with the metaphors – and answered the comprehension questions based on this second reading. If the metaphor was necessary and sufficient to answer the comprehension questions correctly, the adult native speakers should be at chance when answering the questions after the first reading (where they did not have access to the information from the metaphor), but should be above chance when answering the questions after reading the stories that included metaphors. The results matched this expectation: participants performed indistinguishably from chance in the first round (M = 60% correct, SD = 23.1%), and significantly above chance in the second round (M = 95.4% correct, SD = 4.7%; based on cut-off of 78.3% for performance significantly above chance, determined from a 95% confidence interval around chance performance).

Children's performance on the three measures was scored as either 1 = correct or 0 = incorrect. Multiple-choice questions were scored as 1 if the child had chosen the correct answer. Recall was scored as 1 if the child had reproduced the metaphor from the story (e.g., "hours flash by"), a plausible variant of it (e.g., "because hours fly quickly" for "hours flash by"), or a plausible literal equivalent of the metaphor meaning (e.g.,

Figure 1. Example administration of one study in the metaphor comprehension test.

"because time was so short" for "hours flash by"). Reasoning was scored as 1 if the child justified their choice to the multiple-choice question by directly referring to the metaphor, again either by reproducing it, producing a plausible figurative variant, or a literal equivalent. Ten percent of the scoring was also done by a second rater. We again checked inter-rater reliability by calculating Cohen's kappa using the psych package and R and found that the other rater's scores substantially overlapped with our codings (84%, Cohen's kappa of .65). All mismatches between codings could be resolved in a discussion of the questionable items.

We checked the validity of our measures for internal consistency using the alpha function in the psych package in R. Cronbach's α showed low internal consistency ($\alpha = .34$) for the multiple-choice tasks, but a good consistency for the recall ($\alpha = .70$) and reasoning tasks ($\alpha = .70$). The low alpha values in multiple-choice is likely due to the ceiling effect, with monolingual children scoring near to the maximum on the multiple-choice task ($M = 86.08$, $SD = 11.84$), an effect that indicates that multiple-choice fails to efficiently distinguish between low and high achievers in the monolingual group. However, there was far less of a ceiling effect for the children with EAL ($M = 77.67$, $SD = 12.28$), suggesting that the multiple-choice task discriminates well at lower ability levels.

Results

All analyses were conducted using R (R Core Team, 2016). This section describes the results by research question. Children in year 1 and year 2 as well as children with EAL and monolinguals were compared on their receptive vocabulary using a two-way Analysis of Variance (ANOVA). The ANOVA was calculated using the lm function from the MASS package (Venables & Ripley, 2002). Metaphor comprehension scores contained repeated measurements and could thus be analysed using a mixed effect model (MEM). We chose MEMs since they allowed us to model continuous and categorical predictors, and to control for random effects on both participant and item level (Baayen, 2008). Since metaphor comprehension is a binary outcome variable, the MEM was fitted as a binomial model using the glmer command in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and the optimx package to support convergence (Nash, 2014). The MEM was fitted with all relevant predictors and their interactions as fixed effects (language group, year group, measure, metaphor type, and frequency), and with crossed random effects. Contrasts for categorical predictors were defined using effect coding. Frequency was entered as a re-scaled and centered continuous predictor. To check model fit, the residuals were checked for normality using the dharma package (Hartig, 2018). It was further checked that no outliers had biased the model results by checking the robustness of all reported beta coefficients and p-values using the influence.ME package (Nieuwenhuis & Grotenhuis, 2012). All checks confirmed that the model is reliable and fits the data well. For the ANOVA, we report the effect size Cohen's d. There is no conventionalized way of reporting standardized effect sizes for MEMs given the difficulties of determining degrees of freedom, and it has been suggested that they could be calculated from raw data (Nakagawa & Schielzeth, 2013). We chose to report odds ratios (OR) which indicate the relative likelihood across groups of an event happening (here, getting the answer right). OR have the advantage of avoiding the difficulty of degrees of freedom, and are a conventional way of reporting effect sizes for logistic regressions (of which the binomial MEMs we ran are a variant) (Field, 2009). Table 2 shows the results of the model.

Do children in year 2 show a larger receptive vocabulary and better metaphor comprehension than children in year 1?

Children in year 2 had larger receptive vocabularies [(M = 66.3, SD = 14.4) in year 1, and (M = 78.1, SD = 12.3) in year 2], and stronger metaphor comprehension than children in year 1 [(M = 42.2, SD = 49.4) in year 1, and (M = 55.3, SD = 49.7) in year 2, d = .88)]. The main effect on metaphor comprehension, however, interacted with other factors. To explore these interactions, we calculated predicted means in a cell means model and compared them using the glht function from the multcomp package (Hothorn, Bretz, & Westfall, 2008). When investigating the effect of the continuous frequency measure, we compared the predicted means for ±1 SD of frequency, which will also be reported in the following figures (marked as low- and high-frequency). The results of these contrasts indicated that the advantages of children in year 2 was significant on low-frequency metaphors (b = 0.53 (0.13), z = –4.2, p < .0001, OR = 0.57), but not on high-frequency metaphors (n.s., see Figure 2). The interaction was further qualified by a three-way interaction with metaphor type. Children in year 2 scored higher on low-frequency nominal than

Table 2. Mixed Effect Model Results Predicting Metaphor Comprehension from Children's Year Group and Language Group, Metaphor Measures, and Metaphor Type and Frequency, as well as All Significant Interactions

Main effects:	b(se)	z	p
Language group	0.21 (.11)	1.91	.55
Year group	0.38 (.11)	−3.52	<.001
Measure	1.82 (.08)	21.94	<.0001
Metaphor type	−0.10 (.11)	−0.93	.36
Metaphor frequency	−0.08 (.28)	−0.28	.78
Significant interactions:			
Language group × Measure	0.15 (.07)	2.19	<.05
Year group × Measure	0.17 (.08)	2.17	<.05
Measure × Metaphor type	−0.22 (.08)	−2.77	<.01
Year group × Frequency	0.31 (.15)	2.10	<.05
Measure × Frequency	−0.88 (.22)	−4.10	<.0001
Year group × Metaphor type × Frequency	−0.31 (.15)	2.10	<.05
Measure × Metaphor type × Frequency	0.46 (.22)	2.11	<.05
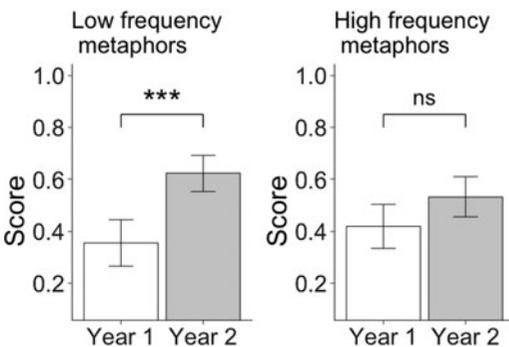
Figure 2. Performance of children in year 1 and year 2 on the comprehension of low- and high-frequency metaphors (*** $p < .001$).

low-frequency verbal metaphors ($b = 0.20$ (0.09), $z = 2.2$, $p < .05$, OR = 0.73), a difference that did not show in year 1 or in high-frequency metaphors (all contrasts n.s.). Further interactions with metaphor measure indicated that the difference was only significant in the multiple-choice measure ($b = 1.28$ (0.44), $z = −2.9$, $p < .01$, OR = 0.87, all other contrasts n.s.).

Do children with EAL lag behind their monolingual peers in their receptive vocabulary and metaphor comprehension?

Children with EAL scored lower than their monolingual peers in receptive vocabulary [(M = 67.3, SD = 14.3) for EAL, and M = 75.6, SD = 14.0) for monolinguals, $d = .59$)] and marginally in metaphor comprehension [(M = 0.45, SD = 0.50) for EAL, and M
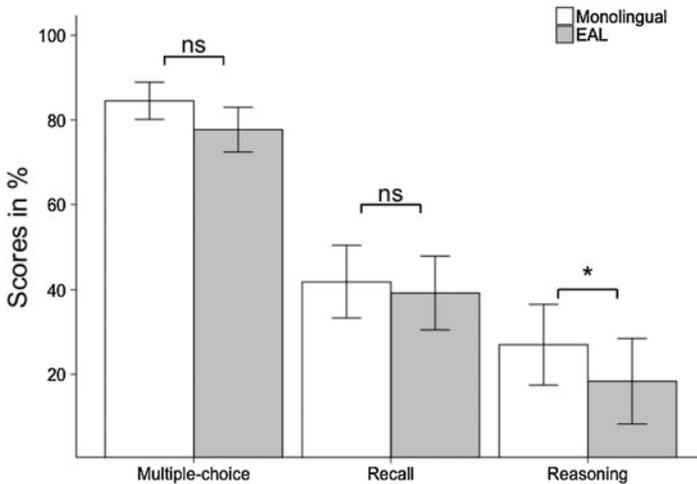
Figure 3. Performance of children with EAL and monolinguals on metaphor comprehension across different measures (* *p* < .05).

= 0.51, SD = 0.50) for monolinguals, *p* = .055, OR = 0.88)]. The language group differences on metaphor comprehension furthermore interacted with the measure (see Figure 3). Planned contrasts showed that children with EAL scored significantly lower on the reasoning measure only [(*b* = 0.62 (0.27), *z* = 2.3, *p* < .05, OR = 0.60, other contrasts numerically in the same direction, but n.s.)], indicating that EAL children were only noticeably weaker on the most demanding of the three tasks. A main effect of tasks furthermore showed that the three measures were of different difficulty, with multiple-choice being easiest, followed by recall, with reasoning as the most difficult measure. There was no interaction with year group or metaphor type, indicating that the gap between monolinguals and children with EAL was small but persistent across years and phrase types.

Discussion

We found that children in year 2 had larger vocabularies and stronger metaphor comprehension than their younger peers. The advantage on metaphor comprehension was carried by the older children's higher ability with low-frequency metaphors. These findings are in line with previous research that reports a steady increase in metaphor comprehension over the years (Gardner *et al.*, 1975; Özçalışkan, 2007). What, then, might children in year 2 have that helps them understand metaphors better than their younger peers? For one thing, we know from the comparison on the BPVS that they had larger – and probably also deeper – vocabulary knowledge. Their stronger metaphor comprehension plausibly stems from more well-developed knowledge of words, and their extended meanings and usage. Particularly for the systematic verbal metaphors, older children's stronger comprehension may be supported by implicit knowledge of the structures underlying them, and be a first indication of the 'burst of multi-word vocabulary' that has been reported for children in years 4–5 (Smith & Murphy, 2015). However, we also know that metaphor comprehension is more than

recalling word meanings. The low-frequency metaphors in our study especially were novel even to most adults. The majority of our young participants likely heard them for the first time when doing our test. What allows all speakers to make sense of such novel phrases is the ability to make inferences about meanings, and to build plausible interpretations by drawing analogies (Frisson & Pickering, 2007; Oakhill, Cain, & Nesi, 2016). These skills develop over time and are likely to have been larger in our older participants, which would have supported their comprehension of novel metaphors (Johnson, 1991).

The interaction of year group with frequency was further qualified by interactions with the measure and metaphor type. These interactions indicated that the advantage of children in year 2 on low-frequency metaphors was driven by their performance on nominal metaphors in the multiple-choice task. There are several candidate explanations for these results: The low-frequency nominal metaphors may have been particularly salient to the children because they always described the main character's emotional event (instead of an abstract idea like time passing, as the verbal metaphors). Metaphors on people and emotional states are learned early on (Winner *et al.*, 1976). Verbs are also more complex and carry more grammatical information than nouns, and are known to be learned later than nouns (Waxman, Fu, Arunachalam, Leddon, Geraghty, & Song, 2013). The grammatical structure may have contributed to older children's advantage on nominal metaphors. The difference also showed most on the multiple-choice task, the only receptive task. Nominal metaphors seem thus to only be easier to understand, but not necessarily to reproduce and explain. An advantage for the production of verbal metaphors may, for example, be that they follow a conventionalised structure which may help to produce them, or variants of them (an advantage that nominal metaphors do not have). All these explanations are, however, tentative at this point and would need to be tested in further research.

We further found that, in line with past research, children with EAL in our study had smaller receptive vocabularies than their monolingual peers (Bialystok *et al.*, 2009; Pearson *et al.*, 1993). To our knowledge, this research is the first to replicate the gap in English receptive vocabulary of children with EAL in a UK sample at the age studied here (students in year 1 and year 2). A comparison with studies in the US context shows that, for similar age ranges, the mean differences between monolinguals and children with EAL on the standardised vocabulary scores are very similar (14.68 in our study and 16.68 in Morales, Calvo, & Bialystok, 2012). This overlap across educational contexts indicates the, at least partial, applicability of research findings on vocabulary development to the UK context, where less research is available into minority language learners.

We further found that children with EAL had lower metaphor comprehension than their monolingual peers. An interaction effect showed that EAL learners had overall lower metaphor comprehension because of their weaker performance on the reasoning task. To understand why this is the case, it is first worth thinking about the differences between reasoning and the other two tasks. Previous research has found that children tend to find explaining metaphors harder than recalling them, or answering multiple-choice questions on them (Johnson, 1991; Özçalışkan, 2007). Similarly, in this study, multiple-choice yielded the highest scores, followed by recall and then reasoning. Reasoning was the hardest task, and the reason for this is likely to lie in the number of skills it requires. To successfully do the reasoning task, the children had to remember their answer on the multiple-choice question, remember the metaphor, and link the two in their explanation. The reasoning task requires both receptive memory, reasoning,

and free production. Reasoning is a higher-order task compared to recall and multiple-choice. The increased demands of higher-order tasks have been found to limit a speaker's ability to produce or comprehend linguistic features that they might show mastery of on an easier task (Robinson, 2001). The same may be true for children with EAL on the reasoning measure. They were competent with easier tasks on the metaphors, but struggled when the task demands were raised. For teaching, these findings suggest that teachers can expect difficulties in metaphor comprehension to show most strongly in more challenging tasks; the more demanding a task becomes, the more difference it will make if language comprehension is more effortful or incomplete. Importantly, the resulting differences in performance between monolinguals and children with EAL can easily be confused with academic difficulties. It is crucial that practitioners are aware of such language features that their students are likely to struggle with in classroom interactions and teaching materials. The finding highlights the importance of looking at vocabulary beyond single words to understand how vocabulary items are learned and processed in the company of other words. Previous research has already noted the importance of including multi-word vocabulary in the investigation of EAL comprehension research (Kan, 2014; McKendry, 2014; McKendry & Murphy, 2011; Smith & Murphy, 2015). The current study adds to this literature by highlighting the differential development of metaphor comprehension across tasks.

When discussing the language skills in students with EAL, we also need to acknowledge that any group findings will only ever represent overarching trends in a heterogeneous group. As we know from previous research, different students with EAL vary greatly in the languages they speak at home, their English language skills, and their overall educational attainment (NALDIC, 2011; Strand & Demie, 2005; Strand et al., 2015). This heterogeneity is also refleted in our sample, where children varied greatly in their exposure to the home language, which is likely to be linked to different levels of L1 and L2 proficiency. Given the lack of previous research, these mixed proficiency levels might best represent the overall UK population. This means for our findings that, despite the overarching trend for children with EAL to underperform on metaphor comprehension, this finding will not hold for the more skilled English speakers amongst them. Furthermore, we tested children from different L1 backgrounds. Due to the small number of children in each language group, L1 groups were not considered as variables in the analyses. Grouping children together with different L1s is problematic in so far as potentially interesting L1 effects will be lost in the analyses. On the other hand, the L1 diversity also renders the study more ecologically valid, as language diversity is an inherent aspect of the UK EAL population. Different from what is commonly done in EAL research, we also decided against selecting only children with particular language problems, but instead tested the whole range of students. While our sampling method may introduce potential noise and complexities to the data, it also makes our sample a true sample of the highly heterogeneous EAL population in the UK. Our participants are a realistic representation of the whole range of learners of EAL that teachers will encounter in the classroom. This is particularly important in a population where little research has been done so far.

Limitations

Our reflections on developmental trajectories have to be interpreted with some care since we based them on cross-sectional data (where ideally it would be longitudinal). Language group differences were interpreted as differences in language backgrounds, but may be influenced by other factors as well, such as socioeconomic background.

At the same time, the students with EAL and monolinguals were all each other's peers, being schooled in the same classrooms and living in the same neighbourhoods, and matched on age and non-verbal reasoning, which indicates a substantial, albeit not perfect, match between the groups. External validity might, however, be further limited by sample size (roughly n = 20 per group). Arguably, another possible limitation of our research could be that the data were collected using a researcher-designed measure, resulting in limited reliability and validity statistics. For example, the task's reliability may be limited by irregularities in the metaphor spacing, the choice to insert two metaphors in one story, the clarity of picture stimuli, or the degree to which inferencing skills were involved in understanding nominal and verbal metaphors in the story context. Nonetheless, as described in the 'Method' section, we took several measures to ensure the test's reliability, and our findings chime with similar results in the literature, which should serve to somewhat mollify this concern.

Conclusion

We explored the vocabulary knowledge and metaphor comprehension of monolinguals and learners of EAL in year 1 and 2. Our study builds on research highlighting the early school years as an important time for children to develop their vocabulary depth. It is also a time where many children with EAL lag behind their monolingual peers considerably, in both language skills and educational attainment overall. We thus conducted a study to investigate different facets of metaphor comprehension in these groups.

The results showed that children in year 2 had a considerable advantage over their younger peers in metaphor comprehension overall, and particularly on low-frequency or novel metaphors. We suggest that this age difference is due to two simultaneous processes: the increase in vocabulary knowledge (as we could also see in older children's larger receptive vocabulary), and the development of children's reasoning skills with age. We also saw that children with EAL lag behind their monolingual peers in overall vocabulary size and in metaphor comprehension. Their difficulties with metaphors seem to be particularly relevant in the higher-order task of reasoning that requires making and explaining inferences about the metaphors. This finding is in line with the idea that weaker language skills limit particularly the ability with higher-order tasks. The link between vocabulary gaps and task performance indicates ways in which an EAL vocabulary gap may affect their overall educational performance in the classroom. Of course, not all children with EAL will lag behind their monolingual peers – we know that there is a large heterogeneity in the EAL population and in the sample of our study (Strand *et al.*, 2015). For those who do have comprehension delays, however, comprehension difficulties on demanding tasks are likely to negatively impact their educational attainment overall (Strand & Demie, 2005). Our findings highlight the need to support the specific learning needs of students with EAL in order to militate against any delays in vocabulary development; delays which could impact their ability to engage in the multiple demands of a classroom setting.

The current study provides evidence on vocabulary items worth including in educational interventions that focus on vocabulary. Continuing professional development courses could further aim at raising teachers' awareness for potential comprehension difficulties of younger children and those with EAL. Beyond the applied interest of this research, this study has for the first time explored different facets of metaphor comprehension within one study. Our study illustrates, therefore, how developmental research can enrich both educational practice and research on figurative language.

References

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

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

Banerji, N., Gupta, V., Kilgarriff, A., & Tugwell, D. (2013). *The Oxford Children's Corpus: a corpus of children's writing, reading and education*. Retrieved from <https://www.sketchengine.eu/wp-content/uploads/Beebox_2013.pdf>.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beaty, R. E., & Silvia, P. J. (2013). Metaphorically speaking: cognitive abilities and the production of figurative language. *Memory & Cognition*, *41*(2), 255–67.

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–50.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2009). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–31.

Billow, R. M. (1975). Cognitive developmental study of metaphor comprehension. *Developmental Psychology*, *11*(4), 415–23.

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'année Psychologique*, *114*, 647–62.

Camac, M. K., & Glucksberg, S. (1984). Metaphors do not use associations between concepts, they are used to create them. *Journal of Psycholinguistic Research*, *13*(6), 443–55.

Cambridge idioms dictionary, **2nd ed.** (2006). Cambridge University Press.

Cameron, L., & Besser, S. (2004). *Writing in English as an additional language at key stage 2*. Retrieved from <https://www.naldic.org.uk/Resources/NALDIC/Research%20and%20Information/Documents/RR586.pdf>.

Castles, S. (2013). The forces driving global migration. *Journal of Intercultural Studies*, *34*(2), 122–40.

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

Cobb, T. (2013). Web Vocabprofile, an adaptation of Heatley, Nation & Coxhead's (2002) Range. Retrieved from <http://www.lextutor.ca/vp/eng/>.

Department of Education (2013). Schools, pupils, and their characteristics, January 2013. Retrieved from <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/207670/Main_text-_SFR21_2013.pdf>.

Dunn, L. M., Dunn, D. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale II: BPVS II*, 2nd ed. Windsor: NFER-Nelson.

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.

Frisson, S., & Pickering, M. (2007). The processing of familiar and novel senses of a word: why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, *22*(4), 595–613.

Gardner, H., Kircher, M., Winner, E., & Perkins, D. (1975). Children's metaphoric productions and preferences. *Journal of Child Language*, *2*(1), 125–41.

Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *The analogical mind: perspectives from cognitive science.* (pp. 199–253). Boston, MA: MIT Press.

Hartig, F. (2018). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.2.0. Retrieved from <https://CRAN.R-project.org/package=DHARMa>.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–63.

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

Johnson, J. (1991). Developmental versus language-based factors in metaphor interpretation. *Journal of Educational Psychology*, *83*(4), 470–83.

Kan, R. (2014). *Effects of frequency and idiomaticity on second language reading comprehension in children*. Unpublished Master's thesis, University of Oxford.

Kogan, N., Connor, K., Gross, A., & Fava, D. (1980). Understanding visual metaphors: developmental and individual differences. *Monographs of the Society for Research in Child Development*, *45*(1), 1–78.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press.

Levorato, M. C., & Cacciari, C. (1992). Children' s comprehension and production of idioms: the role of context and familiarity. *Journal of Child Language*, *19*(2), 415–33.

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*(2), 267–90.

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

McKendry, M., & Murphy, V. A. (2011). A comparative study of listening comprehension measures in English as an additional language and native English-speaking primary school children. *Evaluation & Research in Education*, *24*(1), 17–40.

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, *140*(2), 409–33.

Morales, J., Calvo, A., & Bialystok, E. (2012). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*(2), 187–202.

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

Murphy, V. A. (2018). Literacy development in linguistically diverse pupils. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: the state of the science across its subfields* (Studies in Bilingualism, 54) (pp. 315–23). Amsterdam: John Benjamins.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–42.

NALDIC (National Association of Language Development in the Curriculum) (2011). *EAL and ethnicity in schools nationally and by LA*. Report published by the National Association of Language Development in the Curriculum (NALDIC), January 2011.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*(2), 1–14.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nieuwenhuis, R., & Grotenhuis, M. (2012). influence.ME : tools for detecting influential data in mixed effects models. *R Journal*, *4*(2), 38–47.

Oakhill, J., Cain, K., & Nesi, B. (2016). Understanding of idiomatic expressions in context in skilled and less skilled comprehenders: online processing and interpretation. *Scientific Studies of Reading*, *20*(2), 124–39.

OECD (2006). Where immigrant students succeed: a comparative review of performance and engagement in PISA 2003. *Intercultural Education*, *17*(5), 507–16.

Özçaliskan, Ş. (2003). Metaphorical motion in crosslinguistic perspective: a comparison of English and Turkish. *Metaphor and Symbolic Activity*, *18*(3), 189–228.

Özçaliskan, Ş. (2005a). Metaphor meets typology: ways of moving metaphorically in English and Turkish. *Cognitive Linguistics*, *16*(1), 207–46.

Özçaliskan, Ş. (2005b). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language*, *32*(2), 291–318.

Özçalışkan, Ş. (2007). Metaphors we move by: children's developing understanding of metaphorical motion in typologically distinct languages. *Metaphor and Symbol*, *22*(2), 147–68.

Pearson, B. Z., Fernandez, S., & Oller, D. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Language Learning*, *43*(1), 93–120.

R Core Team (2016). R: a language and environment for statistical computing. Retrieved from <http://www.r-project.org/>.

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. Evanston, IL: Northwestern University.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influence on SLA. In P. Robinson (Ed.), *Cognition and second langauge instruction* (pp. 287–318). Cambridge University Press.

Smith, S. A., & Murphy, V. A. (2015). Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, *28*(3), 347–69.

Spencer, M., & Wagner, R. K. (2017). The comprehension problems for second-language learners with poor reading comprehension despite adequate decoding: a meta-analysis. *Journal of Research in Reading*, *40*(2), 199–217.

Strand, S., & Demie, F. (2005). English language acquisition and educational attainment at the end of primary school. *Educational Studies*, *31*(3), 275–91.

Strand, S., Malmberg, L., & Hall, J. (2015). English as an Aditional Language (EAL) and educational achievement in England: an analysis of the national pupil database, (January), 1–94. Retrieved from <https://ore.exeter.ac.uk/repository/bitstream/handle/10871/23323/EAL_and_educational_achievement2.pdf?sequence=1>.

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

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*, 4th ed. New York: Springer.

Waggoner, J. E., Messe, M. J., & Palermo, D. S. (1985). Grasping the meaning of metaphor: story recall and comprehension. *Child Development*, *56*(5), 1156–66.

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H. (2013). Are nouns learned before verbs? Infants provide insight into a longstanding debate. *Child Developement*, *7*(3), 1–8.

Wechsler, D. (2011). *WASI -II: Wechsler abbreviated scale of intelligence*, 2nd ed. San Antonio, Tex.: Psychological Corporation.

Whiteside, K. E., Gooch, D., & Norbury, C. F. (2017). English language proficiency and early school attainment among children learning English as an additional language. *Child Development*, *88*(3), 812–27.

Wiig, E. H., & Secord, W. (1992). *Test of word knowledge (TOWK): complete kit*. Cambridge: Pearson.

Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, *12*(4), 289–97.

Appendix A

Language background questionnaire

Name		Code		Date	
School		Class		Teacher	

What languages can you speak?

Who else lives with you at home?

Mum/Caregiver 1   Dad/Caregiver 2   Brothers   Sisters   Anybody else?

Tick: ....

A. INTERPERSONAL INTERACTION	
A1. Which language does your Mum/Caregiver 1 speak at home?	**Score (circle)**
i. Always/Mostly English	1
ii. English and L1 equally	2
iii. Mostly L1	3
Other	
A2. Which language does your Dad/Caregiver 2 speak at home?	
i. Always/Mostly English	1
ii. English and L1 equally	2
iii. Mostly L1	3
Other	
A3. Which language do you use when you talk with your Mum/Caregiver 1?	
i. Always/Mostly English	1
ii. English and L1 equally	2
iii. Mostly L1	3
Other	
A4. Which language do you use when you talk with your Dad/Caregiver 2?	
i. Always/Mostly English	1
ii. English and L1 equally	2
iii. Mostly L1	3
Other	
(If any other adults reported to live with the child answer A5. If not, go to Question A6).	
A5. Which language do you use when you talk with any other adults who live at home?	
i. Always/Mostly English	1
ii. English and L1 equally	2
iii. Mostly L1	3
Other	

A6. What language do you use when you talk with your brothers and sisters?		
i.	Always/Mostly English	1
ii.	English and L1 equally	2
iii.	Mostly L1	3
Other		

A7. Overall, what language do you hear the most in your home?		
i.	Always/Mostly English	1
ii.	English and L1 equally	2
iii.	Mostly L1	3
Other		

A8. Are there any other children in this school who speak (insert L1)?	
Yes No	
(If Yes, complete Question A8a. If No, go to Question A8b).	

A8a. Which language do you use in the playground with other children who speak (insert L1)?		
i.	Always/Mostly English	1
ii.	English and L1 equally	2
iii.	Mostly L1	3
Other		

A8b. Do you know any other children outside your family who speak (insert L1)?	
Yes No	
(If Yes, complete Question A8c. If No, go to Section B).	

A8c. Which language do you speak with these children?		
i.	Always/Mostly English	1
ii.	English and L1 equally	2
iii.	Mostly L1	3
Other		

Section A Total: Total Score / Number of Times Points Awarded =	

Appendix B

Full list of stimuli

(Target phrases are underlined, correct answers marked with an asterisk)

Verbal metaphors

1. This is Jack. Jack is doing his maths homework so that afterwards, he can play with his friends. Jack does not like maths and does not know how to do today's homework. His dad is good at maths, and offers to help him. Jack's dad explains how to do the calculations. Jack wants to listen, but his favourite story <u>crosses his mind</u>. All the funny stories keep <u>passing through his mind</u>. When his dad asks him to finish the homework, Jack does not know how to do it.

Why couldn't Jack finish his math homework?
A: Because Jack couldn't understand what his dad explained to him.
*B: Because Jack started daydreaming instead of listening to his dad.

2. This is Amy. Amy wants to go out and play with her friends. She asks her mother 'Mum, can I go out?' Her mum says 'Yes, you can go out.' Amy happily goes out to play. Meanwhile her mother puts dinner in the oven and goes out to work in the garden. <u>Time flies</u> away very quickly. Amy does not realise how many <u>hours pass</u> by. When Amy comes home, she can't find any dinner.

Why can't Amy find dinner when she comes home?
*A: Amy plays for a long time and forgets to go home for dinner.
 B: Her mum gets busy with other things, forgets the food and it burns.
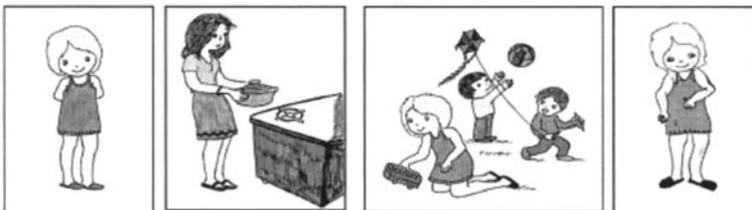
3. This is Sarah. Sarah goes outside to play with her friends in the sun. It is a very sunny summer day, and the birds sing. Sarah really likes to be outside in the sunshine. Sarah's friends said they were going to come to play with her soon. She starts playing with a cute little kitten until her friends would arrive. The <u>minutes drag on. Hours crawl by</u>. She decides to go inside.

Why does Sarah decide to go inside?
 A: Because the kitten ran away, and Sarah could not play with it anymore.
*B: Because Sarah had waited for her friends for a very long time.
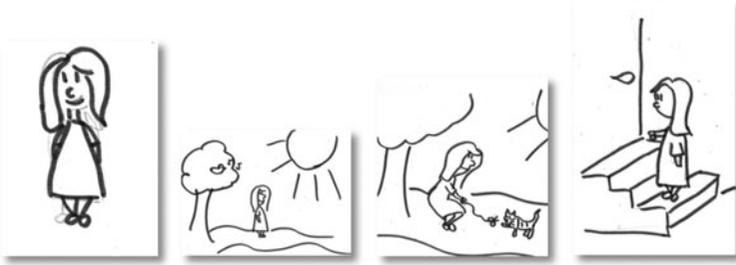
4. This is Lucy and this is her mum. Her mum gives Lucy some money and asks her to buy milk, bread, and eggs from the nearby grocery store. While Lucy is on her way to the store, she sees a candy seller on the street. She thinks how delicious the candy looks. Then she sees a balloon man and thinks how colourful the balloons are. Even though she arrives at the store, a lot of ideas wander in Lucy's mind now. The things that her mum wanted escape from her mind. Lucy returns home without buying the groceries her mum wanted.

Why didn't Lucy buy the groceries her mum wanted?
*A: Because she forgot what her mum told her to buy.
 B: Because she bought candies with the money.

5. This is John. John really loves reading, and he got a new story book for his birthday. John wants to read one story in his new book before he has to go to lunch. He starts reading an exciting story about dragons. He really wants to know how the story finishes. But minutes hurry along. Hours flash by. When his mum calls him for lunch, John still has not finished the story.

Why didn't Tom finish his story before lunch?
*A: Because there was not enough time until lunch.
 B: Because he got bored halfway through.

6. This is Sam and this is his grandma. Sam loves his grandma very much, but he does not get to see much of her, because she lives in a different city. This is Sam's mum. His mother tells him that he

might go see grandma when school closes. But <u>time drips by. Days slowly crawl by</u>. Sam can't even study his lessons and gets bad grades.

Why can't Sam study his lessons and gets bad grades?
  A: Because Sam is worried that his mum will go see his grandma without him.
*B: Because there is too much time before school closes and Sam gets impatient.
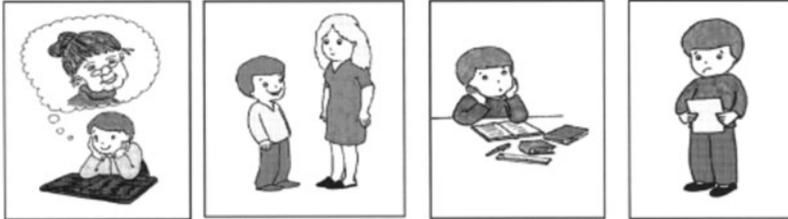
Nominal metaphors

1. This is Betty and this is her father. Betty goes to the fair with her father. Betty sees a big stuffed animal she wants her Dad to win for her. All he has to do is to knock over the bottles with three baseballs. She begs him to try. He pays the man for three balls. Betty watches him throw and hopes he can do it. She would feel so sad if she couldn't take the animal home and so happy if the animal were hers. After her dad has thrown the third ball, Betty is <u>on top of the world</u>. She <u>is on cloud nine</u>.

How did Betty feel after her dad had thrown the third ball?
*A: Betty was happy.
  B: Betty was sad.

2. This is Johnny with his puppy. One day when Johnny is playing with the puppy, it runs into some woods where lots of snakes live. Johnny goes into the woods to look for it. Johnny wonders whether he will find the puppy or see the snake. He sees something moving a bush and moves closer to see what it is. If he saw a snake, he would be sad, but if he found the puppy he would be happy. When he looks behind the bush, <u>his heart sinks</u>. He <u>has a big lump in his throat</u>.

How did Johnny feel when he looked behind the bush?
*A: Johnny was sad.
  B: Johnny was happy.
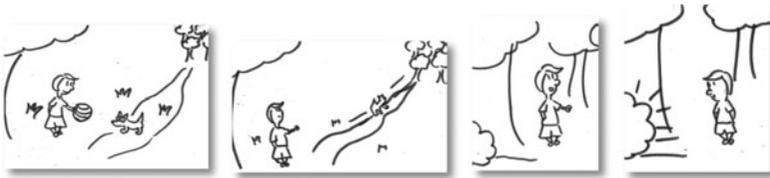
3. This is Rosemary. Rosemary was walking down a dark street on her way home. The moon was bright, and there were lots of shadows moving across her path. As she got closer to home, she saw a shadow that seemed to be following her. She did not know what to do. If it was someone she didn't know she would be afraid, but if it was her brother trying to scare her she would be angry. Finally she turned around to see who it was and it made her blood run cold. Rosemary was shaking like a leaf.

How did Rosemary feel when she turned around?
*A: Rosemary was afraid.
 B: Rosemary was angry.

4. This is Tom. Tom is a really good football player. His team plays in a really important match against another team. Tom and his friends really want to win. There are only two minutes left. Tom's team needs only one more goal to win. Tom is standing just in front of the goal and he shoots. He would be so sad if he missed, and he would be so happy if he scored the goal. When he sees what happened, he is a big smile. He feels like a colourful rainbow.

How did Tom feel after he had kicked the ball?
 A: Tom was sad.
*B: Tom was happy.

5. This is Susie. Susie has got a pet dog that she loves very much. Since last week, her little dog has been really sick. Susie and her mum take the dog to the vet to know if he can heal the dog. After examining her dog, the vet steps out to Susie who is waiting for him in the waiting room. She

would be so happy if he told her that her dog could be healed, and so sad if the dog would have to die. When she heard the vet's words, Susie felt like a sinking ship. She was a cold stone statue.

How did Susie feel when she heard the vet's words?
 A: Susie was happy.
*B: Susie was sad.
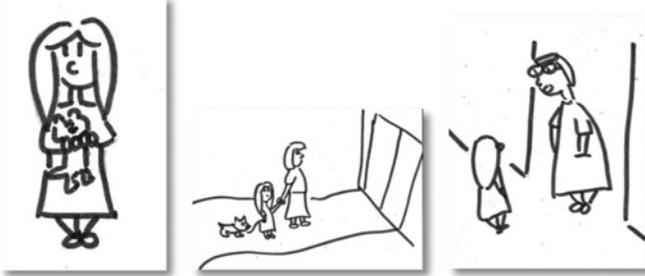
6. This is Michael. After a long nice day of playing in the sun, Michael went to his room and to bed. When he was almost asleep, he heard someone outside his room. He decided to go see who it was. He would be really scared if it was a stranger, and he would be really happy if it was his grandma coming to see him. When Michael opened the door, he looked at the person outside and he became a shivering animal. He was a bundle of nerves.

How did Michael feel when he saw the person in the corridor?
 A: Michael was happy.
*B: Michael was afraid.