

This is a repository copy of *Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: it'll be alright on the night?*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/160724/>

Version: Published Version

Article:

James, Emma orcid.org/0000-0002-5214-0035, Gaskell, M Gareth orcid.org/0000-0001-8325-1427 and Henderson, Lisa M orcid.org/0000-0003-3635-2481 (2020) Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: it'll be alright on the night? *Journal of Child Psychology & Psychiatry*. ISSN 1469-7610

<https://doi.org/10.1111/jcpp.13253>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: it'll be alright on the night?

Emma James,  M. Gareth Gaskell, and Lisa M. Henderson

Department of Psychology, University of York, York, UK

Background: Vocabulary is crucial for an array of life outcomes and is frequently impaired in developmental disorders. Notably, 'poor comprehenders' (children with reading comprehension deficits but intact word reading) often have vocabulary deficits, but underlying mechanisms remain unclear. Prior research suggests intact encoding but difficulties consolidating new word knowledge. We test the hypothesis that poor comprehenders' sleep-associated vocabulary consolidation is compromised by their impoverished lexical-semantic knowledge. **Methods:** Memory for new words was tracked across wake and sleep to assess encoding and consolidation in 8-to-12-year-old good and poor comprehenders. Each child participated in two sets of sessions, one beginning in the morning (AM-encoding) and the other in the evening (PM-encoding). In each case, they were taught 12 words and were trained on a spatial memory task. Memory was assessed immediately, 12- and 24-hr later via stem-completion, picture-naming, and definition tasks to probe different aspects of word knowledge. Long-term retention was assessed 1–2 months later. **Results:** Recall of word-forms improved over sleep and postsleep wake, as measured in both stem-completion and picture-naming tasks. Counter to hypotheses, deficits for poor comprehenders were not observed in consolidation but instead were seen across measures and throughout testing, suggesting a deficit from encoding. Variability in vocabulary knowledge across the whole sample predicted sleep-associated consolidation, but only when words were learned early in the day and not when sleep followed soon after learning. **Conclusions:** Poor comprehenders showed weaker memory for new words than good comprehenders, but sleep-associated consolidation benefits were comparable between groups. Sleeping soon after learning had long-lasting benefits for memory and may be especially beneficial for children with weaker vocabulary. These results provide new insights into the breadth of poor comprehenders' vocabulary weaknesses, and ways in which learning might be better timed to remediate vocabulary difficulties. **Keywords:** Poor comprehenders; vocabulary; word learning; sleep; memory consolidation.

Introduction

Good vocabulary knowledge is a key contributor to comprehension success (Perfetti, 2007) and – in turn – successful comprehension permits the acquisition of new word knowledge (Verhoeven, van Leeuwe, & Vermeer, 2011). Yet even in the context of explicit vocabulary instruction, there lies considerable variability in the ease with which children learn new vocabulary, with vocabulary deficits being a prominent and cross-cutting characteristic of developmental disorders (Ricketts, 2011). To understand individual differences in vocabulary acquisition, we must consider both how to successfully encode a new word representation in memory, and the factors that enable consolidation of this initial representation into longer-term vocabulary. Understanding variability in both processes is critical for better targeting robust and long-lasting vocabulary instruction. One possible source of variation is in children's existing semantic knowledge, proposed to bolster the consolidation of new words (James, Gaskell, Weighall, & Henderson, 2017). In the present study, we sought to understand these processes by comparing the learning and consolidation of new

spoken vocabulary in children with good versus poor reading comprehension, who typically differ in lexical-semantic knowledge.

Vocabulary ability of poor comprehenders

Children with specific reading comprehension difficulties can be classified under DSM-5 as having 'Specific Learning Disorder with impairment in reading' (American Psychiatric Association, 2013). More commonly described as 'poor comprehenders', these children have at least age-appropriate phonological and reading accuracy skills, but show relative weaknesses in accessing meaning from language (Nation & Snowling, 1998). An estimated ~5% of children show such difficulties (Nation, 2019), and these comprehension problems frequently co-occur with poor oral language skills (Catts, Adlof, & Weismer, 2006). Although there are many putative causes of poor comprehension, a wealth of evidence points to weaker performance on standardised tests of vocabulary in poor comprehenders than typically developing peers, with this performance gap widening across the school years (Cain & Oakhill, 2011). Studies of lexical processing highlight specific weaknesses in lexical-semantic rather than phonological components of word knowledge for this group (Landi

Conflict of interest statement: No conflicts declared.

& Ryherd, 2017), and intervention studies support a causal role for vocabulary weaknesses in reading comprehension difficulties (Clarke, Snowling, Trulove, & Hulme, 2010). Poor comprehenders' vocabulary weaknesses are often apparent in receptive vocabulary tasks that capture breadth of word knowledge (Cain & Oakhill, 2011). This deficit is perhaps more consistently observed for tasks that require expression of vocabulary knowledge (Ricketts, Sperring, & Nation, 2014), which are more strongly predictive of reading comprehension than receptive measures (Ouellette, 2006).

Other studies have explored the mechanisms underlying poor comprehenders' vocabulary acquisition weaknesses. In line with relative deficits in semantic processing, Nation, Snowling, and Clarke (2007) found that poor comprehenders showed weaker expressive recall of new word meanings – but not new word-forms – than reading accuracy-matched control children when tested immediately after training. Interestingly though, even word-form knowledge was not retained over time, with poor comprehenders recalling fewer words than control children one week later. A similar pattern was found by Ricketts, Bishop, and Nation (2008), suggesting that poor comprehenders may have weaknesses in consolidating new lexical knowledge into long-term memory.

Models of lexical consolidation

One novel theoretical account of poor comprehenders' retention weaknesses is that their poor lexical-semantic knowledge constrains consolidation of new words. This account is embedded in the Complementary Learning Systems (CLS) account of new word acquisition (Davis & Gaskell, 2009). According to this model, two neural systems are engaged in the process of acquiring new vocabulary: the hippocampal system supports an initial representation, whilst a neocortex-based system slowly integrates the new word into existing vocabulary knowledge. The CLS account proposes that this slower learning can happen as the hippocampus replays memory traces to the neocortex, gradually reducing hippocampal involvement in retrieving new words via systems consolidation. This replay can occur 'off-line', during sleep, facilitating overnight improvements in word knowledge (Henderson, Weighall, Brown, & Gaskell, 2012) that can be predicted by neural activity during sleep (Smith et al., 2018). However, only recently have researchers begun to consider the factors that might influence the consolidation process (Stickgold & Walker, 2013).

One factor proposed to support the consolidation of new word-forms is the abundance of associated semantic information, which allows for an enriched lexical representation with many potential connections to existing knowledge (James et al., 2017). For example, Henderson, Weighall, and Gaskell (2013) showed that children who were taught meanings of new words outperformed a group taught only word-

forms when tested on a form-recall task one week after training – but not within 24-hr of learning. This late-emerging difference is strikingly similar to the pattern of weaknesses seen in poor comprehenders. More recently, Henderson and James (2018) also showed that an abundance of semantic knowledge was only beneficial for children with more extensive vocabulary knowledge to capitalise upon. This highlights that the variability of semantic support in consolidating new vocabulary can also come from the learner as well as the learning environment (James et al., 2017), which may be key to understanding word learning differences in poor comprehenders.

A lexical consolidation deficit?

In this study, we examined whether poor comprehenders have specific difficulties in consolidating new words, as would be predicted on account of their more limited lexical-semantic knowledge. Indeed, two other studies have produced findings broadly consistent with this hypothesis. Henderson, Snowling, and Clarke (2013) found that poor comprehenders had explicit knowledge of less frequent homonym meanings (e.g. *bank-river* vs. *bank-money*) but did not access them in speeded semantic tasks, suggesting they were not well-integrated into the neocortical vocabulary system. Furthermore, a neuroimaging study by Cutting et al. (2013) found that adolescent poor comprehenders showed abnormal hippocampal engagement during a simple lexical decision task. One explanation for this finding was that poor comprehenders have difficulty with consolidating word representations into cortical structures. We take the first step in examining this hypothesis using a behavioural experiment of learning and sleep-associated consolidation processes in good versus poor comprehenders.

We taught children new spoken words in the morning or the evening and tested their memory immediately, 12- and 24-hr later, enabling us to isolate memory changes in relation to sleep-associated consolidation processes. Three tasks were designed to probe different aspects of word knowledge: a stem-completion task to assess memory of the new forms, a picture-naming task to assess the form-meaning mapping and a definitions task to probe the richness of newly acquired semantic knowledge. These tasks enabled us to test the preregistered hypotheses that poor comprehenders would show poorer semantic learning than good comprehenders – in keeping with their anticipated weaknesses with expressive vocabulary – but that their relative impairments would broaden to other aspects of word knowledge after a period of sleep-associated consolidation (<https://osf.io/4frxd>). A declarative spatial memory task also provided a test of the hypothesis that any weaknesses were specific to linguistic information. More broadly, this study contributes to a growing literature on the importance of sleep for learning in development.

Methods

Participants

Fifteen poor and 15 good comprehenders participated, meeting the following criteria: 8–12 years old; native English speakers; no reported learning, neurological, or sleep disorders; reading accuracy score ≥ 95 on the Phonemic Decoding Efficiency subtest of the Test of Word Reading Efficiency (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012). Poor comprehenders had a reading comprehension score on the York Assessment for Reading Comprehension (YARC; Snowling et al., 2009; Stothard, Hulme, Clarke, Barnby, & Snowling, 2010) that was < 100 and ≥ 10 standard score points below their reading accuracy. Good comprehenders had a reading comprehension score > 100 , and at least as good as their accuracy score (see Table 1 and full recruitment details in Appendix S1). Parents gave informed consent, and the study was approved by the University of York Psychology Ethics Committee. Children received a gift voucher for participating.

Design and procedure

Each child took part in two sets of sessions, separated by at least one week (median = 7.4; range 6.4–21.4 days) of sleep monitoring (Motionlogger Actigraph; Ambulatory Monitoring, Inc.)¹. Each set represented one of two within-subjects encoding-time conditions (AM-encoding, PM-encoding), the order of which was counterbalanced across participants (with no difference in training performance, $p = .15$). For the AM-encoding condition, the child completed the initial encoding session (~45 min) as early as possible in the morning (median = 08:56, range: 08:35–10:09); for the PM-encoding condition, the session was completed as close as possible to their bedtime (median = 19:34, range: 17:55–21:25). For each set, memory tasks were administered immediately, ~12- and ~24-hr later (Figure 1). The morning sessions typically took place in school, whereas the evening sessions typically took place in the child's home. All tasks were presented via headphones to reduce issues of noise and avoid parental engagement, and the test environment was closely monitored by a single researcher.

A delayed follow-up session for all memory tasks was administered 1–2 months after the second set of sessions. Although scheduling issues resulted in substantial variability in delay (4.09–10.77 weeks), the difference in delay was not statistically significant between comprehension groups.

Length of delay was not associated with change in performance for any task.

Word stimuli

We created two lists of 12 living things that were unlikely to be known to the children. Each list containing three exemplars from four different categories (e.g. three birds, three trees, etc.), designed to promote in-depth semantic learning. The lists were matched on syllable number, phoneme length and biphone probability (CLEARPOND, Marian, Bartolotti, Chabal, & Shook, 2012).

Illustrations of each item were sourced using a Web-based search and presented on a white background during training. We also created three sets of photographs (matched on rated similarity to the training illustration) for the picture-naming task, enabling a different photograph to be named at each test point. The order of the three lists was counterbalanced across participants, and a fourth separate list used for the follow-up.

Word exposure phase

Learning and test tasks were run on a laptop using OpenSesame (v.3.1.9; Mathôt, Schreij, & Theeuwes, 2012), with a headset for audio presentation and vocal response recording. Participants heard each new spoken word 19 times (13 alongside its image) across five training tasks, administered in the order below. Item order was randomised within training and test tasks.

Familiarity check. Children heard each word and were asked whether they had heard it before. Eight children provided a relevant definition for one item from either ($n = 7$)/both ($n = 1$) lists, and the corresponding observations were removed from analysis.

Form-repetition. Children heard each word and repeated it aloud.

Picture-naming. As form-repetition, but with the illustration presented (two rounds).

Multiple-choice tasks. Children were asked to select which of two pictures matched a spoken word (rounds 1, 3), or which of two spoken words matched a picture (rounds 2, 4).

Table 1 Selection and background measures summarised by comprehension group

| | Poor comprehenders (7m, 8f) | | Good comprehenders (8m, 7f) | | <i>t</i> | <i>p</i> |
|---|--------------------------------|-----------|--------------------------------|-----------|----------|----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| Age (years; months) | 10; 04 | 1; 07 | 11; 0 | 1; 09 | 1.13 | .269 |
| TOWRE – Sight word ^a | 102.20 | 10.37 | 106.40 | 8.77 | 1.20 | .242 |
| TOWRE – Phonemic decoding ^a | 102.87 | 26.16 | 107.60 | 6.65 | 0.68 | .507 |
| YARC Accuracy ^{a,b} | 104.82 | 9.53 | 113.17 | 4.88 | 2.39 | .030 |
| YARC Rate ^a | 106.07 | 13.28 | 113.13 | 7.12 | 1.82 | .083 |
| YARC Comprehension ^a | 92.67 | 5.70 | 114.13 | 5.37 | 10.61 | <.001 |
| WASI Matrix Reasoning ^c | 47.33 | 9.71 | 53.20 | 8.23 | 1.79 | .085 |
| WASI Vocabulary ^{c,d} | 49.93 | 10.15 | 61.43 | 5.56 | 3.82 | <.001 |
| Total sleep time (AM-encoding) ^e | 481.79 | 85.30 | 461.07 | 40.86 | 0.82 | .423 |
| Total sleep time (PM-encoding) ^e | 499.62 | 44.31 | 458.62 | 72.12 | 1.75 | .096 |

^aStandardised score ($M = 100$, $SD = 15$).

^bOnly relevant for Primary edition, data from 6 GCs and 11 PCs only.

^cT-score ($M = 50$, $SD = 10$).

^dData missing from one GC (time constraints).

^eMinutes sleep for the night within the test set, as recorded by the Actigraph.

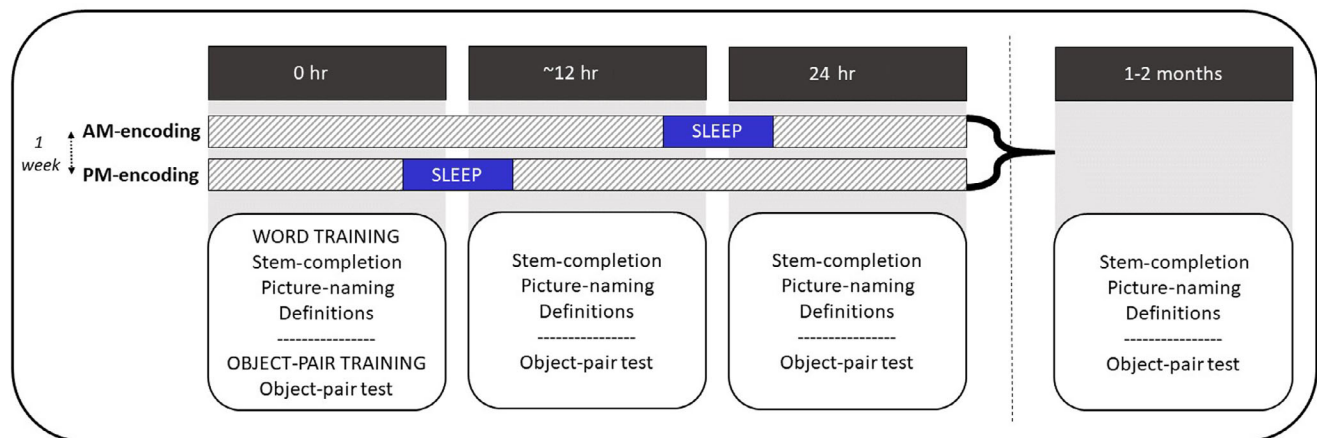


Figure 1 Schematic of experimental design. All participants completed both AM- and PM-encoding conditions, with the order of administration counterbalanced across participants

The incorrect option was a trained item from either a different (rounds 1–2) or the same (rounds 3–4) semantic category. Feedback provided the correct response.

Delayed picture-naming. Children heard each word and were instructed to think of the picture. The correct picture appeared after 2.5 s, and children repeated the word-form aloud.

Word test phase

Children rated their sleepiness (1–10) at the start of each test session and completed test tasks in the order below. These tests all required production of new word knowledge for two reasons: first, expressive vocabulary knowledge appears most consistently impaired in poor comprehenders (Ricketts et al., 2014); and second, tasks that require explicit recall (vs. recognition) of new knowledge are more sensitive to sleep-associated improvements (Diekelmann, Wilhelm, & Born, 2009), and thus provide a good behavioural measure of consolidation. There were two sessions of missing data (one technical failure, one absence).

Stem-completion. To assess word-form memory, children heard the first consonant and vowel of each word and were asked to say the full word (e.g. *ko-*, *komondor*). Each response was voice-recorded and scored off-line for accuracy using CheckVocal (Protopapas, 2007), blind to encoding condition.

Picture-naming. To assess memory for the form-meaning mapping, children named a previously unseen photograph (first round) or training image (second round) aloud as quickly as possible. Voice recordings were scored for accuracy and response time (RT).

Definitions. To probe explicit semantic knowledge, children were asked to tell the experimenter about each item. Responses were transcribed and scored by an independent scorer (blind to conditions) for semantic category (e.g. *tree*) and distinctive feature (e.g. *rainbow bark*; maximum two points/item). Where only one of these was provided, or the feature was generic to more than one item, the experimenter probed once for further information.

Object-location task

A separate task was used to assess declarative memory for spatial locations, which did not place demands on verbal learning. We created two versions of the object-location task

from Henderson et al. (2012). In each, ten object-pairs were presented across two locations on a 4 × 5 grid, and children had to remember the locations of each pair. The stimuli were colour illustrations of easily nameable animals/objects, each with monosyllabic high-frequency names (e.g. *drum*, *sheep*).

Learning phase. In the first block, children viewed each of the 10 pairs on the grid. For each pair, the first picture emerged at a grid location, followed by its matching picture 1,000 ms later. Both pictures remained for 3,000 ms, before a 3,000 ms interval. A second learning block involved testing with feedback: one object appeared at its grid location, and the child clicked on the square where they thought the matching picture was. A sound played to indicate accuracy, and the correct pair location was displayed for 1,000 ms (followed by 1,000 ms interval).

Test phase. As the second learning block, except without feedback. A sound played to register their response, and the next trial started after 2,000 ms.

Analyses

We used *lme4* (Bates, Maechler, Bolker, & Walker, 2015) and *ordinal* (Christensen, 2015) to fit mixed-effects models for each dependent variable. For the main analyses, we entered group (poor/good comprehenders), encoding-time (AM/PM) and test session (0-/12-/ 24-hr) as fixed effects, alongside all interactions. The three-level factor of test session was coded to contrast 0–12 hr and 12–24 hr tests, enabling direct interpretation of interactions with encoding-time; separate models contrasted the 0-hr and delayed follow-up scores. For the picture-naming task, a fixed effect of picture-type (novel/trained) also revealed a consistent benefit for trained items, but is of limited theoretical interest in the absence of further interactions (as trained items always reflected a second retrieval attempt). As such, these effects are not reported in the main text.

We pruned higher-order interactions that did not contribute to model fit ($p > .2$) to enable a more parsimonious model, and incorporated random slopes using the same criteria. We report only significant predictors in the text; full model tables are presented in the Supporting Information materials. Data and analysis scripts are available at <https://osf.io/nyat5>.

Results

We first test the hypothesis that poor comprehenders show weak semantic learning – as has been found in

previous studies – before examining consolidation-based deficits across broader aspects of word knowledge.

Definitions

Supporting the hypothesis that poor comprehenders would show weak semantic learning, poor comprehenders averaged significantly lower definition scores per item ($M = 0.98$, $SD = 0.83$) than good comprehenders ($M = 1.29$, $SD = 0.79$; $\beta = .48$, $SE = 0.2$, $Z = 2.42$, $p = .016$). This effect was stable across test sessions, regardless of encoding-time (Table S1; Figure S1).

The group difference remained at the delayed follow-up test ($\beta = .45$, $SE = 0.21$, $Z = 2.11$, $p = .034$; poor comprehenders $M = 0.78$, $SD = 0.84$, good comprehenders $M = 1.02$, $SD = 0.88$). There was a decline in performance between initial training ($M = 1.09$, $SD = 0.82$) and the follow-up ($M = 0.90$, $SD = 0.87$; $\beta = -.30$, $SE = 0.10$, $Z = -2.85$, $p = .004$), which did not differ between groups (Table S2).

Picture-naming

Accuracy. Poor comprehenders were less accurate than good comprehenders ($\beta = .56$, $SE = .25$, $Z = 2.20$, $p = .028$; Figure 2). Performance improved across test sessions (0–12 hr: $\beta = .48$, $SE = 0.11$, $Z = 4.35$, $p < .001$; 12–24 hr: $\beta = .49$, $SE = 0.10$, $Z = 4.71$, $p < .001$) and, importantly, interacted with encoding-time during the 0–12-hr period. Consistent with the hypothesis that sleep is beneficial for off-line consolidation, there was a larger improvement between the first and second test for PM-encoded items that featured sleep than for AM-encoded items

that featured a day awake ($\beta = .62$, $SE = 0.11$, $Z = 5.78$, $p < .001$). No higher-order interactions were significant (Table S3).

The group difference remained at the follow-up test (Table S4; $\beta = .56$, $SE = 0.24$, $Z = 2.30$, $p = .021$) with no significant change in accuracy across the delay. There was a significant interaction between encoding-time and test session ($\beta = .24$, $SE = 0.06$, $Z = 4.23$, $p < .001$) that suggested a long-term benefit for learning closer to sleep: performance improved from PM-encoding to the delayed test, whereas there was a decline in performance from AM-encoding to the delayed test.

RT. We analysed only accurate responses and removed trials with prolonged onsets or vocalisations indicating earlier retrieval ($n = 56$); skewed residuals were remediated by a Box-Cox transform (Table S3; raw scores reported for interpretation). RTs decreased across all three test sessions (0–12 hr: $\beta = -.80$, $SE = 0.17$, $t = -4.82$, $p < .001$; 12–24 hr: $\beta = -.74$, $SE = 0.15$, $t = -5.09$, $p < .001$) and interacted with encoding-time (Figure 3): there was a greater reduction between 0 and 12 hr following PM-encoding than AM-encoding ($\beta = -.66$, $SE = 0.16$, $t = -4.10$, $p < .001$), and vice versa for 12–24 hr ($\beta = .46$, $SE = 0.14$, $t = 3.19$, $p = .001$). Across both encoding conditions therefore, periods of sleep always facilitated retrieval ($M = -477$ ms) more than periods awake ($M = 40$ ms); leaving PM-encoded items recalled slightly more quickly overall ($\beta = -.21$, $SE = 0.10$, $t = -2.11$, $p = .047$). Good and poor comprehenders did not differ.

There was weak statistical evidence for a decline in RTs from the 0-hr to the follow-up ($\beta = -.37$, $SE = 0.19$, $t = -1.92$, $p = .070$), in the context of an interaction with comprehension group ($\beta = .31$,

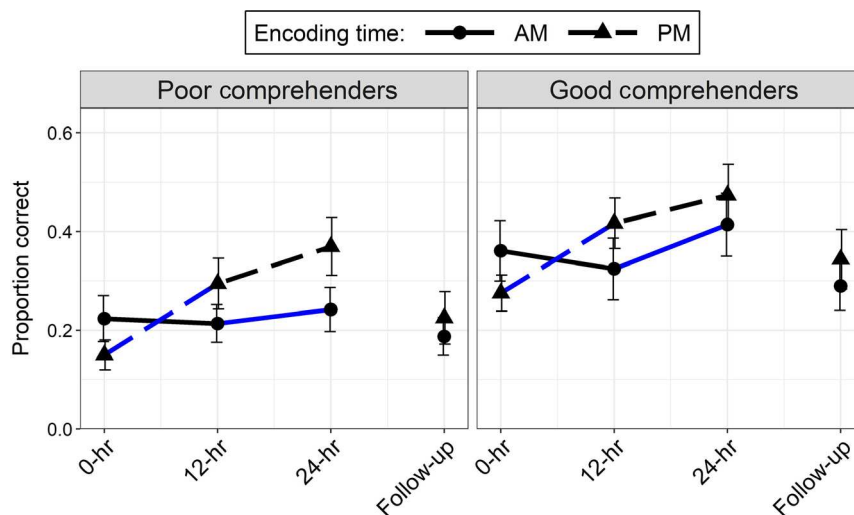


Figure 2 Mean picture-naming accuracy (averaged across picture types) at each test following AM/PM encoding separately, for each comprehension group. Blue lines highlight changes in performance associated with sleep; error bars mark standard error

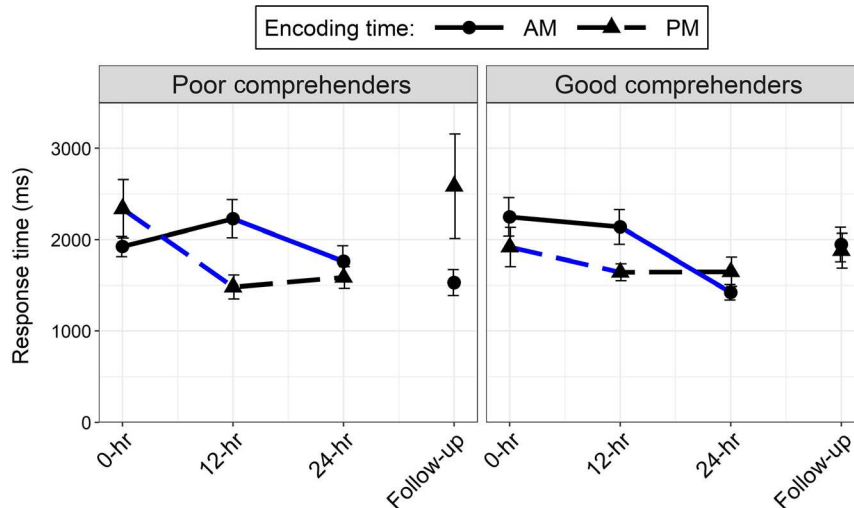


Figure 3 Mean picture-naming response times (averaged across picture types) at each test following AM/PM encoding separately, for each comprehension group. Blue lines highlight changes in performance associated with sleep; error bars mark standard error

$SE = 0.13$, $t = 2.43$, $p = .015$). In contrast to hypotheses, poor comprehenders showed *bigger* reductions in RTs ($M = -255$ ms) than good comprehenders ($M = 42$ ms). There was also a significant interaction between group and encoding-time ($\beta = -.40$, $SE = 0.13$, $t = -3.01$, $p = .003$), with poor comprehenders faster to respond in the AM- versus PM-encoding condition, and the opposite trend for good comprehenders. However, note that poor comprehenders contributed fewer trials to these analyses (due to their lower accuracy), and so estimates may be less reliable.

Stem-completion

As with other tasks, recall of word-forms improved across test sessions (0–12 hr: $\beta = .29$, $SE = 0.14$, $Z = 2.07$, $p = .038$; 12–24 hr: $\beta = .56$, $SE = 0.14$, $Z = 4.06$, $p < .001$) and interacted with encoding-time

for the 0–12 hr tests ($\beta = .73$, $SE = 0.14$, $Z = 5.11$, $p < .001$): PM-encoded items improved more between the first two sessions than AM-encoded items (Figure 4). The data did not support the hypothesis that poor comprehenders would show broadening impairments with consolidation on this task: poor comprehenders showed weaker recall than good comprehenders ($\beta = .47$, $SE = 0.22$, $Z = 2.17$, $p = .030$), but there were no further interactions (Table S5).

At the follow-up test, there remained an overall group difference in recall ($\beta = .39$, $SE = .16$, $Z = 2.38$, $p = .017$), but there was no significant change in performance over time (Table S6). There was again an interaction between encoding-time and test session ($\beta = .19$, $SE = 0.07$, $Z = 2.78$, $p = .005$): stem-completion performance was poorer following PM-encoding but improved by the follow-up, whereas the higher performance following AM-encoding showed a slight decline by the follow-up.

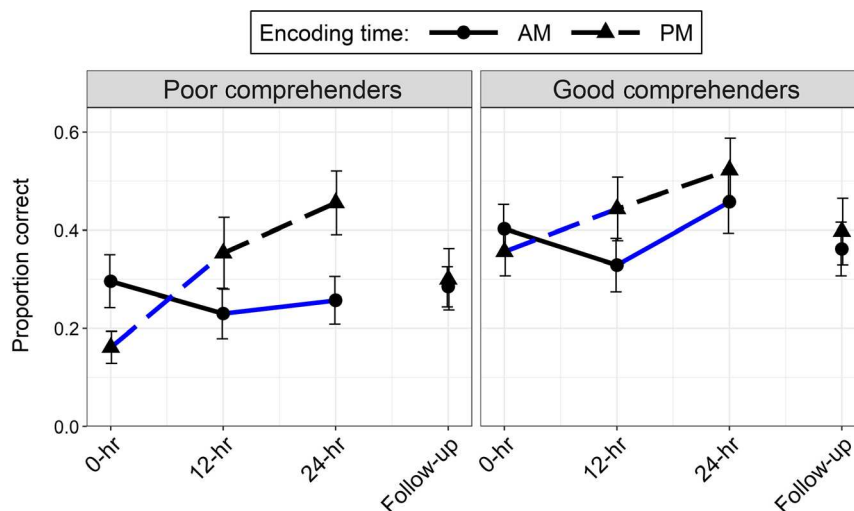


Figure 4 Mean stem-completion accuracy at each test following AM/PM encoding separately, for each comprehension group. Blue lines highlight changes in performance associated with sleep; error bars mark standard error

Object-location task

In contrast to hypotheses that poor comprehenders' difficulties would be language-specific, poor comprehenders also performed more poorly on the object-location task than good comprehenders ($\beta = .30$, $SE = 0.11$, $Z = 2.67$, $p = .008$; Figure 5). There was a general deterioration in performance between 0 and 12 hr ($\beta = -1.61$, $SE = 0.14$, $Z = -11.16$, $p < .001$), which interacted with encoding-time ($\beta = .31$, $SE = 0.14$, $Z = 2.16$, $p = .031$): there was a smaller decline following PM-encoding that featured sleep between the 0- and 12-hr tests than there was following AM-encoding. However, there was no change in performance between 12 and 24 hr nor an interaction with encoding-time, suggesting no further benefits for postsleep wake or for sleep to recover information lost from morning (Table S7).

Although participants showed a steep decline in performance by the follow-up ($\beta = -1.44$, $SE = 0.15$, $Z = -9.45$, $p < .001$), the comprehension group difference was maintained ($\beta = .24$, $SE = 0.11$, $Z = 2.28$, $p = .023$; Table S8). However, there also emerged a three-way interaction between group, encoding-time and test session ($\beta = -.27$, $SE = 0.08$, $Z = -3.32$, $p < .001$): poor comprehenders were poorer at learning in the evening but declined less by the follow-up than when they learned in the morning. Good comprehenders did not show such large immediate differences between AM-encoding and PM-encoding, with both declining similarly by the follow-up.

Exploring individual differences in vocabulary knowledge as a predictor of consolidation

The group contrasts were one way of examining the hypothesis that weak semantic knowledge may constrain later consolidation of new word-forms, in line

with previous studies that had indicated a retention deficit for poor comprehenders. However, our poor comprehenders did not have as weak comprehension skills as previous samples, and there was substantial overlap in the two groups' standardised vocabulary scores (good comprehenders: 48–70; poor comprehenders: 36–76). Given we proposed weaknesses in lexical-semantic knowledge to be the most influential in poor comprehenders' consolidation difficulties, we additionally analysed whether expressive vocabulary scores might better predict differences in consolidating new word-form knowledge. We focused on stem-completion to maximise comparability with previous consolidation studies (Henderson, Devine, Weighall, & Gaskell, 2015).

One child was missing a vocabulary score, but four additional children were included who had not met our comprehension group criteria (total $n = 33$). We entered vocabulary score as a fixed-effect alongside encoding-time (AM/PM), test session (0-/12-/24-hr) and all interactions. Vocabulary ability was a highly significant predictor of overall performance ($\beta = .66$, $SE = 0.15$, $Z = 4.48$, $p < .001$). Most interestingly, there was a three-way interaction between vocabulary ability, encoding-time and 12–24 test session ($\beta = -.34$, $SE = 0.13$, $Z = -2.57$, $p = .010$). Children's prior vocabulary knowledge better predicted improvements in recall over sleep (AM-encoded) than wake (PM-encoded) during this 12–24 hr period (Figure 6). Although in a similar direction for the relative sleep and wake comparisons, there was no statistical evidence for a similar interaction with vocabulary ability across the 0–12-hr sessions ($p = .82$; Table S9).

Discussion

This study sought to understand the impact that semantic knowledge has on the learner's ability to

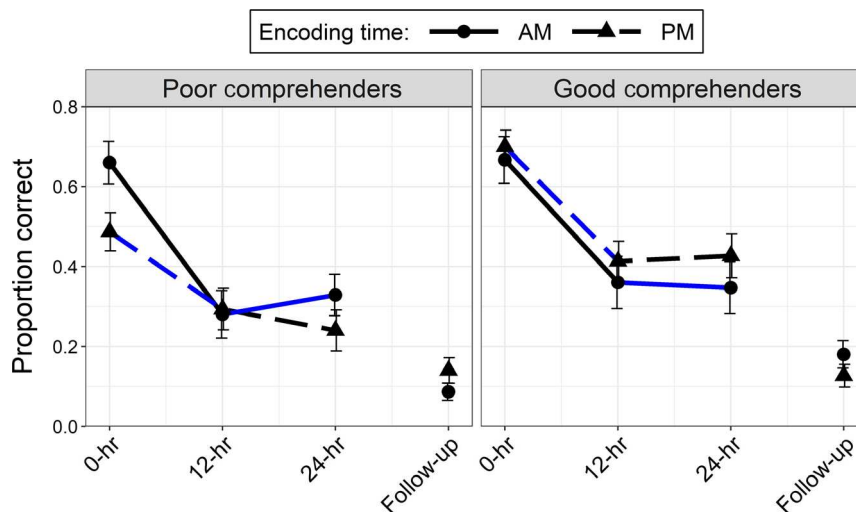


Figure 5 Mean object-pair accuracy at each test following AM/PM encoding separately, for each comprehension group. Blue lines highlight changes in performance associated with sleep; error bars mark standard error

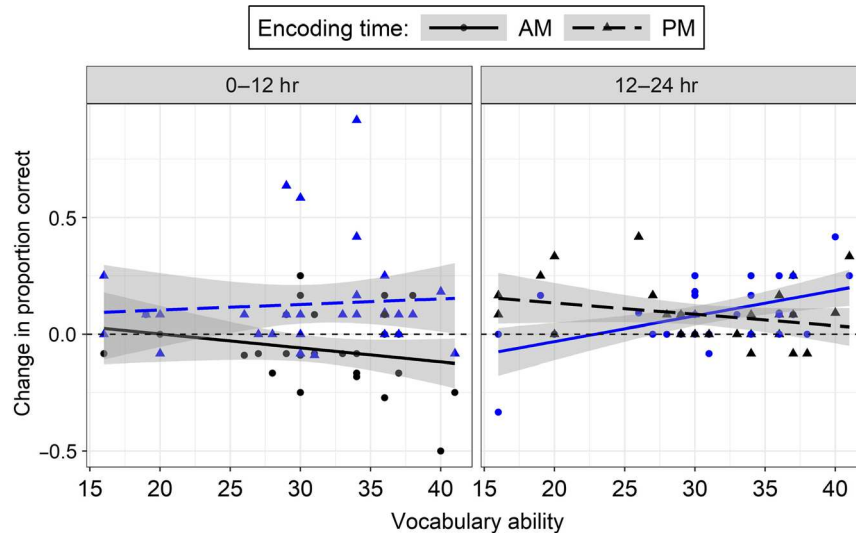


Figure 6 Change in stem-completion accuracy across 12-hr periods for each of the AM/PM-encoding conditions, plotted against participants' vocabulary scores. The zero line indicates no change in performance; sleep intervals are marked in blue to aid interpretation. For the 0–12 hr contrast (featuring sleep for PM-encoded items), there were no interactions with vocabulary ability. For the 12–24 hr contrast (featuring sleep for AM-encoded items), vocabulary is differently associated with change across the two encoding conditions

encode, consolidate and retrieve new vocabulary, by carrying out the first investigation of whether poor comprehenders show encoding or sleep-associated consolidation problems. Poor comprehenders were relatively impaired on tasks assessing memory for new vocabulary compared to good comprehenders and – in contrast to previous studies – this weakness was general to semantic *and* form-based aspects, and extended to object-location memory. Strikingly, there was no indication that poor comprehenders had weaknesses in consolidation: their relative impairments were apparent immediately after encoding and were not exacerbated by periods of sleep, nor a 1-to-2-month delay. On the contrary, there were clear sleep-associated benefits for performance across *both* comprehension groups, and these were long-lasting when sleep could occur soon after learning. When a day of wake intervened before opportunities to consolidate, an exploratory analysis (pooling across comprehension groups) suggested that expressive vocabulary ability may be a better predictor of vocabulary consolidation than the more heterogeneous comprehension-decoding profiles. These findings suggest that children with weak vocabulary knowledge may be better able to consolidate new words when given opportunities to do so immediately, with important implications for timing remediation to maximise success.

Learning and consolidation in poor comprehenders

Previous literature had suggested that poor comprehenders' encoding weaknesses are specific to semantic aspects of word learning (Nation et al., 2007; Ricketts et al., 2008), with phonological skills a relative strength for these children (Nation & Snowling, 1998). However, we found that poor

comprehenders' difficulties extend beyond semantics to phonological aspects of word learning, and also into declarative spatial memory – an ability that has not been examined in this population. The training and testing demands of the present experiment likely enabled us to capture these weaknesses not detected by previous studies: we taught children significantly more words than Nation et al. (2007) and assessed explicit recall of the new words as opposed to recognition measures used by Ricketts et al. (2008). Our tasks were demanding not only on the children's knowledge of the words, but their ability to access and produce the new material. Although it is not possible to fully dissociate whether poor comprehenders' difficulties arise at encoding or retrieving the information within these tasks, it is worth noting that group differences were observed in picture-naming accuracy but not retrieval time, suggesting that poor comprehenders did not struggle to access the information they had learned. Poor comprehenders also showed lower accuracy in the object-location task – which did not require expressive recall – and in the multiple-choice tasks at training (although this difference was not statistically significant, $p = .065$; Table S10). Together, these findings suggest that poor comprehenders' difficulties likely arise at encoding rather than solely in expressing their new knowledge and that previous studies may not have been sufficiently powered and/or challenging to capture the breadth of poor comprehenders' encoding deficit. Closely monitoring processes during learning will better inform our understanding of these encoding versus retrieval difficulties.

A key question is generated by the present findings: if poor comprehenders show weaker encoding across all tasks, what is the underlying nature of this

difficulty? We had predicted that poor comprehenders would show equivalent performance to good comprehenders on the object-location task, as this task was designed to place minimal demands on verbal processes during learning. Two related explanations are possible here, which are not mutually exclusive. First, poor comprehenders' difficulties may be best characterised as a learning deficit that extends across domains (and could plausibly account for language difficulties from an early age). Indeed, performance across the stem completion and object-location tasks was correlated ($r(28) = .45$, $p = .012$), suggesting a 'learning ability' element common to both tasks. Alternatively, it may be that individuals use verbal strategies across a wide variety of tasks, including to remember spatial locations in the object-location task. Speaking to this, vocabulary ability did strongly predict performance in this task ($r(27) = .52$, $p = .004$), almost to the same extent that it predicted word-form learning ($r(27) = .64$, $p < .001$). It seems likely that comprehension weaknesses impact performance across domains (Pimperton & Nation, 2010), and the present study highlights that weaknesses cannot be considered specific to language across development.

To our knowledge, the present study is the first to isolate processes of initial learning, wake-based forgetting and sleep-associated consolidation that might underlie poor comprehenders' weaker vocabulary acquisition. The data did not support our prediction that poor comprehenders would show weaker consolidation of new vocabulary in the context of their poorer semantic knowledge (James et al., 2017): although poor comprehenders showed broad weaknesses immediately after encoding, their consolidation profile was similar to that of good comprehenders for both sleep-associated changes and longer-term retention. There was a slight indication of weaker overnight consolidation of word-forms when poor comprehenders learned in the morning (Figures 2 and 4), but this difference was not statistically significant (possibly a consequence of the small sample size). However, our exploratory analysis of individual differences was more strongly indicative of this pattern: from 12 to 24 hr, vocabulary was a more positive predictor of recall improvements following AM-encoding (i.e. overnight) than following PM-encoding. It thus seems likely that vocabulary differences better capture differences in consolidation than comprehension profiles, which have heterogeneous aetiologies that likely vary within and between samples. Indeed, a limitation of this study is that we do not have broader language measures to better characterise the strengths and weaknesses of children in our sample.

Predictors of successful vocabulary consolidation

This study contributes to a broader literature supporting a benefit for sleep in learning new vocabulary

and highlights the value of examining individual differences to further inform models of vocabulary consolidation. For both stem-completion and picture-naming, we observed clear benefits for sleep in the first 12-hr after learning which boosted recall between the first two test sessions. Memory also improved across the 12–24-hr period for these tasks regardless of encoding-time, suggesting that wake is less detrimental to memory *after* versus *before* a period of sleep. This finding is consistent with proposals that wake-based decay of hippocampal representations is less detrimental to retrieval accuracy after sleep (Hardt, Nader, & Nadel, 2013) and that these more stable representations may better benefit from retrieval practice to continue processes of consolidation (Antony, Ferreira, Norman, & Wimber, 2017). Sleep within the first 12-hr was also more beneficial to memory than sleep following an intervening day awake. Whilst benefits for immediate sleep have been seen in previous studies (Gais, Lucas, & Born, 2006), we showed that that these extend to the longer-term retention of new information, with benefits for PM-encoded information still apparent 4–10 weeks later. In contrast, a day's wakefulness before opportunities to consolidate risks longer-term forgetting of new information. This timing benefit may also help to explain why frequent napping better predicts vocabulary development in young children than overnight sleep does (Horváth & Plunkett, 2016).

Our goal was to better understand individual differences in consolidating new vocabulary, in line with models proposing a role for prior knowledge in supporting this process (James et al., 2017). An exploratory analysis using expressive vocabulary knowledge as a predictor of word-form recall suggested that sleep soon after learning may be especially beneficial for children with weak vocabulary knowledge: existing vocabulary did not predict changes in memory during the first 12-hr of learning (i.e. there was only an overall benefit for sleep), whereas children with poorer existing knowledge were less able to benefit from sleep during the 12–24-hr period. Interestingly, it did not appear as if children with weaker vocabulary ability had simply forgotten more items during the course of the day (Figure 6), an explanation considered by Walker et al. (2020). As such, we propose that these differences reflect the multiple ways in which new information may be 'tagged' for memory consolidation (Stickgold & Walker, 2013): all children may benefit from the saliency of learning information immediately before bed, whereas superior vocabulary knowledge affords more robust connections to prior knowledge that can facilitate consolidation regardless of delay. However, it is important to remember that this finding resulted from exploratory analyses and thus requires replication and further examination. Furthermore, it will be important to determine whether vocabulary knowledge remains the best predictor over and above other aspects of language

ability that were not measured in the present study (e.g. morphological skills).

Conclusions and implications

Theoretical models of language consolidation have value to inform – and in turn be informed by – our understanding of individual differences in vocabulary learning across development. This study showed that children with reading comprehension difficulties have a lower capacity for vocabulary learning than children with good comprehension and that this relative impairment is apparent even when new vocabulary is taught directly (i.e. not reliant on text comprehension). The study also provides clear evidence that sleep soon after learning can have long-lasting benefits for memory, regardless of language ability. When learning was followed by a day awake, new words were less likely to be retained for the longer term, and this was particularly the case for children with poorer existing vocabulary knowledge. Importantly then, our data support the view that defining literacy disorders on the basis of skill discrepancies (i.e. between decoding and comprehension) may have limited use in understanding a child's ongoing difficulties, especially in complex domains like reading comprehension. Although our research questions were derived from previous studies of poor comprehenders, profiling their vocabulary ability on a continuous scale proved more useful for capturing weaknesses in vocabulary consolidation and the potential role for timing in understanding this relationship. Given that literacy instruction typically features in the morning in the UK education system, this finding – if supported by future studies – has important practical implications for how vocabulary instruction can be better timed to support struggling learners.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Recruitment information.

Figure S1. Mean definition accuracy at each test point following AM/PM encoding separately, for each comprehension group.

Table S1. Predictors of definition accuracy in the main 24-hr analysis.

Table S2. Predictors of definition accuracy in the delayed follow-up analysis.

Table S3. Predictors of picture naming performance for the main 24-hr analyses.

Table S4. Predictors of picture naming performance for the delayed follow-up analyses.

Table S5. Predictors of stem-completion accuracy in the main 24-hr analysis.

Table S6. Predictors of stem-completion accuracy in the delayed follow-up analysis.

Table S7. Predictors of object-pair accuracy in the main 24-hr analysis.

Table S8. Predictors of object-pair accuracy in the delayed follow-up analysis.

Table S9. An exploratory analysis using vocabulary ability to predict performance in the stem completion task.

Table S10. An exploratory analysis looking at predictors of performance in the multiple-choice training tasks.

Acknowledgements

The authors wish to express their gratitude to the children, families and schools who dedicated their time to the project. The research was funded by ESRC grant ES/N009924/1 awarded to L.M.H and M.G.G.. E.J. was supported by an ESRC 1 + 3 Studentship in conducting the research, and ESRC Postdoctoral Fellowship (ES/T007524/1) in preparation for publication. The authors are also grateful to Dr Fay Fletcher and Dr Sarah Walker for their assistance with data collection and scoring. The authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Emma James, Department of Psychology, University of York, YO10 5DD York, UK; Email: emma.james@york.ac.uk

Key points

- Previous studies show that poor comprehenders have difficulties in learning new word meanings and poor retention of new words, which may be attributable to sleep-associated consolidation mechanisms.
- We present new evidence that poor comprehenders' encoding difficulties extend beyond word meanings and into the phonological domain; but that consolidation mechanisms remain intact.
- Existing vocabulary knowledge better predicted consolidation of new word-forms when words were learned in the morning, whereas this relationship was not apparent when sleep followed soon after learning.
- Future research should therefore thoroughly examine the hypothesis that learning close to bedtime may boost consolidation for children with weak vocabulary.
- Given that comprehension difficulties have heterogeneous aetiologies, individual differences in component skills are likely more informative for research and practice.

Note

1. Note that 22/30 children completed the sets one week apart, and differences in set performance for children with longer gaps fell within the range of those who completed the tasks one week apart. Actigraphy data were collected to check for overall group differences in sleep that might account for any differences seen in consolidation (see Table 1), but total sleep time was not hypothesised to predict consolidation itself (with previous studies implicating more specific neural markers during sleep, for example Smith et al. (2018)).

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th edn). Washington, DC: American Psychiatric Pub.
- Antony, J.W., Ferreira, C.S., Norman, K.A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, 21, 573–576.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Cain, K., & Oakhill, J. (2011). Matthew effects in young readers reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, 44, 431–443.
- Catts, H.W., Adlof, S.M., & Weismer, S.E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, 49, 278–293.
- Christensen, R.H.B. (2015). ordinal-Regression Models for Ordinal Data. R package version 2015.6-28.
- Clarke, P.J., Snowling, M.J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, 21, 1106–1116.
- Cutting, L.E., Clements-Stephens, A., Pugh, K.R., Burns, S., Cao, A., Pekar, J.J., ... & Rimrodt, S.L. (2013). Not all reading disabilities are dyslexia: Distinct neurobiology of specific comprehension deficits. *Brain Connectivity*, 3, 199–211.
- 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, 3773–3800.
- Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, 13, 309–321.
- Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning and Memory*, 13, 259–262.
- Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences*, 17, 111–120.
- Henderson, L.M., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodil flew to the intergalactic zoo: Offline consolidation is critical for word learning from stories. *Developmental Psychology*, 51, 406–417.
- Henderson, L.M., & James, E. (2018). Consolidating new words from repetitive versus multiple stories: Prior knowledge matters. *Journal of Experimental Child Psychology*, 166, 465–484.
- Henderson, L., Snowling, M., & Clarke, P. (2013). Accessing, integrating, and inhibiting word meaning in poor comprehenders. *Scientific Studies of Reading*, 17, 177–198.
- Henderson, L.M., Weighall, A.R., Brown, H., & Gaskell, M.G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, 15, 674–687.
- Henderson, L.M., 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, 572–592.
- Horváth, K., & Plunkett, K. (2016). Frequent daytime naps predict vocabulary growth in early childhood. *Journal of Child Psychology and Psychiatry*, 57, 1008–1017.
- James, E., Gaskell, M.G., Weighall, A., & Henderson, L. (2017). Consolidation of vocabulary during sleep: The rich get richer? *Neuroscience and Biobehavioral Reviews*, 77, 1–13.
- Landi, N., & Ryherd, K. (2017). Understanding specific reading comprehension deficit: A review. *Language and Linguistics Compass*, 11, e12234.
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, 7, e43230.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44, 314–324.
- Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties*, 24, 47–73.
- Nation, K., & Snowling, M.J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, 39, 85–101.
- Nation, K., Snowling, M.J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Advances in Speech Language Pathology*, 9, 131–139.
- Ouellette, G.P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98, 554.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383.
- Pimperton, H., & Nation, K. (2010). Understanding words, understanding numbers: An exploration of the mathematical profiles of poor comprehenders. *British Journal of Educational Psychology*, 80, 255–268.
- Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39, 859–862.
- Ricketts, J. (2011). Research review: Reading comprehension in developmental disorders of language and communication. *Journal of Child Psychology and Psychiatry*, 52, 1111–1123.
- Ricketts, J., Bishop, D.V., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading*, 31, 117–135.
- Ricketts, J., Sperring, R., & Nation, K. (2014). Educational attainment in poor comprehenders. *Frontiers in Psychology*, 5, 445.
- Smith, F.R., Gaskell, M.G., Weighall, A.R., Warmington, M., Reid, A.M., & Henderson, L.M. (2018). Consolidation of vocabulary is associated with sleep in typically developing children, but not in children with dyslexia. *Developmental Science*, 21, e12639.
- Snowling, M.J., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., ... & Hulme, C. (2009). *York assessment of reading for comprehension (YARC) passage reading*. London: GL.
- Stickgold, R., & Walker, M.P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. *Nature Neuroscience*, 16, 139.

- Stothard, S., Hulme, C., Clarke, P., Barnby, P., & Snowling, M. (2010). *York assessment of reading for comprehension (YARC): Passage reading secondary*. London: GL Assessment.
- Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (2012). *TOWRE-2 examiner's manual*. Austin, TX: Pro-Ed.
- Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*, 8–25.
- Walker, S., Gaskell, M.G., Knowland, V.C.P., Fletcher, F.E., Cairney, S.A., & Henderson, L.M. (2020). Growing up with interfering neighbours: the influence of time of learning and vocabulary knowledge on written word learning in children. *Royal Society Open Science, 7*(3), 191597.

Accepted for publication: 1 April 2020