Consolidating new words from repetitive versus multiple stories: The role of prior knowledge

Lisa Henderson & Emma James

*Journal of Experimental Child Psychology*

<https://doi.org/10.1016/j.jecp.2017.09.017>

**Abstract**

Prior knowledge is proposed to support the consolidation of newly acquired material. The present study examined whether children with superior vocabulary knowledge show enhanced overnight consolidation, particularly when new words are encountered in varying stories. Children aged 10-11-years-old (*n* 42) were exposed to two sets of 8 spoken novel words (e.g., “crocodol”), with one set embedded in the same story presented twice and the other presented in two different stories. Children with superior vocabulary knowledge showed larger overnight gains in explicit phonological and semantic knowledge when novel words had been encountered in multiple stories. However, when novel words had been encountered in repetitive stories, existing knowledge exerted no influence on the consolidation of explicit phonological knowledge and had a negative impact on the consolidation of semantic knowledge. Twenty-four hours after story exposure, only very weak evidence of lexical integration (i.e., slower animacy decisions towards the basewords e.g., “crocodile” than control words) was observed for novel words learned via repetitive (but not multiple) stories. These data suggest that whilst the consolidation of explicit new word knowledge learned through multiple contexts is supported by prior knowledge, lexical integration might benefit more from repetition.

**Keywords**

Vocabulary; memory consolidation; lexical integration; oral language; lexical competition; prior knowledge

Oral storytelling throughout childhood is important for socio-emotional development (Baker, 2013) and confers significant educational advantages (Scarborough & Dobrich, 1994) particularly for literacy (Whitehurst & Lonigan, 1998; Ardoin et al., 2008; Lonigan et al., 2008). However, the most commonly reported effect of early storybook exposure is on children’s vocabulary knowledge (Farrant & Zubrich, 2013; Hammer, Farkas & MacZuga, 2010; Hepburn, Egan & Flynn, 2010; Senechal, 1997; Senechal, Pagan, Lever & Ouellette, 2008). It has even been argued that rich home literacy experiences may mitigate the risk posed by low socio-economic status for children’s vocabulary development (Payne, Whitehurst & Angell, 1994). Nonetheless, children’s ability to garner new words from listening to stories is highly variable (Biemiller & Boote, 2006; Mol, Bus & de Jong, 2009), and associated with children’s existing vocabulary knowledge (Henderson, Devine, Weighall & Gaskell, 2015; Karweit & Wasik, 1996; Wilkinson & Houston-Price, 2013). For pedagogical reasons it is important that we understand how best to utilise stories to promote vocabulary learning, and examine how individual differences in the existing knowledge base may impact optimal conditions for word learning.

Studies of spoken word learning that utilise explicit training regimes have consistently demonstrated that word learning is a long-winded process, both in developing and adult learners (see James, Gaskell, Weighall & Henderson, 2017, for a review). A distinction has been made between ‘lexical configuration’ (i.e., the rapid acquisition of explicit word knowledge, e.g., as measured by recall tasks) and ‘lexical engagement’ (i.e., the slower emergence of a novel word becoming ‘integrated’ and interacting with existing lexical knowledge) (Leach & Samuel, 2007). This protracted process can be explained by a complementary learning systems (CLS) account (Davis & Gaskell, 2009; McClelland, McNaughton & O’Reilly, 1995), whereby initial exposure to novel words gives rise to sparse representations in the short-term hippocampal system, with consolidation required for a long-term representation to be strengthened in neocortical memory. Studies of children and adults have shown that novel word forms can be more accurately recalled and begin to compete for recognition with existing words after a period of off-line consolidation (Gaskell & Dumay, 2003; Brown, Weighall, Henderson & Gaskell, 2012; Henderson, Weighall, Brown & Gaskell, 2013), over and above the effects of repeat testing (Henderson, Weighall & Gaskell, 2013). Consistent with a substantial body evidence suggesting an active role for sleep in supporting memory consolidation (see Spencer, Walker & Stickgold, 2017, for a review), improvements in novel word knowledge are particularly apparent after sleep in adults (Dumay & Gaskell, 2007) and throughout childhood (Gais, Lucas & Born, 2006; Henderson, Weighall, Brown & Gaskell, 2012; Huang, Deshpande, Yeo, Lo, Chee, & Gooley, 2016; Axelsson, Williams & Horst, 2016; James et al., 2017). Furthermore, specific parameters of slow-wave sleep (e.g., slow oscillations <1Hz and sleep spindles 12-15Hz) are associated with overnight changes in lexical competition and explicit phonological memory for novel words, in adults (Weighall, Henderson, Barr, Cairney & Gaskell, 2016; Tamminen, Payne, Stickgold, Wamsley & Gaskell, 2010) and children (Smith, Gaskell, Weighall, Warmington, Reid & Henderson, under revision). These EEG events are temporally synchronised with hippocampal ripples (Staresina et al., 2015) and together have been proposed to coordinate the reactivation of newly learned information stored in hippocampal networks and its subsequent integration in to neocortical systems (Rasch & Born, 2013).

Whilst the importance of off-line consolidation is well established for phonological aspects of word learning, there is less consistency within the literature on semantic aspects of learning. For example, Henderson et al (2013b) found overnight improvements in children’s ability to recall definitions of previously unfamiliar science words. However, Weighall et al (2016) reported no consolidation advantage for children’s ability to make true/false judgements on newly learned novel word-object pairs, and Tamminen and Gaskell (2013) reported a decline in adults’ ability to recall nonword meanings on the day after learning.

Off-line consolidation also appears to support vocabulary learning when words are encountered incidentally via listening to stories. For example, Henderson et al. (2015) exposed children to novel competitor words (e.g., “daffodat” derived from “daffodil”) that were embedded in a spoken fictional story. Consistent with the CLS account, children’s ability to recall the phonological forms of the novel words improved after 24 hours, and ‘lexical competition’ was not demonstrated immediately after exposure, but did emerge 24 hours later. Lexical competition was measured via a pause detection task (Mattys & Clarke, 2002), in which children showed slower pause detection responses (i.e., is a 200ms pause present/absent?) to the basewords (e.g., “daffo\_dil”) of the novel words heard in the story (e.g., “daffodat”) than to control words which had no trained competitor. Tamura, Castles and Nation (2017) have recently reported a similar protracted time course of lexical integration when children read unfamiliar written words in fictional texts.

An important question is how vocabulary consolidation can be optimised from story encounters. One characteristic of successful vocabulary learning from stories is repetition (e.g., Blachowicz, Fisher, Ogle, & Watts-Taffe, 2006). In line with this, Horst and colleagues (e.g., Horst, Parsons & Bryan, 2011; Williams, Horst & Oakhill, 2011) have demonstrated that pre-schoolers’ word learning via stories benefits from repetitive encounters with the same story. Horst et al (2011) read 3-year-olds nine bespoke storybooks in three sessions over a one-week period. Within each session, children either listened to three different stories, or heard the same story repeated three times. Importantly, each of the nine storybooks contained two novel-nonword pairs that were associated with an illustrated novel object. Repeated-contexts led to better novel word comprehension (measured via a picture matching task), both immediately after the final story and one week later. Williams and Horst (2014) replicated this effect in 3-year-olds, demonstrating that although a nap soon after exposure improved performance for both story conditions, the repeated-context advantage maintained after one week.

Horst and colleagues attributed this repetition advantage to ‘contextual cueing’; namely, that repetition of stories works to reduce attentional demands and permit more resources for word learning (Horst, 2013). It follows that the same benefit of repetitive story readings may not be found in older children and/or children with better vocabulary knowledge. Such children may have more resources for word learning if they can reap support from oral language skills (Horst, 2013). Indeed, Karweit & Wasik (1996) found that children with poorer vocabulary demonstrate increased involvement with and comprehension of a story with each repetition, whereas children with greater vocabulary ability show the opposite pattern. Word learning through multiple contexts has been central to dominant approaches in vocabulary instruction in school-aged children (e.g., Beck, McKeown & Kucan, 2013). Theoretically, it has been argued that flexible and automatically retrieved representations of meaning, which allow for words to take on different meanings in different contexts, are crucial for language comprehension (Perfetti, 2007). In adults, it has also been shown that contextual diversity leads to more accurate/faster semantic decisions towards words (Ferreira & Ellis, 2016) and better learning of abstract words (Bolger, Balass, Landen & Perfetti, 2008).

To examine whether multiple story contexts are more advantageous for word learning in older children than repetition, and whether vocabulary skills underlie this effect, Wilkinson & Houston-Price (2013) presented 165 children aged 6-9 years with three fictional stories that contained a set of 8 unfamiliar (but real) words on three different days. Children either heard the same story repeated across three days or they heard three different stories. Children’s comprehension of the new words (measured via a picture matching task) was tested prior to them hearing the stories, within 24 hours after the final exposure to the story and again two weeks later. Children showed gradual improvements in their comprehension of the newly learned words across the three test points. There was no overall effect of hearing multiple versus repeated stories on new word comprehension at any test point. Although there was an age x context interaction after the final exposure (i.e., 8-9-year-olds showed similar word learning for words learned in both context conditions, but 6-7-year-olds showed poorer word learning following multiple than repetitive story encounters) this interaction was not observed two weeks later. Thus, whilst the data are clear in suggesting that repetition of the same story does not lead to a word learning advantage in school-aged children, neither did exposure to different stories exert a reliable long-term benefit.

Wilkinson and Houston-Price (2013) also found that existing receptive vocabulary knowledge accounted for more than 20% of variance in comprehension of the new words, both 24 hours after exposure to the stories and two weeks later (see also Penno, Wilkinson & Moore, 2002). This is consistent with the existence of a *Matthew effect* in literacy (Stanovich, 1986; Scarborough, Catts & Kamhi, 2005) and even more consistently in vocabulary learning (Cain & Oakhill, 2011). That is, better vocabulary skills allow for superior vocabulary learning over time, resulting in a faster rate of vocabulary growth (e.g., Webb & Chang, 2015) and a vocabulary knowledge gap that widens across the school years (Cain & Oakhill, 2011). Prior vocabulary knowledge is also an important predictor of success in vocabulary intervention studies, with children who begin interventions with low vocabulary tending to show smaller gains (Coyne, McCoach, Loftus, Zipoli & Kapp, 2009; Coyne et al., 2004; Penno et al., 2002).

Counter to their predictions, Wilkinson and Houston-Price (2013) *did not* find that children with better vocabulary knowledge benefited more from multiple contexts. However, the three stories were administered over three weeks, with the ‘immediate’ post-story test of new word knowledge administered between 0 and 24 hours after the final reading. Arguably, any benefits of learning from multiple story contexts could have been contaminated by these spaced learning conditions (e.g., Lindsay & Gaskell, 2013; Vlach, Sandhofer & Kornell, 2008) which may have allowed for consolidation between sessions. Furthermore, it cannot be determined from this design how story contexts and existing vocabulary knowledge may interact during the consolidation process, because no measures of word learning were taken pre/post off-line consolidation.

In a recent review, James et al. (2017) posed that existing vocabulary plays an important role in scaffolding the integration of new words. This is consistent with the most recent description of the CLS model which emphasises that neocortical learning is dependent upon prior knowledge (McClelland, 2013), and with studies that demonstrate enhanced consolidation of new information that is consistent with existing knowledge (Lewis & Durrant, 2011; Tse et al., 2007; Wilhelm et al., 2012). Indeed, in Henderson et al (2015) overnight improvements in recall and overnight increases in lexical competition following exposure to novel words in stories were positively correlated with children’s existing expressive vocabulary knowledge (see also Horvath, Myers, Foster & Plunkett, 2015). Arguably, overnight consolidation of new vocabulary may be even more dependent upon existing vocabulary knowledge when children are learning new words from multiple (as opposed to repetitive) stories, given the additional demands on comprehending and integrating information between different contexts.

*The present study*

The over-arching aim of the present study was to uncover optimal conditions for new word learning from stories. Children aged 10-11-years-old were exposed to two sets of 8 novel words by listening to fictitious stories. One set were presented in the same story twice (the repetitive context condition) and the other set were presented in two different stories (the multiple context condition). Testing the predictions of Horst et al (2011), we examined whether the overnight consolidation of these novel words would benefit from encountering the words in multiple, rather than repetitive, stories in older school aged children, particularly when children have support from a rich body of existing vocabulary knowledge. In order to untangle the influence of exposure context on initial levels of learning and the subsequent consolidation of that learned material, we exposed all children to multiple and repetitive story conditions in a single session and measured their learning immediately after exposure and again 24 hours later. A standardised measure of expressive vocabulary (i.e., a definitions task) was used to capture existing vocabulary knowledge, following previous studies that report an association between the consolidation of new vocabulary and the existing knowledge base (e.g., Henderson et al., 2015; Horvath et al., 2015).

Three measures of *explicit lexical knowledge* were administered to measure phonological (cued recall) and semantic aspects of novel word learning (definitions and animacy judgement). Children learned novel competitors (e.g., “dolpheg”) derived from familiar English words (e.g., “dolphin”), allowing us to also measure *lexical integration* within the animacy task. More specifically, children made animacy judgements for both the novel words (e.g., is “dolpheg” an animal?), which measured children’s ability to retrieve explicit semantic knowledge, and their baseword counterparts (e.g., is “dolphin” an animal?) to measure lexical integration, intermixed within the same task. According to models of spoken word recognition (Gaskell & Marslen-Wilson, 2002; Grainger & Jacobs, 1996; Marslen-Wilson, 1989; McClelland & Elman, 1986; Norris, 1994), once “dolpheg” has been integrated with the lexicon, hearing the speech input “dolphin” should momentarily lead to co-activation of “dolpheg” and “dolphin” and subsequently slow down children’s ability to decide whether “dolphin” is an animal. This ‘lexical competition effect’, present in childhood, has been argued to provide a marker of lexical integration and permit fast retrieval of stored lexical knowledge for the purposes of automatic word recognition (Henderson, Weighall, Brown & Gaskell, 2013). Previous studies with children have relied on measuring lexical competition via pause detection (Henderson, Weighall, Brown & Gaskell, 2013; Henderson et al., 2012, 2015) or lexical decision tasks (Brown et al., 2012) which require children to make a non-semantic decision towards the basewords. The present animacy task allowed us to examine whether the same lexical competition effects emerge when children make semantic judgements towards the basewords (e.g., does learning “dolpheg” interfere with children’s ability to retrieve semantic information about “dolphin”?). A similar task was utilised by Bowers, Davis & Hanley (2005), in an examination of lexical competition effects following orthographic word learning.

*Hypotheses*. Consistent with a CLS account of vocabulary learning (Davis & Gaskell, 2009) and theories that newly formed memories are strengthened off-line through reactivations during sleep (see James et al., 2017, for a review), children’s ability to recall the phonological forms of the novel words and make animacy judgements towards them should improve 24hrs after hearing the stories. Although, it should be noted that previous literature is inconsistent regarding overnight improvements in explicit semantic knowledge (e.g., Weighall et al., 2016; Tamminen & Gaskell, 2013). Also consistent with a CLS account (Davis & Gaskell, 2009), animacy judgements to the basewords (e.g., “dolphin”) should slow down relative to responses to control words after 24 hours, if the novel competitors (e.g., “dolpheg”) begin to compete for recognition with the basewords, as reflected by a lexical competition effect. Based on the most recent description of the CLS model (McClelland, 2013) and findings of a general association between existing vocabulary knowledge and overnight consolidation effects (see James et al., 2017), we predicted an interaction between vocabulary knowledge and session, such that larger overnight increases in recall would be expected for individuals with better vocabulary knowledge, consistent with a Matthew effect in vocabulary consolidation. Based on Horst et al (2013), one could hypothesise enhanced learning and/or consolidation and integration for words encountered in different (versus repeated) stories within the present age group (although this would be counter to the findings of Wilkinson & Houston-Price, 2013, who found no overall effect of context). Importantly however, and in accordance with the hypotheses of Horst (2013), we predicted that any advantage of learning in different contexts (for learning and/or consolidation) would be positively associated with existing expressive vocabulary knowledge (i.e., the advantage would be more apparent for children with superior vocabulary). For the consolidation of explicit phonological memory (as measured by cued recall), this would manifest as an interaction between session (immediate test, 24-hour test), context condition (repetitive, multiple) and vocabulary.

**Method**

*Participants*

Forty-two children aged 10-11 years old took part (mean age = 11.04 years; SD = 0.37 years; 17 males), recruited from three mainstream primary schools representing a range of socio-economic backgrounds. The age range of 10-11 years was selected to facilitate comparison with previous studies (e.g., Wilkinson & Houston-Price, 2013), and by this age, vocabulary knowledge has a demonstrated impact on text and discourse comprehension (Clarke, Henderson & Truelove, 2010). All children were monolingual, native English speakers with normal or corrected to normal vision and hearing. All parents provided informed consent, and the study was approved by the Department of Psychology Ethics Committee at the University of York.

Children completed standardised tests to measure nonverbal ability (Matrix Reasoning, Weschler Abbreviated Scales of Intelligence 2, WASI-2; Weschler & Hsaio-Pin, 2011), expressive vocabulary (Vocabulary, WASI-2) and word and nonword reading ability (Test of Word Reading Efficiency – 2, TOWRE-2; Torgesen, Wagner & Rashotte, 1999). Notably, the Vocabulary subtest captured existing expressive vocabulary knowledge and required children to produce definitions of words which increased in difficulty. Confirming this sample was representative of the age-group under investigation with regards to relevant cognitive and literacy abilities, mean scores spanned the average range for all measures (mean nonverbal ability T score = 49.76, SD = 9.42; mean expressive vocabulary T score = 52.98, SD = 10.44; mean word reading standard score = 98.43, SD = 6.84; mean nonword reading standard score = 107.17, SD = 12.26) and children showed a normal distribution of ability on all measures (all Kolmogorov-Smirnov z scores were >.76).

*Stimuli*

All stimuli (words, novel words, stories and illustrations) are provided as Supplementary Material.

*Novel words*. Two lists (List 1, List 2) of 8 stimulus pairs were taken from Henderson et al. (2012). Each pair comprised a spoken novel word (e.g., “crocodol”) and a familiar spoken baseword (e.g., “crocodile”). All basewords were picturable nouns that have been used in multiple previous studies and have been deemed to be highly familiar to children of the present age range. The two novel word and baseword lists were matched on length and Celex frequency. For each child, one list of novel words was assigned to the ‘repetitive contexts’ condition, and the other list was assigned to the ‘multiple contexts’ condition, with list assignment counterbalanced across children. To allow the measurement of lexical competition in the animacy task, two additional lists of 8 control basewords were also taken from Henderson et al (2012) so that responses to the basewords (e.g., “crocodile”) could be compared to responses to control basewords (e.g., “dolphin”) for which no new competitors had been encountered. These two control baseword lists were also matched on length and frequency to the competitor baseword lists (see Supplementary Material). Half of the basewords in each list (i.e., the two experimental baseword lists, and the two control baseword lists) were non-living objects and half were animals. The initial fragments of all competitor/control basewords (e.g., “croco…”) had no other possible completions.

Each novel word was paired with a novel object/animal (taken from Weighall et al., 2016). To increase the likelihood of observing lexical competition on the animacy task and avoid the risk of children responding correctly to the novel words simply because they were activating the correct response for the basewords, the novel competitors always required the alternative response (e.g., if the baseword was an animal e.g., “crocodile”, then the novel competitor e.g., “crocodol” would be an object, and vice versa). All objects/animals were coloured clipart-type images (see Weighall et al., 2016 for more detail). To ensure that 50% of the responses on the animacy task would be “yes”, half of the novel words in each list were assigned object meanings, and half were assigned animal meanings. Each novel word had between 3 and 5 semantic features. The majority of these semantic features were similar to or compatible with a familiar object/animal but crucially one or two of the features was novel. For example, a “flamingist” had the familiar features of *being like a scooter*, *going fast*, and *being pink and brown*, and the novel feature of *being able to fly*.

*Stories*. Stories were presented audio-visually via iPads. Headphones were used to limit disruption from background noise. Four fictional stories were written by the experimenters: Two were themed around “Lucy’s trips to an alien zoo” and contained the List 1 novel words, and two were themed around “Jack’s adventures on the moon” and contained the List 2 novel words. A single version of each story was recorded by a native English female speaker. Efforts were made to record the stories using similar intonation, pitch, pace and prosody, and they were matched on length (i.e., the zoo stories were 548 and 557 words in length; the moon stories were 555 and 556 words in length). Each novel word occurred three times at distributed positions throughout the story, and no novel word occurred in more than two consecutive sentences. All words used in the stories had an age of acquisition of 7 years or younger (Kuperman et al., 2012). No novel word meanings were provided explicitly; rather, in order to extract meaning of the novel words an inference had to be made (e.g., “She ran downstairs to unwrap the best present ever – a brand new flaminigist! Lucy couldn’t wait to fly to school on it!”) (following Henderson et al., 2015). We asked three local teachers who taught children of the target age range to read the stories. They all agreed that the stories were gender neutral and would be of interest and appropriate to the age range.

*Illustrations*. Each story was accompanied by four original colour illustrations. Each novel animal/object featured in the illustrations twice per story. The association between the novel word and novel object occurred via mutual exclusivity (Markman, 1990) and/or via process of elimination (Halberda, 2006). Each illustration was assigned to a recorded segment of the story (see Supplementary material). The illustration was present on the iPad screen throughout the segment, and at the start of the next segment the next illustration appeared.

*Design and Procedure*

A within-subjects design was adopted, such that all children completed both multiple and repetitive story conditions in a single session. The order of context condition, and the assignment of story themes to context conditions (Lucy at the zoo/Jack on the moon), were fully counterbalanced across participants. Children were tested in a quiet room in their schools. They first listened to the stories and then completed the measures of word learning in the following fixed order.

 **Story exposure**: Children were told that they were going to listen to some stories about “Lucy and her trips to the zoo and Jack’s adventures on the moon”. They were told that “there may be some words that you have not heard before but please keep listening to the stories and avoid asking questions so that you do not miss any of the story.” Children listened to four stories via headphones in total: one story repeated twice (from either the zoo/moon theme) and two different stories (from either the zoo/moon theme). Each story was split into four recorded segments that each corresponded to an illustration that was displayed throughout the segment and automatically replaced by the next illustration at the start of the next segment (see Supplementary Material). Children were only exposed to the novel words in the stories, which is a notable departure from Henderson et al. (2015) where children were first exposed to the list of novel words in isolation and asked to repeat each novel word aloud. Henderson et al (2015) adopted this procedure as a consequence of a pilot study where levels of learning were very low when the novel words were encountered only in the stories. Here, we opted to include supportive illustrations and decrease the number of novel words encountered in each story, in an effort to preserve the implicit nature of the instruction but observe reliable levels of learning.

 After hearing the two repetitions of the story in the repetitive contexts condition, and after hearing each story in the multiple contexts condition, children answered three questions to check their comprehension. Each question was scored as 1/0 (maximum 3). For the multiple contexts condition the score for each story was averaged (i.e., to produce an average score out of 3 for purposes of comparison with the repetitive contexts condition). Children scored at ceiling for both conditions, suggesting they were attending to and successfully comprehending the stories, and there was no significant difference in the mean comprehension score for repetitive versus multiple contexts (repetitive contexts mean 2.79, SD = 1.32; multiple contexts mean 2. 59, SD = 1.41; *t*<1, *p*>.05).

 **Animacy task (semantic retrieval / lexical integration)**. An animacy task (also programmed and delivered via iPad) captured children’s explicit semantic knowledge of the novel words (e.g., “crocodol”) as well as providing an index of lexical integration (i.e., by measuring responses to the basewords e.g., “crocodile”, versus control basewords that had no new competitor). On each trial children heard a spoken novel / baseword / control baseword (recorded by the same native speaker who recorded the stories) and were asked to decide if the word was an animal or not by pressing one of two “yes” and “no” images on the iPad as quickly and as accurately as possible. The stimuli comprised 16 novel words (8 each from the repetitive/multiple context conditions), 16 basewords (8 each from the repetitive/multiple context conditions), and 16 control basewords (8 assigned to each condition). An additional 16 filler words were included to balance the number of words that had been encountered in the story or were competitors of these words with words that had not been encountered or were not competitors. The filler words were matched on spoken frequency to the control / competitor basewords. Half of the 64 words required “yes” responses (i.e., it is an animal). A single version of the task was used, with the item order randomised for each participant. RT and accuracy were measured from word onset.

**Cued phonological recall**. A cued recall task measured children’s ability to retrieve the phonological forms of the words. Using the same procedure as Henderson et al (2012), children heard the first syllable of each novel word (e.g., “daff”) and were asked to provide the whole word using one of the new words they heard in the story. Children scored one point for a correct production of each novel word, and 0 for an incorrect response. This task was presented via the same iPad that was used to present the stories.

**Definitions**. A definitions task was used to measure children’s ability to express the meanings of each word. For each word, researchers instructed children to “Tell me what XXX means. Imagine that you are explaining the word to someone who has never heard it before.” Each definition was scored out of 3 (maximum score 48). Each novel word was assigned 3-5 features that could be learned from either of the two stories within each theme (i.e., all features were present in both stories). However, additional features could also have been learned from the illustrations. To give precedence to information acquired from the story, no more than one point could be earned for features arising from the illustrations. Furthermore, responses that were correct but associated with specific story events were not credited, since this would have given the multiple context condition an immediate advantage. For example, for the item “crocodol”, features that were included in both stories included: (1) belongs to zookeeper, (2) used to call animals / get animals back to cages, (3) blow through it, (4) makes a noise / makes a paroop. In addition, a feature that was not provided in the text, but was provided in the illustration was “coloured silver”. However, features specific to a single story (e.g., “zookeeper threw it when he was happy” or “used to get the escaped napkig”) were not awarded points. Two raters independently scored all of the definitions data. Their initial level of agreement (i.e., the percent of responses that were given the same score) was high, at 91%. Any discrepant items were discussed and agreed upon for the final analysis.

The cued recall and animacy tasks were administered immediately after listening to all four stories, and again 24 hours later. The definitions measure was administered at the end of the 24 hour test only (following Henderson et al., 2015).

*Data Analysis*

Data were analysed using R (R Core Team, 2015). Models were fitted using packages *lme4* (Bates, Maechler, Bolker, & Walker, 2014) or *ordinal* (Christensen, 2015), and figures were made using *ggplot2* (Wickham, 2009). For each dependent variable, a mixed effects model was fitted, with fixed effects of context condition (repeated, multiple), expressive vocabulary score, and - where relevant - test session (same day, next day). As appropriate for the different types of data collected, logistic regression models were used to model binary outcomes (cued phonological recall, animacy task accuracy), cumulative link models for tasks with more than two categorical outcomes (definitions), and linear models for continuous outcome data (animacy task reaction time). In each instance, categorical predictors were coded using deviance coding for comparability to traditional ANOVA analyses, providing an assessment of each main effect and interaction independently of other predictors in the model. Vocabulary score was scaled and centred for entering into each model.

Our aim was to include maximally specified random effects structures with random slopes for participants and items for all effects, as has been recommended for factorial experimental designs (Barr, Levy, Scheepers, & Tily, 2013). However, this caused well-documented convergence problems during model fitting (Bates, Kliegl, Vasishth, & Baayen, 2015). Given the experiment was designed to test specific hypotheses regarding the fixed effects, our approach was to fit the model using random intercepts in the first instance, and where possible use this to prune fixed effects to create the most parsimonious model (Bates et al., 2015). The random effects structure was subsequently built up using a “forward best path” approach (Barr et al., 2013), using pairwise comparisons between simple and increasingly complex models and retaining only those random effects justified by the data. This involved first testing whether each slope improved the fit of the model independently (under a liberal criterion of *p* <.2; Barr et al., 2013), retaining the slope that led to the biggest model improvement, and then repeating this process to test the inclusion of additional slopes.

 The best fitting model is described for each analysis below.The models are interpreted as standard regression models: a positive *b* coefficient means that the independent variable showed a positive relationship with task performance. *P*-values were calculated from Wald’s statistic where provided by model output, or by *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2016). It should also be noted that the pattern of significant predictors remained consistent throughout the modelling process.

**Results**

*Cued phonological recall*

 A mixed effects logistic regression model was fitted to the recall accuracy data (0, 1), and included by-item random slopes for the effect of vocabulary, and by-participant intercepts only (Table 1). The model revealed that the test session was a significant predictor of cued recall performance (*p* <.001): accuracy was better the next day (Session 2 *M* = .28, *SD* = .45) than on the day of learning (Session 1 *M* = .14, *SD* = .34; Figure 1). Vocabulary knowledge also significantly influenced overall performance (*p* < .001). The context manipulation was not an independent predictor of cued recall performance, but did contribute to a three-way interaction alongside test session and vocabulary ability (*p* = .01). We further explored this interaction by computing separate, maximally specified models for each context condition separately. Whilst test session and vocabulary remained significant predictors in both models (all *p*s <.01), their interaction was significant only in the multiple contexts condition ($b$ = 0.79, *SE* = 0.32, $z $= 2.47, *p* = .014) and not the repetitive contexts condition ($b$ = -.33, *SE* = 0.30, $z $= -1.08, *p* = .28). That is, vocabulary ability was a positive predictor of improvements in performance at the second test session only after learning in the multiple contexts condition (Figure 2).

Table 1

*Predictors of cued phonological recall performance*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed effects | $$b$$ | *SE* | $$z$$ | $$p$$ |
| (intercept) | -2.00 | 0.32 | -6.35 | <.001 |
| **test session** | **1.11** | **0.17** | **6.41** | **<.001** |
| context | -0.05 | 0.18 | -0.26 | .792 |
| **vocabulary** | **0.74** | **0.19** | **3.88** | **<.001** |
| test session x context | 0.32 | 0.34 | 0.93 | .351 |
| test session x vocabulary | 0.24 | 0.17 | 1.36 | .173 |
| context x vocabulary | 0.04 | 0.18 | 0.23 | .819 |
| **test session x context x vocabulary** | **0.89** | **0.34** | **2.59** | **.010** |
| Random effects | Variance | *SD* |  |
| participant: (intercept) | 0.9 | 0.95 |  |  |
| item: (intercept) | 1.05 | 1.02 |  |  |
| item: vocabulary (slope) | 0.06 | 0.24 |  |  |

*Note.* Model formed from 1344 observations, collected from 42 participants across 16 items. Random effects included in the model were selected using a data-driven approach.



*Figure 1.* Mean proportion of words recalled in the cued phonological recall task for each context and time condition separately. Error bars represent +/-1 SE of the mean, adjusted for within-subject variance (Morey, 2008).

*Definitions*



*Figure 2.* Overnight change in proportion recalled ([next day – same day]) in the cued phonological recall task as a function of expressive vocabulary ability, for words learned in the repetitive (solid line) and multiple (dashed line) story contexts conditions separately. Grey bars indicate 95% confidence intervals.

*Definitions task*

Children could often produce some relevant information in the definition task, but this was often limited to a single core feature, as demonstrated by the relatively low mean item-level performance in this task (*M* = 0.83, *SD* = 0.39; maximum score of 3). The definitions task was only administered on the day after children encountered the stories, enabling a multinomial model that converged with a fully specified random effects structure (see Supplementary Material). The context manipulation did not influence the meanings learned from the stories ($b$ = 0.02, *SE* = 0.2, $z $= 0.13, *p* = .9), with mean performance in the multiple contexts condition (*M* = 0.85, *SD* = 0.93) no different from the repetitive contexts condition (*M* = 0.82, *SD* = 0.91). Furthermore, there was no effect of vocabulary ($b$ = 0.04, *SE* = 0.19, $z$ = 0.21, *p* = .83) or vocabulary-by-context interaction ($b$ = 0.17, *SE* = 0.19,$ z$ = 0.92, *p* = .36).

*Animacy task*

Data from this task was lost (due to a technical error) from three participants at the 0 hr test and two particpants at the 24 hr test, meaning that 1264 observations contributed to the models (rather than 1344).

**Novel words accuracy.** The best fitting model included by-item random slopes for the effect of context, whereas the inclusion of by-participant random effects was not supported by the data (Table 2). The accuracy of speeded judgements to the novel words improved slightly between Session 1 (*M* = .43, *SD* = .50) and Session 2 (*M* = .48, *SD* = .50), although test session did not quite reach significance as an overall predictor in the model (*p* = .067, Figure 3). Session, context, and vocabulary ability presented a significant three-way interaction (*p* = .022), whereas no other fixed effects were significant predictors. When further analysed in separate models for each context condition, neither session nor vocabulary predicted performance for the repetitive contexts condition (all *p*s > .3). However, consistent with the cued phonological recall analyses, words learned in the multiple contexts condition improved over night ($b$ = 0.40, *SE* = 0.18, $z $= 2.17, *p* = .03), and more so for children with higher levels of existing vocabulary knowledge ($b$ = 0.37, *SE* = 0.18, $z $= 2.04, *p* = .042) (Figure 4). However, it should be noted that mean performance remains at chance during this task. As such, reaction time analyses were not appropriate for these data.

Table 2

*Predictors of animacy decision accuracy for novel words*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed effects | $$b$$ | *SE* | $$z$$ | $$p$$ |
| (intercept) | -0.20 | 0.14 | -1.40 | .163 |
| test session | 0.22 | 0.12 | 1.83 | .067 |
| context | 0.12 | 0.19 | 0.64 | .520 |
| vocabulary | 0.11 | 0.06 | 1.77 | .077 |
| test session x context | 0.29 | 0.24 | 1.22 | .223 |
| test session x vocabulary | 0.11 | 0.12 | 0.94 | .349 |
| context x vocabulary | 0.07 | 0.12 | 0.58 | .559 |
| **test session x context x vocabulary** | **0.55** | **0.24** | **2.29** | **.022** |
| Random effects | Variance | *SD* |  |
| item: (intercept) | 0.28 | 0.53 |  |  |
| item: context (slope) | 0.33 | 0.58 |  |  |

*Note.* Model formed from 1264 observations, collected from 42 participants across 16 items. Random effects included in the model were selected using a data-driven approach.



*Figure 3.* Mean animacy decision accuracy for novel words, for each context and time condition separately. Error bars represent +/-1 SE of the mean, adjusted for within-subject variance (Morey, 2008). Chance = 0.5.



*Figure 4.* Overnight change in mean accuracy ([next day – same day]) in the animacy decision task as a function of expressive vocabulary ability, for words learned in the repetitive (solid line) and multiple (dashed line) story context conditions separately. Grey bars indicate 95% confidence intervals.

**Basewords.** To analyse the effect of learning a novel word on speeded judgements to known competing words (baseword), responses to known words were also analysed for accuracy and reaction time. Word type (baseword vs. control) was added as an additional fixed effect into each of the models, to enable an assessment of competitor effects over general practice effects.

*Reaction times*. Only reaction times to correct responses were analysed, and responses below 200ms or 2.5 standard deviations above/below each participant’s condition mean were removed as outliers. A positive skew was improved using a log transformation of the reaction time data prior to entry into the model. The four-way interaction between test session, context condition, word type and vocabulary was below our liberal threshold for model inclusion (*p* = .14), and thus all fixed effects were retained in the model (Table 3). Test session was a negative predictor of reaction times overall (*p* = .03), highlighting speeded judgements with task practice (Session1: *M* = 1535.03, *SD* = 397.41; Session 2: *M* = 1500.43, *SD* = 458.37). No other predictors in the model could offer strong support for lexical competition effects, as would be indicated by a significant slowing of reaction times to basewords relative to control words, predicted to emerge at Session2. Counter to our hypotheses and the previous analyses there was only marginally significant slowing of responses to basewords relative to control words at Session 2 only weakly suggestive of lexical competition only for the repetitive contexts condition (*p* = .055; Figure 5). This change in lexical competition introduced by learning novel words in the repeated context condition was weakly (but not significantly) related to vocabulary performance (*p* = .073; Figure 6).

Table 3

*Predictors of animacy decision reaction times for known words*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed effects | $$b$$ | *SE* | $$t$$ | $$p$$ |
| (intercept) | 7.30 | 0.02 | 360.94 | <.001 |
| **test session** | **-0.03** | **0.01** | **-2.20** | **.034** |
| context | 0.01 | 0.01 | 1.80 | .072 |
| word type | -0.01 | 0.02 | -0.53 | .603 |
| vocabulary | 0.00 | 0.02 | 0.06 | .934 |
| test session x context | -0.01 | 0.02 | -0.52 | .603 |
| test session x word type | 0.02 | 0.02 | 1.58 | .113 |
| context x word type | 0.01 | 0.02 | 0.48 | .631 |
| test session x vocabulary | 0.01 | 0.01 | 0.45 | .656 |
| context x vocabulary | 0.00 | 0.01 | 0.45 | .652 |
| word type x vocabulary | 0.00 | 0.01 | 0.45 | .651 |
| test session x context x word type | -0.06 | 0.03 | -1.92 | .055 |
| test session x context x vocabulary | 0.01 | 0.02 | 0.77 | .441 |
| test session x word type x vocabulary | 0.01 | 0.02 | 0.54 | .589 |
| context x word type x vocabulary | 0.03 | 0.02 | 1.80 | .073 |
| test session x context x word type x vocabulary | -0.05 | 0.03 | -1.48 | .139 |
| Random effects | Variance | *SD* |  |
| participant: (intercept) | 0.01 | 0.10 |  |  |
| participant: test session (slope) | 0.003 | 0.06 |  |  |
| participant: word type (slope) | <0.001 | 0.02 |  |  |
| item: (intercept) | 0.004 | 0.07 |  |  |

*Note.* Model formed from 2315 observations, collected from 42 participants across 32 items. Random effects included in the model were selected using a data-driven approach. Reaction times underwent a log transformation before being entered into the model.



*Figure 5.* Animacy decision times as an indicator of lexical competition ([competitor ms – control ms]) resulting from novel words learned in each condition. A positive value indicates slowed decision times linked to the introduction of a novel word competitor. Error bars represent +/-1 SE of the mean, adjusted for within-subject variance (Morey, 2008).



*Figure 6.* Overnight change in lexical competition ([competitor RT – control RT]) in the animacy decision task, resulting from novel words learned in each context condition (repetitive: solid line; multiple: dashed line), as a function of expressive vocabulary ability. Positive values indicate slowed decision times linked to the introduction of a novel word competitor. Grey bars indicate 95% confidence intervals.

*Accuracy.* In fitting the model, the four-way interaction between session, context condition, word type and vocabulary was not significant (*p* = .42), justifying its removal from the model. We continued to prune all interactions that included the effect of session (*p*s > .3), and session as a predictor itself. In doing so, we encountered no reduction in model fit (*p* = .94). The final model is presented in Table 4. As one would expect for known words, none of the variables significantly predicted accurate decision performance. The three-way interaction between context, word type and vocabulary approached significance (*p* = .07), but such effects were very small with high variability across participants (see Supplementary Material for Figures).

Table 4

*Predictors of animacy decision accuracy for known words*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed effects | $$b$$ | *SE* | $$z$$ | $$p$$ |
| (intercept) | 2.97 | 0.21 | 14.10 | 0.000 |
| context | -0.16 | 0.16 | -1.03 | 0.304 |
| word type | -0.21 | 0.34 | -0.60 | 0.546 |
| vocabulary | 0.04 | 0.14 | 0.27 | 0.786 |
| context x word type | -0.26 | 0.31 | -0.83 | 0.409 |
| context x vocabulary | 0.00 | 0.16 | -0.01 | 0.994 |
| word type x vocabulary | -0.07 | 0.16 | -0.46 | 0.646 |
| context x word type x vocabulary | 0.56 | 0.31 | 1.79 | 0.073 |
| Random effects | Variance | *SD* |  |
| participant: (intercept) | 0.46 | 0.68 |  |  |
| item: (intercept) | 0.70 | 0.84 |  |  |

*Note.* Model formed from 2527 observations, collected from 42 participants across 32 items. Random effects included in the model were selected using a data-driven approach.

**Discussion**

This study examined how children aged 10-11 years acquire and consolidate new vocabulary from passively listening to fictional stories. A particular focus was to uncover optimal conditions for vocabulary consolidation, with an emphasis on factors both intrinsic (i.e., existing vocabulary knowledge) and extrinsic (i.e., story repetition) to the child. Consistent with evidence that off-line consolidation is important for supporting the acquisition of new vocabulary (James et al., 2017), children showed overnight improvements in their ability to recall the novel phonological forms. In keeping with Wilkinson & Houston-Price (2013), there was no overall word learning advantage when words were encountered in two different stories (multiple contexts). However, consistent with our hypotheses and with the predictions of Horst (2013), children with better existing vocabulary knowledge showed superior phonological recall, as well as larger overnight gains, when encountering words in multiple contexts. Vocabulary exerted no impact on the overnight consolidation of explicit phonological knowledge of novel words learned via listening to the same story twice (repetitive contexts) (Figure 2). A similar interaction was observed for children’s ability to make accurate judgments about the animacy of the novel words, although levels of semantic learning were very low overall. There was no strong evidence of lexical integration; weak lexical competition effects were only observed at the 24 hour test for words learned via repeated (but not multiple) stories. Together, these data suggest that existing vocabulary knowledge supports the consolidation of explicit novel word knowledge encountered in stories, particularly when information is acquired across different contexts.

*Phonological and semantic consolidation following story exposure*

Previous studies converge on the view that off-line consolidation (particularly during sleep) is crucial for strengthening memory of new word forms and embedding them into the developing lexicon (Henderson et al., 2012; 2015; James et al., 2017; Smith et al., under revision). Consistent with this, the 10-11-year-olds in this study showed significant overnight improvements in phonological recall 24 hours after hearing novel words in stories. This pattern concurs with a CLS account of vocabulary learning (e.g., Davis & Gaskell, 2009) and systems consolidation theories (e.g., Diekelmann & Born, 2010; Rasch & Born, 2013). Such models claim that, during sleep, newly encoded information is transferred from temporary memory stores in hippocampal structures into more robust long term memory stores that are distributed across the neocortex. Similar to Henderson et al., (2015), the present data suggest that the same protracted time course of word learning occurs when words are learned incidentally via stories.

A drawback of the present design is that we did not control for the influence of repeat testing: Children were tested with the same materials immediately after encoding and a day later. This raises questions about whether the overnight improvements in recall were due to strengthening through retrieval or through off-line consolidation (Bauml, Holterman & Able, 2014; Fritz, Morris, Nolan & Singleton, 2007; Goosens, Camp, Verkoeijen & Tabbers, 2013; Roediger & Karpicke, 2006). Whilst previous developmental studies have demonstrated that sleep-associated improvements in cued recall are not simply a product of repeat exposure (e.g., Henderson et al., 2012; Henderson, Weighall & Gaskell, 2013), such repeat testing effects should be examined in the context of word learning via story exposure. However, it is not clear why repeat testing would benefit recall for the multiple context condition and not the repetitive context condition or why this effect would interact with vocabulary knowledge.

Children’s learning and consolidation of novel words was notably weaker for semantic aspects of novel word knowledge. Upon being asked to define the novel words at the 24 hour test, the majority of children were only able to identify a single feature for a minority of items. Furthermore, children showed mean levels of performance on the animacy task (i.e., in which children decided whether the novel words e.g., “crocodol” were animals or not) that were below chance at both test points, with only a trend for overnight improvements in accuracy. This chance-level performance could have been a consequence of task difficulty, especially since correct responses to the novel words (e.g., *dolpheg = not animal)* were always in opposition to correct responses to basewords (e.g., *dolphin = animal*). Alternatively, it is plausible that overnight consolidation of novel word meanings may only occur when richer (or explicit) information is provided about word meaning. For instance, Henderson, Weighall & Gaskell (2013) trained unfamiliar science words via explicit definition instruction and reported overnight improvements in children’s definitions of the words. In contrast, Weighall et al (2016) trained novel word-object pairs via a two alternative forced choice procedure (without explicit instruction on meaning) and did not report an overnight advantage for novel words learned the previous day on a novel word-picture matching task. These possibilities need testing directly in future studies.

In addition to measuring how children acquire and consolidate explicit semantic and phonological knowledge following exposure to novel words in stories, we also set out to examine how these novel words become integrated with existing knowledge and compete for recognition with known words. When pooling across the multiple and repetitive context conditions, there was no strong support for the emergence of lexical competition 24 hours after the novel words were encountered, counter to numerous studies with children of a similar age range (see James et al., 2017, for a review). These results seemingly conflict with Henderson et al (2015) in which children learned novel words from spoken stories and showed clear lexical competition effects on the pause detection task after 24 hours, despite similarly low levels of novel word recall for both phonological and semantic tasks. However, there a number of reasons why lexical competition effects might have been weaker here. Henderson et al (2015) included an explicit exposure to each of the novel words in isolation, prior to presenting the stories which may have supported overnight lexical integration. Alternatively, the demands of the animacy task may have been greater than the demands of the pause detection task which may have contaminated the effects. Making a semantic decision is arguably more conceptually demanding than judging whether a pause is present or absent, and as discussed above, the animacy of the novel words was always in conflict with the animacy of the basewords. It is also possible that the animacy task was under-powered given the relatively small number of items contributing to the analysis.

*The effect of context and prior knowledge on consolidation*

 Overall, hearing new vocabulary in varying contexts (versus hearing the same story twice) did not appear to confer a general advantage for explicit aspects of word learning and/or consolidation in children aged 10-11 years (similar to Wilkinson & Houston-Price, 2013, who studied 6-9 year-olds). However, there were three-way interactions in the phonological recall and novel word animacy (accuracy) data. Whilst children, regardless of vocabulary level, were able to consolidate new explicit phonological and semantic information from repetitive stories, consolidation following multiple stories was associated with vocabulary knowledge, such that children with better vocabulary knowledge showed enhanced consolidation. Thus, acquiring new explicit lexical knowledge from multiple contexts may draw more extensively on prior knowledge. Children with a richer body of existing semantic knowledge may have been better equipped to create a large network of associations between the new word and related lexical concepts across the two stories, and this network of associations may have bolstered consolidation of explicit lexical knowledge (Bartlett, 1932; Collins & Loftus, 1975; James et al., 2017; Lewis & Durrant, 2011; McClelland, 2013; Tse et al., 2007; van Kesteren, Fernandez, Norris & Hermans, 2010). According to Lewis and Durrant (2011), prior knowledge supports consolidation through facilitating connections between new and existing knowledge. Applying this to the present study, novel words encountered in different contexts may benefit from more connections, but only for children with better vocabulary knowledge who are able to capitalise on these connections.

In contrast to the finding that consolidation from multiple contexts may be prior knowledge dependent, existing vocabulary had *no* impact on the overnight consolidation of explicit phonological knowledge of novel words learned via repetitive stories. This is confirmed in Figure 2, which shows that a similar degree of consolidation occurred regardless of vocabulary ability. This implies that repetition may go some way to narrowing the gap between children with good and poor vocabulary knowledge. Lending further support to this, children with poorer vocabulary knowledge showed bigger overnight improvements when making semantic judgements towards the novel words if they learned through repetition (see Figure 4). In contrast however, children with better vocabulary knowledge showed bigger overnight improvements if they learned through multiple contexts, intuitively suggesting that too much repetition may be detrimental for individuals with greater oral language proficiency. Therefore, in keeping with the hypotheses of Horst (2013), repetition and multiple contexts may be differentially advantageous, depending on children’s underlying language ability.

Wilkinson and Houston-Price (2013) reported a somewhat similar interaction, such that older children aged 8-9 years showed similar levels of word learning on an explicit recognition measure for words learned in both repetitive and multiple context conditions, in contrast to younger 6-7 year olds who showed an advantage for learning via repetition. Similar to the present data, this suggests that younger children presumably with a more limited knowledge base may not learn so effectively through hearing new words in varying contexts. However, it is important to note that the interaction reported by Wilkinson and Houston-Price disappeared two weeks later. An important extension of the present study, therefore, will be to examine whether any benefit of learning through multiple contexts for children with good vocabulary knowledge maintains in the longer-term.

 An important question for future research is whether the importance of having good vocabulary knowledge for word learning through multiple stories diminishes when explicit instruction about word meaning is incorporated (see Bolger et al., 2008). It has been argued that explicit definition instruction during storybook readings may enhance word learning (Wilkinson & Houston-Price, 2013) and also narrow the gap between children with good and poor vocabulary knowledge (Coyne et al., 2004). Justice, Meyer & Walpole (2005) went as far to argue that word learning from stories cannot occur without elaborated exposure, particularly for children who have impoverished vocabulary (see also Dockrell et al., 2007; Marulis & Neuman, 2010).

There was no strong evidence that the novel words had been integrated with existing lexical knowledge following story encounters in the present study. However, there was some weak evidence to suggest that lexical integration might be more likely following repetitive story exposure. Namely, there was a slight slowing of responses to basewords relative to control words 24 hours after story exposure that was restricted to the repeated context condition (as reflected by a marginal Context x Session x Word Type interaction, *p* = .055). Numerically, this pattern was mirrored in the accuracy data, such that children with superior expressive vocabulary knowledge showed larger lexical competition effects but only for the repeated context condition (with this interaction only approaching significance, *p* = .073). Although these results are weak and clearly require replication, they could suggest that repetition may exert an advantage for novel word lexical integration. Lexical competition effects have been argued to rely upon distributed lexical representations that are automatically accessed on-line during spoken word recognition (Gaskell & Dumay, 2003; Mattys & Clark, 2002). Such automated phonological retrieval may benefit from more repetitive training, akin to a reactivation of the memory trace. In contrast, tasks that involve explicit retrieval of phonological or semantic knowledge may benefit more so from richer links with semantic knowledge (garnered from multiple contexts) that could be utilised as cues for retrieval, at least for children with proficient vocabulary knowledge. This explanation may also account for why evidence of lexical integration was stronger following story exposure in Henderson et al (2015), when children were additionally required to repeat the novel words aloud in isolation.

*Conclusion*

 The present data confirm that offline consolidation is necessary for the strengthening of new vocabulary acquired through listening to stories in children. Nonetheless, children showed limited ability to acquire semantic knowledge about the novel words, and evidence that the novel words had become integrated with existing knowledge was very weak. Thus, more extensive exposures and/or incorporating explicit instruction alongside story encounters may be necessary for robust long-term learning, consistent with concerns that merely exposing children to story books is not enough to boost children’s oral language development. Most notably, prior knowledge was shown to be important for acquiring new vocabulary across different contexts, perhaps because children with superior prior knowledge capitalised on richer connections between new and existing knowledge (e.g., Lewis & Durrant, 2011). In contrast, prior knowledge had no positive impact on the consolidation of explicit knowledge acquired through repetitive stories, suggesting that for children with weaker vocabulary, repetitive exposure may go some way to reducing the gap between children with good and poor vocabulary. The present study focussed only on a very narrow age range of children aged 10-11 years: It will be important for future work to examine the trajectory of how existing vocabulary influences vocabulary consolidation across the school years. Understanding this developmental trajectory will be crucial to furthering our understanding of the mechanism by which the ‘rich get richer’ with regards to vocabulary knowledge.

**References**

Ardoin, S. P., Eckert, T. L., and Cole, C. A. S. (2008). Promoting generalization of reading: a comparison of two fluency-based interventions for improving general education student’s oral reading rate. *J. Behav. Educ.*17, 237–252.

Axelsson, E. L., Williams, S. E., & Horst, J. S. (2016). The Effect of Sleep on Children's Word Retention and Generalization. *Frontiers in psychology*, *7*.

Baker, C. E. (2013). Fathers' and mothers' home literacy involvement and children's cognitive and social emotional development: Implications for family literacy programs. *Applied Developmental Science*, *17*(4), 184-197.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bartlett, F. C. (1932). Remembering: An experimental and social study. *Cambridge: Cambridge University*.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7).

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, *98*(1), 44.

Blachowicz, C. L., Fisher, P. J., Ogle, D., & Watts‐Taffe, S. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly*, *41*(4), 524-539.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, *45*(2), 122-159.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. Cognition, 97(3), B45-B54

Brown, H., Weighall, A., Henderson, L. M., & Gaskell, M. G. (2012). Enhanced recognition and recall of new words in 7-and 12-year-olds following a period of offline consolidation. *Journal of experimental child psychology*, *112*(1), 56-72.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of learning disabilities*, *44*(5), 431-443.

Clarke, P. J., Henderson, L. M., & Truelove, E. (2010). The poor comprehender profile: Understanding and supporting individuals who have difficulties extracting meaning from text. *Advances in Child Development and Behaviour*, *39*, 79-129.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli Jr, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal*, *110*(1), 1-18.

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, *12*(3), 145-162.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3773-3800.

Dockrell, J. E., Braisby, N., & Best, R. M. (2007). Children's acquisition of science terms: Simple exposure is insufficient. *Learning and Instruction*, *17*(6), 577-594.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114-126.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35-39.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory*, *13*(3), 259-262.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105-132.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive psychology*, *45*(2), 220-266.

Goossens, N. A., Camp, G., Verkoeijen, P. P., & Tabbers, H. K. (2014). The effect of retrieval practice in primary school vocabulary learning. *Applied Cognitive Psychology*, *28*(1), 135-142.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological review*, *103*(3), 518.

Farrant, B. M., & Zubrick, S. R. (2013). Parent–child book reading across early childhood and child vocabulary in the early school years: Findings from the Longitudinal Study of Australian Children. *First Language*, *33*(3), 280-293.

Ferreira, R. A., & Ellis, A. W. (2016). Effects of contextual diversity on semantic decision and reading aloud: evidence from a word learning study in English as a second language/Efectos de la diversidad contextual en la decisión semántica y la lectura en voz alta: evidencia de un estudio de aprendizaje de palabras en inglés como segundo idioma. *Estudios de Psicología*, *37*(1), 162-182.

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *The Quarterly Journal of Experimental Psychology*, **60**, 991–1004.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive psychology*, *53*(4), 310-344.

Hammer, C. S., Farkas, G., & Maczuga, S. (2010). The language and literacy development of Head Start children: A study using the Family and Child Experiences Survey database. *Language, Speech, and Hearing Services in Schools*, *41*(1), 70-83.

Henderson, L., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental psychology*, *51*(3), 406.

Henderson, L. M., Weighall, A. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental science*, *15*(5), 674-687.

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child development*, *84*(5), 1668-1685.

Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of experimental child psychology*, *116*(3), 572-592.

Hepburn, E., Egan, B., & Flynn, N. (2010). Vocabulary acquisition in young children: The role of the story. *journal of early childhood literacy*, *10*(2), 159-182.

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in psychology*, *4*, 149.

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 17.

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of sleep research*, *24*(5), 503-509.

Huang, S., Deshpande, A., Yeo, S. C., Lo, J. C., Chee, M. W., & Gooley, J. J. (2016). Sleep restriction impairs vocabulary learning when adolescents cram for exams: the Need for Sleep Study. *Sleep*, *39*(9), 1681.

James, E., Gaskell, M. G., Weighall, A., & Henderson, L. (2017). Consolidation of vocabulary during sleep: the rich get richer?. *Neuroscience & Biobehavioral Reviews*, *77*, 1-13.

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning New Words From StorybooksAn Efficacy Study With At-Risk Kindergartners. *Language, speech, and hearing services in schools*, *36*(1), 17-32.

Karweit, N., & Wasik, B. A. (1996). The effects of story reading programs on literacy and language development of disadvantaged preschoolers. *Journal of Education for Students Placed at Risk*, *1*(4), 319-348.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978-990.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. *R package version*, *2*.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive psychology*, *55*(4), 306-353.

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in cognitive sciences*, *15*(8), 343-351.

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608.

Lonigan, C. J., Shanahan, T., and Cunningham, A. (2008). “Impact of shared-reading intervention on young children’s early literacy skills,” in *Developing Early Literacy: Report of the National Early Literacy Panel* (Washington, DC: National Institute for Literacy), 153–171

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57-77.

Marslen-Wilson, W. (1989, November). Access and integration: Projecting sound onto meaning. In *Lexical representation and process* (pp. 3-24). MIT Press.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children’s word learning: A meta-analysis. *Review of educational research*, *80*(3), 300-335.

Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, *47*(3), 343-359.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

McClelland, J. L., McNaughton, B. L., & O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419.

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*(2), 979-1007.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234.

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly*, *9*(3-4), 427-440.

Penno, J. F., Wilkinson, I. A., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect?. *Journal of Educational Psychology*, *94*(1), 23.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, *11*(4), 357-383.

R Core Team (2015). R: a language and environment for statistical computing. Vienna, Austria; 2015.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological reviews*, *93*(2), 681-766.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, **1**, 181–210.

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental review*, *14*(3), 245-302.

Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child language*, *24*(01), 123-138.

Sénéchal, M., Pagan, S., Lever, R., & Ouellette, G. P. (2008). Relations among the frequency of shared reading and 4-year-old children's vocabulary, morphological and syntax comprehension, and narrative skills. *Early Education and Development*, *19*(1), 27-44.

Smith, F., Gaskell, M.G., Weighall, A.R., Warmington, M., Reid, A., & Henderson, L.M. (under revision). A different role for sleep in the consolidation of new vocabulary in children with and without dyslexia. *Developmental Science*.

Spencer, R. M., Walker, M. P., & Stickgold, R. (2017). Sleep and Memory Consolidation. In *Sleep Disorders Medicine* (pp. 205-223). Springer New York.

Staresina, B. P., Ole Bergmann, T., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., … Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*, *18*(11), 1679–1686.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading research quarterly*, 360-407.

Scarborough, H. S., Catts, H. W., & Kamhi, A. G. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. *The connections between language and reading disabilities*, 3-24.

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology*, *66*(5), 1001-1025.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, *30*(43), 14356-14360.

Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition*, *163*, 93-102.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *TOWRE: Test of word reading efficiency*. Psychological Corporation.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., ... & Morris, R. G. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76-82.

van Kesteren, M. T., Fernández, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences*, *107*(16), 7550-7555.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children’s memory and category induction. *Cognition*, *109*(1), 163-167.

Webb, S., & Chang, A. C. S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program?. *Studies in Second Language Acquisition*, *37*(04), 651-675.

Wechsler, D., & Hsiao-pin, C. (2011). *WASI-II: Wechsler abbreviated scale of intelligence*. Pearson.

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time‐course of novel word learning and lexical competition in adults and children. *Brain and language*, *167*, 13-27.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. *Springer New York*, *1*(2), 3.

Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess...: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics*, *34*(03), 591-613.

Wilhelm, I., Prehn-Kristensen, A., & Born, J. (2012). Sleep-dependent memory consolidation–what can be learnt from children?. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1718-1728.

Williams, S.E., & Horst, J.S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology, 5*, 184.

Williams, S. E., Horst, J. S., and Oakhill, J. (2011). “The same old story–contextual cueing facilitates word learning in shared storybook reading,” *Paper Presented at the 2011 British Psychological Society, Developmental Section Conference,*Newcastle, UK.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child development*, *69*(3), 848-872.