Enhanced recognition and recall of new words in 7- and 12-year old children following a period of offline consolidation

Helen Brown[a] (h.brown@psych.york.ac.uk)

Anna Weighall[b] (A.R.Weighall@shu.ac.uk)

Lisa M. Henderson[a] (l.henderson@psych.york.ac.uk)

M. Gareth Gaskell[a] (g.gaskell@psych.york.ac.uk)

[a]Department of Psychology, University of York, York, UK, YO15 5DD

[b]Psychology Research Group, Sheffield Hallam University, Sheffield, UK, S10 2BP

Address for correspondence:

Gareth Gaskell,

Department of Psychology,

University of York,

York,

YO10 5DD.

Email: g.gaskell@psych.york.ac.uk

Abstract

Recent studies of adults have found evidence for consolidation effects in the acquisition of novel words (Dumay & Gaskell, 2007), but little is known about whether such effects are found developmentally. In two experiments we familiarized children with novel nonwords (*e.g., biscal*) and tested their recognition and recall of these items. In Experiment 1 7-year old children were then retested either the same or the following day to examine changes in performance after a short delay compared with a longer delay including sleep. Experiment 2 used two age groups (7- and 12-year old children), with all participants retested 24 hours later. Twelve-year-old children accurately recognized the novel nonwords immediately after exposure, as did 7-year-olds in Experiment 2 (but not 1), suggesting generally good initial rates of learning. Experiment 1 revealed improved recognition of the novel nonwords after both short 3-4 hour and longer 24 hour delays. In contrast, recall was initially poor but showed improvements only when children were re-tested 24-hours later, not after a 3-4 hour delay. Similar improvements were observed in both age groups despite better overall performance in 12-year-olds. We argue that children, like adults, exhibit offline consolidation effects on the formation of novel phonological representations.

Keywords: Language acquisition, word learning, phonology, lexical access, spoken word recognition

Word count – 8991 (text) + 1505 (references) = 10,496

Word learning is a vital component of language acquisition. Children typically acquire their first words between 10 and 14 months of age (Horst, McMurray, & Samuelson, 2006). By 2-years-old, a child will know approximately 300 words, with this number increasing to over 14,000 by the age of 6 (Carey, 1978), and reaching approximately 30,000 by adulthood (Altmann, 1997). Recently a line of research has indicated that, for adults, offline consolidation plays an important role in word learning (Dumay & Gaskell, 2007). We will first review these adult studies before turning to the potential for consolidation effects during development.

Word Learning in Adults

In order to acquire a new spoken word one must be able to both recognize that word, and distinguish it from similar-sounding words (Davis, Di Betta, Macdonald, & Gaskell, 2009). To achieve the first component it is necessary to form an accurate phonological representation of the new word. Studies of word learning in adults show that novel nonwords can be recognized accurately immediately after training (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b) indicating that new phonological representations are formed very rapidly. However, free recall of novel nonwords has been shown to increase after a period of offline consolidation, (Dumay & Gaskell, 2007; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010), suggesting that consolidation may play a role in stabilizing and enhancing novel phonological representations.

To achieve the second of these components the new word must be integrated into the existing lexicon. Lexical competition has been used to investigate the time-course of this process in adults; it is assumed that a new word can only engage in lexical competition during spoken word recognition once it has been integrated with existing lexical representations (Gaskell & Dumay, 2003b). According to the Cohort model of spoken word recognition (Marslen-Wilson & Zwitserlood, 1989) lexical competition occurs between phonologically

similar words up to the point at which only one word in the lexicon matches the speech input (the uniqueness point). Previous experiments have taught adults novel nonwords (*e.g., cathedruke*) that differed from their existing basewords (*e.g., cathedral*) only after the normal uniqueness point, thus shifting the uniqueness point of the baseword towards its offset (Davis et al., 2009; Dumay, Gaskell, & Feng, 2004; Gaskell & Dumay, 2003b; Tamminen & Gaskell, 2008; Tamminen et al., 2010). Thus, once a novel nonword becomes integrated into the lexicon there should be a greater amount of lexical competition when the baseword is heard, resulting in longer recognition latencies. These experiments revealed that the same novel nonwords that rapidly generated novel phonological representations did not engage in lexical competition with their basewords until the following day (Dumay et al., 2004; Gaskell & Dumay, 2003b). The emergence of lexical competition between novel nonwords and their basewords has been linked specifically to sleep (Tamminen et al., 2010), typically emerging only after a period that includes sleep, not an equal period of time awake (Dumay & Gaskell, 2007).

The temporal dissociation between recognition of novel spoken words and subsequent improvements in recall and engagement in lexical competition has been interpreted within a dual-systems framework of memory and learning, such as the *complementary learning systems (CLS)* model (McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; O'Reilly & Norman, 2002; O'Reilly & Rudy, 2001). The CLS account proposes that newly encoded memory representations are initially mediated via a hippocampal network, but that during offline periods these representations are replayed, resulting in the strengthening of neocortical representations and the integration of new memory representations with information already held in long-term memory. This offline replay allows novel phonological representations to begin engaging in lexical competition with phonologically similar existing words.

In support of this framework a recent fMRI study demonstrated different neural responses to novel nonwords learnt one day prior to the fMRI scan (consolidated nonwords), novel words learned on the day of the scan (unconsolidated nonwords), and untrained nonwords (Davis et al., 2009). Presentation of untrained nonwords resulted in elevated hippocampal responses relative to unconsolidated nonwords, indicating that the hippocampus was involved in the formation of new phonological representations. In comparison, levels of cortical activity in the superior temporal gyrus were equivalent for unconsolidated and untrained nonwords, whereas the activation for consolidated nonwords was lower, closer to the level of activation for existing words. This suggests that novel phonological representations are integrated with pre-existing lexical knowledge in neocortical areas only after offline consolidation. These results provide strong evidence in support of the CLS framework.

Word Learning in Children

A key area of uncertainty pertains to whether offline consolidation plays the same role during word learning in children as it does in adults. Previous research has often focused on the immediate aspects of word learning in children, such as rapid learning of novel phonological forms (Church & Fischer, 1998; Houston, Tincoff, & Jusczyk, 2001; Jusczyk & Aslin, 1995; Spiegel & Halberda, 2011). This research followed from Carey's (1978) theory of fast-mapping, which proposed that children were able to learn new words after only minimal exposure. Yet Carey did not assume that fast-mapping was the end point of word learning. Rather, she claimed that fast-mapping must allow the child to create an initial representation of a word containing sufficient information to allow this representation to be maintained within the lexicon until a more stable and complete representation develops through further experience with that word (see also Horst et al., 2006; Horst & Samuelson, 2008).

Despite this focus on the immediate effects of learning a number of developmental researchers have included delayed measures of word learning. Jusczyk and Hohne (1997) demonstrated that 8-month-old infants could accurately recognize novel nonwords up to two weeks after initial exposure indicating that rapidly formed phonological representations are long-lasting, as has also been demonstrated in adults (Dumay et al., 2004; Tamminen & Gaskell, 2008). Recognition of novel words has also been found to improve when older children are tested after a delay of one (Storkel, 2001; 3-6 year-olds), four (Frazier Norbury, Griffiths, & Nation, 2010; 7-year-olds) or 24 weeks (Dockrell, Braisby, & Best, 2007; 6-7 year-olds) even with no additional exposure to the novel words during this period. Stabilizing effects of consolidation on novel phonological forms have also been demonstrated in studies examining the effects of gesture (McGregor, Rohlfing, Bean, & Marschner, 2009) and causal information (Booth, 2009) on word learning in infants and toddlers. Together these studies provide evidence suggesting that phonological representations are stabilized and enhanced over time in children, as also appears to be the case in adults (Dumay & Gaskell, 2007).

Such similarities in word learning for school-aged children and adults, particularly the improvements observed in children's recognition of newly learned words at delayed test points, support Church and Fischer's (1998) *continuity of learning hypothesis*, which states that there is continuity across development of the basic word learning mechanisms. This hypothesis implies that offline consolidation should play a role in word learning in children, just as it does in adults. There is already evidence that sleep enhances declarative memory in 6-12 year-old children (Backhaus, Hoeckesfeld, Born, Hohagen, & Junghanns, 2008; Wilhelm, Diekelmann, & Born, 2008), as it does in adults (Plihal & Born, 1999). However, studies of procedural memory have failed to demonstrate the beneficial effects of sleep on procedural memory consolidation in 6-11 year-olds that are typically observed in adults (Fischer, Wilhelm, & Born, 2007; Prehn-Kristensen et al., 2009; Wilhelm et al., 2008),

suggesting that for certain types of memory representations the time-course of offline consolidation may differ between children and adults.

Furthermore, some evidence suggests that that the processing and representation of spoken words may not be fully adult-like even in school-age children. Although 2-year-old children have shown incremental processing of spoken words (Swingley, Pinto, & Fernald, 1999), and children as young as four have been found to exhibit effects of neighborhood density and phonotactic probability (Storkel & Hoover, 2011; Storkel & Lee, 2011), the extent to which children show fully adult-like lexical competition effects between phonologically-similar items remains unclear (Sekerina & Brooks, 2007). Evidence from studies using lexical decision, repetition, and gating tasks suggests that lexical competition effects emerge over development, with the size of the competition effects depending on chronological age and the age at which the lexical items have been acquired (Garlock, Walley, & Metsala, 2001; Metsala, Stavrinos, & Walley, 2009; Metsala & Walley, 1998; Munson, Swenson, & Manthei, 2005; Walley, 1993). In addition, there is evidence suggesting that phonological (Storkel & Rogers, 2000) and orthographic (Castles, Davis, Cavalot, & Forster, 2007) representations are more loosely specified in children than adults even by 7-8 years of age. Finally, word recognition appears to be less automatic in 7-year-old children than it is in adults (Ojima, Matsuba-Kurita, Nakamura, & Hagiwara, 2011). Given these representational differences it remains feasible that 7-year-old children will show differences in the time-course of word learning and onset of lexical competition effects compared to adults. However, by around 10 years of age word recognition and lexical representations appear to be more adult-like (Ojima et al., 2011; Storkel & Rogers, 2000), suggesting that older children should show a more adult-like effect of consolidation on word learning.

The Current Study

The two experiments reported here investigated whether 7- and/or 12-year-old children show the same time-course as adults in both the formation of novel phonological representation and their engagement in lexical competition. The children in this study were tested on modified versions of the adult tasks used by Dumay, Gaskell, and colleagues, allowing relatively direct comparison between performance in children and adults. Children were familiarized with novel nonwords (*e.g., biscal*), derived from existing basewords (*e.g., biscuit*), through repeated exposure in phoneme monitoring and phoneme isolation tasks. In Experiment 1 7-year-old children completed tests of recognition and explicit recall of nonwords immediately after training, and again after a delay of either 3-4 or 24 hours to examine the precision of children's phonological representations of the novel words. Speeded lexical decisions to the basewords were also examined at both time points to determine whether familiarization with the novel nonwords slowed processing of the basewords, indicating that the novel nonwords were engaging in lexical competition. It is important to note that the design of Experiment 1 did not specifically examine the role of sleep in the formation and lexical integration of novel words. Rather it considered the time-course of offline consolidation, comparing a short daytime retention interval of 3 to 4 hours with a more extended interval of 24 hours that included sleep.

Based on prior research demonstrating that children form new phonological representations very rapidly (*e.g.,* Spiegel & Halberda, 2011) and show long-term retention of these new representations (*e.g.,* Jusczyk & Hohne, 1997) it was predicted that children, like adults, would show accurate recognition of the novel nonwords both immediately after exposure and 24 hours later. It was also predicted that recall of the novel nonwords would be poorer than recognition, but that recall would improve after a period of offline consolidation. Such a result would be consistent with both adult data on word learning (Dumay & Gaskell, 2007; Dumay et al., 2004) and more generally the finding that declarative memories appear to

benefit from sleep-associated consolidation in both children (Backhaus et al., 2008) and adults (Plihal & Born, 1999). Finally, insofar as children show adult-like lexical competition effects (Sekerina & Brooks, 2007), learning the novel words (*e.g., biscal*) should have the eventual effect of slowing down recognition of similar existing basewords (*e.g., biscuit*). However, since spoken word recognition is a highly automatized skill, and thus may be considered as part of procedural rather than declarative memory (Dumay & Gaskell, 2007), it was less clear whether engagement in lexical competition would be influenced by offline consolidation and/or sleep, given the finding that procedural memory skills do not appear to benefit from sleep in children (*e.g.,* Wilhelm et al., 2008).

Experiment 1

In Experiment 1 6- to 8-year-old children were familiarized with novel nonwords on Day 1, and were re-tested either later the same day following a 3-4 hour delay, or one day later approximately 24 hours after the first session.

Method

Participants

Twenty-one children aged between 6 years 10 months and 7 years 9 months (10 male; *M* age = 7 years 2 months) from two primary schools in Northern England participated in the study. For all children British English was their first language and they had no known hearing, speech, or language impairments. Written consent was obtained from schools and parents prior to the study, with verbal consent obtained from the children prior to each session.

Stimuli

Twenty-six stimulus triplets (Appendix A) consisting of one baseword (*e.g., biscuit*), one novel nonword (*e.g., biscal*), and one foil (*e.g., biscan*) were selected from 68 stimulus triplets used in a previous longitudinal study of adult word learning (Tamminen & Gaskell,

2008). Basewords were mono-morphemic with uniqueness points located at or before the final vowel. Novel nonwords differed from the baseword at the final vowel, and from each other at the final consonant/consonant cluster. All three words had the same stress pattern.

A pilot study using 29 adults (11 male; M age = 28.7 years; SD = 12.1 years) collected age-of-acquisition (AoA) ratings for the basewords. Previous research has shown that AoA ratings provide a more sensitive measure of lexical familiarity in children than adult word frequency or neighborhood density measures (Garlock, Walley, & Metsala, 2001), and have been shown to correspond with more objective AoA measures (Carroll & White, 1973a, 1973b; Lyons, Teer, & Rubensten, 1978; Morrison, Chappell, & Ellis, 1997). Twenty-six words with a rated AoA of 7.5 years or less (M = 6.1 years, $range$ = 3.3 – 7.4 years) were selected. All were bisyllabic (14), or trisyllabic (12), ranging from 6 to 11 phonemes (M = 8.0) in length. CELEX frequency varied between 2 and 15 per million (M = 4.0; Baayen, Piepenbrock, & van Rijn, 1993), and written word frequency, taken from the Educator's Word Frequency Guide varied between 32.3 and 51.5 (M = 43.8). Two lists of 13 stimulus triplets were created with basewords matched as closely as possible for rated AoA, number of syllables, number of phonemes, and frequency. Independent samples t-tests indicated that the lists were matched on all these variables (AoA, p = .58; syllables, p = 1.00; phonemes, p = .89; frequency p = .77), plus the acoustical duration (ms) of the novel words (p = .21), foil words (p = .28), and word-stem cues (p = 1.00).

Fifty-two mono-morphemic nouns (also selected from Tamminen and Gaskell, 2008), were chosen as fillers for the lexical decision task so that, including the basewords, there were 26 mono-syllabic, 26 bi-syllabic, and 26 tri-syllabic words. Seventy-eight nonword fillers, matched on duration, and number of syllables and phonemes, were also selected. Nonwords were created by changing one or two phonemes of existing words. All stimuli were recorded on a Pioneer PDR 509 system by a female native English speaker.

Design and Procedure

Participants were tested individually in a quiet area of their school. Tasks were run on a Dell Latitude laptop computer, using DMDX experimental software (Forster & Forster, 2003), which recorded button-press response-times from stimulus onset. A Creative SD-50 microphone and Sony MZ-N710 mini-disk recorder were used to record verbal responses in phoneme isolation and cued recall tasks. Responses were later transcribed and scored for accuracy. All stimuli were presented at a comfortable listening level, with the child sat directly in front of the computer, approximately 50cm from the screen.

Each child participated in two sessions; for the *am-pm* group Session 1 began at 9-10am of Day 1, with the second 3-4 hours later on the same day. The *am-am* group completed Session 1 around 11am-12pm, and the second approximately 24 hours later on Day 2. Children were randomly assigned to these two groups. The tasks completed in each session are summarized in Figure 1, and described in more detail below. Before each task the instructions were explained and an example was given. Key instructions were also presented on screen. Short breaks were included between tasks to maximize the child's attention throughout the session. No feedback was provided during the experimental trials.

During the *exposure phase* participants completed two phonics-based tasks[1]: phoneme monitoring and phoneme isolation. In *phoneme monitoring* participants listened to the new words and indicated whether a specified phoneme was present or absent in each word. The task began with five real-word practice trials. Additional examples were provided if the child made two or more errors. Experimental trials were split into six blocks with the target phonemes /p/, /t/, /d/, /s/, /m/, and /b/ in this fixed order for all participants, and each novel nonword occurring twice per block. Target phonemes occurred at all positions across the novel words, with the number of target present trials varying between both lists and blocks.

During each block the target phoneme and a picture of a common object beginning with that phoneme were displayed centrally on screen, with images of a happy and sad face displayed in the bottom left and right corners of the screen respectively, above the appropriate response buttons. Items were separated by an inter-trial interval of 500ms, with a maximum response-time of 5s.

In *phoneme isolation* participants listened to the new words and said either the first (Block 1) or last phoneme (Block 2) of each word aloud. Each block began with a written example, which remained on screen throughout the block. Three real-word practice trials with feedback were provided prior to each block in order to ensure that the child understood the task. Each novel word occurred three times per blocks. The task was self-paced, with instructions emphasizing only accuracy. Novel words were heard 18 times during exposure; 12 times in the monitoring task and six times in the isolation task.

The *testing phase* began with an auditory *lexical decision* task during which participants indicated whether or not each spoken item was a real word. A sad and a happy face were presented on the left and right sides of the screen, above the appropriate response keys, to serve as a reminder of the task instructions. The task began with eight practice items, followed by 156 experimental trials during which all 26 basewords, 52 filler-words, and 78 filler nonwords were presented in a randomized order. None of the trained novel nonwords were included in the lexical decision task. Items were separated by an inter-trial interval of 500ms, with a maximum response-time of 5s. Instructions emphasized both speed and accuracy.

The second test task was *cued recall,* during which participants heard the first CVC syllable (*e.g. bis-*) of the 13 novel nonwords from the exposure phase and were instructed to try to complete the cue using one of the new words. Cues could be replayed as many times as required. If a real word was recalled the participant was reminded to try to recall one of the

new words that they had heard during the phonics-based tasks. If no response was given after 5s the child was encouraged to guess before the next trial was initiated.

The final test task was a *two-alternative forced choice task* (*2afc*) in which participants heard both a novel word (*e.g., biscal*) and its corresponding foil (*e.g., biscan*). Participants decided which of the two items had been heard during the phonics tasks. All 13 of the novel words from the training list, and their corresponding foils, were presented during this task. Participants were instructed to listen to both items before responding. The numbers "1" and "2" were presented on the left and right sides of the screen respectively, indicating which response key to press for the first and second word heard. Response-times were measured from the onset of the first item. However, since emphasis was placed on accurate responding, and response-times in this task were very long ($M = 3794$ms), only accuracy data were analyzed. The order of the novel word – foil word pairs was randomized across participants, as was the order of the items within each pair.

Two *control measures* were included at the end of the second session. A *picture-word matching task* was included to assess whether the children were familiar with the basewords of the novel nonwords (*e.g., biscuit*). Each trial began with four pictures of common objects displayed on screen; one target baseword and three distracter images. Five hundred ms after the pictures were displayed the target baseword was played, and participants indicated which of the four pictures corresponded to the word heard. Accuracy was recorded by the experimenter. If the participant paused for more than 3s before responding this was taken to indicate uncertainty about the word's meaning. At the end of the task the child was asked to provide a definition or explanation for items marked as "uncertain" in order to avoid the possibility that the child was eliminating items that they knew, leaving them with the correct baseword image without knowing what the word meant. In all cases children were able to provide a rough definition of these "uncertain" items. Distracter images were matched as

closely as possible in AoA to the basewords (according to the MRC psycholinguistic database, Coltheart, 1981; or the Bristol norms, Stadthagen-Gonzalez & Davis, 2006). The order of the trials was randomized across participants, but the same three distracter images always occurred with the same target, and the position of these four images on screen remained constant. Target pictures were distributed equally across all four quadrants of the screen. Images were selected from *www.fotosearch.com/clip-art.*

The *Wechsler Abbreviated Scale of Intelligence (WASI)* was included in order to ensure that there was a normal range of IQ scores in the sample of children tested. The two-subtest version, involving vocabulary and matrix design sub-scales, was administered rather than the full four-subtest version in order to minimize the length of the second test session.

Results

Eleven participants were tested in the am-pm group (5 male), and 10 in the am-am group (5 male). The groups did not differ significantly in age, t(19) = .20, p = .84, d = .09, (am-pm – M = 7,2 years SD = 5 months; am-am – M = 7,3 years, SD = 5 months) or IQ, t(19) = -2.01, p = .059, d = .92 (am-pm – M = 106.2, SD = 11.6; am-am – M = 102.9, SD = 8.4). Allocation to word list was balanced within each group.

Exposure phase

The mean error rates in the phoneme monitoring and phoneme isolation tasks were reasonably low (19.4% and 11.6% respectively), indicating that participants had attended to the phonological form of the novel items during exposure. Occasionally in the phoneme isolation task participants responded with letter names rather than phonemes. When this occurred the participant was reminded to respond using phonemes. Letter names corresponding to the correct phoneme were marked as correct. One participant in the am-am group had an error score more than 2.5*SD* above the mean in phoneme monitoring. Data from

this participant were removed from subsequent analyses since this higher-than-average error rate indicated that the novel nonwords were not correctly encoded during training.

Test phase

Mean accuracy scores for the *2afc task* are presented in Table 1. Participants recognized the phonological form of the novel nonwords at a level significantly above chance after a delay of both 3-4 hours, $t(10) = 2.36$, $p = .040$, $d = 1.49$, and 24 hours, $t(8) = 2.83$, $p =.022$, $d = 2.00$,[2] but not immediately after exposure to the novel words in Session 1, $t(19) = 1.67$, $p = .11$, $d = .77$. Percentage correct scores from the 2afc tasks were subject to an arcsin transform in order to better meet the assumptions of normality. In this and all subsequent ANOVAs list (1 *vs.* 2) was included as a dummy variable in order to reduce the estimate of random variation (Pollatsek & Well, 1995). Main effects of list and interactions involving list will not be reported. A 2 (session) x 2 (group) repeated-measures ANOVA revealed that the overall difference between sessions was significant, $F_1(1,16) = 9.59$, $p = .007$, $\eta_p^2 = .38$, $F_2(1,24) = 13.78$, $p = .001$, $\eta_p^2 = .37$. In comparison, the main effect of group was not significant, $F_1(1,16) = .26$, $p = .62$, $\eta_p^2 = .02$, $F_2(1,24) = 2.61$, $p = .12$, $\eta_p^2 = .10$, nor interaction between session and group, $F_1(1,16) = .06$, $p = .82$, $\eta_p^2 = .00$, $F_2(1,24) = .12$, $p = .73$, $\eta_p^2 = .01$. These latter findings indicate that the am-pm and am-am groups showed similar amounts of improvement in recognition between sessions.

In *cued recall* scoring focused on the final syllable of the word since the first syllable was provided as the cue, and the second syllable of the tri-syllabic nonwords was identical to that of the baseword. Responses received zero if no word was recalled, if only a real word was recalled, or if the final syllable was incorrect, one point if the first syllable (plus the middle syllable for trisyllables) was correct and either the final vowel or final consonant/consonant cluster was also correct, and two points if the response was completely correct. Accuracy scores, converted into percentages, are reported in Table 1.

A 2 (session) x 2 (group) repeated-measures ANOVA revealed that there was a significant main effect of session, $F_1(1,16) = 11.63$, $p = .004$, $\eta_p^2 = .42$, $F_2(1,24) = 20.04$, $p = .000$, $\eta_p^2 = .46$, with more nonwords recalled overall in Session 2. There was also a significant main effect of group, $F_1(1,16) = 6.62$, $p = .020$, $\eta_p^2 = .29$, $F_2(1,24) = 12.46$, $p = .002$, $\eta_p^2 = .34$, as well as a significant interaction between session and group, $F_1(1,16) = 7.55$, $p = .014$, $\eta_p^2 = .32$, $F_2(1,24) = 11.61$, $p = .002$, $\eta_p^2 = .33$, indicating that the number of novel nonwords recalled increased significantly across sessions only in the am-am group, $t(8) = -2.71$, $p = .027$, $d = 1.12$, not in the am-pm group, $t(10) = -1.49$, $p = .17$, $d = .19$. In support of these analyses six out of the nine children in the am-am group but only 4 out of 11 in the am-pm group showed improved performance between Session 1 and 2. In both groups one child showed poorer performance in Session 2, whilst the remaining children (2 x am-am; 6 x am-pm) showed no change in performance between sessions. Despite these differences, the two groups did not differ significantly in cued recall accuracy in Session 1, $t(18) = .94$, $p = .36$, $d = .44$.

In *lexical decision* participants performed fairly accurately across all items (*M* error = 20.5%; *SD* = 11.8%). Only data from the 26 basewords were included in the main analysis, allowing comparison between words that had a novel competitor (*test*) and words that did not (*control*). Data from one participant in the am-pm condition were excluded from the analysis due to technical failure and loss of lexical decision data in Session 2. Before analysis, incorrect responses were removed from the baseword data set, as were correct lexical decision responses with response-times under 200ms, or more than 2.5*SD* above the mean for each participant in each session. Finally responses to basewords were removed if the child did not correctly identify the item in the picture-word matching task. An independent samples t-test showed that the number of errors made in the picture-word matching task did not differ between groups, $t(19) = -.37$, $p = .71$, $d = .17$ (am-pm = 2.1%; am-am = 1.5%). One

participant had an error score more than 2.5*SD* above the grand mean, and another had a mean response-time, calculated across both sessions combined, more than 2.5*SD* above the grand mean. Data from both of these participants were removed, as were data points corresponding to one item (*partridge*) that produced an error score more than 2.5*SD* above the mean in the by-items analysis. Overall participants responded correctly to 83.9% (*SD* = 6.0%) of the basewords. An independent samples t-test indicated that the final error rate did not differ between groups, t(15) = .02, p = .99, d = .01 (am-pm = 83.6%, am-am = 83.5%).

A 2 (session) x 2 (group) x 2 (baseword-type; test *vs.* control) repeated-measures ANOVA for the response-time data revealed only a significant interaction between baseword type and group, $F_1(1,13) = 5.57$, p = .035, $\eta_p^2 = .31$, $F_2(1,23) = 4.21$, p = .052, $\eta_p^2 = .16$. Post hoc analyses indicated that the main effect of baseword type was significant only for the am-am group, $F_1(1,6) = 8.09$, p = .03, $\eta_p^2 = .57$, $F_2(1,23) = 6.79$, p = .16, $\eta_p^2 = .23$ (Figure 2a), suggesting that this group showed lexical competition between the novel nonwords and their phonologically similar basewords. However, given that the two groups differed in lexical decision performance immediately after exposure to the novel nonwords, as indicated by the marginal interaction between group and baseword type in Session 1, $F_1(1,13) = 4.39$, p = .056, $\eta_p^2 = .25$, $F_2(1,23) = 3.24$, p = .085, $\eta_p^2 = .12$, it is difficult to draw any firm conclusions about group differences in lexical competition in Session 2.

<p style="text-align:center">*Discussion*</p>

Overall, Experiment 1 provides evidence that declarative representations of novel words are enhanced following periods of offline consolidation in 7-year-old children. Most importantly improved performance in cued recall was observed only in the group who had the longer delay and an opportunity for nocturnal sleep between training and the second testing session. No such improvement in recall was found for the group tested twice within the same day, although this group did show improved recognition of the novel nonwords after as little

as three to four hours. These findings support the suggestion that offline consolidation plays a role in the formation and stabilization of new declarative representations of novel words.

Whilst it could be argued that re-exposure to the novel nonwords during the 2afc task in Session 1 may have provided an additional instance of learning, thus allowing improved recognition and recall of the novel nonwords in Session 2, cued recall of the novel nonwords did not improve in the am-pm group, implying that the significant improvement in the am-am group was unrelated to re-exposure. Moreover, Dumay and Gaskell (2007) examined the effects of re-exposure in adults by giving half of their participants the 2afc task only at the end of the Session 2 whilst the other half completed the task in both sessions. This manipulation did not affect performance measures in any of their other experimental tasks. As such, re-exposure to the novel nonwords during the 2afc task in Session 1 is unlikely to have been causally related to any improvements seen in Session 2 of the current experiment.

Data from the lexical decision task were less clear. Given that participants in the am-am group showed slower responses to test basewords compared to control words immediately after exposure to the novel nonwords, but the am-pm group did not, it is difficult to interpret the pattern of performance in Session 2. It is possible that the large variation in response-times, as indicated by the standard error bars in Figure 2a, may have masked the true pattern of data. This variation may have been due to a loss of attention during the long lexical decision task. Experiment 2 addresses these possibilities by shortening the lexical decision task and dividing it into multiple blocks to maximize attention throughout the task. Alternatively, time of testing may have had an effect since participants in the am-pm group typically began their first session early in the morning at around 9-10 am whereas the am-am group completed the first session at approximately 11am -12 pm due to restrictions of testing within the school day. Therefore, it is possible that differences in circadian rhythms may have affected performance across the two groups in Session 1 (Schmidt, Collette, Cajochen, &

Peigneux, 2007). This confound is addressed in Experiment 2 where children were trained

and tested at all time-points across the school day using the am-am design.

Experiment 2

Experiment 2 examined whether overnight improvements in recognition and recall of

novel words were quantitatively similar across two different age-groups of children; 6-8 and

11-13 year-olds (referred to as the 7- and 12-year-old groups respectively). Given research

suggesting that representation and processing of lexical items does not appear to be fully

adult-like until around 10 years of age (Ojima et al., 2011; Storkel & Rogers, 2000) it is

possible that different patterns of recognition, recall, and lexical competition may be

observed during word learning in these two age-groups. The fact that retesting the same day

did not induce any improvement in cued recall in Experiment 1 allowed us to switch to a

fully within-participants design in which participants were tested immediately after training

and re-tested 24 hours later. Data from Experiment 1 rule out the possibility that

improvements in cued recall observed 24 hours after exposure to the novel nonwords are due

simply to additional exposure to the novel words during the first test session. Instead, this

finding may be attributed to more extended consolidation (more than a few hours) and/or

sleep. An additional advantage of using a fully within-participants design is that it was

possible to address the potential confound of circadian rhythms since participants completed

Session 1 at all time-points across the school day.

In an attempt to clarify the pattern of lexical competition effects in Experiment 2 the

lexical decision task was adapted to make it more "child-friendly". The number of filler items

was reduced, thus shortening the task, and the task was split into three short 3- to 4-minute

blocks in an attempt to maximize attention throughout the task. The inclusion of an older

group of children may also provide more reliable information about the time-course of lexical

competition effects in children since previous research suggests that these effects may depend

on chronological age, and the age at which words have been acquired (Garlock, et al., 2001; Metsala, et al., 2009; Metsala & Walley, 1998; Munson, et al., 2005; Walley, 1993). Notably previous research has found that children aged between 5 and 8 years-old showed lexical competition effects only for early acquired words (*i.e.* those words that are already firmly established in the mental lexicon; Garlock, et al., 2001). Given that the rated AoAs of most of the basewords in the present study were very close to the chronological age of the 7-year-olds it is likely that these words will have only recently been acquired. Consequently, Garlock et al.'s findings may be problematic not only for the novel nonwords, which are only weakly established within the lexicon, but also for the basewords themselves. By 12-years of age the basewords should be firmly established in the lexicon, and thus it should be possible to observe lexical competition effects for these items.

<div align="center">

Method

</div>

Participants

Thirty children aged between 6 years 7 months and 8 years 4 months (12 male; *M age* = 7 years 7 months) and thirty-four children aged between 11 years 3 months to 13 years 11 months (11 male, *M age* = 12 years 8 months) participated in the experiment. Children were recruited from primary and secondary schools in Northern England. For all children British English was their first language, and they had no known hearing, speech, or language impairments. As in Experiment 1, written consent was obtained from both schools and parents prior to the experiment, and verbal consent was obtained from the children at the beginning of each test session.

Stimuli and Design

The stimuli were identical to those used in Experiment 1. However, a number of changes were made to the design. First, all participants completed the two experimental sessions on two consecutive days. This design enabled us to rule out the possible confound of

circadian factors having affected performance of the two groups in Experiment 1 (Schmidt, et al., 2007). In Experiment 2 participants were re-tested approximately 24 hours after Session 1, resulting in approximately half of the participants completing both sessions in the morning, and half completing both sessions in the afternoon.

A number of changes were also made to the lexical decision task in an attempt to reduce the large variation in response-times. Monosyllabic word and nonword filler items were removed from the stimulus set in order to reduce the length of the task. The remaining 26 basewords, 26 word fillers and 52 nonword fillers were split into three blocks, two of 35 items, and one of 34 items, each containing similar numbers of test basewords, control words, word fillers, and nonword fillers. The order of these three blocks was fixed, but the order of items within the blocks was randomized. Three filler items were fixed at the beginning of each block in order to control for start-up effects. Generic feedback (e.g. "*Super!"* and "*Good work*") was provided at the end of each block of lexical decision to motivate participants to keep responding both quickly and accurately throughout the task. Additionally, rather than using the left and right shift keys on the keyboard to respond a 850F Vibraforce Feedback sightfighter game-pad was used for all button-press responses in Experiment 2.

Two further changes were made to the design. In the phoneme monitoring task items were randomized in blocks of 13 within each phoneme monitoring block to minimize the chance of the same word occurring twice in a row. Finally in the picture-word matching task all 26 basewords (13 test and 13 control) were included rather than just the test basewords, as was the case in Experiment 1. Note, only the 7-year-old group completed this task.

Results

Thirty 7-year-old and 34 12-year-old children completed the experiment. Two additional 7-year-olds and 8 additional 12-year-olds were tested, but were removed from the

data set due to absence from school on the day of the second session. Of the 64 children who completed both experimental sessions, 15 7-year-olds (4 male) and 16 12-year-olds (5 male) were exposed to List 1 and 15 7-year-olds (8 male) and 18 12-year-olds (6 male) were exposed to List 2. The groups of children exposed to each list did not differ in either age (7-years – List 1, $M = 7,7$ years, $SD = 7$ months, List 2, $M = 7,8$ years, $SD = 7$ months, $t(28) = -.51$, $p = .61$, $d = .19$; 12-years – List 1, $M = 12,9$ years, $SD = 9$ months, List 2, $M = 12,8$ years, $SD = 8$ months, $t(32) = .20$, $p = .84$, $d = .07$) or IQ, according to the WASI (7-years – List 1, $M = 106.9$, $SD = 15.0$, List 2, $M = 103.7$, $SD = 12.6$, $t(28) = .65$, $p = .52$, $d = .25$; 12-years – List 1, $M = 105.3$, $SD = 10.8$, List 2, $M = 106.0$, $SD = 9.9$, $t(32) = -.19$, $p = .85$, $d = .07$).

Exposure Phase

Mean error rates in phoneme monitoring and phoneme isolation were relatively low (7 years – 12.3% and 11.6%; 12 years – 7.2% and 3.13% respectively). One participant in the 12-year old group had an error score more than $2.5SD$ above the mean in the phoneme monitoring task. Data from this participant were removed from all subsequent analyses.

Test Phase

In the *2afc task* Session 1 data were missing for 4 participants (one 7-year-old and three 12-year-olds) due to a technical error. Overall participants recognized the novel nonwords significantly above chance level both immediately after the exposure phase (7-years – $t(28) = -6.40$, $p < .001$, $d = 2.42$; 12-years – $t(28) = -27.40$, $p < .001$, $d = 10.36$), and one day later (7-years – $t(29) = -10.19$, $p < .001$, $d = 3.78$; 12-years – $t(31) = -58.92$, $p < .001$, $d = 21.16$), when required to make a familiarity judgment between two phonologically similar items (*e.g. biscal* and *biscan*). Data from one participant in the 12-year-old group were removed due to the participant producing an error score more than $2.5SD$ above the mean for that age group. Mean accuracy for each age group in each session is reported in

Table 2. As in Experiment 1, the 2afc data were subject to an arcsin transform in order to better meet the assumptions of normality.

A 2 (session) x 2 (age) repeated-measures ANOVA revealed a main effect age, $F_1(1,54) = 21.21$, $p = .000$, $\eta_p^2 = .28$, $F_2(1,24) = 74.45$, $p = .000$, $\eta_p^2 = .76$, with 12-year-olds scoring higher than 7-year-olds in both sessions. The main effect of session also approached significance by-participants, $F_1(1,54) = 3.37$, $p = .072$, $\eta_p^2 = .06$, and was significant by-items, $F_2(1,24) = 16.45$, $p = .000$, $\eta_p^2 = .41$, although further analyses revealed that this main effect of session approached significance only for the 7-year-olds, $F_1(1,27) = 3.29$, $p = .081$, $\eta_p^2 = .11$, $F_2(1,24) = 10.93$, $p = .003$, $\eta_p^2 = .31$, not for 12-year-olds, $F_1(1,27) = .58$, $p = .45$, $\eta_p^2 = .02$, $F_2(1,24) = 3.29$, $p = .082$, $\eta_p^2 = .12$, most likely due to the 12-year-olds performing closer to ceiling in Session 1, leaving less room for potential improvement.

Participants performed at a lower level in *cued recall* than in 2afc (see Table 2). Nonetheless, there was an increase in recall accuracy between Sessions 1 and 2, $F_1(1,59) = 133.34$, $p = .000$, $\eta_p^2 = .69$, $F_2(1,24) = 142.43$, $p = .000$, $\eta_p^2 = .86$. The data also revealed a significant main effect of age, $F_1(1,59) = 19.82$, $p = .000$, $\eta_p^2 = .25$, $F_2(1,24) = 56.44$, $p = .000$, $\eta_p^2 = .70$, with 12-year-olds showing more accurate recall than the 7-year-olds in both sessions. Nevertheless, the interaction between session and age was non-significant, $F_1(1,59) = .05$, $p = .83$, $\eta_p^2 = .00$, $F_2(1,24) = .01$, $p = .91$, $\eta_p^2 = .00$, indicating that both age groups showed equivalent improvement after a 24-hour delay. Notably, of the 64 children tested only two (both 12-year olds) showed worse performance in Session 2 compared to Session 1, with an additional five children (two 7-year-olds and three 12-year-olds) showing no change between sessions. As such, it appears that improved recall of novel nonwords one day post-exposure to the items is very consistent across both within and between age groups.

Additional analyses that included time of testing (am *vs.* pm) as an additional between-participants variable (or a within-items variable) in the repeated-measures ANOVA

described above revealed that there was no significant main effect of time of testing, $F_1(1,55)$ = .00, p = 95, η_p^2 = .00, $F_2(1,24)$ = .93, p = .34, η_p^2 = .04, and no significant interaction between session and time of testing, $F_1(1,55)$ = .15, p = .70, η_p^2 = .00, $F_2(1,24)$ = .79, p = .38, η_p^2 = .03, allowing us to rule out the possibility that the am-pm group in Experiment 1 failed to show improvements in cued recall due to the potential influence of circadian rhythms at time of retest. Thus, the most likely explanation for improvements in cued recall 24 hours after initial exposure to the novel words is that an extended period of offline consolidation (including sleep) was required to stabilize the novel phonological representations.

In the *lexical decision* task participants performed fairly accurately across all items, with an overall error score of 14.6% (*SD* = 8.6%) in the 7-year-old group and 5.9% (*SD* = 3.4%) in the 12-year-old group. All response-time data were filtered as described in Experiment 1. Four participants (two 7-year-olds and two 12-year-olds) had error scores more than 2.5*SD* above the mean, and three (two 7-year-olds and one 12-year-old) had mean response-times more than 2.5*SD* above the grand mean for their age group. Data from these participants were removed prior to analysis, as were data points corresponding to one item (*partridge*) that produced a mean error score more than 2.5*SD* above the mean in the by-items analysis. Overall 7-year olds responded correctly to 81.1% (*SD* = 9.5%) of the basewords and 12-year olds responded correctly to 90.8% (*SD* = 4.9%).

For the response-time latencies (Figure 2b) a 2 (session) x 2 (baseword type) x 2 (age) repeated-measures ANOVA revealed main effects of session, $F_1(1,52)$ = 4.77, p = .034, η_p^2 = .08, $F_2(1,23)$ = 9.57, p = .005, η_p^2 = .29, baseword type, $F_1(1,52)$ = 15.77, p = .000, η_p^2 = .24, $F_2(1.23)$ = 13.41, p = .001, η_p^2 = .37, and age, $F_1(1,52)$ = 30.71, p = .000, η_p^2 = .37, $F_2(1,23)$ = 170.07, p = .000, η_p^2 = .88, reflecting faster response-times in Session 2, for control words and 12-year-old children respectively. The only significant interaction was between session and age, $F_1(1,52)$ = 7.51, p = .008, η_p^2 = .13, $F_2(1,23)$ = 15.96, p = .001, η_p^2 = .41, with 7-

year-olds showing a greater increase in speed of responses in Session 2 than 12-year-olds. The lack of significant interaction between session and baseword type suggests that response-times to test and control items did not change between sessions. In order to determine whether there was a significant difference between response-times to test and control words at each time point the data were analyzed separately for each age group in each session. Although Figure 2b suggests that both age groups show some evidence of lexical competition at both time points, with slower responses to test basewords than to control words, these lexical competition effects were significant only for the 12-year-old group (Session 1 – $F_1(1,28) = 7.96$, $p = .009$, $\eta_p^2 = .22$; $F_2(1,23) = 14.90$, $p = .001$, $\eta_p^2 = .39$; Session 2 – $F_1(1,28) = 6.89$, $p = .014$, $\eta_p^2 = .20$; $F_2(1,23) = 9.16$, $p = .006$, $\eta_p^2 = .29$). It may be that lexical competition effects are more difficult to detect in 7-year-olds due to the larger variability in response-time latencies in this age group.

Discussion

Experiment 2 provides further evidence of delayed declarative enhancement of novel phonological representations, with participants showing improved recall of the novel non-words one day after initial exposure to them. In addition, 7-year-olds once again showed a trend towards improved recognition in the 2afc task in the 24-hour retest. Intriguingly, the lexical decision data suggest that participants showed slower response latencies to basewords with novel nonword competitors both immediately after studying the nonwords as well as 24 hours later, although the difference in response-times to test and control items was statistically significant only in the 12-year-old group. This pattern of data was unexpected based on previous word learning studies in adults (e.g., Dumay & Gaskell, 2007), and will be discussed further below.

General Discussion

The key novel finding in this study was that children showed improved cued recall of the novel nonwords only after an extended period of offline consolidation. Whilst the results from Experiment 1 were potentially confounded by the fact that all participants in the am-pm group completed their second session in the afternoon, during which there is typically a circadian dip (Schmidt, et al., 2007), Experiment 2 ruled out this possibility by testing participants in both the morning and the afternoon, with Session 2 occurring approximately 24 hours after Session 1. As such, the findings from the cued recall task suggest that improvements in the recall of novel nonwords occur only over a period of 24 hours that includes sleep, not a shorter 3 to 4 hour period whilst awake. Whilst we cannot rule out the possibility that improvements in cued recall occurred simply as a function of increased time between exposure and re-testing in the am-am groups compared to the am-pm group, it is plausible that sleep played a role since previous studies have already demonstrated that 6 to 12 year-old children show improvements in declarative memory after sleep, but not after a similar time period whilst awake (Backhaus, et al., 2008; Wilhelm, et al., 2008). Moreover, Dumay & Gaskell (2007) demonstrated that free recall of novel words in adults improved only after a period of 12 hours that included sleep, not after an equal period of time whilst awake.

In comparison to the poor recall of novel nonwords immediately after exposure, children accurately recognized the novel nonwords immediately post-exposure in Experiment 2 (although this was not the case in Experiment 1). This is consistent with developmental studies showing that infants are able to learn novel phonological forms rapidly from a very early age (Church & Fischer, 1998; Jusczyk & Aslin, 1995), and also with adults studies that have demonstrated good immediate recognition of novel nonwords (Dumay & Gaskell, 2007; Dumay, et al., 2004; Gaskell & Dumay, 2003a, 2003b; Tamminen & Gaskell, 2008). Similarly, accurate recognition of novel words after a delay of up to 24 hours has previously

been demonstrated in both children (Houston, et al., 2001; Jusczyk & Hohne, 1997) and adults (Dumay & Gaskell, 2007; Dumay, et al., 2004; Gaskell & Dumay, 2003b). Interestingly, recognition of the novel nonwords improved to a similar extent three to four hours, as compared to 24 hours later in Experiment 1. In Experiment 2 both age groups showed a numerical increase in recognition accuracy 24 hours after initial exposure to the novel words, although this effect approached significance only for the 7-year-olds, most likely due to a ceiling effect for the 12-year olds. In adult studies recognition immediately after training is also typically near-ceiling (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b), potentially masking improvement in recognition in Session 2. Nevertheless, adult studies that have made the 2afc task more difficult, either by reducing the number of initial exposures to the novel items (Davis, et al., 2009) or by adding filler words that were not derived from real words (*e.g. trogist*) to the exposure and recognition phases of the experiment (Dumay, et al., 2004), have demonstrated overnight improvements in recognition of novel nonwords, similar to the effect for the 7-year-old children here.

How might we explain the fact that novel phonological representations appear to be strengthened after as little as 3 to 4 hours in the 2afc task but only after 24 hours in the cued recall task? One explanation may be that explicit recall of novel words requires retention and retrieval of highly-specified phonological representations, whereas simple recognition of novel words does not, particularly when recognition occurs amongst a limited number of alternatives (Newman, 2008). In such instances recognition may simply involve listening preferentially for a familiar acoustic form (Werker, Fennell, Corcoran, & Stager, 2002). Thus, whilst some consolidation processes may occur immediately after encountering a novel stimulus, supporting the small improvements observed in the 2afc task only 3 to 4 hours post-exposure to the novel words, there appears to be a longer period of refinement and

consolidation, allowing participants to develop more highly-specified representations over time that can then be used to support recall as well as recognition.

In comparison, whilst 12-year-old children showed lexical competition between the newly-learned nonwords and their phonologically similar basewords immediately after studying the nonwords there did not appear to be any change in the size of these effects as a result of offline consolidation. Numerical trends in the 7-year-old data also followed a similar pattern. These findings are potentially consistent with developmental sleep studies that have demonstrated sleep-associated strengthening of declarative memories in children, but have failed to show overnight improvements in either the speed or accuracy of procedural memory tasks in children of the same age (e.g., Wilhelm et al., 2008). As noted in the introduction, since spoken word recognition is such a highly automatized skill, lexical competition between phonologically similar items might rely more on implicit (procedural) memory processes rather than explicit (declarative) memory (Dumay & Gaskell, 2007). Thus, overnight increases in lexical competition effects may not be expected in children. This conclusion, however, should be treated with caution in the absence of further developmental data. In particular, Dumay and Gaskell (2007) used pause detection rather than lexical decision as their measure of lexical activity, and this task may be more sensitive to sleep associated consolidation (*cf.* Tamminen et al., 2010).

Although differences between children and adults in the observed patterns of lexical competition effects might be a consequence of differences in activation, connectivity, and myelination of the medial temporal and neocortical networks across development (Menon, Boyett-Anderson, and Reiss, 2005; Sowell, Thompson, Holmes, Jernigan, & Toga, 1999), a simpler explanation is that auditory lexical decision may not be the optimal task for measuring lexical competition effects in children. Firstly, as is clear from Figure 2, there was great variability in response-times. Moreover, response-times in the current study were

approximately 500ms and 200ms longer (for 7- and 12-year-olds respectively) than those reported in previous studies of adult word learning (e.g., Gaskell & Dumay, 2003b). It will be important for further studies to investigate the time-course with which new words engage in lexical competition using different tasks such as pause detection (Mattys & Clark, 2002) or eye-tracking, both of which provide more online measures of lexical competition than the auditory lexical decision task. Neither require a metalinguistic judgment to be made, and as such should be less subject to strategic biases (*e.g.,* explicit awareness of the overlap between basewords and novel competitors) than auditory lexical decision.

Conclusions

The two experiments reported here indicate that the establishment of representations of new words in 7- and 12-year-old children is not complete immediately after encountering the novel word; offline consolidation appears to play an important role in the stabilization and enhancement of novel phonological representation, as demonstrated by the overnight improvements in cued recall. Most importantly, whilst recognition of novel nonwords appears to improve after as little as 3-4 hours post-exposure, improvements in cued recall are only observed when the children are tested one day later. Whilst the design used in these two experiments cannot disentangle the roles of sleep and time in the process of offline consolidation we have argued, based on previous research showing the beneficial influence on sleep on declarative memory in children (Backhaus, et al., 2008; Wilhelm, et al., 2008), and research in adults demonstrating a key role for sleep in word learning (Davis, et al., 2009; Dumay & Gaskell, 2007; Tamminen et al., 2010), that sleep may have an important role to play in the offline consolidation of novel phonological representations in children.

Appendix A

List 1

Basewords	Novel word	Foil word	Rated AoA	No. of syllables	No. of phonemes	Celex Frequency	EWFG Frequency	Length (ms)	Cue Length (ms)
apricot	aprickel	apricken	6.17	3	9	3	39.0	845	223
baboon	babeel	babeen	6.76	2	6	4	41.2	715	248
blossom	blossail	blossain	6.41	2	7	2	46.2	744	284
bramble	brambooce	bramboof	6.48	2	7	2	40.2	782	300
caravan	caravoth	caravel	6.00	3	9	3	44.9	871	300
cathedral	cathedruke	cathedruce	7.28	3	10	3	47.8	922	238
crocodile	crocodiss	crocodin	4.48	3	10	6	46.4	875	234
dolphin	dolpheg	dolphess	5.93	2	7	3	48.0	797	242
fountain	fountel	founted	5.59	2	7	12	50.0	948	319
lantern	lantobe	lantoke	6.55	2	7	2	48.8	853	302
octopus	octopoth	octopol	5.55	3	9	2	45.6	1002	295
partridge	partred	partren	7.24	2	7	10	41.8	954	235
skeleton	skeletobe	skeletope	5.76	3	9	3	52.1	994	317

List 2

Basewords	Novel word	Foil word	Rated AoA	No. of syllables	No. of phonemes	Celex Frequency	EWFG Frequency	Length (ms)	Cue Length (ms)
badminton	badmintel	badmintet	7.38	3	11	3	32.2	887	288
biscuit	biscal	biscan	3.28	2	7	15	44.4	811	244
cardigan	cardigite	cardigile	4.69	3	9	3	35.2	814	325
daffodil	daffadat	daffadan	4.86	3	9	3	36.3	801	235
dungeon	dungeill	dungeic	5.86	2	7	2	41.3	770	263
napkin	napkem	napkess	6.83	2	7	2	44.7	762	286
ornament	ornameast	ornameab	7.07	3	9	3	36.4	816	242
parachute	parasheff	parashen	6.45	3	9	3	46.5	898	272
pyramid	pyramon	pyramotch	7.10	3	9	3	50.0	759	272
siren	siridge	sirit	6.17	2	8	5	45.4	832	353
squirrel	squirrome	squirrope	5.21	2	7	2	51.7	852	332
tulip	tulode	tulome	5.72	2	7	2	42.3	856	270
yoghurt	yogem	yogell	6.72	2	6	4	41.5	759	276

References

Altmann, G. T. M. (1997). *The ascent of Babel: An exploration of language, mind, and understanding*. Oxford: Oxford University Press.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Philadelphia.

Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F., & Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiology of Learning and Memory, 89*(1), 76-80.

Booth, A. E. (2009). Causal supports for early word learning. *Child Development, 80*(4), 1243-1250.

Carey, S. (1978). The child as a word learner. In M. Halle, J. Bresnan & A. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 264-293). Cambridge, MA: MIT Press.

Carroll, J. B., & White, M. N. (1973). Age-of-acquisition norms for 220 pictureable nouns. *Journal of Verbal Learning and Verbal Behaviour, 12*, 563-576.

Castles, A., Davis, C., Cavalot, P., & Forster, K. I. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology, 97*, 165-182.

Church, B. A., & Fischer, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language, 39*, 523-542.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*.

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

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

Dumay, N., Gaskell, M. G., & Feng, X. (2004). *A day in the life of a spoken word.* Paper presented at the Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society.

Fischer, S., Wilhelm, I., & Born, J. (2007). Developmental differences in sleep's role for implicit off-line learning: Comparing children with adults. *Journal of Cognitive Neuroscience, 19*(2), 214-227.

Forster, J. C., & Forster, K. I. (2003). DMDX: A Windows display program with millisecond accuracy. *Behaviour Research Methods, Instruments, & Computers, 35*, 116-124.

Frazier Norbury, C., Griffiths, H., & Nation, K. (2010). Sound before meaning: Word learning in autistic disorders. *Neuropsychologia, 48,* 4012-4019.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences, 3*(4), 128-135.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*, 468-492.

Gaskell, M. G., & Dumay, N. (2003a). *Effects of vocabulary acquisition on lexical competition in speech perception and production.* Paper presented at the Proceedings of the 5th ICPhS Conference.

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

Horst, J. S., McMurray, B., & Samuelson, L. K. (2006). *Online processing is essential for learning: Understanding fast mapping and word learning in a dynamic connectionist architecture.* Paper presented at the Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*(2), 128-157.

Houston, D. M., Tincoff, R., & Jusczyk, P. W. (2001). 7.5-month olds' memory for words after a 1-week delay. In D. M. Houston, A. Seidl, G. Hollich, E. Johnson & A. Jusczyk (Eds.), Jusczyk Lab Final Report

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sounds patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science, 277*, 1984-1986.

Lyons, A. W., Teer, P., & Rubensten, H. (1978). Age-of-acquisition and word recognition. *Journal of Psycholinguistic Research, 7*, 1790187.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. . *Journal of Experimental Psychology-Human Perception and Performance, 15*(3), 576-585.

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., 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), 419-457.

McGregor, K. K., Rohlfing, K. J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of *under. Journal of Child Language, 36,* 807-828.

Menon, V., Boyett-Anderson, J. M., & Reiss, A. L. (2005). Maturation of medial temporal lobe response and connectivity during memory encoding. *Cognitive Brain Research, 25*(1), 379-385.

Metsala, J. L., Stavrinos, D., & Walley, A. C. (2009). Children's spoken word recognition and contributions to phonological awareness and nonword repetition: A 1-year follow-up. *Applied Psycholinguistics, 30*(1), 101-121.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89-120). Mahwah, NJ: Erlbaum.

Morrison, C. A., Chappell, T. D., & Ellis, A. (1997). Age-of-acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology, 50*, 528-559.

Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech Language and Hearing Research, 48*(1), 108-124.

Newman, R. S. (2008). The level of detail in infants' word learning. *Current Directions in Psychological Science, 17*(3), 229-232.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110*(4), 611-646.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences, 6*(12), 505-510.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review, 108*(2), 311-345.

Ojima, S., Matsuba-Kurita, H., Nakamura, N., & Hagiwara, H. (2011). The acceleration of spoken-word processing in children's native-language acquistion: an ERP cohort study. *Neuropsychologia, 49*(5), 790-799.

Plihal, W., & Born, J. (1999). Effects of early and late nocturnal sleep on priming and spatial memory. *Psychophysiology, 36*(5), 571-582.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. [Article]. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*(3), 785-794.

Prehn-Kristensen, A., Göder, R., Chirobeja, S., Breßmann, I., Ferstl, R., & Baving, L. (2009). Sleep in children enhances preferentially emotional declarative but not procedural memories. *Journal of Experimental Psychology, 104*, 132-139.

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology, 24*, 755-789.

Sekerina, I. A., & Brooks, P. J. (2007). Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology, 98*, 20-45.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience, 2*(10), 859-861.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology, 109*(1), 132-140.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598-605.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech Language and Hearing Research, 44*(6), 1321-1337.

Storkel, H. L., & Hoover, J. R. (2011). The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *Journal of Child Language, 38*, 628-643.

Storkel, H. L., & Lee, S. Y. (2011). The independent effects of phonotactic probability and neighborhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26*(2), 191-211.

Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics & Phonetics, 14*, 407-425.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*(2), 73-108.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology, 61*(3), 361-371.

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.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13*, 286-350.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*, 1-30.

Wilhelm, I., Diekelmann, S., & Born, J. (2008). Sleep in children improves memory performance on declarative but not procedural tasks. *Learning & Memory, 15*(5), 373-377.

Footnotes

[1] The exposure phase was more varied than in previous adult studies, in which only phoneme monitoring was used (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b; Tamminen & Gaskell, 2008; Tamminen et al., 2011). This change was made in an effort to make the exposure blocks shorter, and more engaging for the children.

[2] Cohen's d is reported as a measure of effect size for t-test comparisons ('small' effect = .2, 'medium' effect = .5, 'large' effect = .8) whilst partial eta squared (η_p^2) is reported as the measure of effect size for main effects and interactions in ANOVAs ('small' effect = .01, 'medium' effect = .09, 'large' effect = .25).

Table 1.

Mean accuracy (%) in the 2afc and cued recall tasks in Experiments 1. Standard error scores are provided in parentheses.

	2afc		Cued recall	
	am-pm	am-am	am-pm	am-am
Session 1	55.9 (7.0)	60.7 (6.6)	4.6 (1.9)	7.7 (2.9)
Session 2	67.8 (7.5)	72.7 (7.9)	5.9 (2.5)	19.7 (4.1)

Table 2.

Mean accuracy (%) in the 2afc and cued recall tasks in Experiments 2. Standard error scores are provided in parentheses.

	2afc		Cued recall	
	7 years	12 years	7 years	12 years
Session 1	86.1 (2.3)	93.6 (0.9)	8.9 (2.1)	28.7 (3.2)
Session 2	90.4 (1.9)	96.1 (0.5)	23.0 (3.6)	50.9 (3.9)

Figure 1.

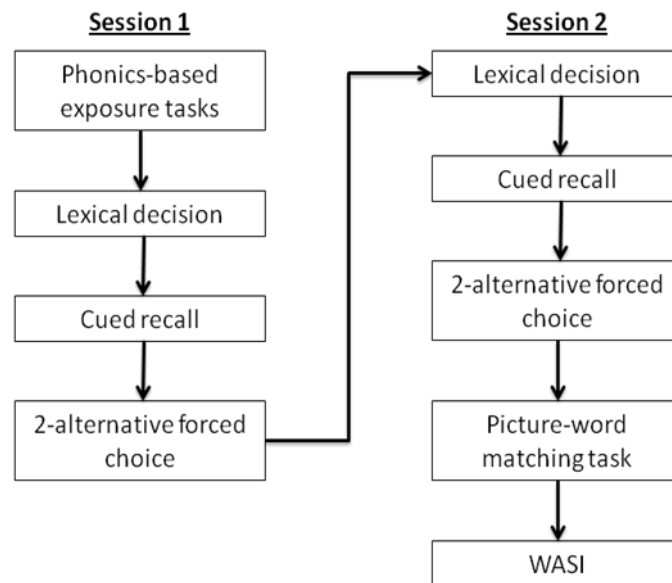Overview of the experimental tasks completed in each session.

Figure 2.

Mean difference between response-times to control (no novel competitor) and test (novel competitor) basewords in the lexical decision task in Experiment 1. Negative values reflect slower response-times for test compared to control basewords, and are indicative of increased lexical competition for the test basewords as a result of learning a novel nonword competitor. Error bars indicate ± standard error.