Published in final edited form as:
J Speech Lang Hear Res. 2014 October 1; 57(5): 1842–1850. doi:10.1044/2014_JSLHR-L-13-0273.

What a Difference a Day Makes: Change in Memory for Newly Learned Word Forms Over 24 Hours

Karla K. McGregor[a]
[a]The University of Iowa

Abstract

Purpose—This study explored the role of time and retrieval experience in the consolidation of word forms.

Method—Participants were 106 adults trained on 16 novel word-referent pairs, then tested immediately and 24 hr later for recognition and recall of word forms. In the interim, tests were repeated 2 hr or 12 hr after training, or not at all, thus varying the amount and timing of retrieval experience.

Results—Recognition accuracy was stable and speed improved over the 24-hr period. But these manifestations of consolidation did not depend on interim retrieval experience; in fact, the 2-hr interim test interfered with improvements in speed. In contrast, the number of word forms recalled increased only with interim retrieval experiences, and the 12-hr interim test was more advantageous to recall than the 2-hr test.

Conclusions—After a word form is encoded, it can become stronger with time. Retrieval experience can also strengthen the trace, but, if retrieval occurs when the memory is still labile, it can be disruptive. This complex interplay between retrieval experience and time holds implications for measuring learning outcomes and for scheduling practice in classrooms and clinics.

Keywords

language; memory; development

Learners of all ages take part in vocabulary instruction. In the early years of formal schooling, children receive vocabulary lessons to promote reading comprehension, spelling ability, and general content knowledge. An emphasis on vocabulary learning in support of field-specific content knowledge and foreign language learning continues into postsecondary education and professional settings. Moreover, for children and adults with communication or memory disorders, vocabulary instruction is a frequent aspect of clinical intervention.

But learning a new word is no small feat. A complete lexical configuration can be viewed as a set of factual information that the word learner has stored in memory— the phonological

Correspondence to Karla K. McGregor: karla-mcgregor@uiowa.edu.
Disclosure: The author has declared that no competing interests existed at the time of publication.

and orthographic form of the word as well as its syntactic class and meaning and the pragmatic constraints on its usage (Leach & Samuel, 2007). The complete configuration is built over time and over multiple experiences with that word (Bloom, 2000).

In the current article we ask how memory for one aspect of the lexical configuration, phonological form, changes over the course of a day given the mere passage of time and the opportunity for additional experience with the word. Ultimately, we seek to understand the conditions that promote the establishment of spoken word form information in the lexicon in hopes of determining the processes at play and the practical implications for vocabulary instruction. Below we review the current understanding of word form learning in response to overt experience as well as change over time in the absence of these experiences.

The Role of Experience in Learning and Remembering Words

In the simplest terms, a learner must be exposed to a word to encode it in memory. Adults speak (and therefore presumably hear) 16,000 or 17,000 words per day (Hart & Risley, 1995; Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007). Estimates for children ages 5 through 15 years are even higher (Wagner, 1985). Exposure to words occurs incidentally as learners read, participate in, and overhear conversations, attend lectures, and listen to media on the computer, television, or radio (Gupta & Tisdale, 2009a). Exposure also occurs in didactic settings as learners engage in instruction specifically aimed at teaching words.

In classic fast-mapping experiments in which the learner is presented with a novel word and a novel object in the presence of one or more familiar objects, child and adult learners readily infer the link between the novel word form and the novel object (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Halberda, 2006), and they can recognize the form and link after a minimal delay (Carey, 1978; Dollaghan, 1985; Golinkoff et al., 1992). The phenomenon, in fact, is termed *fast mapping* because only one or two exposures are necessary. Nevertheless, fast-mapped lexical configurations are fragile, and recognition tests administered even 5 min after the initial exposure reveal that they tend not to be retained (Horst & Samuelson, 2008).

In contexts that are more didactic, for example, when the object referent is held up by the examiner and named repeatedly, recognition performance after a delay improves. Nevertheless, the resulting memory trace is still rarely strong enough to support production (Booth, McGregor, & Rohlfing, 2008; Munro, Baker, McGregor, Docking, & Arculi, 2012). Instead, multiple exposures over time appear necessary. Under didactic conditions, 8-year-olds who received 16 exposures to each of 20 foreign words during each of two weekly lessons could recall only eight of those words, on average, during a referent-naming test 1 week later and only nine when the test was repeated 1 month later (McGregor, Sheng, & Ball, 2007).

Although multiple exposures are necessary, not all exposures are equally advantageous to the learner. In didactic contexts, test exposures enhance retention of newly encoded information more than additional training exposures even when the tests do not include feedback (see the review in Roediger & Karpicke, 2006). There are a number of reasons that testing might help. First, testing almost always constitutes additional exposure to relevant information. One cannot measure the speed of a participant's word recognition or the accuracy of her word definitions without presenting the word itself, for example. But of course this cannot explain the advantage of testing over training exposures. A more relevant explanation is that, to the extent that the processes involved in testing as a means of review and the processes involved in testing for final retention overlap, test reviews improve final performance by enabling transfer-appropriate processing (Karpicke & Roediger, 2007). Simply put, practicing retrieval improves retrieval. There is a related explanation to consider as well: Successful retrieval activates the young memory trace, causing it to become destabilized or malleable. In this destabilized state, the memory trace may be strengthened (or degraded) (Walker & Stickgold, 2006). We shall return to this point below.

The Role of Time in Learning and Remembering Words

Although incidental and didactic experiences drive the initial encoding of the lexical configuration in memory and enable encoding of additional aspects of the lexical configuration over time, once encoded, memory for words can also change in the absence of these experiences, that is, without re-exposure or retrieval in response to overt testing of the newly learned information. Sometimes this change is characterized by forgetting. As time passes, memories decay; however, the proportional rate of decay *decreases* over time (Wixted, 2004). That is, the longer a memory survives, the less likely it is to be forgotten. Consolidation is proposed as the counterforce that slows the rate of forgetting (Wixted, 2004). Importantly, consolidation, like forgetting, unfolds over time.

Dudai (2004, p. 51) describes consolidation as a “progressive post acquisition” event in which new memories are strengthened. Consolidation is progressive in that it plays out over time in two overlapping phases operating in parallel (McGaugh, 2000). The first is synaptic consolidation that takes place within minutes or hours after encoding (Walker, 2005). As a result of modification and reorganization of synaptic proteins, the memory trace becomes more stable, that is, more resistant to interference. The second phase, system consolidation, requires a longer time course of hours, days, or even weeks. During system consolidation, resistance to interference continues to grow (Dudai & Eisenberg, 2004). Activation of the hippocampal trace sends synaptic messages to neocortical circuits, and these messages trigger local synaptic consolidation there (Dudai, 2004). One hypothesis is that the transfer allows for rapid learning via the hippocampus and gradual strengthening, formation of retrieval links, and integration of new information into the existing memory network at the level of the neocortex (Gupta & Tisdale, 2009b; McClelland, McNaughton, & O'Reilly, 1995; Walker, 2005).

Traditionally, system consolidation was thought to depend upon automatic unintentional processing of the newly learned material; conscious, behavioral rehearsal or other overt experiences were not considered germane (Walker, 2005). For example, unintentional processing can occur during sleep as we replay memories of the day, and there is evidence that sleep facilitates consolidation of declarative memory (Ellenbogen, Payne, & Stickgold, 2006), including memory for newly learned words (Dumay & Gaskell, 2007; Gais, Lucas, & Born, 2006).

More recently, debates on whether a memory can "reconsolidate" led to the recognition that overt experiences with the newly learned material, though not necessary, can trigger additional system consolidation (Stickgold & Walker, 2005; Suzuki et al., 2004). Reactivation of a (partially) consolidated memory trace can result in destabilization (Nader, Schafe, & Le Doux, 2000), which can, in turn, lead to degradation or reconsolidation (Stickgold & Walker, 2005). Consider an early experiment by Izquierdo and Chaves (1988). Their adult participants read a text on the 1954 World Cup and were tested 48 hr later on the accuracy of their recall. The participants were assigned to one of four conditions: factual text only or the factual text plus a nonfactual comment shown immediately, 3 hr, or 6 hr after the text. Participants shown a negative comment about the quality of play at the World Cup immediately or 3 hr after reading the text mistakenly recalled this information as part of the text; those shown the comment 6 hr after reading the text did not, thus demonstrating that reactivation of newly learned information too early in the consolidation process can be deleterious.

In summary, we have considered three memory processes that support the learning of words: encoding, consolidation, and retrieval. Encoding is the experience-dependent process via which a new memory is formed. Encoding is followed by consolidation, the slower time-dependent process in which the fragile newly encoded memory is strengthened. Retrieval is a window onto the outcomes of the encoding and consolidation processes. However, retrieval is itself a process of memory. Another way to think about this is to consider that, despite the train-then-test approaches that we take in the laboratory, classroom, and clinic, the brain can and does respond to both sorts of activities as learning opportunities. Testing prompts retrieval of the newly formed memory trace, and, once reactivated, that trace can be degraded or updated and further strengthened. These outcomes may depend upon the extent of consolidation achieved when the retrieval takes place.

The Current Study

In the current experiment, we examined evidence of consolidation of newly learned word forms over 24 hr and then asked whether and to what extent interim test experience improved word form configuration across that interval. The speed of retrieval in recognition tests and the accuracy of retrieval in recall tests were the dependent variables. We predicted that interim test experience with the word forms would enhance recognition of the word form and would increase the number of word forms recalled (Roediger & Karpicke, 2006).

We explored the optimal timing of the interim test experience by comparing the effect of one administered only 2 hr after training to the effect of another administered 12 hr after training. We predicted that the 2-hr interim test would be less advantageous than the 12-hr interim test to the establishment of the word form configuration because (a) it involved a shorter interval between initial and interim test—hence, the young memory trace would be less stabilized and more prone to degradation upon retrieval (Walker, 2005); and (b) it involved a longer retention interval between interim and final test, hence a greater opportunity for forgetting (Wixted, 2004).

Method

Participants

Participants were 106 English speakers (58 females, 48 males) whose ages ranged from 18 to 25 years and whose years of education ranged from 12 to 20. These participants were recruited and tested according to procedures approved by the institutional review board at the University of Iowa. All were students at the University of Iowa or nearby colleges with no history or current diagnosis of language learning impairments. They were paid hourly for their participation.

All participants were trained on new words and tested for retention 24 hr later. Participants were assigned to one of three conditions that varied with regard to interim test exposure between the training and 24-hr retention test (see Figure 1). Fifty-four participants received interim test exposures to the target words 2 hr after training; 37 received interim test exposures to the target words 12 hr after training; and 15 received no interim test exposure. The uneven number per condition reflects the fact that the data were collected at different points in time and for different purposes (although with identical procedures). Those in the 12-hr interim condition were part of a larger study reported in McGregor et al. (2013).

Prior to the experiment, all participants took the Peabody Picture Vocabulary Test–IV (Dunn & Dunn, 2007). Mean standard scores (and standard error) were 113 (1.54) for those in the 2-hr test condition, 114 (1.86) for those in the 12-hr test condition, and 115 (2.93) for those in the no-test condition. The participants in the three conditions did not differ by standard score, $F(2, 103) < 1$, or raw score, $F(2, 103) < 1$.

Stimuli

The stimuli were 32 pairs of novel words that diverged from each other at the final syllable (e.g., *armo* and *armu*). In each pair, one novel word was trained and the other was an untrained foil in a two-alternative forced-choice (2AFC) task. Each novel word to be trained was randomly assigned to a novel referent created by combining two objects (e.g., a pony and a snake). The 32 novel words and referents were divided into two sets of 16, one of which was trained with selection of training set counterbalanced across participants. Each novel word was recorded twice by a female native speaker of American English and digitized at a sampling rate of 44,110 kilobytes per second. One recording was used for training, the other for testing. A complete list of the words and referents is available in McGregor et al. (2013).

Procedure

Each participant was seen for a training visit during which they were trained on the new words and referents and then tested immediately. Each participant in the 2-hr interim condition was tested (but not trained) again 2 hr after training. Each participant in the 12-hr interim condition was tested (but not trained) again 12 hr after training. Participants in all three conditions were seen 24 hr after training and the tests were repeated.

The participants in the 12-hr interim condition were part of a larger study of the role of sleep on verbal memory consolidation (McGregor et al., 2013). In that study, to manipulate sleep, half of the participants were scheduled for their training visit in the morning and their 12-hr interim visit in the evening and were instructed not to sleep in the interim. The other half were scheduled for their training visit in the evening and their 12-hr interim visit the next morning and were instructed to sleep in the interim as per usual. Crucially, there were no effects of morning/no sleep vs. evening/sleep on word form recognition or recall; so, in the current study, these two groups were collapsed to allow investigation of the effect of timing of interim test exposure (rather than sleep) on memory for word form. That said, to isolate the effect of timing, the 2-hr and 12-hr interim experiences should be comparable. And they were. Participants assigned to the 2-hr test condition remained in the lab during the interim, where half of them played computer games and half of them napped. Polysomnography verified that every napper did sleep. This procedure allowed comparison of interim testing at 2 and 12 hr while controlling for sleep exposure. We did subsequently compare the performance of nappers and gamers, but, as in McGregor et al. (2013), we found no effect of sleep on word form recognition or recall. In the current study, these two groups are collapsed.

One limitation of this post hoc analysis of three extant data sets is that participants were scheduled for their training visits at varying times of day. Participants in the no-test condition were scheduled for their training visit in the evening; half of those in the 12-hr condition were scheduled in the morning and half in the evening; and those in the 2-hr condition were scheduled in the midafternoon. We already know from McGregor et al. (2013) that there were no differences in the performance of participants scheduled in the morning versus the evening. However, we cannot assume that performance in the afternoon would not differ from morning or evening given that performance in a range of cognitive tasks does fluctuate with time of day (Schmidt, Collette, Cajochen, & Peigneux, 2007). In other words, any difference between the 2-hr and 12-hr conditions could reflect the hypothesized effects of the time interval between training and interim test or it could reflect effects of circadian rhythms on performance. In the analysis below, we attempt to tease apart these confounds by comparing performance immediately after training and the change in performance from immediate testing to 24 hr. If circadian rhythms affect performance, we should see differences at both points; if, as hypothesized, the length of the interval between training and interim test affects performance, we should see differences in change between the points only.

Training—Participants were trained via a computerized script. In each of 12 blocks, the 16 novel referents were displayed on the screen one at a time for 7 s and the associated novel word was presented via digital audio-tape four times within a two-sentence description. To ensure that the participant was actively engaged, a question about either the sounds in the novel word (e.g., “does it start with /p /?”) or the physical features of its referent (e.g., “is it edible?”) was presented after a third of all items. No feedback on accuracy was provided. The complete training script resulted in 12 exposures to each referent and verbal description and 48 exposures to each word form. Order of presentation was randomized within blocks and order of blocks was counterbalanced across participants.

Tests—All participants were given a battery of tests that required recognition of form and meaning, lexical decision judgments, word associations, and free recall (see McGregor et al., 2013, for details). Thus, the tests involved a number of factors known to enhance development of the lexical configuration, including exposure, effort, and production practice (Leach & Samuel, 2007). Because the current article focuses exclusively on memory for word forms, we will present the results from only two of these probes. This does not mean that exposure to and retrieval of information on the other tests did not affect performance; but, critically, the test battery was identical for all participants in all conditions, so this fact cannot account for differences between conditions.

A 2AFC test measured auditory recognition of the newly learned word forms. Each trained word (e.g., *armo*) was presented along with its untrained lexical neighbor (e.g., *armu*) separated by an interstimulus interval of 500 ms. The participants pressed a button to indicate whether the first or second form was familiar. Versions of the probe varied in order of item presentation with assignment of versions to testing period counterbalanced across participants. Responses were scored for accuracy and also for reaction time (RT).

In the free recall test, participants orally recalled as many of the trained words as possible in a 2-min period. The dependent variable was number of words recalled. Words only needed to be recognizable as one of the trained words to be tallied; 100% accuracy was not required. Two investigators, blind to condition assignment, independently transcribed 20% of the data from digital audio files. Their point-to-point agreement on which word was being attempted was 100%. After establishing this level of agreement, one of the investigators transcribed the remaining data and coded it for words recalled.

Results

2AFC

Accuracy of recognition on the 2AFC task was at ceiling, falling at or above 96% for all conditions at all test points (see Table 1). Therefore, we turned to reaction time (median RT for accurate responses only) for a more sensitive measure of variation according to condition and test time. First, to examine possible effects of circadian rhythms, we compared RT immediately after training and found no significant differences between participants in the three interim test conditions whose training sessions were scheduled at different times of day (2-hr test: $M = 513$, $SE = 31.8$; 12-hr test: $M = 602$, $SE = 38.5$; no-test: $M = 550$, $SE = 60.4$), $F(2, 103) = 1.57$, $p = .21$.

We were most interested in change over time relative to the immediate posttest. To determine whether interim test exposures served to improve word recognition across the 24-hr interval, we applied a one-way analysis of variance (ANOVA) and post hoc Tukey's honestly significant difference (HSD) tests for unequal Ns. The dependent variable was RT difference scores (24-hr performance – immediate performance). Change in speed of recognition is presented in Figure 2. There was a significant effect of condition, $F(2, 103) = 15.45$, $p < .0001$, partial $\eta^2 = .23$. Change in RT in the 2-hr test condition differed from change in the 12-hr test condition, $p = .0001$, and the no-test condition, $p = .05$. The 12-hr and no-test conditions did not differ. Change in the 12-hr condition was significantly different than 0, $t = -8.83$, $df = 36$, $p < .0001$, as was change in the no-test condition, $t = -4.13$, $df = 14$, $p = .001$, with change in both cases involving decreased RTs. Change in the 2-hr condition was not significantly different than 0, $t = 0.44$, $df = 53$, $p = .66$.

To better interpret the disadvantage of testing 2 hr after training, we compared recognition RT at each of the three testing points in the interim conditions using a mixed ANOVA with interim condition (12 hr, 2 hr) as a between-subjects variable and with testing point (immediate, interim, final) as a within-subject variable. There was a significant effect of testing point, $F(2, 178) = 8.45$, $p = .0003$, partial $\eta^2 = .09$, qualified by a Testing Point × Interim Condition interaction, $F(2, 178) = 11.26$, $p < .0001$, partial $\eta^2 = .11$ (Figure 3). A Bonferroni post hoc test revealed that the 12-hr interim test group was significantly faster to respond at the final test than at either of the previous testing points, $p \leq .03$. The two previous testing points did not differ. In contrast, the 2-hr interim test group did not differ in RT at any testing point relative to another.

Word Form Recall

First, to examine possible effects of circadian rhythms, we compared the number of words recalled immediately after training and found no significant differences between participants in the three interim test conditions whose training sessions were scheduled at different times of day (2-hr test: $M = 5.8$, $SE = .42$; 12-hr test: $M = 6.1$, $SE = .51$; no-test: $M = 6.5$, $SE = .80$), $F(2, 103) < 1$.

We determined whether interim test exposures served to improve word form recall over the 24-hr interval by applying a one-way ANOVA and post hoc HSD tests for unequal Ns. The dependent variable was a difference score calculated as the number of words recalled at 24 hr minus the number recalled immediately after training. These difference scores are presented in Figure 4. There was a main effect of condition, $F(2, 103) = 8.65$, $p = .0003$, partial $\eta^2 = .14$, such that those in the 12-hr test condition differed significantly from those in the no-test condition, $p = .003$, and marginally from those in the 2-hr test condition, $p = .054$. The no-test and 2-hr test conditions did not differ. The change in the 12-hr condition was significantly greater than 0, $t = 5.4$, $df = 36$, $p < .0001$, as was the change in the 2-hr condition, $t = 2.96$, $df = 53$, $p = .005$; in both cases the change involved increased recall. The change in the no-test condition was not significant, $t = -1.5$, $df = 14$, $p = .16$.

To better interpret the advantage of testing at 12 hr, we compared number of words recalled by participants in the 2- and 12-hr interim condition at each of the three testing points. A mixed ANOVA with interim condition (2 hr, 12 hr) as a between-subjects variable and testing point (immediate, interim, final) as a within-subject variable yielded a significant effect of testing point, $F(2, 178) = 33.92$, $p < .0001$, partial $\eta^2 = .28$, qualified by a Testing Point × Interim Condition interaction, $F(2, 178) = 4.05$, $p = .02$, partial $\eta^2 = .04$ (Figure 5). A Bonferroni post hoc test revealed that the 12-hr interim test group recalled significantly more words at the final test than at either of the previous testing points, $p < .0001$. The two previous testing points did not differ. In contrast, the 2-hr interim test group recalled significantly fewer words at the interim testing point than at the immediate and final testing points, $p \leq .01$. There was a marginally significant gain in number of words recalled from the immediate to the final testing point, $p = .08$.

Discussion

In this study we explored the role of time and retrieval experience in the consolidation of newly learned word forms. We found the recognition of word forms to consolidate over the 24 hr following training in the absence of overt experience involving retrieval. This consolidation was manifested as stable accuracy and increased speed of recognition. Overt experience requiring retrieval did not facilitate this increase; in fact when offered too close to training, it was deleterious. Recall revealed a different pattern of change. In the absence of overt retrieval experience during the interim, consolidation was manifested as stability. The same number of words was recalled at immediate and 24-hr posttests. But with overt retrieval experience in the interim, word recall improved by one or two words, on average, from immediate to 24-hr posttest depending on the timing of the interim test. Below we interpret these findings as they relate to current conceptions of consolidation and reconsolidation.

Recognition

Accuracy of recognition was stable over the 24 hr of the experiment. The experience provided by an interim test did not change the accuracy of word recognition, for better or worse. Because accuracy was at ceiling, the pattern of change in speed of recognition was more revealing. For participants in the 12-hr interim test condition, RT decreased significantly over the 24-hr period. However, this was not the result of interim testing, as participants in the no-interim-test condition evinced decreases of similar magnitude. We conclude instead that the faster speed of response reflected consolidation over time. The pattern of results bears similarity to results reported in Davis, Di Betta, Macdonald, and Gaskell (2008). They taught university students 18 novel words on Day 1; taught them another 18 on Day 2; then, also on Day 2, tested performance on both sets. On a task requiring the students to repeat the word forms as quickly as possible, the researchers found that words trained on the first day were repeated more quickly than words trained on the second day. They concluded that the difference reflected overnight changes in the representation of novel word forms. In a subsequent experiment, they found different neural signatures for Day 1 and Day 2 words as measured by fMRI.

A different, and seemingly contradictory, pattern of change characterized the participants in the 2-hr interim condition. Participants in the 2-hr interim condition demonstrated nearly identical response times at the initial, interim, and final test. If 24 hr brought decreased response times for the other participants, why not for them? We maintain that the answer lies in the difference in timing of the interim test experience relative to training. By forcing retrieval only 2 hr posttraining, the still labile memory trace was disrupted. We shall explore the temporal sensitivity of consolidation in more detail after considering the word form recall findings.

Recall

Unlike recognition, recall was far from ceiling immediately after training. Of the 16 words that were trained, participants recalled only five or six, on average. In the no-interim-test condition, there was no significant change in performance from immediately after training to

24 hr later. Therefore, given only the passage of time, consolidation served to stabilize the memory trace.

But additional testing yielded additional learning. There are at least two likely reasons. First, the interim test session provided additional exposure to the correct word forms. Recall that word association and lexical decision tasks were part of the complete protocol. Thus, they received exposures to the word forms during these tasks as well as during the 2AFC task. Second, the interim and final tests were identical, thus introducing the possibility of transfer-appropriate processing (Karpicke & Roediger, 2007). One potential curiosity about the effectiveness of the test experience is that the examiner provided no feedback on the participant's test performance. Although testing without feedback can be a very effective way to boost retention (Roediger & Karpicke, 2006), this is typically the case only after initial training that results in high levels of performance (Pyc & Rawson, 2009). Here, performance on word recall was low, with fewer than half of the trained words recalled at interim test. Difficult retrieval that results in failure is not conducive to memory; instead, difficult but *successful* defines the construct of "desirable difficulty" (Pyc & Rawson, 2009). However, note that correct exposures from other test probes together with the participants' excellent recognition of the word forms constituted implicit feedback on their recall efforts.

Although testing 2 hr after training was preferable to not testing at all, it was marginally less useful than testing 12 hr after training. Those in the 2-hr condition gained one word over the 24-hr interval; those in the 12-hr condition gained two. Cepeda and colleagues reviewed 14 studies that compared the effect of practicing newly learned information closer to or further from initial encoding. In every one, a gap of 1 day or more between encoding and practice was better than a gap of 3 hr or less in promoting accuracy on a final retention test (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). In these 14 studies, the interval between practice and retention test was held constant, so the effect was necessarily attributable to the different intervals between encoding and practice. But in the current study, participants in the 2-hr condition had to retain the word forms for 22 hr while those in the 12-hr condition had to retain for only 12 hr. But an interpretation based on the length of the final retention period seems unsatisfactory here given that both those in the 12-hr and those in the 2-hr condition demonstrated gains indicative of recon-solidation during that interval. Instead, the disadvantage of testing 2 hr after training was more evident between training and interim test. Those in the 12-hr condition demonstrated stable performance from immediate to interim tests (much as those in the no-test condition did over the entire 24-hr interval), but those in the 2-hr condition demonstrated a decline in performance. To explain, we again appeal to the temporal characteristics of consolidation. Apparently 12 hr, but not 2 hr, was enough to stabilize the memory trace.

Animal models are highly informative to understanding the dynamic interplay between time and retrieval in support of memory. For example, memory for a conditioned fear in mice is disrupted by protein synthesis inhibitors when administered immediately after training and also when administered later after retrieval (Lattal & Abel, 2004). Importantly, the disruption is more likely when retrieval follows closer rather than further from training (Suzuki et al., 2004). In other words, new memories are weak and therefore more prone to interference than older memories. Alberini (2005) puts it this way: "each training and reactivation event contributes to a gradient of stabilization that gradually increases and eventually results in a fully consolidated trace that is insensitive to disruption. However, during the gradient, the still-unstable trace can become labile if activated by a modulation event" (p. 55).

Conclusions

Over the course of a single day, the ability to recognize newly learned word forms remained stable while speed of recognition improved. This consolidation was fueled solely by the passage of time. Interim exposure and associated retrieval experience did not contribute to the effect; in fact, enhancements in speed of recognition were blocked when interim testing occurred soon after training. This pattern stands in contrast to the pattern of change in word form recall. With the passage of time alone, word form recall remained stable. With additional experience via interim test exposure, word form recall improved; but, when the interim test occurred too soon after initial training, that improvement was minimal. Although not predicted, the different outcomes for recognition and recall are not surprising because the two tasks involve different retrieval cues, response types, and scales of measurement. Moreover, these differences are compatible with current dual-process theories of memory that posit that a sense of familiarity can enable recognition but not recall (Yonelinas, 2002). Distinct neuroanatomical support for recognition and recall is also in evidence (O'Reilly & Norman, 2002). These differences are germane to understanding why word learners typically need more exposure to produce new words than to recognize them (Booth et al., 2008; Munro et al., 2012).

A replication of these results with an a priori design in which participants are randomly assigned to conditions is called for. Although we found no indication of differences in performance that can be attributed to circadian rhythms, elimination of the confound between time of day in which training occurs and the interval between training and interim test is also needed. Finally, an important next step toward applying these findings is to determine how the dynamics of word learning and memory unfold over much longer, more relevant time periods. For now, the conclusions are more general: The lexicon is not a passive memory bank. Instead, lexical configurations are dynamic. We offer preliminary evidence that they change via a complex interplay between experience and time. The same experience that can help learning and memory can also disrupt learning and memory, depending on *when* it occurs. These findings hold implications for measuring word learning outcomes and, ultimately, for scheduling practice and review of verbal material learned in classrooms and clinics.

Acknowledgments

Thanks goes to Nichole Eden for data collection, Tim Arbisi-Kelm for data extraction, Mark Eric Dyken for interpretation of the polysomnography recordings, and Prahlad Gupta for comments on an earlier draft. This work was supported by Grant R21 DC009292-02S1 and an American Recovery and Reinvestment Act of 2009 extension, both from the National Institute on Deafness and Other Communication Disorders.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

References

Alberini CM. Mechanisms of memory stabilization: Are consolidation and reconsolidation similar or distinct processes? Trends in Neurosciences. 2005; 28:51–56. [PubMed: 15626497]

Bloom, P. How children learn the meanings of words. Cambridge, MA: MIT Press; 2000.

Booth AE, McGregor KK, Rohlfing KJ. Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. Language Learning and Development. 2008; 4:179–202.

Carey, S. The child as word learner. In: Halle, M.; Bresnan, J.; Miller, GA., editors. Linguistic theory and psychological reality. Cambridge, MA: MIT Press; 1978. p. 264-293.

Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychological Bulletin. 2006; 132:354–380. [PubMed: 16719566]

Davis MH, Di Betta AM, Macdonald MJE, Gaskell MG. Learning and consolidation of novel spoken words. Journal of Cognitive Neuroscience. 2008; 21:803–820. [PubMed: 18578598]

Dollaghan C. Child meets word: Fast mapping in preschool children. Journal of Speech and Hearing Research. 1985; 28:449–454. [PubMed: 4046586]

Dudai Y. The neurobiology of consolidations, or, how stable is the engram? Annual Review of Psychology. 2004; 55:51–86.

Dudai Y, Eisenberg M. Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. Neuron. 2004; 44:93–100. [PubMed: 15450162]

Dumay N, Gaskell MG. Sleep-associated changes in the mental representation of spoken words. Psychological Science. 2007; 18:35–39. [PubMed: 17362375]

Dunn, LM.; Dunn, DM. Peabody Picture Vocabulary Test—Fourth Edition. San Antonio, TX: Pearson; 2007.

Ellenbogen JM, Payne JD, Stickgold R. The role of sleep in declarative memory consolidation: Passive, permissive, active or none? Current Opinion in Neurobiology. 2006; 16:716–722. [PubMed: 17085038]

Gais S, Lucas B, Born J. Sleep after learning aids memory recall. Learning & Memory. 2006; 13:259–262. [PubMed: 16741280]

Golinkoff RM, Hirsh-Pasek K, Bailey LM, Wenger RN. Young children and adults use lexical principles to learn new nouns. Developmental Psychology. 1992; 28:99–108.

Gupta P, Tisdale J. Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. Journal of Memory and Language. 2009a; 61:481–502.

Gupta P, Tisdale J. Word learning, phonological short-term memory, phonotactic probability and long-term memory: Towards an integrated framework. Philosophical Transactions of the Royal Society B: Biological Sciences. 2009b; 364:3755–3771.

Halberda J. Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. Cognitive Psychology. 2006; 53:310–344. [PubMed: 16875685]

Hart, B.; Risley, T. Meaningful differences in the everyday experience of young American children. Baltimore, MD: Brookes; 1995.

Horst JS, Samuelson L. Fast mapping but poor retention in 24-month-old infants. Infancy. 2008; 13:128–157.

Izquierdo I, Chaves MLF. The effect of non-factual post-training negative comment on the recall of verbal information. Journal of Psychiatric Research. 1988; 22:165–169. [PubMed: 3225787]

Karpicke JD, Roediger HL. Repeated retrieval during learning is the key to long term retention. Journal of Memory and Language. 2007; 57:151–162.

Lattal KM, Abel T. Behavioral impairments caused by injections of the protein synthesis inhibitor anisomycin after contextual retrieval reverse with time. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:4667–4672. [PubMed: 15070775]

Leach L, Samuel AG. Lexical configuration and lexical engagement: When adults learn new words. Cognitive Psychology. 2007; 55:306–353. [PubMed: 17367775]

McClelland JL, McNaughton BL, O'Reilly RC. 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. 1995; 102:419–457. [PubMed: 7624455]

McGaugh JL. Memory: A century of consolidation. Science. 2000 Jan 14.14:248–251. [PubMed: 10634773]

McGregor KK, Licandro U, Arenas R, Eden N, Stiles D, Bean A, Walker E. Why words are hard for adults with developmental language impairments. Journal of Speech, Language, and Hearing Research. 2013; 56:1845–1856.

McGregor KK, Sheng L, Ball T. Complexities of expressive word learning over time. Language, Speech, and Hearing Services in Schools. 2007; 38:353–364.

Mehl MR, Vazire S, Ramirez-Esparza N, Slatcher RB, Pennebaker JW. Are women really more talkative than men? Science. 2007 Jun 6.317:82. [PubMed: 17615349]

Munro N, Baker E, McGregor KK, Docking K, Arculi J. Why word learning is not fast. Frontiers in Developmental Psychology. 2012; 3:41.

Nader K, Schafe GE, Le Doux J. Reply— Reconsolidation: The labile nature of consolidation theory. Nature Reviews Neuroscience. 2000; 1:216–219.

O'Reilly R, Norman KA. Hippocampal and neocor-tical contributions to memory: Advances in the complementary learning systems framework. Trends in Cognitive Sciences. 2002; 6:505–510. [PubMed: 12475710]

Pyc MA, Rawson KA. Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? Journal of Memory and Language. 2009; 60:437–447.

Roediger HL, Karpicke JD III. The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science. 2006; 1:181–210.

Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: Circadian rhythms in human cognition. Cognitive Neuropsychology. 2007; 24:755–789. [PubMed: 18066734]

Stickgold R, Walker MP. Memory consolidation and reconsolidation: What is the role of sleep? Trends in Neurosciences. 2005; 28:408–415. [PubMed: 15979164]

Suzuki A, Josselyn SA, Frankland PW, Masushige S, Silva AJ, Kida S. Memory reconsolidation and extinction have distinct temporal and biochemical signatures. Journal of Neuroscience. 2004; 24:4787–4795. [PubMed: 15152039]

Wagner K. How much do children say in a day? Journal of Child Language. 1985; 12:475–487. [PubMed: 4019616]

Walker MP. A refined model of sleep and the time course of memory formation. Behavior and Brain Sciences. 2005; 28:51–64.

Walker MP, Stickgold R. Sleep, memory, and plasticity. Annual Review of Psychology. 2006; 57:139–166.

Wixted JT. The psychology and neuroscience of forgetting. Annual Review of Psychology. 2004; 55:235–269.

Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language. 2002; 46:441–517.

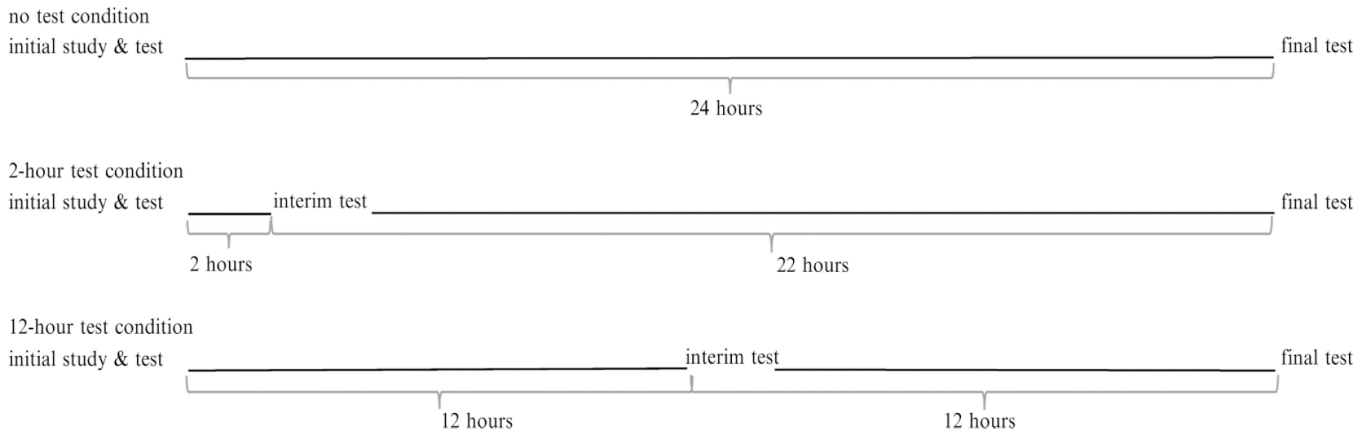

Figure 1.

Time intervals between tests in each of the three interim conditions.

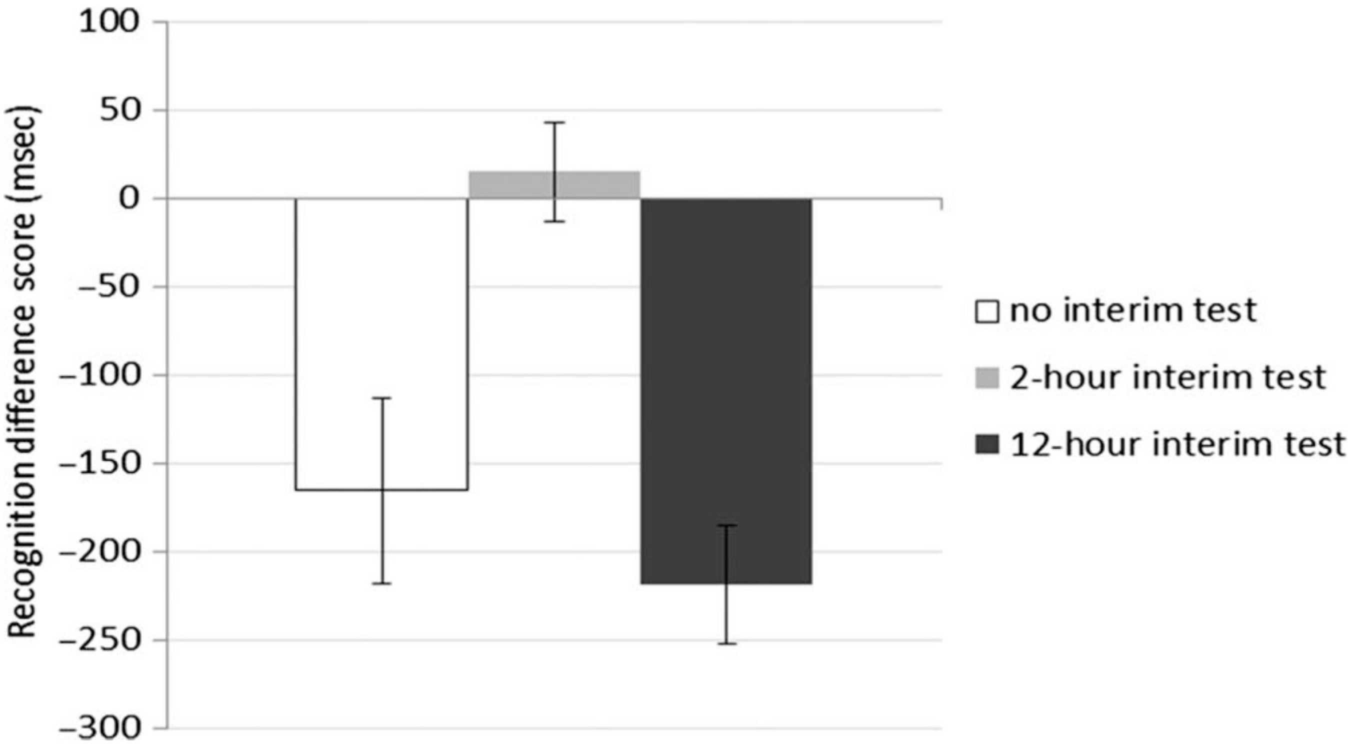

Figure 2.
Change in reaction time for word recognition over 24 hr in each of the three interim conditions expressed as means and standard errors.

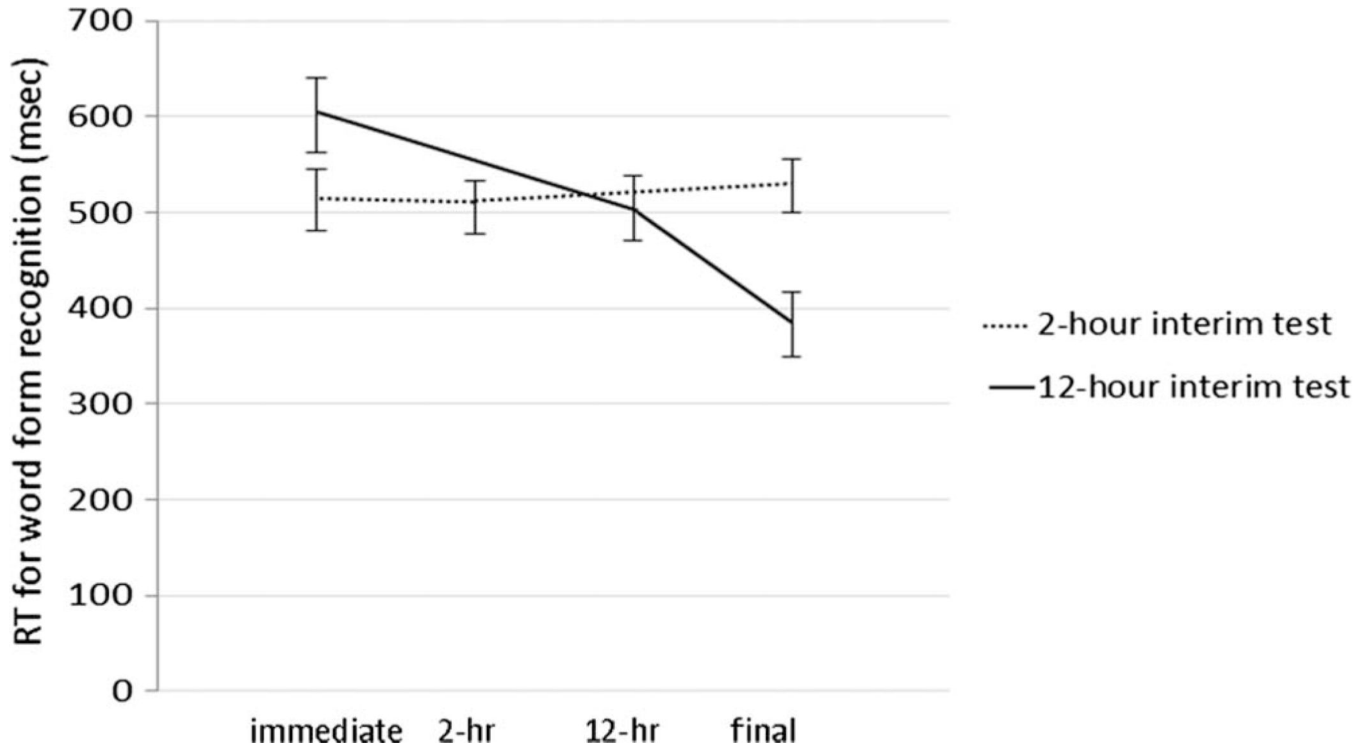

Figure 3.
Change in reaction time (RT) for word form recognition at each testing point for the 2- and 12-hr interim conditions expressed as means and standard errors.

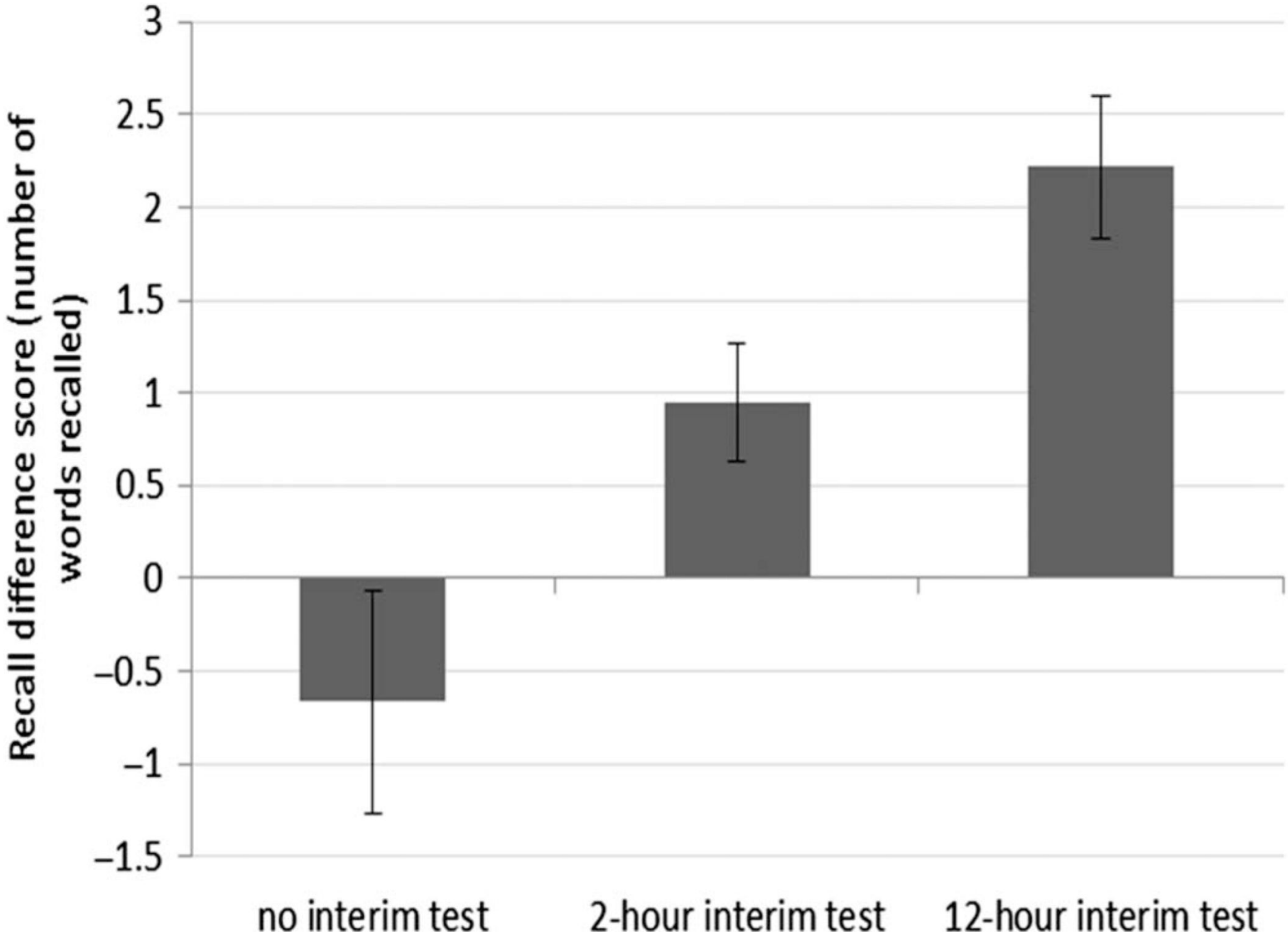

Figure 4.

Change in number of word forms recalled over 24 hr in each of the interim conditions expressed as means and standard errors.

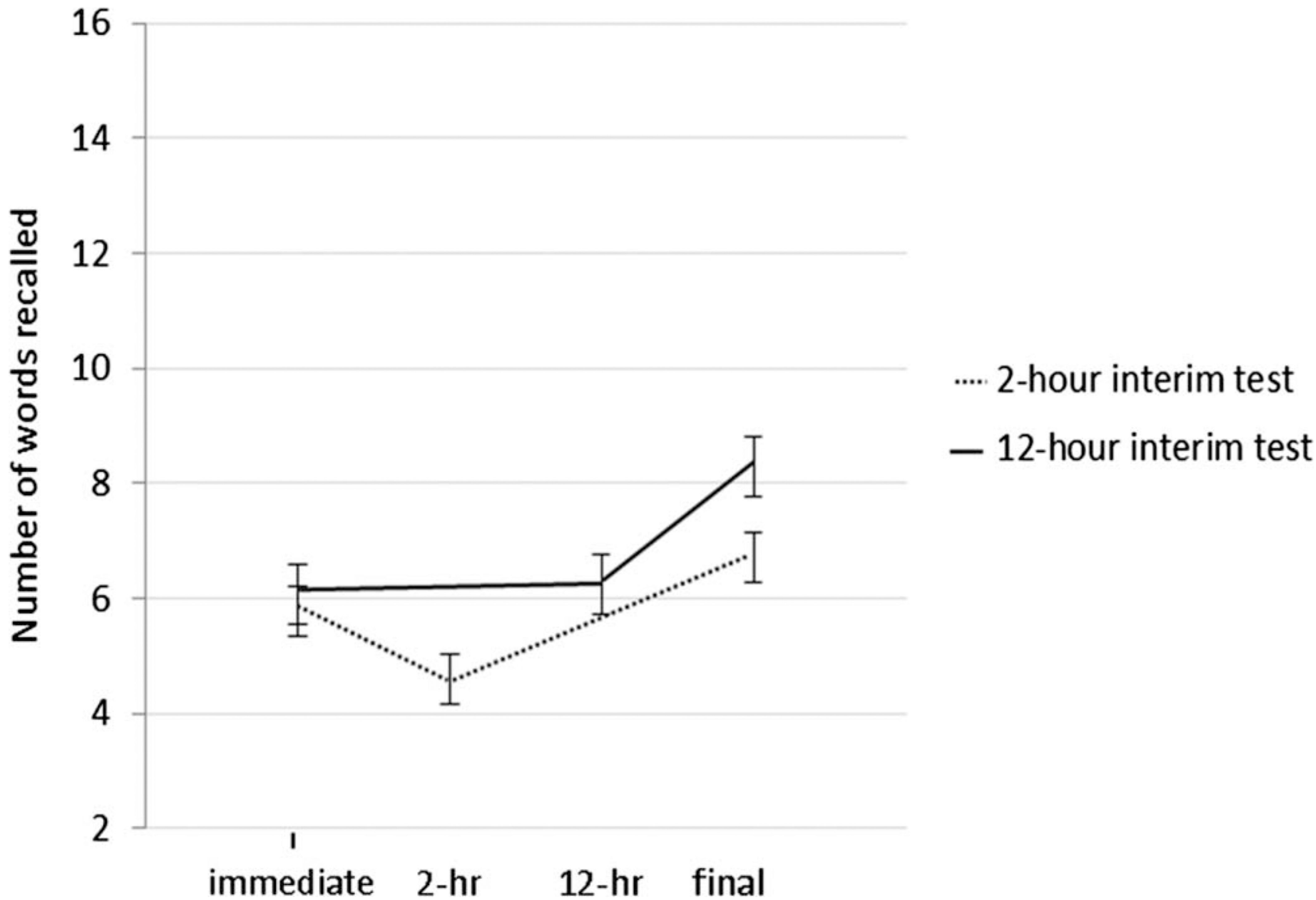

Figure 5.
Number of word forms recalled at each testing point for the 2- and 12-hr interim conditions expressed as means and standard errors.

Table 1

Accuracy and speed of word recognition on the 2AFC task by condition and test expressed as mean proportion of responses correct (and standard error).

Interim condition	Immediate	Interim	24-hr post
2-hr test	.97 (.005)	.96 (.007)	.96 (.008)
12-hr test	.99 (.006)	.97 (.008)	.97 (.009)
No-test	.99 (.009)		.97 (.014)

Note. 2AFC = two-alternative fixed choice.