

Research Article

Towards a Theory of Learning for Naming Rehabilitation: Retrieval Practice and Spacing Effects

Erica L. Middleton,^a Myrna F. Schwartz,^a Katherine A. Rawson,^b
Hilary Traut,^a and Jay Verkuilen^c

Purpose: The purpose of this article was to examine how different types of learning experiences affect naming impairment in aphasia.

Methods: In 4 people with aphasia with naming impairment, we compared the benefits of naming treatment that emphasized *retrieval practice* (practice retrieving target names from long-term memory) with *errorless learning* (repetition training, which preempts retrieval practice) according to different schedules of learning. The design was within subjects. Items were administered for multiple training trials for retrieval practice or repetition in a *spaced* schedule (an item's trials were separated by multiple unrelated trials) or *massed* schedule (1 trial intervened between an item's trials). In the spaced condition, we

studied 3 magnitudes of spacing to evaluate the impact of effortful retrieval during training on the ultimate benefits conferred by retrieval practice naming treatment. The primary outcome was performance on a retention test of naming after 1 day, with a follow-up test after 1 week.

Results: Group analyses revealed that retrieval practice outperformed errorless learning, and spaced learning outperformed massed learning at retention test and at follow-up. Increases in spacing in the retrieval practice condition yielded more robust learning of retrieved information.

Conclusion: This study delineates the importance of retrieval practice and spacing for treating naming impairment in aphasia.

For individuals with acquired language disorders such as aphasia from stroke, *naming impairment* (i.e., difficulty retrieving and producing words for familiar objects and entities) is ubiquitous and can be a profound impediment to effective communication. The challenge set before the clinician tasked with treating naming impairment is to adopt an intervention that maximizes and sustains improvements in performance. However, how to achieve this goal is anything but straightforward. To start, the clinician must decide how to prioritize use of various available treatment methods (e.g., repetition training, picture–word matching, confrontation or cued naming practice) and how to administer the treatment to maximize benefit. Furthermore, the efficacy of different treatments may vary with the functional locus of naming impairment,

which can involve one or several stages of processing in the course of word retrieval (i.e., object recognition/categorization, word selection, word form retrieval, post-lexical/articulatory operations). This echoes a prevalent assumption in naming treatment research that the devised treatment should target the disrupted process(es) implicated in the impairment (e.g., Abel, Willmes, & Huber, 2007; Nickels, 2002). However, in addition to an analysis of the underlying functional impairment, we assert, as have others (Baddeley, 1993; Stark, 2005), that a model of how the damaged system changes with experience (a *theory of learning*) is critical for optimizing rehabilitation.

The present study builds on our prior work (Middleton, Schwartz, Rawson, & Garvey, 2015) to establish an empirical foundation for a theory of learning for naming rehabilitation that is centered on *retrieval practice* and *spacing*. A wealth of basic psychological research demonstrates that retrieval practice (the act of retrieval from long-term memory) confers powerful and persistent learning, particularly when retrieval is effortful. The *spacing effect* refers to the ubiquitous advantage to learning when repeated training opportunities for individual items are distributed over time, rather than massed. Middleton et al.

^aMoss Rehabilitation Research Institute, Elkins Park, PA

^bKent State University, OH

^cCity University of New York

Correspondence to Erica L. Middleton: middleer@einstein.edu

Editor: Rhea Paul

Associate Editor: Robert Marshall

Received August 28, 2015

Revision received December 23, 2015

Accepted February 23, 2016

DOI: 10.1044/2016_JSLHR-L-15-0303

Disclosure: The authors have declared that no competing interests existed at the time of publication.

(2015) reported initial evidence that retrieval practice persistently improved naming performance in people with chronic aphasia (PWA). In the present study, we build on our prior work by: (a) investigating the impact of retrieval practice in a new performance domain (proper noun naming); (b) determining whether improvement is greater with spaced versus massed training. In our work, the experimental tasks we use are based on naming treatment methods (i.e., confrontation naming, errorless learning treatment) that are used widely in the clinic. However, we manipulate those treatments systematically in relation to retrieval practice and spaced learning in order to investigate the relevance of such mechanisms for optimizing existing treatments.

Aphasia interventions typically take one of two general approaches, compensatory and restorative. Compensatory approaches strive to achieve functionality by working around the deficit, such as training a strategy or use of an assistive device. In such approaches, generalization to untreated items is a key goal. In contrast, our work exemplifies a *restorative* approach, where the goal is to promote access to affected vocabulary through (re-)learning techniques. Here, the expected benefits are primarily item-specific, with less focus on potential generalization to untreated items. To maximize functional impact, such treatments aim to promote long-term improvements in treated vocabulary and to develop tools that facilitate self-administration of treatment. The goal of the present work is to delineate which learning experiences are most effective for enhancing the ability of PWA to retrieve treated vocabulary in a persistent fashion.

Learning Theory in Naming Treatment Research

In response to the need for models of learning to inform naming treatment research, a growing body of work has focused on the potentially deleterious impact of *error learning* wherein naming errors committed by PWA during treatment may be learned and negatively affect treatment efficacy. This possibility has been the impetus for numerous studies on *errorless learning*, where the avoidance of retrieval errors in the course of naming treatment is prioritized (for reviews, see Fillingham, Hodgson, Sage, & Lambon Ralph, 2003; Middleton & Schwartz, 2012). The typical form of errorless learning treatment for naming impairment involves *repetition training* where at picture onset, the correct (i.e., target) name is heard/seen and the participant repeats the name. In keeping with Hebbian learning (often summarized as “cells that fire together, wire together”), errorless learning treatment may be desirable because it preempts naming attempts (thus avoiding error), and instead only strengthens the association between the target name and the depicted object (for discussion, see Fillingham et al., 2003). However, a name that is produced via repetition is not retrieved in a top-down fashion from long-term semantic memory; rather, the name is activated from input phonology either directly or with lexical mediation (Nozari, Kittredge, Dell, & Schwartz, 2010). Circumventing retrieval from long-term memory may limit the efficacy of errorless learning

treatment, in light of the benefits of retrieval practice, described next.

Retrieval Practice

Over 100 years ago in *Principles of Psychology*, William James (1890) noted that “recollect[ing] by an effort from within” (p. 646) was important for learning. Since this observation, countless studies on retrieval practice have confirmed the basic notion that retrieval from long-term memory persistently changes the accessibility of the retrieved information (for reviews, see Rawson & Dunlosky, 2011; Roediger, Putnam, & Smith, 2011). In research demonstrating the importance of retrieval practice for learning, the standard method begins with an initial study period to familiarize learners with the to-be-learned information (i.e., target information). Initial study is followed either by further study opportunities (i.e., *restudy*) or by tests in which participants attempt to retrieve target information from long-term memory (i.e., retrieval practice). A *retrieval practice effect* is demonstrated when performance on a final test is greater following training involving retrieval practice versus training involving restudy. The contribution of feedback to the retrieval practice effect has also been investigated, e.g., by withholding correct-answer feedback (Roediger & Karpicke, 2006; Wheeler, Ewers, & Buonanno, 2003). Absent feedback, the benefits of retrieval practice are driven by correct trials in which the target information is successfully retrieved; failed retrieval attempts unaccompanied by corrective feedback confer minimal learning of the target information (Pashler, Cepeda, Wixted, & Rohrer, 2005). On the other hand, failed retrieval attempts followed by corrective feedback do show a learning advantage, relative to restudy (e.g., Kornell, Hays, & Bjork, 2009, but see Vaughn & Rawson, 2012). Retrieval practice is a potent learning technique of broad applicability—retrieval practice effects have been established in a wide variety of learning contexts including paired associate learning (e.g., Carrier & Pashler, 1992), learning text (e.g., Roediger & Karpicke, 2006), statistics (e.g., Lyle & Crawford, 2011), unfamiliar visual symbols (e.g., Kang, 2010), second language learning (e.g., Barcroft, 2007), and proper name learning (e.g., Carpenter & DeLosh, 2005), to list just a few.

Given the applicability of retrieval practice for enhancing many different types of knowledge and skill acquisition, Middleton et al. (2015) sought to establish the clinical relevance of retrieval practice for treating naming impairment in PWA by comparing a naming treatment featuring retrieval practice to errorless learning treatment. As language deficits in aphasia are typically reflective of dysfunction in retrieving and assembling language-based representations rather than a loss of linguistic knowledge, it was of empirical and theoretical interest whether retrieval practice—primarily studied in the context of knowledge acquisition—would be relevant in aphasia rehabilitation. Middleton et al. (2015) studied PWA whose naming impairment implicated an access deficit in naming (i.e., difficulty reliably mapping from semantics to a known word, or from the word to its form).

Prior to the experiment, Middleton et al. (2015) had participants name 615 pictures of common objects to identify participant-specific errorful items. These items were assigned into training conditions while matching for several psycholinguistic variables. In the experiment, each picture was first presented with its name for one exposure trial, to parallel initial study in the standard retrieval practice method. After 5 minutes, each item underwent one training trial of either errorless learning treatment (i.e., repetition training, in which the name was seen and heard at picture onset and the participant repeated the name) or retrieval practice, which involved either cued naming (first sound and letter of the word were presented with the picture to facilitate retrieval of the name) or noncued naming (only the picture was presented). Trials in all training conditions terminated with feedback in which the name was presented and repeated by the participant. During training, the rate of production of target names in the repetition condition was nearly perfect and superior to the rate of target production in the retrieval practice conditions. However, despite this disadvantage at training, both retrieval practice conditions outperformed the repetition condition on a posttraining retention test administered after 1 day. The advantage for the cued retrieval practice condition over repetition also persisted on a retention test 1 week later. Thus, the experiment demonstrated that retrieval practice is a potent learning experience for facilitating persistent access to known words in PWA. The merits of a treatment centered on retrieval from long-term memory far surpassed any potential decrement from error learning, countering the motivation for the errorless approach.

Spaced Learning

In Middleton et al. (2015), a single training trial of retrieval practice versus errorless learning produced superior retention test performance in the retrieval practice conditions. An important next step is to investigate how the proposed learning mechanisms affect performance when there are multiple training trials per item, which will increase the similarity of our experimental method to clinical practice. Equally important, multiple training trials per item enables us to investigate how the trials should be scheduled over time (i.e., the schedule of learning) to maximize benefit. In a *spaced* schedule of learning, repeated training opportunities for an item are separated by enough time or intervening material to exceed the limits of short-term memory. In massed practice, training trials for an item are administered close in time so that the item is still accessible in short-term memory on each trial. The spacing effect refers to the advantage at later test for material trained in spaced versus massed schedules. Explanations of the spacing effect commonly appeal to the notion that in spaced schedules, linking current and past experiences with an item across multiple trials requires *reactivation* (i.e., referencing prior instances of that item in long-term memory, which strengthens the item's representation). In massed schedules, there is less reactivation because the item remains in an activated state across trials (e.g., Benjamin & Tullis,

2010; Braun & Rubin, 1998). Spacing benefits various types of learning (e.g., episodic memory, concept acquisition, procedural skill learning) in learners across the age span and even from different species (e.g., honeybees; Menzel, Manz, Menzel, & Greggers, 2001; for a recent review, see Toppino & Gerbier, 2014).

In the present study, we evaluated the impact of spaced and massed schedules of learning in naming in PWA. The massed condition involved a *lag* of one trial (i.e., lag = number of trials for other items intervening between repeated training trials for a given item). In a recent review of spacing effects, Toppino and Gerbier (2014) operationalized massing as 15 seconds (or less) between repeated trials for an item, roughly corresponding to the limits of short-term memory. In the present experiment, a lag of one trial is comfortably within these limits, given the timing of a trial (i.e., 8 seconds for attempted naming or repetition plus a few seconds for feedback; see Method). Though it is not uncommon in the spacing literature to use a lag of zero for the massed condition, we were concerned this would disadvantage the massed condition, with contiguous presentation of the four trials for an item operating functionally like one long trial.

The spaced condition included different magnitudes of spacing (i.e., lag 5, lag 15, lag 30). This secondary manipulation was included to address two issues concerning the retrieval practice condition, one theoretical, the other practical. The theoretical issue concerns the role of effort in potentially enhancing the benefits conferred by retrieval practice naming treatment. Numerous studies have shown that successful retrieval that requires effort is more potent than successful retrieval that is easy (e.g., Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2007; Pyc & Rawson, 2009). For example, Pyc and Rawson (2009) found long (versus short) lag was associated with greater effort during retrieval practice (as revealed by greater time to successfully retrieve target information), which conferred better performance on final outcome measures. Thus, in our retrieval practice condition, successful retrievals may be associated with better retention as lag increases. However, the memory literature also suggests that increased spacing can lead to increased retrieval failures during training (Pashler, Zarow, & Triplett, 2003), which can diminish the net benefit from retrieval practice training (Pashler et al., 2005). It is an empirical question how our selected lags are situated with regards to this costs–benefits tradeoff. Generally speaking, evidence in support of the retrieval effort hypothesis would involve a demonstration that though the rate of retrieval success decreases with increasing lag at training, at retention test, the relationship between lag and accuracy displays a different function (i.e., increasing accuracy or stable levels of accuracy with increasing lag; see Pashler et al. 2003, for evidence along these lines). The lag manipulation also addresses a practical concern about a potential boundary in the retrieval practice naming treatment beyond which (at the lags investigated here) an increase in spacing becomes suboptimal.

Overview of the Current Research

We recruited PWA for the current work from the sample studied in Middleton et al. (2015) because they were already well characterized and evinced a cognitive–linguistic profile consistent with an access deficit (difficulty consistently and fluently retrieving known words) underlying their naming impairment. Training materials involved entities named with proper nouns (e.g., famous people). Starting with a large corpus, we selected items for treatment for each participant for whom the entity and name were known, but the participant experienced difficulty naming (a *personalized item set*). Thus, we could attribute any benefits of the treatment to improved access to existing linguistic representations rather than acquisition of new lexical knowledge or concepts.

Each participant's personalized item set was divided equally into the conditions of a two-level factor of *type of training* (retrieval practice versus errorless learning, i.e., repetition) crossed with a spacing factor, which included a massed condition (lag 1) and three spaced conditions (lags 5, 15, and 30). Lag corresponded to the number of training trials for other items that intervened between repeated training trials for an item.

On a repetition training trial, the participant repeated the name in the presence of the picture, and on a retrieval practice trial, the participant attempted to name the picture without any cueing. Feedback was provided at the end of all trials. The primary outcome was naming test performance 1 day following training, with a follow-up naming test administered after 1 week to assess the persistence of effects at a longer interval. With this design we evaluated the following predictions: (a) consistent with Middleton et al. (2015), retention test performance will be greater for retrieval practice training versus repetition training; (b) retention test performance will be greater for spaced training (collapsing across the three lag conditions) versus massed training (lag 1 condition); (c) increasing lag will decrease retrieval success during retrieval practice training but will yield better retention test performance. However, it is possible that our highest lags will be suboptimal if the rate of retrieval failure overshadows the benefits from increasing the effort required for successful retrieval practice.

Method

A challenge to experimental investigations with neuropsychological populations is the frequent heterogeneity of cognitive–linguistic deficits even among those with the same diagnosis, which can inflate variability and decrease power. Our design addressed this issue in the following manner: (a) our PWA sample was relatively homogeneous in severity and type of naming impairment; (b) the study was designed to maximize the number of observations per participant per condition, thus increasing power to detect effects within as well as across participants. To maximize the number of observations, the design required a minimum of 22 sessions for each participant to complete the entire experimental protocol. Because of the substantial resources

required for data collection and processing for each PWA, we tested four participants.

Participants

Participants gave informed consent under a protocol approved by the institutional review board of Einstein Healthcare Network, and were reimbursed \$15 for each hour of participation. The four participants (2 men, 2 women) were right-handed with chronic aphasia secondary to left-hemisphere stroke with mean education level of 13.5 years ($SD = 1.9$; range, 4 years) and mean age of 62.5 ($SD = 11.4$; range, 24). Three were diagnosed with fluent aphasia of the anomic subtype; the fourth participant was nonfluent with transcortical motor aphasia. The participants demonstrated mild-to-moderate naming impairment in independent testing of oral naming ability of common everyday objects (see the online supplemental materials, Supplemental Material S2). The naming impairment was principally attributable to failure to reliably and fluently retrieve known words (i.e., lexical access deficit). Alternative sources of difficulty were ruled out by background tests and other data. For example, given their generally good performance on tests of nonverbal semantics and word comprehension (see the online supplemental materials, Supplemental Material S2), their naming impairment is unlikely to reflect a central semantic deficit, which can compromise the semantic input to word retrieval. Their generally good word repetition (see the online supplemental materials, Supplemental Material S2) weighs against phonological-phonetic dysfunction as a major contributor to our participants' naming impairment. We also required good word repetition as a recruitment criterion because errorless learning (i.e., repetition training) is assumed to confer learning by scaffolding correct production of target names during training. Errorless learning should be most beneficial for participants with intact repetition abilities. Thus, studying patients with good repetition worked against our prediction of observing a retrieval practice effect. For additional discussion of the participant sample, see Middleton et al. (2015); the current sample included participants S1, S3, S7, and S8 from that study.

Materials

A picture corpus containing 700 entities named with proper nouns was collected from the Internet. The corpus was administered in the item selection phase (described below) to select items for each participant to be used in the main experiment. An additional 49 nonexperimental pictures were collected to fulfill other purposes. The 700-item corpus and nonexperimental items included a variety of entities including famous people (e.g., movie stars, politicians, historical figures), fictional characters (e.g., Popeye), and films with iconic movie posters (e.g., Casablanca). Ratings from six neurotypical older adults confirmed that these items were generally familiar to people in the age range of our participants. Between 7 and 10 additional neurotypical older adults named the 700-item picture corpus.

If half or more naming responses for an item included a phonological (e.g., /z/ instead of /s/ sound in Presley) or lexical variation (e.g., Jimmy/James Stewart), that variation was considered correct if produced in the main experiment. Approximately 3.4% of the 700-item corpus had alternative names.

Response Coding

Naming responses in aphasia can be coded in different ways for different purposes. For example, a typical goal in naming treatments is to promote retrieval of the correct name for an object. Here, it is common to define a binary outcome measure (correct/incorrect) that accepts as correct any production that contains most of the target phonemes. The willingness to accept minor deviations from perfect production is based on the idea that such deviations can arise from errant phonological-phonetic encoding after the correct name has been retrieved. The final product of the present coding system, described below, is a binary response accuracy score (correct/incorrect), where responses with most or all of the target phonemes are coded as correct.

In the current study, the production targets are proper names, the majority of which are composed of multiple morphemes. Hence, coding proceeded with first mapping each of the target name's constituents (i.e., first name constituent, last name constituent) to the "best" *response constituent* from among all nonfragment constituents produced on that trial. Determination of best was based on a *phonological overlap* formula (Lecours & Lhermitte, 1969), a continuous measure of phonological similarity between a response and target standardized across different word lengths (Formula 1). Shared phonemes were identified independent of position and credit was assigned only once if a response had two instances of a single target phoneme.

phonological overlap =

$$\frac{\# \text{ shared phonemes in target and response}}{\sum \text{ phonemes in target and response}} \times 2 \quad (1)$$

Supplemental Material S1 and Supplemental Material S3 (see the online supplemental materials) provide full details of how best response constituents were identified, but we provide a summary here. A response constituent was identified as best if it had the highest overlap with a target constituent and overlap exceeded .50 (we adopted this threshold as the minimum standard of evidence that a response constituent was an attempt on a target constituent). Once a response constituent was deemed best for one target constituent (e.g., last name), it could not be used as best for the other target constituent (first name). In some cases, this could result in no best constituent being identified for a target constituent (see Case 3 in the online supplemental materials, Supplemental Material S3). After mapping each target constituent to either its best response constituent or no response constituent, we calculated

phonological overlap between identified best response constituents and the target constituents as a whole (Column 6 in the online supplemental materials, Supplemental Material S3). A binary variable of *response accuracy* was derived from the phonological overlap measure for the whole name. Phonological overlap $\geq .75$ was defined as correct; $< .75$ was defined as incorrect (Column 7 in the online supplemental materials, Supplemental Material S3). Response accuracy was the main dependent variable for the analyses.

Procedure

Item Selection Phase

Given that a goal of the present work was to investigate the learning experiences that best promote recovered access to known words in PWA, the procedure in the item selection phase was designed to isolate a subset of the 700-item corpus with which the PWA was familiar but experienced difficulty naming. Such items comprised the participant's personalized item set, which was distributed into the experimental conditions. The first task in the item selection phase involved presenting the 700-item corpus in random order for name-to-picture matching. Participants were asked to select the correct name for each picture from among five written options. Three foils were the names for similar entities or people (e.g., for Cameron Diaz, foils were Claire Danes, Charlize Theron, and Cate Blanchett), and the fourth foil was a "None of the above" option. Drawing from the nonexperimental pictures, we constructed 42 filler items to be presented among the 700-items in which the correct answer was "None of the above," to prevent participants from ignoring this option.

Over multiple sessions in separate weeks following the name-to-picture matching task, the 700-item corpus was presented twice in random order for confrontation naming followed by recognition judgments. On a naming trial, the picture was shown and the participant was provided up to 20 seconds to produce the name. They were instructed to let the experimenter know when they had given their final answer by pointing to a paper with a "thumbs up" graphic, after which the experimenter advanced the trial. This procedure avoided experimenter-provided feedback of any kind. Each naming trial was followed by a prompt for recognition of the entity ("Do you recognize this person or thing? 1 = *yes*, 2 = *not sure*, 3 = *no*") followed by recognition of the name ("Even if you can't think of the name right now, would you recognize the name if you saw it? 1 = *yes*, 2 = *not sure*, 3 = *no*"). This enabled us to isolate items for each PWA that met the following criteria: (a) the item yielded a naming response that was not an exact match to the name (all the target phonemes in the correct order) on at least one of two administrations of the item selection naming test. Note, we prioritized items for each participant for which the response deviated from correct on both administrations of the item selection test (i.e., response accuracy was low for selected items prior to training; see the online supplemental materials, Supplemental

Material S4); (b) the item elicited a correct response in the name-to-picture matching task; (c) the PWA recognized the entity and the name as indicated by a “2” or “1” response on at least one administration of each task. Each participant’s personal item set was divided evenly into the conditions, matching for the variables reported in the online supplemental materials (see Supplemental Material S4). This procedure yielded between 36 and 72 observations per each of the eight conditions per participant.

Main Experiment

The protocol was divided into multiple cycles with a cycle composed of three sessions: (a) a training session; (b) a session held the next day, in which a retention test of naming was administered for the items trained in the previous session; (c) a third session held 7 days after training, in which a follow-up test of naming was administered for the same items. The follow-up test always completed a cycle before the next cycle was initiated (typically the following week). There were either three or four cycles per training condition (naming or repetition), depending on the number of available items in the participant’s personalized item set. Either 60 or 72 experimental items were trained in each training session (with 15–18 items assigned to each of the four lag conditions). The order of repetition versus naming cycles was counterbalanced across participants with the two types of cycles administered in interleaved order (e.g., for participant A, the cycles would be repetition, naming, repetition, naming, etc.; for participant B, the cycles would be naming, repetition, naming, repetition, etc.).

A training session began with 10 practice trials followed by three blocks of 120 trials each. Participants were encouraged to rest between blocks. Each block began and ended with five filler trials to avoid privileging memory for experimental items appearing at the beginning or end of the list. Each experimental item was presented four times according to its assigned lag. The average ordinal position within a block was equated for the items in the four lag conditions. Fillers were used as needed to complete the sequence.

In the repetition condition, an item was presented for four repetition trials. On a repetition trial, the written name and auditory form of the word were presented at picture onset (the auditory form was played once, whereas the written name was displayed for the duration of the trial).¹ The participant was instructed to repeat the name

¹Presentation of the target for the entirety of the repetition trial was adopted to parallel the restudy condition in the standard retrieval practice paradigm. In the testing literature, this practice is generally understood to set the retrieval practice condition at a disadvantage relative to restudy (exposure to the target on a retrieval practice trial is limited to whether and when during the trial the target is successfully retrieved). Despite this, retrieval practice conditions prevail, illustrating the potency of retrieval from long-term memory for learning. Note, retrieval practice effects are not likely attributable to superficial processing of the target in the restudy condition because when deep, elaborative encoding is encouraged during restudy, retrieval practice still confers superior learning (Karpicke & Blunt, 2011).

once. After 8 s the auditory form of the name was played again, and the participant repeated the name (i.e., this was *feedback*). In the naming condition, the first trial for an item was a repetition trial to prime the association between the entity and its name (analogous to initial study in the typical retrieval practice paradigm). The subsequent three trials were naming trials. On a naming trial, the picture was presented, the participant attempted to name the picture, and after 8 seconds, feedback was provided. Thus, all trials in both the naming and repetition condition ended in feedback, and the timing of the trials was identical across conditions.

The form of feedback used in the present study—repetition of the target name—was adopted to parallel standard practice in retrieval practice studies. However, this practice does not include explicit information of the correctness of an utterance, a form of feedback often used in the clinic. We assume one function of our feedback procedure was to assist participants in calculating the accuracy of their naming attempt, because the items selected for training had low name accessibility but were known to the participant. In future studies it will be important to evaluate how our feedback procedure may be supplemented with explicit feedback regarding accuracy.

The procedure for the retention naming test and follow-up naming test was identical to the naming test during the item selection phase (i.e., there was no feedback). Each item selection session and training session required approximately two hours. Retention test and follow-up test sessions required approximately 30 minutes.

Analyses

The data were analyzed with mixed-effects regression using Stata 13 (StataCorp, 2013) statistical software (for an introduction to mixed-effects models, see Baayen, Davidson, & Bates, 2008). We would have adopted a crossed random effects structure, with items and participants as random effects (Baayen et al., 2008). However, because the number of participants was small and these individuals were selected from a larger pool with specific inclusion criteria, participants were dummy coded as a fixed effect in all models (see Tables 1 and 2). This approach enabled estimation of coefficients for the experimental factors while accounting for participant-specific effects (Park, Goral, Verkuilen, & Kempler, 2013). Items were treated as random in all models except in the models reported last in the results section (two models measuring the effect of lag in the naming condition reported in Table 2, middle and bottom panels). In those models, we used simple logistic regression because it was not appropriate to treat items as a random effect (i.e., very few items were administered to more than one participant). In all other models, we included a random by-items intercept because in each case its inclusion improved model quality (by Akaike information criterion) and the intraclass correlation was nonnegligible.

In each model, to inspect whether the experimental factors interacted with the participants’ factor (indicating

Table 1. Mixed model results—retrieval practice and spacing effects.

Dependent Variable: Response Accuracy at Retention Test				
Fixed Effects	Coefficient	SE	Z	p
Intercept	-2.24	0.21		
Retrieval practice effect				
Repetition^a	-0.44	0.13	-3.29	.001
Spacing effect				
Spaced^b	0.71	0.16	4.46	<.001
Subject effect				
P1 ^c	2.13	0.19	11.01	<.001
P2 ^c	2.65	0.21	12.69	<.001
P3 ^c	1.47	0.18	8.08	<.001
Random effect	<i>s</i> ²			
Items	1.09			
Dependent Variable: Response Accuracy at Follow-Up Test				
Fixed Effects	Coefficient	SE	Z	p
Intercept	-2.11	0.22		
Retrieval practice effect				
Repetition^a	-0.35	0.14	-2.55	.01
Spacing effect				
Spaced^b	0.45	0.16	2.78	.006
Subject effect				
P1 ^c	1.50	0.19	7.82	<.001
P2 ^c	2.25	0.21	10.72	<.001
P3 ^c	0.82	0.19	4.29	<.001
Random effect	<i>s</i> ²			
Items	1.18			

Note. Excluding the intercepts, Coefficient = model estimation of the change in response accuracy (in log odds) from the reference category for each fixed effect; SE = standard error of the estimate; Z = Wald Z test statistic; *s*² = random effect variance.

^aReference is naming condition. ^bReference is massed condition.

^cReference is participant 4 (P4).

differences across participants in their response to the conditions), we used a model comparison procedure testing nested higher-order and lower-order interactions between the fixed effects. A main effects model was compared with a model with all two-way interactions possible, and the all two-way interactions model was compared with a three-way interaction model. We report the more complex model if model selection criteria (Bayesian information criterion and Akaike information criterion) indicated the added complexity was warranted and the more complex model demonstrated better fit (compared with the simpler model) by chi-square deviance in model log likelihoods. In only one analysis (a secondary analysis concerning the effects of lag in the retrieval practice condition) was a more complex model warranted (see Table 2, top). In that case, there was a two-way interaction between participants and retrieval practice training performance, indicating that retrieval success for the final training trial in the naming condition tended to be lower for one participant (P4) compared with the others. As this outcome has little bearing on the interpretation, it is not discussed further. Fixed effects relevant to evaluation of the study hypotheses are shown in bold font in Tables 1 and 2, and in each case, the coefficient corresponds to the estimated effect across participants (i.e., across the models,

Table 2. Mixed model results—effects of retrieval effort.

Dependent Variable: Response Accuracy at Retention Test and Final Training Trial (Naming Condition)				
Fixed Effects	Coefficient	SE	Z	p
Intercept	-2.65	0.68		
Log of lag	0.73	0.56	1.30	.19
Effect of event				
Training ^a	3.85	0.68	5.66	<.001
Interaction of lag and event				
Log of Lag × Training^a	-2.07	0.54	-3.82	<.001
Subject effect				
P1 ^b	3.97	0.94	4.21	<.001
P2 ^b	4.27	1.18	3.63	<.001
P3 ^b	2.23	0.91	2.46	.014
Interaction of subject and lag				
Log of Lag × P1 ^b	-1.14	0.77	-1.48	.14
Log of Lag × P2 ^b	-0.87	0.98	-0.89	.38
Log of Lag × P3 ^b	-0.42	0.76	-0.56	.58
Interaction of subject and event				
Training ^a × P1 ^b	-0.50	0.41	-1.22	.22
Training ^a × P2 ^b	1.21	0.72	1.69	.092
Training ^a × P3 ^b	0.74	0.42	1.77	.076
Random effect	<i>s</i> ²			
Items	1.42			
Dependent Variable: Response Accuracy at Retention Test (Naming Condition)				
Fixed Effects	Coefficient	SE	Z	p
Intercept	-3.29	0.47		
Log of lag	0.82	0.32	2.58	.01
No. successful retrievals at training	0.71	0.10	6.73	<.001
Subject effect				
P1 ^b	1.58	0.27	5.86	<.001
P2 ^b	1.65	0.32	5.11	<.001
P3 ^b	0.74	0.27	2.80	.005
Dependent Variable: Response Accuracy at Follow-Up Test (Naming Condition)				
Fixed Effects	Coefficient	SE	Z	p
Intercept	-3.59	0.58	-6.18	<.001
Log of lag	1.06	0.34	3.16	.002
No. successful retrievals at training	1.03	0.13	7.83	<.001
Subject effect				
P2 ^c	0.21	0.29	0.72	.47
P3 ^c	-0.71	0.27	-2.58	.01
P4 ^c	-0.58	0.30	-1.94	.05

Note. Excluding the intercepts, Coefficient = model estimation of the change in response accuracy (in log odds) from the reference category for each fixed effect; SE = standard error of the estimate; Z = Wald Z test statistic; *s*² = random effect variance.

^aReference is retention test. ^bReference is participant 4 (P4).

^cReference is participant 1 (P1).

the participants' factor did not interact with the fixed effects of primary interest).

To evaluate whether we confirmed the retrieval practice effect of Middleton et al. (2015) and to examine whether spacing affected the benefits of treatment, we used mixed logistic regression to model response accuracy (correct/incorrect response) on the retention test with a

two-level factor of training type (naming versus repetition) and a two-level factor of spacing (massed versus spaced, with spaced collapsed across lags 5, 15, and 30). To inspect whether these effects persisted after 1 week, a parallel model was applied to the dependent variable of follow-up test accuracy. Additional models were constructed to evaluate how varying retrieval effort in the naming condition affected learning, discussed in greater detail below.

Results

Figure 1 shows response accuracy in the naming and repetition conditions (left panel) and massed and spaced conditions (right panel) at retention test and follow-up. Refer to the online supplemental materials (see Supplemental Materials S5 and S6) for average response accuracy as a function of condition at training, retention test, and follow-up for the group and for each participant. Tables 1 and 2 report inferential statistics.

Retrieval Practice

Concerning Prediction #1, the naming condition outperformed the repetition condition on both the retention test ($Coefficient = -0.44$, $SE = 0.13$, $Z = -3.29$, $p = .001$) and the follow-up test ($Coefficient = -0.35$, $SE = 0.14$, $Z = -2.55$, $p = .01$; see Table 1). This pattern aligns with the findings of Middleton et al. (2015) who reported an advantage for naming treatment involving retrieval practice over errorless learning at similar retention intervals.

Spacing Effects

Concerning Prediction #2, we observed a spacing effect with superior performance for spaced versus massed training on the retention test ($Coefficient = 0.71$, $SE = 0.16$, $Z = 4.46$, $p < .001$) and the follow-up test ($Coefficient = 0.45$, $SE = 0.16$, $Z = 2.78$, $p = .006$; see Table 1).

To summarize, we observed enhanced benefits for retrieval practice compared with errorless learning treatment,

and spaced training compared with massed, after 1 day and after 1 week. Thus, retrieval practice and spacing facilitate persistent access to treated names, providing two empirical pillars for our theory of (re)-learning of access to names. In secondary analyses, we analyzed the effects of lag in the retrieval practice condition. Figure 2 plots mean accuracy on the final training trial in the naming condition (corresponding to the ultimate level of mastery achieved by the end of training) and retention test performance as a function of lag. As Figure 2 shows, training performance decreased with increasing lag, whereas performance at retention test was stable with increasing lag. To evaluate whether the slopes across lags for accuracy at the end of training versus retention test were different, we modeled accuracy as a function of lag (the log base 10 of lag) entered as a numerical fixed effect, a two-level factor for event (end of training versus retention test), and their interaction. The interaction of lag and event was significant ($Coefficient = -2.07$, $SE = 0.54$, $Z = -3.82$, $p < .001$; see Table 2, top). The finding that retention performance was similar across lags despite the decreasing rate of retrieval success during training as lag increases suggests that successful retrieval at longer lags—because of greater effort required—is more potent than successful retrieval at lower lags. We conducted a follow-up analysis to evaluate this claim more directly.

In the follow-up analysis, we examined whether increasing lag would be associated with *increasing* retention test performance if the number of times an item was successfully retrieved during training was statistically controlled. Controlling for training performance was deemed important because the more often an item is successfully retrieved, the better it is learned, and the rate of successful retrieval decreased with increasing lag. Thus, we modeled retention test performance by log of lag with number of accurate responses during training per item (0, 1, 2, or 3) entered as a numerical covariate. The model revealed a significant positive effect of lag on retention test performance ($Coefficient = 0.82$, $SE = 0.32$, $Z = 2.58$, $p = .01$; see Table 2, middle), consistent with the retrieval effort hypothesis. The effect of lag persisted at the longer retention

Figure 1. Mean response accuracy (with standard errors) as a function of training condition (left panel) and spacing condition (right panel) at retention test and follow-up.

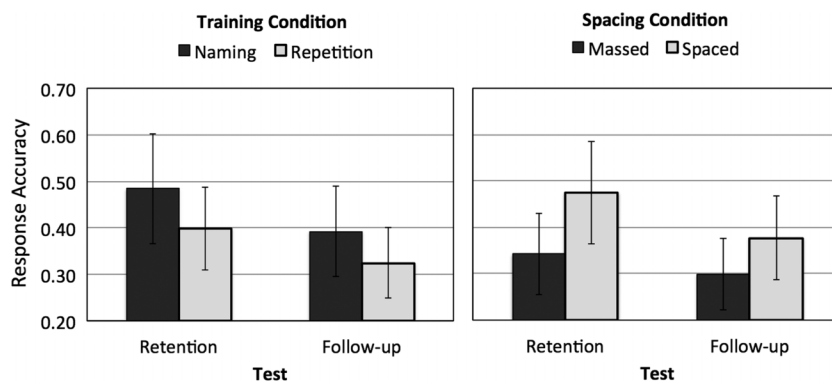
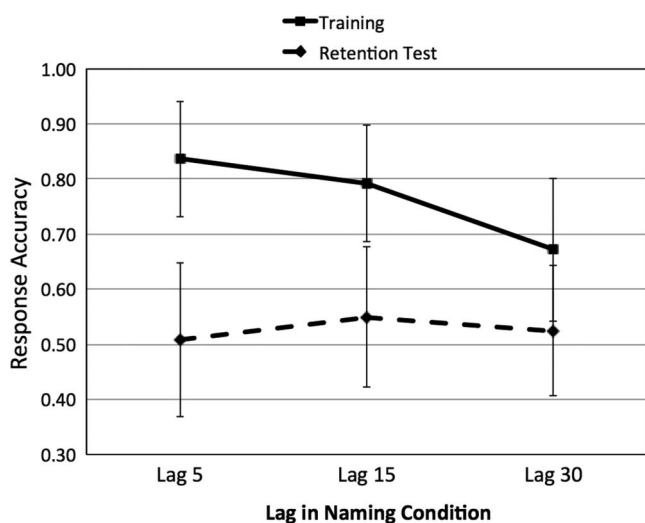


Figure 2. Mean response accuracy (with standard errors) at training and retention test as a function of lag in the naming condition. Training performance corresponds to average accuracy of the final training trial per item.



interval: in an identical model with follow-up (1-week) test performance entered as the dependent variable, the effect of lag was significant ($Coefficient = 1.06, SE = 0.34, Z = 3.16, p = .002$; Table 2, bottom). In summary, the degree of effort expended in the course of successful retrieval affects the benefits derived from retrieval practice.

Discussion

Middleton et al. (2015) demonstrated that retrieval practice confers greater persistent improvement in PWA's ability to name common objects compared with errorless learning treatment. The present study built on that work by investigating the impact of retrieval practice in a new domain of naming (i.e., proper nouns) as well as factors that affect the influence of treatment—spaced learning and retrieval effort. Treatment that provided retrieval practice opportunities (naming condition) outperformed treatment that minimized retrieval practice (repetition condition) at retention test, with the effects persisting after 1 week. Spaced training outperformed massed training at the retention test, an effect that also persisted after 1 week. Secondary analyses provided evidence that not all retrievals confer equivalent learning—successful retrievals that are difficult (operationalized as retrievals at longer lags) are more potent than easy retrievals. The spaced lags studied generally conferred the same net benefit at retention test, despite decreased rate of successful retrievals at training with longer lags. Thus, for clinicians interested in treating a similar population with retrieval practice, a (tentative) prescriptive recommendation is to choose from among the higher range of lags examined here to capitalize on the benefits from effortful retrieval.

We contend that naming ability improved in the current study because of increased accessibility of treated lexical entries and/or their forms. An alternative is that performance improved due to knowledge acquisition, (i.e., development of semantic representations of the proper noun entities or lexical representations of the names). However, this explanation is unlikely, given that we isolated items for treatment that the participant could successfully match to the name of the person and that they recognized. Another possibility is that the locus of learning was improved postlexical encoding and/or articulation, processes that operate subsequent to word form retrieval. This also is unlikely in the current study. On average, 77% of experimental items elicited naming responses that had zero phonological overlap with the target name one or more times during the item selection phase. Postlexical encoding/articulation deficits would be expected to lead to a preponderance of naming responses that are recognizable as the target name but with phonological or phonetic distortions.

Thus, the results suggest retrieval practice and spacing are particularly effective for bolstering the accessibility of existing lexical representations in a persistent fashion. How should these learning effects be explained? The retrieval practice effects are consistent with the class of “effort” theories in the test-enhanced learning literature (e.g., Bjork, 1975; Pyc & Rawson, 2009), which contend that the benefits of retrieval practice over restudy reflect the depth or complexity of processing required to retrieve target information from long-term memory. This descriptive framework has motivated the development of more mechanistic accounts. For example, according to the elaborative retrieval hypothesis (e.g., Carpenter, 2009) the advantage of retrieval practice over restudy reflects the activation and encoding of related elaborative information during the search through memory for a target given a cue. To illustrate, a retrieval cue (*cucumber* - _____) initiates a search of long-term memory that along with the target (*frog*) may also activate other information related to the cue (*green, smooth*). If the target is successfully retrieved, this additional information may be encoded along with the cue and target to yield an “elaborated” retrieval structure. This structure provides additional retrieval routes for subsequently accessing the target when given the cue (e.g., *cucumber* → *green* → *frog*). However, semantic elaboration accounts are not well suited for explaining how retrieval practice effects are obtained in tasks where the cue and/or target are meaningless to learners (e.g., learning unfamiliar visual symbols such as Chinese characters; Kang, 2010). Likewise, semantic elaboration is an unlikely mechanism underlying the retrieval practice and effort effects observed in the present study because (aside from gender features) the constituents of proper names are largely semantically opaque.

Our effects may be better understood—and perhaps implemented computationally—by drawing on existing models of lexical retrieval and use-dependent language change. An emerging view in psycholinguistics is that each occasion of language use constitutes a learning experience, even in the adult language system. This *incremental*

learning view of language processing is supported by a variety of psycholinguistic phenomena, ranging from structural priming to semantic blocking effects (e.g., Chang, Dell, & Bock, 2006; Oppenheim, Dell, & Schwartz, 2010). Oppenheim et al. (2010) captured a broad range of incremental learning effects in naming in a computational model where subsequent to naming, a learning algorithm strengthened retrieval connections (weights) between meaning-based representations and the lemma that was retrieved, and weakened weights to nontarget lemmas that had been activated because of shared semantics with the target. The weight-strengthening component of the learning algorithm implemented repetition priming (i.e., persistent increase in retrievability of a target lemma), whereas the weight-weakening component implemented semantic blocking (persistent decrease in retrievability of competitor lemmas). In such a framework, the retrieval practice and retrieval effort effects could be captured by assuming (for example) a greater degree of strengthening from meaning to lemma when (a) the lemma is selected in a top-down fashion from meaning (i.e., retrieval practice) rather than mapping from input phonology (i.e., repetition); and (b) selection from meaning requires more versus less effort.

How might the spacing effect be implemented within such a framework? A consistent theme in the memory literature is that differential involvement of long- versus short-term memory across multiple training trials for an item underlies spacing effects (for review, see Toppino & Gerbier, 2014). In the incremental learning framework we outlined, it may be necessary to invoke an additional mechanism or parameter corresponding to whether a word is referenced in long-term memory versus continuously activated in short-term memory on training trials for an item, with concomitant consequences for differential learning to explain spacing effects in word retrieval. There is a great need for further research to expand existing theories or develop new frameworks in order to explain retrieval practice, retrieval effort, and spacing effects in word retrieval.

Turning to the clinical implications of the present work, next steps for advancing our theory of learning will involve investigating how retrieval practice and spacing factors play out when items are trained in dosage levels more representative of clinical practice (e.g., multiple training trials in multiple sessions) and with retention intervals of greater functional significance (e.g., weeks, months). We anticipate encouraging results, given that studies of skill and knowledge acquisition show that these learning principles do scale up. For example, comparison of multiple sessions of retrieval practice versus restudy have shown a reliable advantage for retrieval practice at final test (Cull, 2000; Morris & Fritz, 2002; Rawson, Dunlosky, Scartelli, 2013), even after relatively long retention intervals (2 months; Morris & Fritz, 2002).

Likewise, greater spacing has been shown to confer greater learning with retention intervals on the order of years (Bahrnick, Bahrnick, Bahrnick, & Bahrnick, 1993; Rohrer,

Taylor, Pashler, Wixted, & Cepeda, 2005). For example, on tests administered after retention intervals spanning 1 to 5 years, Bahrnick et al. (1993) found 13 sessions of training administered every 56 days produced the same level of performance as 26 sessions of training administered every 14 days. This and similar observations from the memory literature, coupled with the current demonstration that naming performance in aphasia is subject to spacing principles, are provocative in light of growing interest in *intensive therapies in aphasia*. Intensive therapy involves many treatment sessions administered in a relatively short period of time (e.g., over a few weeks, rather than months). However, it is not clear which aspects of intensive therapy are critical for conferring persistent functional change—intensive therapies typically involve more time in treatment relative to control treatments (Bhagal, Teasell, & Speechley, 2003; Harnish, Neils-Strunjas, Lamy, & Eliassen, 2008; Sage, Snell, & Ralph, 2011), and more therapy typically confers better outcomes (e.g., Bhagal et al., 2003; Cherney, Patterson, Raymer, Frymark, & Schooling, 2008). Thus, whether massing is critical to the success of intensive therapies is not clear. Indeed, in a recent study on intensity focusing on naming impairment in PWA (Sage et al., 2011) there was some indication that *greater* spacing of sessions may confer superior benefit. In that study, matched sets of items were trained intensively (10 sessions of naming therapy distributed over 2 weeks) or in a nonintensive schedule (10 sessions of naming therapy distributed over 5 weeks). Shortly after treatment, performance in the two conditions was similar, but a significant advantage was found for the nonintensive condition relative to intensive 1 month later. This pattern aligns with the well-documented spacing by retention interval interaction, in which the relative benefit of greater (versus lesser) spacing during training increases at longer retention intervals (for review, see Toppino & Gerbier, 2014). Our results and those of Sage et al. (2011) suggest it may be important in future studies on intensity to delineate the contribution of amount and type of therapy separate from schedule of administration in promoting persistent improvement in language function.

In conclusion, the current study demonstrates that naming rehabilitation is subject to the same learning principles that are proving to dramatically affect outcomes in other domains of cognitive rehabilitation (e.g., Sumowski, Chiaravalloti, & DeLuca, 2010) and in education (for review, see Dunlosky et al., 2013). We hope the empirical foundation our work provides will stimulate future research aimed at exploring how such learning principles may be used to develop maximally effective, long-lasting, and efficient treatments of aphasia.

Acknowledgments

This work was supported by National Institutes of Health research grants R01-DC000191, awarded to Myrna F. Schwartz, and R03-DC012426, awarded to Erica L. Middleton. A portion of this work was presented at the International Workshop on Language Production, Geneva, Switzerland.

References

- Abel, S., Willmes, K., & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology, 21*, 411–447.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.
- Baddeley, A. (1993). A theory of rehabilitation without a model of learning is a vehicle without an engine: A comment on Caramazza and Hillis. *Neuropsychological Rehabilitation, 3*, 235–244.
- Bahrnick, H. P., Bahrnick, L. E., Bahrnick, A. S., & Bahrnick, P. E. (1993). Maintenance of foreign-language vocabulary and the spacing effect. *Psychological Science, 4*, 316–321.
- Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57*, 35–56.
- Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*, 228–247.
- Bhagal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34*, 987–992.
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.
- Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: Evidence from once-presented words. *Memory, 6*, 37–65.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1563–1569.
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619–636.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633–642.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234–272.
- Cherney, L. R., Patterson, J. P., Raymer, A., Frymark, T., & Schooling, T. (2008). Evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of Speech, Language, and Hearing Research, 51*, 1282–1299.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215–235.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4–58.
- Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation, 13*, 337–363.
- Harnish, S. M., Neils-Strunjas, J., Lamy, M., & Eliassen, J. (2008). Use of fMRI in the study of chronic aphasia recovery after therapy: A case study. *Topics in Stroke Rehabilitation, 15*, 468–483.
- James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt and Company.
- Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition, 38*, 1009–1017.
- Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1250–1257.
- Karpicke, J. D., & Blunt, J. R. (2011, February 11). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772–775.
- Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704–719.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989–998.
- Lecours, A. R., & Lhermitte, F. (1969). Phonemic paraphasias: Linguistic structures and tentative hypotheses. *Cortex, 5*, 193–228.
- Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performances on statistics exams. *Teaching of Psychology, 38*, 94–97.
- Menzel, R., Manz, G., Menzel, R., & Greggers, U. (2001). Massed and spaced learning in honeybees: The role of CS, US, the intertribal interval, and the test interval. *Learning and Memory, 8*, 198–208.
- Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation, 22*, 138–168.
- Middleton, E. L., Schwartz, M. F., Rawson, K. A., & Garvey, K. (2015). Test-enhanced learning versus errorless learning in aphasia rehabilitation: Testing competing psychological principles. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1253–1261.
- Morris, P. E., & Fritz, C. O. (2002). The improved name game: Better use of expanding retrieval practice. *Memory, 10*, 259–266.
- Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology, 16*, 935–979.
- Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language, 63*, 541–559.
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*, 227–252.
- Park, Y., Goral, M., Verkuilen, J., & Kempler, D. (2013). Effects of noun-verb conceptual/phonological relatedness on verb production changes in Broca's aphasia. *Aphasiology, 27*, 811–827.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3–8.
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1051–1057.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437–447.
- Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much

-
- is enough? *Journal of Experimental Psychology: General*, *140*, 283–302.
- Rawson, K. A., Dunlosky, J., & Sciartelli, S. M.** (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychological Review*, *25*, 523–548.
- Roediger, H. L., & Karpicke, J. D.** (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Roediger, H. L., Putnam, A. L., & Smith, M. A.** (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, *44*, 1–36.
- Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J.** (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology*, *19*, 361–374.
- Sage, K., Snell, C., & Ralph, M. A. L.** (2011). How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation*, *21*, 26–41.
- Stark, J. A.** (2005). Analyzing the language therapy process: The implicit role of learning and memory. *Aphasiology*, *19*, 1074–1089.
- StataCorp.** (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.
- Sumowski, J. F., Chiaravalloti, N., & DeLuca, J.** (2010). Retrieval practice improves memory in multiple sclerosis: Clinical application of the testing effect. *Neuropsychology*, *24*, 267–272.
- Toppino, T. C., & Gerbier, E.** (2014). About practice: Repetition, spacing, and abstraction. *Psychology of Learning and Motivation*, *60*, 113–189.
- Vaughn, K. E., & Rawson, K. A.** (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin and Review*, *19*, 899–905.
- Wheeler, M. A., Ewers, M., & Buonanno, J. F.** (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571–580.