

Comprehension monitoring during reading: an eye-tracking study with children learning English as an additional language

Annina K. Hessel^{1,3}, Kate Nation² and Victoria A. Murphy¹

¹Department of Education, University of Oxford

²Department of Experimental Psychology, University of Oxford

³Department of Educational Psychology, University of Goettingen

This article has been accepted for publication in *Scientific Studies of Reading*.

Author Note:

Correspondences concerning this article should be addressed to Annina K. Hessel,
University of Goettingen, Waldweg 26, 37073 Göttingen, Germany; e-mail:
annina.hessel@psych.uni-goettingen.de

Abstract

This experiment investigated comprehension monitoring in children learning English as an additional language (EAL) compared to monolinguals. Sixty-three 9–10-year-old children read texts containing an internal inconsistency (e.g. a barking kitten vs. barking puppy) while their eye movements were monitored. Standardized tests measured word reading fluency and vocabulary size and the children completed a questionnaire tapping rereading behavior. There was no overall difference between EAL and monolingual children. Regardless of EAL status, children with larger vocabularies were more efficient in their re-analysis of inconsistent information, as revealed by regressive eye movements. However, rereading of inconsistent vs consistent words in the eye movement record was not related to children's self-reported rereading, not providing any support for deliberate rereading. Our findings indicate that successful online monitoring relies on strong word knowledge leading to efficient processing of texts, both for bilingual and monolingual children, and beyond deliberate rereading.

Comprehension monitoring during reading: an eye-tracking study with children learning English as an additional language

Learning to read is a complex matter, drawing on a range of cognitive and linguistic operations (Castles, Rastle, & Nation, 2018). It is perhaps not surprising that it can pose particular challenges for children who are learning English as an additional language (EAL) while using a minority language at home (Spencer & Wagner, 2017). Since EAL children's English proficiency is highly predictive of their educational success and personal well-being (Strand & Demie, 2005; Strand & Hessel, 2018; Whiteside, Gooch, & Norbury, 2017), it is important to understand these challenges. EAL learners often lag behind monolinguals in reading comprehension and this is linked to relative deficits in vocabulary size and depth, but not reading fluency (Bowyer-Crane, Fricke, Schaefer, Lervåg, & Hulme, 2017; D'Angiulli, Siegel, & Serra, 2001; Hessel & Murphy, 2019; Lesaux, Rupp, & Siegel, 2007; Spätgens & Schoonen, 2018; Spencer & Wagner, 2017; Verhoeven, 2000). There is some evidence that EAL learners are less adept at comprehension monitoring, that is, with checking and regulating their understanding when reading figurative and multi-word phrases (Burgoyne, Whiteley, & Hutchinson, 2013; Kan & Murphy, *in press*). In this paper, we took a novel approach to investigating comprehension monitoring in EAL learners (and a monolingual comparison group) by measuring children's eye movements as they read sentences that contained semantic inconsistencies. We asked whether there are systematic differences in reading that might reflect underlying differences in comprehension monitoring associated with being an EAL learner.

Comprehension monitoring is at work when a person notices and reacts to mismatches between (possibly erroneous) background knowledge and textual information as they read (van den Broek & Kendeou, 2008), or between their interpretation of earlier and later text segments (Carpenter & Daneman, 1981; Christianson et al., 2017; von der Malsburg & Vasishth, 2012). Comprehension monitoring has traditionally been studied *offline*, by asking people to reflect explicitly on their comprehension. A common approach is to insert mismatched information in a text – for example a syntactic mismatch between pronoun and verb, or a semantic mismatch between two pieces of information in the story – and ask participants to underline or verbalize the mismatch (Baker, 1984; Kim & Phillips, 2014; Meyers, Lytle, Palladino, Devenpeck, & Green, 1990; Oakhill, Hartt, & Samols, 2005). Studies like these have demonstrated a close link between reading comprehension ability and comprehension monitoring performance (Kim & Phillips, 2014): older children (and skilled comprehenders) show better comprehension monitoring than younger children (and less-skilled comprehenders) (Baker, 1984; Oakhill et al., 2005; Zabrocky & Ratner, 1989). Similarly, EAL learners are more likely to overestimate their comprehension on high-frequency collocations such as *by and large* than monolinguals (Kan & Murphy, *in press*), and less likely to notice mismatches between their interpretation of figurative phrases and the story context (Burgoyne et al., 2013). One limitation of this evidence base is that any task that estimates comprehension monitoring using a secondary task comes with additional metacognitive or linguistic demands. It is not clear whether such tasks capture comprehension monitoring as it happens during the course of reading itself.

One methodology that does capture reading processes more directly is to measure people's eye movements as they read text. Eye movements provide *online* data that inform our understanding of the time course of reading and from this we can make inferences about the underlying processes that support reading. If children detect and attempt to correct anomalous text, this should be reflected in the eye movement record and as such, experiments that measure eye movements have been used to index children's comprehension monitoring (Connor et al., 2015; Kim, Vorstius, & Radach, 2018; Kinnunen & Vauras, 1995; van der Schoot, Reijntjes, & van Lieshout, 2012; Vorstius, Radach, Mayer, & Lonigan, 2013). For example, Connor et al. (2015) asked 10–12-year-old students to read short two-sentence stories where the first sentence introduced a scenario (such as an animal that was *barking*) and a second sentence continued in a way that was either consistent with the first sentence (*puppy*), or inconsistent (*kitten*). Of interest was whether children noticed the inconsistency – that is, did they spend longer reading the target word (*puppy/kitten*) in the inconsistent condition? Connor et al. found that all children noticed the mismatch initially, as shown by increased gaze duration on the target word in the inconsistent condition. However, only those children with stronger academic language (a composite of vocabulary, story recall, and academic knowledge) showed longer rereading times for mismatch words, consistent with re-analysis. This interpretation of longer rereading times is consistent with skilled readers rereading more when texts contain difficult content (Rayner, Chace, Slattery, & Ashby, 2006) or anomalous words or structures (Connor et al., 2015; Joseph et al., 2008; Rayner, Warren, Juhasz, & Liversedge, 2004; von der Malsburg & Vasishth, 2011; Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, & Staub, 2017; Zabrocky & Ratner, 1989). Complementary evidence comes from the finding that

comprehension suffers when rereading is made impossible due to masked text (Schotter, Tran, & Rayner, 2014). Thus, rereading time is an important eye movement measure for the study of comprehension monitoring.

The interpretation of increased rereading times is not without ambiguity. It may reflect confusion rather than targeted attention (Murray & Kennedy, 1988); it might also reflect pauses to complete lexical processing, rather than text integration (Mitchell, Shen, Green, & Hodgson, 2008). To help distinguish rereading that is indicative of re-analysis, it is useful to consider spatial measures of rereading alongside rereading time (Murray & Kennedy, 1988). We chose to use regression-in probability as an indicator of targeted re-analysis. This metric captures regressions or returning eye movements directly into the target word. This seems to reflect sensitivity to comprehension monitoring: children who are able to name the inconsistencies they reread tend to show increased regression-in probability (Eilers, Tiffin-Richards, & Schroeder, 2018).

While eye movements have the potential to reveal comprehension monitoring strategies as they happen during the course of reading, it is not clear how this relates to the purposeful and goal-directed strategies captured by offline tasks such as inconsistency detection and think-alouds. Rereading to resolve incoherent information can arguably reflect a whole continuum from automatic to deliberate processes, and the degree of deliberateness may vary according to a child's reading ability, and the reading situation (van den Broek & Helder, 2017). To distinguish deliberate from automatic reading processes, one can ask whether a reader is consciously aware of their strategies (Afflerbach, Pearson, & Paris, 2008). Awareness can be measured by asking participants to verbalize coherence breaks (Meyers et al., 1990; Oakhill et al., 2005) or to think-aloud (Kaakinen, Lehtola, & Paattilammi, 2015) as they read. Unfortunately,

however, such measures do not combine well with eye movement recording. Speaking while reading reduces the quality and accuracy of eye movement data (Kaakinen et al., 2015) and the presence of secondary tasks may alter the naturalness of the reading process (Kaakinen & Hyönä, 2010; Weiss et al., 2017; Yeari, van den Broek, & Oudega, 2015). However, people are able to self-report on their own reading behavior and there is evidence that this relates to eye movement behavior while reading. For example, Hyönä and Nurminen (2006) first asked participants to read a text and then to self-report on their attention to global text features, such as headings vs. main body of a text. They found that people were able to describe their rereading behavior in a way that was consistent with what was seen in the eye movement record, indicating that they were aware of their reading style.

In our experiment, we adapted Connor et al.'s (2015) eye-movement paradigm to examine children's processing of inconsistent information while reading, comparing EAL learners with their monolingual peers. If being an EAL learner is associated with problems with comprehension monitoring, differences in online reading behavior should be apparent. We also tested whether individual differences in vocabulary and reading fluency are associated with differences in online reading behavior. Following Connor et al., we hypothesized that efficient regressions to the target word indicate adaptive reading and as such, should be associated with stronger language ability. Our experiment followed a 2 X 2 design where we compared language group (EAL vs. monolingual, between-subjects) by inconsistency (consistent vs. inconsistent, within-subject) alongside individual differences in vocabulary size and reading fluency. Finally, following Hyönä and Nurminen (2006), we asked the children to reflect and self-report on their reading strategies and tested whether there was an association

between self-report and online reading behavior.

Method

Participants

Sixty-four 9–10-year-olds were recruited from Year 5 classes in three UK primary schools. We secured consent from their legal guardian and assent from the children themselves. Children were classified as EAL or monolingual based on a language background questionnaire (see Appendix A). This asked children whether they interacted with any other language at home or in their community, and if so, how often. If they reported a home language which they had been exposed to in conversations with parents, siblings or friends or in their media use (either in the past or currently), they were classified as EAL (N= 24). Table 1 summarizes the language background of these children. Two EAL children had moved to the UK at age 2 and 8 years respectively and both began learning English by 3 years of age. The other EAL children were born in the UK and had started learning English from birth or, in the case of four children, upon entry to nursery at age 3–4 years. The children varied in their current English language use at home, from one child speaking no English to English being used most of the time by others. They also varied in their media use (reading or watching TV) and conversations with friends, from using English in many to only a few different contexts. Many different languages were spoken at home: six children spoke Arabic, and two spoke Kiswahili. One of the Arabic speakers also spoke French. The other children each spoke one of the following: Italian, Japanese, Malay, Malayalam, Filipino, Serbian, Shona, German, Urdu, Welsh, Zulu, and Hawaiian pidgin. The large variation in amount of English language exposure and home language backgrounds is common in

EAL research (Burgoyne, Kelly, Whiteley, & Spooner, 2009; Hutchinson, Whiteley, Smith, & Connors, 2003; Authors et al., year) and a true reflection of the diversity of the EAL population in the UK (Authors et al., year; NALDIC, 2012). At the same time, this makes our sample extremely varied in their first language abilities and backgrounds which necessarily adds variation to our data (de Cat & Serratrice, 2018).

Given the nature of our experiment, it was important that all children had sufficient word reading ability to permit engagement with the task. We therefore excluded any child (EAL and monolingual) whose word reading ability fell more than one standard deviation below normal range, as assessed by the Test of Word Reading Efficiency 2 (TOWRE-2; Torgesen, Rashotte, & Wagner, 2012). One EAL child was excluded on this basis. The final sample comprised 63 children (EAL N=23, 11 girls; monolinguals N=40, 22 girls). All children had normal or corrected-to-normal vision. EAL and monolingual children were peers: they all lived in the same school catchment area and attended the same classes. Using Welch's *t*-tests to correct for unequal variance across groups (Vasishth, 2014), we found that the two groups did not differ on age or non-verbal reasoning (as measured through the WASI II Matrices subtest), vocabulary size (as measured by the British Picture Vocabulary Scale 2 (BPVS-2, described below) or reading fluency (see Table 2) with both groups performing within normal range.

Materials and Procedure

Comprehension monitoring task

Twenty two-sentence stories were adapted from Connor et al. (2015). An example is shown in Table 3 (Appendix B contains a full list of stimuli). Words specific to

American English were replaced by British English alternatives (e.g., *nail* for *bolt*). The first sentence of each story describes a simple event or action, such as an animal barking. The second sentence continues this action in a way that was either consistent or inconsistent with the first sentence, by inserting a consistent (e.g. *puppy*) or inconsistent target word (e.g. *kitten*).

The target words, which were also the critical region on which eye movements would be compared, were equivalent across inconsistent and consistent conditions on key lexical properties (see Table 4). The stories were similar in length and in placement of the target word and were matched on the number of words and characters preceding and following a target (13 words, or 69-77 letters before target and 6 words, or 27-42 letters afterwards, respectively). The adapted texts were checked for acceptability by two adult speakers of British English.

Stories were divided into two lists following a Latin square design where each child read each of the 20 stories once, 10 containing a consistent target and 10 with an inconsistent target word. To encourage reading for meaning, they were told that some stories would be followed by a comprehension question. These took the form of a forced-choice question (not tapping the target word) and occurred on one third of trials.

The stories were presented via Experiment Builder software (SR Research, 2009b) and eye movements were recorded using an Eyelink 1000 eye tracker with a sampling rate of 1000 Hz. Sentences were presented in black, Courier New font, size 18, with double spacing, and on a grey background. Children read the sentences silently from a 15” Dell monitor, set at a refresh rate of 60 Hz with 1024 x 768 resolution, at a viewing distance of 60 cm. Although the children read with both eyes, only the right eye was monitored. A chin and forehead rest was used to minimize head movements. The

camera was calibrated on a five-point grid and each trial started with drift corrections to make sure calibration was sufficiently precise ($< 0.5^\circ$ of accuracy). Calibration was repeated as necessary.

Self-perceived rereading

To assess children's self-perceived rereading behavior, they were asked to rate four statements using a 3-point scale ("Not like me", "A little like me", or "A lot like me"; they could also answer "Don't know"). The four statements were modelled on a previous study (Hyönä & Nurminen, 2006) but adapted for children by reducing the number of questions from 8 to 4 and reducing the number of scale steps from 6 to 3; the phrasing of the statements was also simplified. After a pilot interview with a 7-year-old, the final four statements were: *I reread words when I don't understand them*; *I reread each sentence before moving to the next*; *I look back at earlier sentences when I need to*; and *I go back to words and sentences that I have found difficult to understand when I first read them*.

Assessment of reading and vocabulary

Vocabulary size was measured using the British Picture Vocabulary Scale, Version 2 (BPVS; Dunn, Dunn, Whetton, & Burley, 1997). This is similar to the Peabody Picture Vocabulary Test (PPVT; Dunn, 1959): children saw four pictures on a page and were asked to choose the picture that fits the word read out to them. *Reading fluency* was tested using the TOWRE-2 where children read aloud lists of words or nonwords, and scores are derived from the number of items read correctly in 45 seconds.

Procedure

The equipment was set up in a quiet room in school. Each child was seen individually twice, separated by 1-3 school days. In the first session, children completed the language background questionnaire and the TOWRE-2 (and a different experiment not reported here). They completed the comprehension monitoring experiment in the second session, followed by self-report, the BPVS, and the WASI. Each session lasted approximately 30 – 40 minutes. The procedure was approved by the Central University Research Ethics Community of the Oxford University Department of Education as conforming to the British Educational Research Association Ethical Guidelines for Educational Research.

Results

Analysis

The eye movement data were cleaned and analyzed using the SR Research Data Viewer software (SR Research, 2009a). Extremely short or long fixations ($< 80\text{ms}$ and $> 1200\text{ms}$) were deleted or merged with adjacent fixations (short fixations at $< 1^\circ$ distance) using the Clean Data function. This removed 3.1% of the data. Gaze drift (fixations recorded outside of interest areas due to eye-tracking inaccuracies) was corrected manually by moving fixations upwards or downwards within a range of 3 pixels. Eye movements on the critical region (which was the target word, e.g. *puppy* or *kitten*) were distilled into early measures (thought to tap word identification) and late measures (thought to reflect text-integration processes, see Table 5).

The data were analyzed in R (R Core Team, 2016), using linear and binomial mixed effect models from the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). Model means of reading times were gained and evaluated in post-hoc tests using cell

means models and single degree of freedom contrasts and implemented via the *glht* function of the *multcomp* package (Hothorn, Bretz, & Westfall, 2008). Reading times were log-transformed since this proved the best transformation to achieve normality, as confirmed through the BoxCox power transformation (Keshtiari & Vasissth, 2013). Significance in the mixed effect models was checked using the Kenward-Rogers approximation implemented using the *pbkrtest* package (Halekoh & Højsgaard, 2009). The mixed effect models included the fixed effects inconsistency (consistent vs. inconsistent), language group (EAL vs. monolingual), reading fluency and vocabulary size (both continuous variables). Crossed random intercepts and slopes by inconsistency and language group were included for participants and items. This random structure was only reduced when high inter-correlation and singularity of random effects caused convergence problems. Predictors were entered with effect coding for the categorical predictors inconsistency and language group (.5 and -.5) and as z-scores for the continuous measures reading fluency and vocabulary size. Collinearity checks confirmed that the continuous measures were not correlated too highly, allowing them to be entered together ($r < 0.5$) (Zuur, Ieno, & Elphick, 2010).

Inconsistency effects by language group, reading fluency, and vocabulary size

We begin by reporting children's reading of consistent and inconsistent target words. Rereading and total times were longer for target words read in the inconsistent condition and they received more rereading and regressions in, as summarized in Figure 1. This general pattern was confirmed by the mixed effects models which showed a main effect of consistency, as summarized in Table 6. There was however no main effect of consistency for gaze duration.

Turning to differences between the EAL learners and monolingual children, there was no main effect of language group on any measure, confirming that the two groups of children did not differ in their online reading of target words.

The absence of an overall group difference between EAL and monolingual learners might reflect large heterogeneity within the EAL group – as discussed earlier, it is common in studies of EAL learners to see considerable variability within the group in terms of the amount of exposure to English. To explore this, we ran additional analyses using a continuous measure of English language use at home, rather than a binary EAL vs. monolingual comparison, summarized in Table 7. For the EAL children, this was the composite score derived from the language background questionnaire. The monolingual children were coded with the highest score, given by definition these children always spoke English at home. These alternative models provided a good fit to the data, compared to the binary models described above (using the *anova* function, model fit parameters were equivalent to or better, depending on the dependent eye-tracking measure). However, these models showed again no main effect of bilingualism when operationalized continuously rather than as a binary variable.

Turning to the continuous predictors, the main effects showed that higher reading fluency was associated with shorter gaze durations as well as rereading and total times, and larger vocabulary size was associated with shorter gaze durations, but not any other reading time measure (see Tables 8-9). Vocabulary was associated with rereading in the inconsistent condition, as evident in the interaction between inconsistency X vocabulary size. We explored this interaction by evaluating predicted reading times on consistent or inconsistent targets at 1SD above or below the sample mean of vocabulary size. These point estimates (in raw BPVS scores, these were 92.8

for -1SD and 119.6 for +1SD; or 95 and 120 in standard scores) were well within the vocabulary range of our sample, both for EAL (range of 78 – 128 in raw BPVS scores) and monolingual children (range of 80 – 135 in raw BVPS scores). This approach allowed us to utilize variation in vocabulary as a continuous variable without having to assign participants to extreme groups on the basis of vocabulary (Halekoh & Højsgaard, 2009) while at the same time still being able to evaluate reading times at different levels of vocabulary. Hereafter, we will refer to these model-based means as the reading times of children with larger (+1SD) or smaller vocabularies (-1SD).

Post-hoc comparisons showed that as predicted (following Connor et al., 2015), children with larger vocabularies were more likely to directly regress to inconsistent targets, and to reread them for longer. These key interactions are shown in Figure 2. First, children with larger vocabularies took less time to reread consistent words compared to those with smaller vocabularies ($z = -3.4, p < .001$), and second, they were more likely to regress directly into implausible words ($z = 2.3, p < .05$). Thus, children with large vocabularies differed from their peers in two ways: first, they read consistent words faster and second, they were more efficient in their rereading of inconsistent words than children with smaller vocabularies.

The models in Table 6 also showed an unexpected three-way interaction between language group, vocabulary size, and consistency on children's gaze duration on target words. This interaction is shown in Figure 3. Post-hoc tests revealed that the interaction was due to EAL children with smaller vocabularies showing longer reading of inconsistent words in gaze duration, relative to consistent words ($z = -4.1, p < .0001$). The same interaction was observed in the alternative models (where bilingualism was operationalized continuously using amount of English language input), both in gaze

duration and in the summed score of total time (see Table 7). This differs from Connor et al.'s finding of a main effect of inconsistency in gaze duration, but no interaction with individual differences in language skills.

Relationship between eye movements and self-reported rereading

We now turn to the relationship between children's eye movements while reading and self-reported rereading. Children's answers to the four questions about their rereading were summarized into a composite score ranging from 4 – 12. Composite scores were created by adding up children's answers to all questions, with the one answer describing non-selective rereading ("I reread each sentence before moving to the next") being reversed before being added. The mean score was 5.6 (SD= 1.4), with higher scores representing more selective rereading. Data was missing from four children who responded "don't know" to one of the questions. Self-reported selective rereading (hereafter: self-report) did not correlate with reading fluency, vocabulary or nonverbal reasoning; all r 's < .01). We ran binomial mixed effect models predicting rereading and regression-in probability on target words with self-report as continuous predictor, alongside language group and target word consistency (N=59). The models are summarized in Table 10.

Children with higher self-report scores reread target words more often and made more direct regressions into them, as shown in Figure 4. Thus, there was an association between children's general rereading and self-report. However, self-report did not interact with whether the target was in the consistent or inconsistent condition, thus giving no indication that the degree of self-reported rereading while reading was specifically related to efficient rereading of inconsistent information during reading itself.

Discussion

This experiment investigated online comprehension monitoring in EAL children and monolingual peers, focusing on their reading patterns on inconsistent words. We also examined the association between online markers of comprehension monitoring and individual differences in reading and vocabulary, and with children's own perceptions of self-reported rereading behavior.

Overall, our paradigm was sensitive to inconsistency. As predicted, children reread inconsistent words more often, and for longer. On the basis of previous research (Burgoyne et al., 2013; Kan & Murphy, *in press*), we had predicted differences in comprehension monitoring between EAL children and their monolingual peers. However, our data are clear in showing no overall group difference. In line with other studies, there was overlapping variation in reading and language between the two groups (de Cat & Serratrice, 2018; Marinova-Todd, Marshall, & Snow, 2000; van den Bosch, Segers, & Verhoeven, 2018); and, variation in reading and language has been shown to capture differences in online reading (Hopp, 2016; van den Bosch, Segers, & Verhoeven, 2018a). This may be particularly true in a sample like ours where the amount and nature of exposure to English within the EAL group varied. Note, however, that a continuous measure of English language exposure also failed to reveal any overall differences in online differences in comprehension monitoring. This further supports the notion that any differences in online reading between bilingual and monolingual samples are better understood as differences in vocabulary knowledge rather than a reflection of categorical differences between first- and second-language speakers (Hopp, 2016). Consistent with this, we found that individual differences in vocabulary size and word reading fluency (as measured by TOWRE-II) predicted differences in

children's online reading: both were associated with generally faster reading in the experiment. Vocabulary also interacted with children's reading of words in the inconsistent condition: children with larger vocabulary size reread inconsistent words more often and for longer than target words in the consistent condition. This interaction mirrors Connor et al.'s (2015) findings and other studies that have reported associations between individual differences in reading abilities and inconsistency effects in children (Eilers et al., 2018; Zabrocky & Ratner, 1989).

We also observed that the shorter rereading times of consistent words went hand in hand with efficient rereading of inconsistent words for children with larger vocabularies, regardless of EAL status. In line with the Reading Systems Framework (Stafura & Perfetti, 2017), this reading pattern might be due to a cascade of effects from strong word knowledge to fluent reading processing which ultimately enable a person to be more selective with their attention. A range of factors is likely to underpin this. First, retrieval of word meanings supports the building of a full mental representation of the text from which to then notice inconsistencies (Yang, Perfetti, & Schmalhofer, 2005). Second, efficient processing in word identification and meaning retrieval frees up processing resources for higher-order processes, such as deciding to reread (Perfetti, 1985; von der Malsburg & Vasishth, 2012). And third, efficient regressions require memory for the location of relevant words, a process that may be easier when general reading runs smoothly (Murray & Kennedy, 1988). Together, these processes can explain how differences in word knowledge could result in targeted re-analysis of inconsistent information. The fact that these reading differences were driven by vocabulary size and not word reading fluency (as measured by the TOWRE-II)

indicates that the ease and quality of meaning retrieval demands word knowledge in addition to fluent word reading.

Although our results show no differences between EAL learners and monolinguals in their comprehension monitoring on short and simple texts, we cannot claim that this is generalizable to the reading of all texts, particularly those containing more challenging vocabulary. Previous offline studies have reported that EAL learners encounter particular difficulties monitoring their comprehension on figurative phrases (Burgoyne et al., 2013; Kan & Murphy, *in press*). It could be that that EAL learners did not differ from monolingual readers in our reading task because the texts contained only simple vocabulary (e.g. *kitten* and *puppy*) and that the groups would have differed when reading texts containing more challenging vocabulary, such as collocations, idioms or metaphors (Hessel & Murphy, 2019; Kan & Murphy, *in press*; McKendry, 2014; Smith & Murphy, 2015). Efficiently processing these phrases would require depth of vocabulary, an area where EAL learners and minority language learners more generally have been found to lag behind monolingual peers (Cremer, Dingshoff, de Beer, & Schoonen, 2011; Cremer & Schoonen, 2013; Hessel & Murphy, 2019; Kan & Murphy, *in press*; McKendry, 2014; Schoonen & Verhallen, 2008; Smith & Murphy, 2015; Spätgens & Schoonen, 2018). Texts containing multi-word and figurative phrases may thus expose processing differences in the online monitoring of EAL and monolingual learners, beyond what we observed in the current study.

Turning to children with smaller vocabularies, it was only children in the EAL group who showed longer gaze durations when reading inconsistent words. Connor et al. (2015) reported a main effect of longer gaze durations which they interpreted as indicating an initial detection of inconsistency that then allows later re-analysis. This

two-step interpretation does not fit our data. Those children who targeted their rereading to inconsistent words (i.e., children with larger vocabularies) were not the same children who showed longer gaze durations on the inconsistent words (i.e., EAL children with smaller vocabularies). Therefore, the longer gaze durations cannot be interpreted as the necessary first step that then allows efficient rereading. Instead, we suggest that these longer gaze durations indicate relative weaknesses with lexical access. Lexical access is slower in readers with smaller word knowledge, which may particularly be the case for lower-vocabulary EAL but not monolingual children as EAL children tend to have gaps in vocabulary depth that go beyond size, as discussed above, which might particularly affect word processing in text (Perfetti, 2007). Possibly, lower-vocabulary EAL children's slower lexical access shows only on the inconsistent words in our experiment since these words were by design less predictable. Predictability plays a stronger role in less skilled comprehenders who seem to rely on context to support lexical access (Ashby, Rayner, & Clifton, 2005). Potentially therefore, the EAL children with smaller vocabularies may have relied more strongly on context to support lexical access than other children; this might be an adaptive strategy in general, but it would fail when a word does not fit the context.

We also took the opportunity to measure the children's self-reported selective rereading behavior and ask whether this was associated with the inconsistency effect, as revealed by the eye movement record. Children who self-reported to reread more selectively were also more likely to reread and regress to target words. However, this was a general effect and not one specific to inconsistent words. Although adults are aware of the degree of selective rereading of global text features (Hyönä & Nurminen, 2006), re-analyzing internal inconsistencies while reading is more automatic than

attention to global text features (van den Broek & Helder, 2017). This is in line with the suggestion that efficient rereading of inconsistent target words was driven by a bottom-up cascade of effects beginning with efficient word processing (Perfetti & Hart, 2002). Another candidate explanation is that children may be less aware of their reading processing than adults, or that our questionnaire was not well-suited for its purpose. For example, an explicit anchoring of our questions in the present reading task might have made it easier for children to describe their rereading (for example by changing “I reread words when I don’t understand them” to “When reading the stories on the screen, I reread the words that I didn’t understand”). Further research is needed to make connections between people’s overt comprehension monitoring (as tapped by self-reports and verbalisations) and online reading.

Conclusion

Larger vocabulary size was associated with fluent and efficient rereading of inconsistent information, a sign of more developed comprehension monitoring. This relationship was the same for EAL and monolingual children, highlighting the importance of vocabulary knowledge for successful comprehension monitoring, regardless of language group and reading fluency. While pinpointing the source of individual differences is never trivial given the many inter-correlations between candidate variables, our data show that vocabulary size matters more for children’s comprehension monitoring than word reading fluency and bilingualism (amount of English language input), two other main candidate explanations for skillful reading. EAL children with smaller vocabularies, however, showed longer gaze durations on inconsistent words, possibly indicating slower lexical access in the inconsistent condition. Finally, self-reported rereading was associated with generally increased rereading in the eye movement

record, but not efficient rereading of inconsistent words. This suggests that there remains a gap between the online comprehension monitoring tapped through eye movements and the overt comprehension monitoring tapped in previous offline studies (Baker, 1984; Kim & Phillips, 2014; Meyers et al., 1990; Oakhill et al., 2005). To bridge this gap, more research is needed that combines measures of online processing with offline self-report and comprehension measures, and that determines where the two coincide or differ. Nevertheless, our findings show the utility of eye movements to understand how differences in online comprehension monitoring arise during reading, and how they are associated with children's language skills and backgrounds.

Funding

This work was supported by the Scatcherd European Scholarship and by the Economic and Social Research Council under Grant 1637966 while the first author was a doctoral student at the University of Oxford.

Acknowledgements

We are grateful to Jane Ashby and one anonymous reviewer for their valuable comments.

Declaration of Conflicting Interests

The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

References

- Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364–373.
<https://doi.org/10.1598/RT.61.5.1>
- Ashby, J., Rayner, K., & Clifton, C. J. (2005). Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Human Experimental Psychology*, 58(6), 1065–1086. <https://doi.org/10.1080/02724980443000476>
- Baker, L. (1984). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology*, 76(4), 588–597.
<https://doi.org/10.1037/0022-0663.76.4.588>
- Bowyer-Crane, C., Fricke, S., Schaefer, B., Lervåg, A., & Hulme, C. (2017). Early literacy and comprehension skills in children learning English as an additional language and monolingual children with language weaknesses. *Reading and Writing*, 30(4), 771–790.
<https://doi.org/10.1007/s11145-016-9699-8>
- Burgoyne, K., Kelly, J. M., Whiteley, H. E., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *The British Journal of Educational Psychology*, 79(Pt 4), 735–747. <https://doi.org/10.1348/000709909X422530>
- Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2013). The role of background knowledge in text comprehension for children learning English as an additional language. *Journal of Research in Reading*, 36(2), 132–148. <https://doi.org/10.1111/j.1467-9817.2011.01493.x>
- Carpenter, P. A., & Daneman, M. (1981). Lexical retrieval and error recovery in reading: a model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior*, 20(2), 137–160. [https://doi.org/10.1016/S0022-5371\(81\)90357-1](https://doi.org/10.1016/S0022-5371(81)90357-1)
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51.
<https://doi.org/10.1177/1529100618772271>
- Christianson, K., Luke, S. G., Hussey, E. K., Wochna, K. L., Christianson, K., Luke, S. G., ... Wochna, K. L. (2017). Why reread? Evidence from garden-path and local coherence structures. *The Quarterly Journal of Experimental Psychology*, 70(7), 1380–1405.
<https://doi.org/10.1080/17470218.2016.1186200>
- Connor, C. M., Radach, R., Vorstius, C., Day, S. L., Mclean, L., Morrison, F. J., & McDonald, C. (2015). Individual differences in fifth graders' literacy and academic language predict

- comprehension monitoring development: an eye-movement study. *Scientific Studies of Reading*, 19(2), 114–134. <https://doi.org/10.1080/10888438.2014.943905>
- Cremer, M., Dingshoff, D., de Beer, M., & Schoonen, R. (2011). Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *International Journal of Bilingualism*, 15(2), 187–204. <https://doi.org/10.1177/1367006910381189>
- Cremer, M., & Schoonen, R. (2013). The role of accessibility of semantic word knowledge in monolingual and bilingual fifth-grade reading. *Applied Psycholinguistics*, 34(6), 1195–1217. <https://doi.org/10.1017/S0142716412000203>
- D'Angiulli, A., Siegel, L. S., & Serra, E. (2001). The development of reading in English and Italian in bilingual children. *Applied Psycholinguistics*, 22(4), 479–507. <https://doi.org/10.1017/S0142716401004015>
- de Cat, C., & Serratrice, L. (n.d.). Predicting language proficiency in bilingual children. *Under Review*.
- Dunn, L. M. (1959). *Peabody picture vocabulary test*. Minneapolis: American Guidance Service.
- Dunn, L. M., Dunn, D. M., Whetton, C., & Burley, J. (1997). *The British picture vocabulary scale II. BPVS II*. Windsor: NFER-Nelson.
- Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2018). Individual differences in children's pronoun processing during reading: detection of incongruence is associated with higher reading fluency and more regressions. *Journal of Experimental Child Psychology*, 173, 250–267. <https://doi.org/S0022096517304770>
- Halekoh, U., & Højsgaard, S. (2009). A Kenward-Roger approximation and parametric bootstrap Mmethods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software*, 10(2), 1–24. <https://doi.org/10.18637/jss.v069.i12>
- Hessel, A. K., & Murphy, V. A. (2019). Understanding how time flies and what it means to be on cloud nine: English as an Additional Language (EAL) learners' metaphor comprehension. *Journal of Child Language*, 46(2), 265–291. <https://doi.org/10.1017/S0305000918000399>
- Hopp, H. (2016). Individual differences in L2 parsing and lexical representations. *Bilingualism: Language and Cognition*, 1–2. <https://doi.org/10.1017/S1366728916000821>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric

- models. *Biometrical Journal*, 50(3), 346–363.
- Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading*, 26(1), 141–423. <https://doi.org/10.1111/1467-9817.261003>
- Hyönä, J., & Nurminen, A.-M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, 97, 31–50. <https://doi.org/10.1348/000712605X53678>
- Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 61(5), 708–723. <https://doi.org/10.1080/17470210701400657>
- Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 36(6), 1561–1566. <https://doi.org/10.1037/a0020693>
- Kaakinen, J. K., Lehtola, A., & Paattilampi, S. (2015). The influence of a reading task on children's eye movements during reading. *Journal of Cognitive Psychology*, 27(5), 640–656. <https://doi.org/10.1080/20445911.2015.1005623>
- Kan, R., & Murphy, V. A. (*in press*). Effects of frequency and idiomaticity on second language reading comprehension in children with English as an additional language. *Studies in Second Language Learning and Teaching*.
- Keshtiari, N., & Vasisht, S. (2013). Reactivation of antecedents by overt versus null pronouns: evidence from Persian. *Journal of Language Modelling*, 1(2), 243–266.
- Kim, Y. S., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly*, 49(3), 269–281. <https://doi.org/10.1002/rrq.74>
- Kim, Y. S., Vorstius, C., & Radach, R. (2018). Does online comprehension monitoring make a unique contribution to reading comprehension in beginning readers? Evidence from eye movements. *Scientific Studies of Reading*, 22(5), 367–383. <https://doi.org/10.1080/10888438.2018.1457680>
- Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction*, 5(2), 143–165. [https://doi.org/10.1016/0959-4752\(95\)00009-R](https://doi.org/10.1016/0959-4752(95)00009-R)
- Lesaux, N. K., Rupp, A. a., & Siegel, L. S. (2007). Growth in reading skills of children from

- diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology*, 99(4), 821–834. <https://doi.org/10.1037/0022-0663.99.4.821>
- Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, 34(1), 9. <https://doi.org/10.2307/3588095>
- McKendry, M. (2014). *Investigating the relationship between reading comprehension and semantic skill in children with English as an Additional Language: A focus on idiom comprehension*. University of Oxford.
- 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. <https://doi.org/10.1080/09500790.2010.531702>
- Meyers, J., Lytle, S., Palladino, D., Devenpeck, G., & Green, M. (1990). Think-aloud protocol analysis: an investigation of reading comprehension strategies in fourth-and fifth-grade students. *Journal of Psychoeducational Assessment*, 8(2), 112–127. <https://doi.org/10.1177/073428299000800201>
- Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: a reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, 59(3), 266–293. <https://doi.org/10.1016/j.jml.2008.06.002>
- Murphy, V. A. (2018). Literacy development in linguistically diverse pupils. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language: The state of the science across its subfields*. (Studies in, Vol. 54, pp. 315–323). Amsterdam: John Benjamins. <https://doi.org/10.1075/btl.128.02gar>
- Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: evidence from eye movement analyses. *The Quarterly Journal of Experimental Psychology*, 40(4), 693–718. <https://doi.org/10.1080/14640748808402294>
- National Association of Language Development in the Curriculum (NALDIC). (2012). Languages in school. Retrieved July 14, 2017, from <http://www.naldic.org.uk/research-and-information/eal-statistics/lang>
- Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, 18(7–9), 657–686. <https://doi.org/10.1007/s11145-005-3355-z>

- Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy*. Amsterdam: John Benjamins.
- R Core Team. (2016). R: a language and environment for statistical computing. Retrieved from <http://www.r-project.org/>
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10(3). <https://doi.org/10.1207/s1532799xssr1003>
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1290–1301. <https://doi.org/10.1037/0278-7393.30.6.1290>
- Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, 25(2), 211–236. <https://doi.org/10.1177/0265532207086782>
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): comprehension is supported by regressions during reading. *Psychological Science*, 25(6), 1218–1226. <https://doi.org/10.1177/0956797614531148>
- Smith, S. A., & Murphy, V. A. (2015). Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, 28(3), 347–369. <https://doi.org/10.1007/s11145-014-9527-y>
- Spätgens, T., & Schoonen, R. (2018). The semantic network, lexical access, and reading comprehension in monolingual and bilingual children: an individual differences study. *Applied Psycholinguistics*, 39(1), 225–256.
- Spencer, M., & Wagner, R. K. (2017). The comprehension problems for second-language learners with poor reading comprehension despite adequate decoding: a meta-analysis. *Journal of Research in Reading*, 40(2), 199–217. <https://doi.org/10.1111/1467-9817.12080>
- SR Research. (2009a). Data Viewer.
- SR Research. (2009b). Experiment Builder.
- Stafura, J. Z., & Perfetti, C. A. (2017). Integrating word processing with text comprehension. In

- Theories of reading development* (pp. 9–32). <https://doi.org/10.1075/swll.15.02sta>
- Strand, S., & Demie, F. (2005). English language acquisition and educational attainment at the end of primary school. *Educational Studies*, 31(3), 275–291.
<https://doi.org/10.1080/03055690500236613>
- Strand, S., & Hessel, A. K. (2018). *English as an additional language (EAL), proficiency in English and pupil's educational achievement: an analysis of local authority data*.
- Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *TOWRE 2: test of word reading efficiency*. Austin, Texas : Pro-Ed.
- van den Bosch, L. J., Segers, E., & Verhoeven, L. (2018a). Online processing of causal relations in beginning first and second language readers. *Learning and Individual Differences*, 61, 59–67. <https://doi.org/10.1016/j.lindif.2017.11.007>
- van den Bosch, L. J., Segers, E., & Verhoeven, L. (2018b). The role of linguistic diversity in the prediction of early reading comprehension: a quantile regression approach. *Scientific Studies of Reading*, 00(00), 1–17. <https://doi.org/10.1080/10888438.2018.1509864>
- van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: passive processes, reader-Initiated processes, and evolving mental representations. *Discourse Processes*, 54(5–6), 360–372. <https://doi.org/10.1080/0163853X.2017.1306677>
- van den Broek, P., & Kendeou, P. (2008). Cognitive Processes in Comprehension of Science Texts: The Role of Co-Activation in Confronting Misconceptions. *Applied Cognitive Psychology*, 22, 335–351. <https://doi.org/10.1002/acp>
- van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, 25(7), 1665–1690.
<https://doi.org/10.1007/s11145-011-9337-4>
- Vasishth, S. (2014). *An introduction to statistical data analysis (Summer 2014) Lecture notes*.
- Verhoeven, L. (2000). Components in Early Second Language Reading and Spelling. *Scientific Studies Of Reading*, 4(4), 313–330. <https://doi.org/10.1207/S1532799XSSR0404>
- von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, 65(2), 109–127.
<https://doi.org/10.1016/j.jml.2011.02.004>
- von der Malsburg, T., & Vasishth, S. (2012). Scanpaths reveal syntactic underspecification and

- reanalysis strategies. *Language and Cognitive Processes*, 28(10), 1–34.
<https://doi.org/10.1080/01690965.2012.728232>
- Vorstius, C., Radach, R., Mayer, M. B., & Lonigan, C. J. (2013). Monitoring local comprehension monitoring in sentence reading. *School Psychology Review*, 42(2), 191–206.
- Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2017). Comprehension demands modulate re-reading, but not first pass reading behavior. *The Quarterly Journal of Experimental Psychology*, 1–37.
<https://doi.org/10.1080/17470218.2017.1307862>
- Whiteside, K. E., Gooch, D., & Norbury, C. F. (2017). English language proficiency and early school attainment among children learning English as an additional language. *Child Development*, 88(3), 812–827. <https://doi.org/10.1111/cdev.12615>
- Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2005). Less skilled comprehenders show sluggish word-to-text integration processes. *Written Language & Literacy*, 8(2), 233–257.
- Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: evidence from eye-movements. *Reading and Writing*, 28(8), 1071–1097. <https://doi.org/10.1007/s11145-015-9561-4>
- Zabracky, K., & Ratner, H. H. (1989). Effects of reading ability on children's comprehension evaluation and regulation. *Journal of Reading Behavior*.
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14.
<https://doi.org/10.1111/j.2041-210X.2009.00001.x>

Appendix

Appendix A: Language Background Questionnaire

Questions asked:

Language use/ EAL or not

"For my experiment, I am also interested in the languages you learn. I know you can speak English because we are speaking it right now."

A Language Learning History

1. ... but do you or your parents speak or hear another language but English at home?

[Not show this just ask.]

Yes	No
1	0

[If Yes: continue. If no: stop here.]

1. Which language do you or your parents speak? _____

[From now on: insert language given by child instead of *your home language*]

2. How old were you when you first lived in the UK? _____

3.

4. How old were you when you started learning English? _____

B Languages Spoken with Family Members

"Please think back over the last month and when you used and heard *your home language* and when you used English. I will tell you times when you might use English or your home language, and I would like you to tell me how often you did this in *your home language*."

"I would like you to tell me if this happened

Never- Less than half the time- More than half the time- All the time"

[Show them response scale and point to it when explaining]

Trial: Tell me how often you had sandwiches for lunch this week/ last week.

Never	Less than half the time	More than half the time	All the time
1	2	3	4

I don't know.

How often did your parents speak in HL to you, for example to ask you a question or tell you about something?	
How often did you talk to your parents in HL, for example to tell them something or respond to a question?	
Do you have brothers or sisters?	
If yes: How often did you talk in HL to your brothers or sisters?	
<i>How often did your brothers or sisters talk in HL to you?</i>	
How often did you use HL with your friends?	

C Media Exposure

"Thank you, that is really helpful to know. Now I would like to ask you about three things that you might also do when you are at home- like watching TV or reading books. I would like you to think back over the last month again and tell me: How often you do this from

Never – Sometimes – Often

And then, how often you do this in English or compared to how often you do things in your HL from

Always in home language- More often in home language than in English -About the same-
More often in English than in home language- Always in English"
[SHOW VISUAL SCALES WHILE TALKING]

Trial: How often did you eat ice-cream over the last two months?

Never	Sometimes	Often
1	2	3

Always in home language	More often in home language than English	About the same	More often in English than in home language	Always in English
1	2	3	4	5

I don't know.

	How often do you do this?	How often do you do this in English compared to your Home language?
1. Watch TV or movies?		
2. Read Books yourself?		
3. Your parents read books to you?		

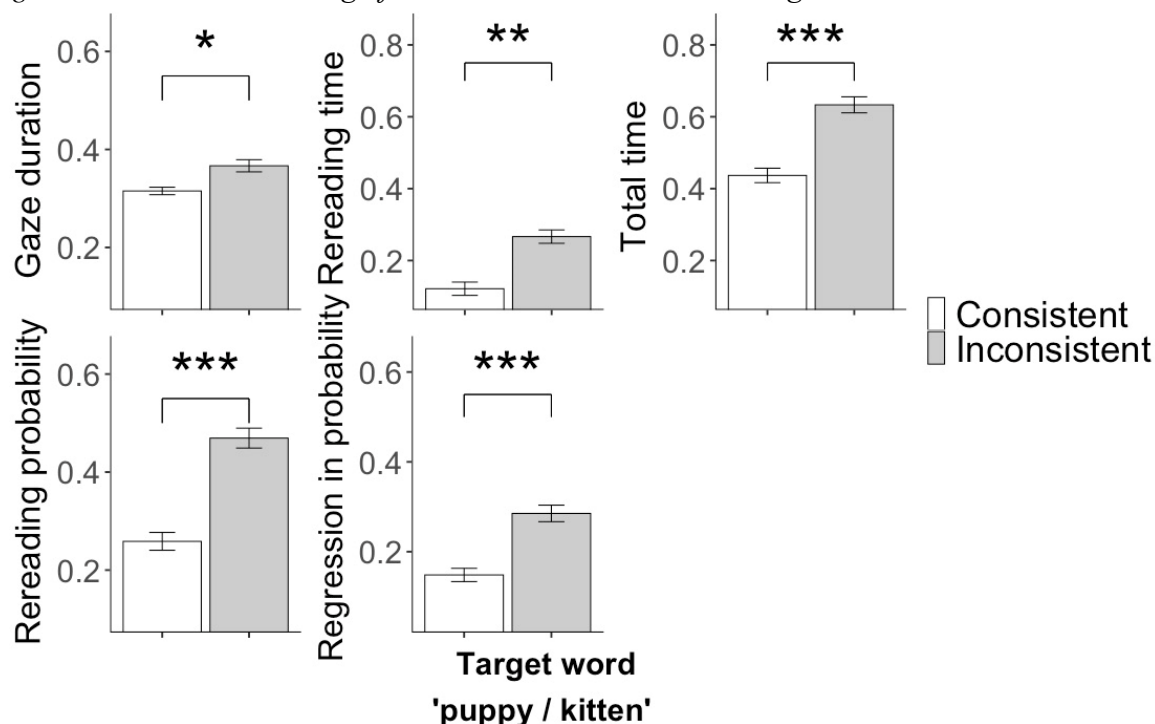
Appendix B: Eye-tracking Stimuli

Inconsistent	Consistent control	Text	Comprehension question	Correct answer	Incorrect answer
plant	dress	For her aunt's wedding in Spring, Linda wore her best outfit. The colourful plant / dress was one of her absolute favourites.			
nail	blade	Jeremy carefully sawed through a large oak tree in his garden. The steel nail /blade became dull after only two minutes.			
truck	plane	Last week William flew to visit his family in another country. The large truck / plane was spacious and quickly transported them.	How long did the journey take?	It was quick.	It was slow.
book	ball	Charles tossed his new toy outside the garden. He became upset when the book / ball went over the blue garden fence.			
broom	sponge	Marcus was washing dishes every night to earn his pocket money. The new broom was great for removing the dirt.	Why does Marcus wash dishes?	To be kind.	To earn his pocket money.
pencil	paper	Before school Jack printed his report for class. He became angry when the pencil / paper got stuck and he couldn't print.			
kitten	puppy	Rover barks at all passing animals on the street. He's the most alert kitten / puppy in the whole neighbourhood in fact.	Does Rover make a lot of noise?	Yes, he does.	No, he is very quiet.
handsaw	hammer	In the evening Nicholas nailed two boards together in his garage. His new handsaw / hammer was proving a really useful tool.			
plate	glass	Tyler cautiously sipped his hot and freshly brewed green tea. He dropped the plate / glass because it was really too hot.			

boots	sheets	Yesterday Joshua slept well for many hours, never once waking up. His new boots / sheets were super cosy and very comfortable.			
saddle	paddle	Thomas and Charlie rowed along the river in the park. Thomas lost his saddle / paddle in the middle of the river.	Where did Thomas and Charlie row?	On the seaside.	In the park.
jogging	football	Freddie threw the ball to his friend during the game. He really liked jogging / football much better than any other sport.			
spanner	needle	Last night Jenny sewed patches onto her old jeans. Sadly, she lost the spanner / needle and could not finish her work.			
hand	foot	When she got annoyed, Charlotte stomped on the floor. It hurt when her hand / foot hit the ground with such force.			
minister	officer	Last night Bob arrested a mean criminal as everyone watched. As an able minister / officer he handled the situation very well.	Was Bob alone when he arrested the criminal?	No, everyone watched.	Yes, he was alone.
movie	novel	Annabelle sat outside and read about a man named Arthur. She loved the movie / novel about ancient times very very much.			
blower	shovel	When planting her tree, Elisabeth dug a hole in the ground. Her sturdy blower / shovel helped make the job much easier.			
shoes	kettle	Michelle poured a fresh cup of coffee for her friend. She dropped the shoes / kettle and got very annoyed with herself.			
kayak	taxi	Today Janet rode with her best friends to the shop nearby. The yellow kayak / taxi quickly got them to the store.	Where did Janet go with her friends?	To school.	To the shop.
bike	boat	They were all really happy as they sailed along the coast. The swift bike / boat raced near the very beautiful beach.			

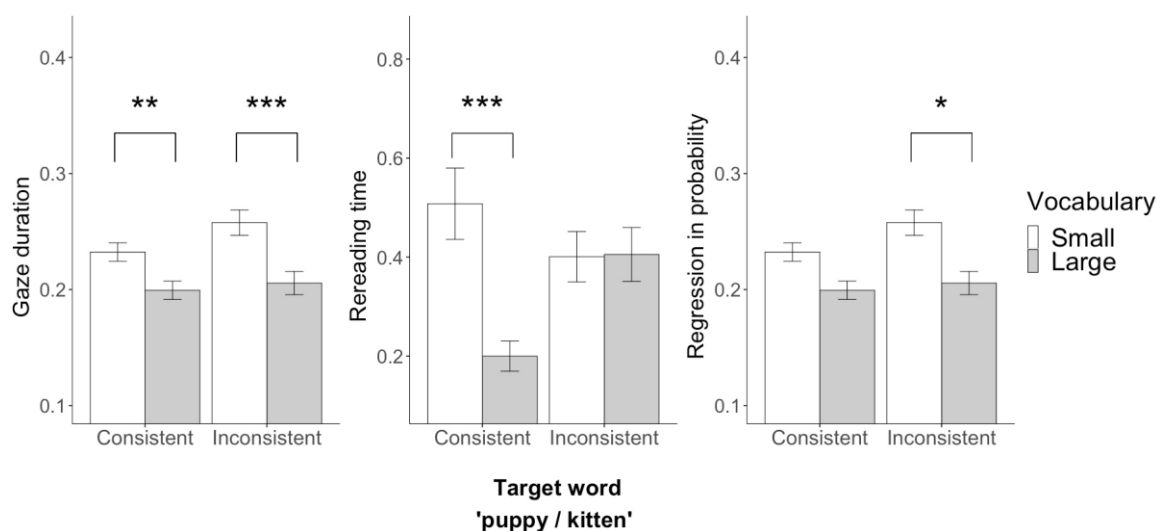
Tables and Figures

Figure 1: Children's reading of consistent and inconsistent target words.



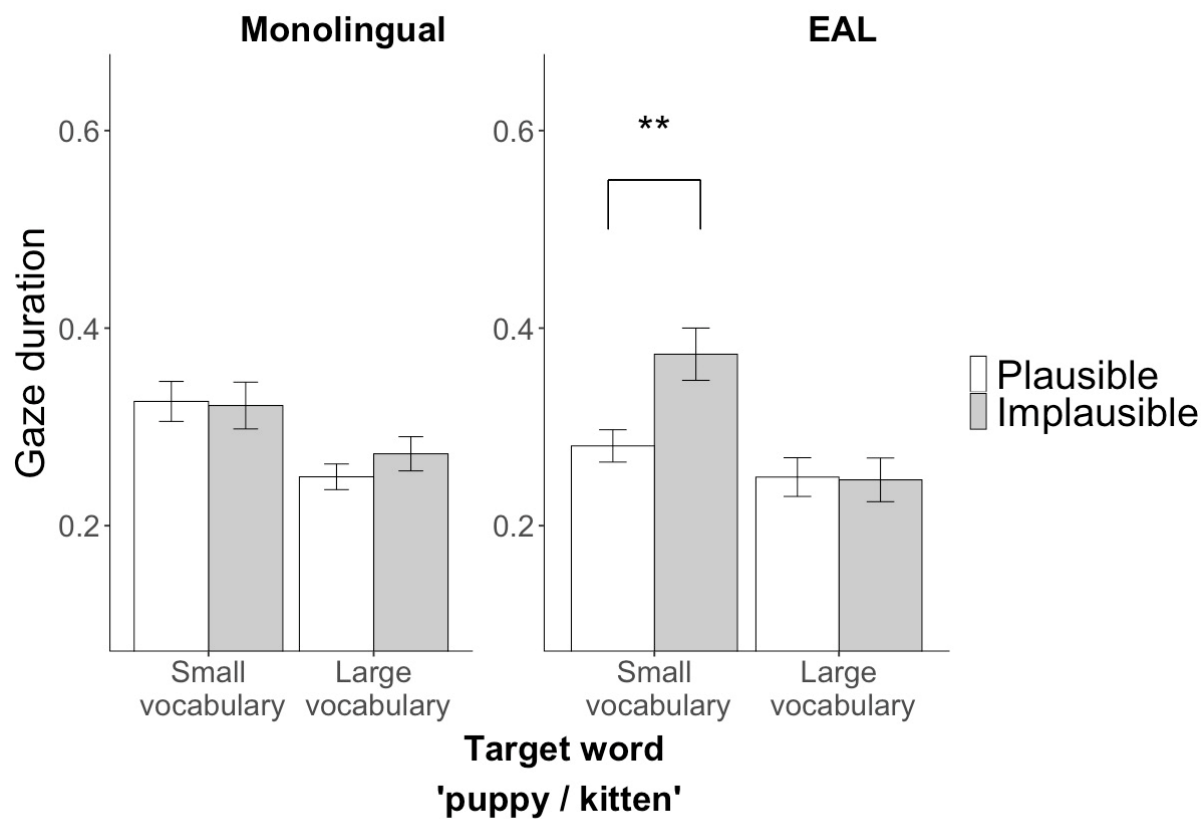
Error bars show the standard error. * $p < .05$ ** $p < .01$ *** $p < .001$.

Figure 2: Reading consistent and inconsistent words as a function of vocabulary. Plotted for predicted means at ± 1 SD of vocabulary size.



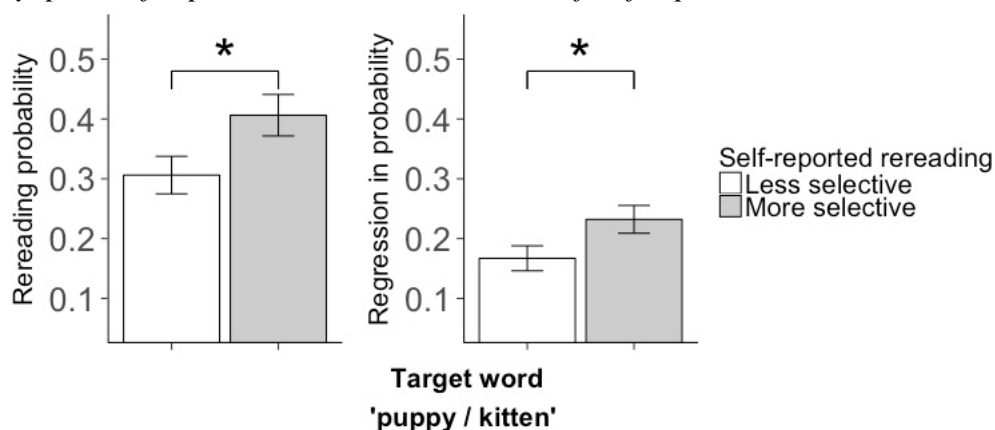
Note: Predicted means for reading times in seconds at ± 1 SD of vocabulary size. Error bars show the standard error. * $p < .05$ ** $p < .01$ *** $p < .001$.

Figure 3: Reading consistent and inconsistent words as a function of vocabulary and language group. Plotted for predicted means at ± 1 SD of vocabulary size.



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

Figure 4: Rereading of target words for children who reported to reread less and more selectively, plotted for predicted means at ± 1 SD of self-report.



Error bars show the standard error. * $p < .05$ ** $p < .01$ *** $p < .001$.

Table 1: English language experience of EAL children.

	Mean (SD)	Range
Age of arrival in UK	0;05 (1;08)	0;00 – 8;00
Age of acquisition of English language	1;01 (1;06)	0;00 – 4;00
Current English language use at home	2.8 (0.7)	0 – 4
Current English language use with media	2.9 (0.6)	2 – 4

'Current English language use at home' and 'Current English language use with media' are both composite scores ranging from 0-5 where 0 indicates no and 5 indicates only English language input.

Table 2: Mean (SD) age and standard scores for non-verbal reasoning, reading fluency and vocabulary size across language groups.

	Monolingual	EAL	Group comparison	Cohen's d
Age	10;02 (0;04)	10;03 (0;04)	$t(52.7) = -1.10, p = .28$	0.28
Non-verbal reasoning	47 (8)	46 (11)	$t(38.0) = 0.30, p = .77$	0.08
Reading fluency	116 (14)	119 (14)	$t(48.0) = -1.0, p = 0.31$	0.3
Vocabulary size	109 (13)	104 (12)	$t(49.3) = 1.5, p = 0.13$	0.4

Non-verbal reasoning scores have a mean of 50 and a standard deviation of 10. Scores for reading fluency and vocabulary size have a mean of 100 and a standard deviation of 15.

Table 3: Example story and question.

Rover barks at all passing animals on the street. He's the most alert *kitten / puppy* in the whole neighbourhood in fact.

Comprehension question: Does Rover make a lot of noise?

A: Yes, he does.

B: No, he is very quiet.

Table 4: Matching of inconsistent and consistent target words.

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

Note: Log frequency based on British National Corpus. Age of acquisition based on norms from Kuperman, Stadthagen-Gonzalez, & Brysbaert (2012).

Table 5: Overview of eye-tracking reading measures used in this study.

Stage of processing	Name	Definition
Early	Gaze duration	Continuous measure of time initially spent fixating a word before exiting it.
Late	Rereading time	Continuous measure of time spent fixating a word on any return to it.
	Rereading probability	Binary measure of whether a reader returned to word during a trial or not, both counting targeted saccades and general second-pass reading.
	Regression in probability	Binary measure of whether a reader returned directly with a leftward-saccade to word during a trial or not.
Both early and late	Total time	Continuous measure of entire time spent reading a word.

Table 6: Results of models using inconsistency, language group, reading fluency and vocabulary size to predict reading of the target word.

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

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

Table 7: Results of models using inconsistency, English language input, reading fluency and vocabulary size to predict reading of the target word.

	Gaze duration		Regression in probability		Rereading probability		Rereading time		Total time	
	<i>B</i>	<i>SE</i>	<i>OR</i>	<i>SE</i>	<i>OR</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
English language input	-0.02	0.03	0.97	0.12	0.97	0.11	0.02	0.05	-0.01	0.03
Inconsistency (Inc)	0.06	0.05	2.22***	0.20	2.81***	0.16	0.25*	0.10	0.32***	0.05
Language Group X Inc	-0.05	0.03	0.79	0.18	0.88	0.17	0.05	0.08	-0.03	0.04
Reading fluency (RF)										
RF Main	-0.30***	0.06	0.89	0.23	0.81	0.23	-0.40***	0.11	-0.37***	0.07
RF X Language Group	0.01	0.06	0.89	0.22	0.87	0.22	0.13	0.11	0.03	0.07
RF X Inc	-0.03	0.07	0.64	0.38	1.16	0.37	-0.05	0.16	-0.04	0.08
RF X Language Group X Inc	-0.07	0.07	1.18	0.37	0.88	0.35	-0.17	0.16	-0.09	0.08
Vocabulary size (Vocab)										
Vocab	-0.16**	0.06	1.23	0.24	1.10	0.24	-0.14	0.11	-0.13	0.08
Vocab X Language Group	0.05	0.06	0.79	0.23	0.97	0.23	0.08	0.11	0.04	0.07
Vocab X Inc	-0.03	0.07	2.71*	0.41	1.23	0.39	0.44**	0.17	0.09	0.08
Vocab X Language Group X Inc	0.23***	0.07	1.14	0.39	1.04	0.36	0.03	0.16	0.22**	0.08

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

Table 8: Model means for children reading the target word by language group, consistency condition and vocabulary size.

		<u>Gaze duration</u>	<u>Regression in Probability</u>	<u>Rereading probability</u>	<u>Rereading time</u>	<u>Total time</u>
Smaller						
Monolingual						
	Consistent	309 (19)	0.18 (0.04)	0.22 (0.05)	396 (53)	387 (30)
	Inconsistent	309 (23)	0.22 (0.05)	0.44 (0.07)	388 (46)	478 (46)
EAL						
	Consistent	280 (16)	0.13 (0.03)	0.24 (0.04)	399 (47)	368 (25)
	Inconsistent	373 (26)	0.21 (0.04)	0.46 (0.06)	404 (43)	567 (53)
Larger						
Monolingual						
	Consistent	255 (13)	0.11 (0.02)	0.24 (0.04)	273 (29)	313 (20)
	Inconsistent	277 (17)	0.29 (0.05)	0.45 (0.06)	459 (43)	467 (37)
EAL						
	Consistent	261 (20)	0.15 (0.04)	0.24 (0.06)	266 (43)	332 (31)
	Inconsistent	258 (24)	0.4 (0.08)	0.55 (0.09)	386 (56)	445 (54)

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

Table 9: Model means for children reading the target word by language group, consistency condition and reading fluency.

	<u>Gaze duration</u>	<u>Regression in Probability</u>	<u>Rereading probability</u>	<u>Rereading time</u>	<u>Total time</u>
Lower					
Monolingual					
Consistent	319 (16)	0.14 (0.03)	0.28 (0.05)	342 (35)	403 (26)
Inconsistent	338 (21)	0.26 (0.05)	0.46 (0.06)	515 (48)	564 (46)
EAL					
Consistent	319 (24)	0.12 (0.03)	0.25 (0.05)	483 (74)	441 (39)
Inconsistent	366 (32)	0.35 (0.07)	0.53 (0.08)	508 (69)	626 (73)
Higher					
Monolingual					
Consistent	248 (16)	0.15 (0.03)	0.18 (0.04)	316 (44)	301 (24)
Inconsistent	253 (19)	0.25 (0.05)	0.44 (0.07)	346 (41)	397 (39)
EAL					
Consistent	229 (13)	0.16 (0.03)	0.23 (0.05)	220 (28)	277 (19)
Inconsistent	263 (19)	0.25 (0.05)	0.48 (0.07)	307 (33)	403 (38)

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

Table 10: Results of models using inconsistency, language group, and self-report to predict rereading of the target word.

	Regression in probability		Rereading probability	
	<i>OR</i>	<i>SE</i>	<i>OR</i>	<i>SE</i>
Language group	0.93	0.20	0.99	0.18
Inconsistency (Inc)	2.34 ***	0.46	2.78 ***	0.45
Self-report	1.23 *	0.12	1.25 *	0.11
Language group X Inc	1.25	0.45	1.27	0.40
Language group X Self-report	1.25	0.25	1.08	0.19
Inc X Self-report	0.73	0.14	0.95	0.15
Inc X Language group X Self-report	0.73	0.27	0.85	0.27

Estimates are odds ratios (OR). * $p < .05$ ** $p < .01$ *** $p < .001$.