Retrieval practice and spacing effects in multi-session treatment of naming impairment in aphasia.

Erica L. Middleton[1], **Katherine A. Rawson**[2], **Jay Verkuilen**[3]

[1]Moss Rehabilitation Research Institute

[2]Kent State University

[3]City University of New York Graduate Center

Abstract

Retrieval practice and spacing are two factors shown to enhance learning in basic psychological research. The present study investigated the clinical applicability of these factors to naming treatment in aphasia. Prior studies have shown that naming treatment that provides retrieval practice (i.e., practice retrieving names for objects from semantic memory) improves later naming performance in people with aphasia (PWA) more so than repetition training. Repetition training is a common form of naming treatment that can support errorless production of names for objects, but it does not provide retrieval practice. Prior work has also demonstrated enhanced naming treatment benefit in PWA when an item's training trials are separated by multiple intervening trials (i.e., spacing) compared to only one intervening trial (i.e., massing). However, in those studies, items were only trained in one session. Also, the effects of the learning factors were probed after one day and one week. The goal of the present study was to examine the effects of retrieval practice and spacing in a more clinically-inspired schedule of delivery and to assess the effects of the learning factors at retention intervals of greater functional significance. Matched sets of errorful items for each of four PWA were presented for multiple trials of retrieval practice or repetition in a spaced or massed schedule in each of multiple training sessions. Mixed regression analyses revealed that retrieval practice outperformed repetition, and spacing outperformed massing, at an initial post-treatment test administered after one week. Furthermore, the advantage for retrieval practice over repetition persisted at a follow-up test administered after one month. The potential clinical relevance of retrieval practice and spacing for multi-session interventions in speech-language treatment is discussed.

1. Introduction

Various domains within cognitive rehabilitation research are demonstrating increasing interest in the application of principles of learning for improving treatment (for reviews see Clare & Jones, 2008; Dignam, Rodriquez, & Copland, 2016; Fillingham, Hodgson, Sage, & Lambon Ralph, 2003; Middleton & Schwartz, 2012; Oren, Willerton, & Small, 2014). This interest in learning principles is due to their potential to illuminate which experiences provided by cognitive rehabilitation interventions are critical to their effect, and to explain how such experiences impact underlying functional deficits. Ultimately, the application of principles of learning may permit theory-guided predictions of how to design treatments to maximize efficacy and efficiency.

A wealth of research in cognitive and educational psychology points to the importance of two learning factors, retrieval practice and spacing, for enhancing skill and knowledge acquisition. *Retrieval practice*, or the act of retrieving information from long-term memory, can powerfully bolster retrieval of that information in the future compared to non-retrieval based forms of training. Likewise, repeated training trials for an item confer more learning when they are separated by multiple intervening trials for other items (i.e., spacing) compared to being presented contiguously or with only one intervening trial (i.e., massing).

In initial studies (Middleton, Schwartz, Rawson, & Garvey, 2015; Middleton, Schwartz, Rawson, Traut, & Verkuilen, 2016), we established the relevance of retrieval practice and spacing to speech-language rehabilitation. In those studies, the impact of the learning factors on oral naming impairment in people with aphasia (PWA) was assessed with a focus on PWAs' persistent ability to name treated vocabulary. These initial studies trained items within one session only. Furthermore, the longest post-training retention interval at which the effects of the learning factors were assessed was one week. Building on this prior work, the present study examined the effects of retrieval practice and spacing during naming treatment that was administered in a more clinically-inspired, multi-session schedule of delivery. Additionally, we examined the effects of these two factors at retention intervals of greater practical importance.

1.1 Retrieval Practice and Spacing Effects in Psychology

In cognitive and educational psychology, the standard paradigm for examining the effects of retrieval practice begins with initial familiarization of the to-be-learned (i.e., target) information. Familiarization is followed by a training phase in which the information is either presented again in its entirety for further study opportunities (i.e., restudy condition), or participants attempt to retrieve the target information from long-term memory (i.e., retrieval practice condition). A *retrieval practice effect* is demonstrated when performance on a post-training test is superior for information assigned to the retrieval practice condition compared to the restudy condition. Retrieval practice effects have been demonstrated in

tasks tapping episodic memory (e.g., Carrier & Pashler, 1992), semantic memory (e.g., Butler & Roediger, 2008; Lyle & Crawford, 2011; Roediger & Karpicke, 2006), procedural learning (Kromann, Jensen, & Ringsted, 2009), and language acquisition (e.g., Barcroft, 2007; Vaughn & Rawson, 2011; for recent reviews see Kornell & Vaughn, 2016; Rawson & Dunlosky, 2011; Roediger & Butler, 2011; Roediger, Putnam & Smith, 2011; Rowland, 2014).

The standard paradigm for examining the effects of spacing also typically begins with familiarization of target items followed by multiple training trials per item. In a spaced schedule, an item's trials are separated by a sufficient number of training trials for other items (i.e., *lag*) so as to exceed the limits of short-term memory. In the massed schedule, an item's trials are presented at Lag 0 (i.e., zero intervening trials) or Lag 1. The consequence is that the item remains accessible in short-term memory across its trials. The *spacing effect* refers to superior post-training test performance following spaced versus massed training schedules. Spacing robustly benefits various types of learning (e.g., episodic memory, concept acquisition, procedural skill learning) in learners across the age span (for recent reviews, see Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Delaney, Verkoeijen, & Spirgel, 2010; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Toppino & Gerbier, 2014). Furthermore, increasing the lag used in the spaced schedule can sometimes enhance spacing effects. For example, a number of studies have found superior post-training performance when trials for specific items are spaced at longer (e.g., 34) versus shorter (e.g., 6) lags (Kornell, 2009; Pyc & Rawson, 2007, 2009, 2012a, 2012b). Other work has shown that when the number of retrievals per item is equated, administering an item's trials across sessions rather than within a session enhances learning (e.g., Kornell, 2009; Rawson, Vaughn, Walsh, & Dunlosky, 2018; Vaughn, Dunlosky, & Rawson, 2016).

1. 2   Retrieval Practice and Spacing Effects in Cognitive Rehabilitation

The degree to which the clinical applicability of retrieval practice and spacing effects has been acknowledged and has influenced treatment research is variable across the domains of cognitive rehabilitation. Some domains show a robust and growing evidence base, such as those dedicated to cognitive rehabilitation of people with traumatic brain injury (e.g., Coyne, Borg, DeLuca, Glass, & Sumowski, 2015; Goverover, Arango-Lasprilla, Hillary, Chiaravalloti, & DeLuca, 2009; Pastötter, Weber, & Bäuml, 2013; Sumowski et al., 2010; Sumowski, Coyne, Cohen, & DeLuca, 2014) and multiple sclerosis (e.g., Goverover, Basso, Wood, Chiaravalloti, & DeLuca, 2011; Goverover, Hillary, Chiaravalloti, Arango-Lasprilla, & DeLuca, 2009; Sumowski, Chiaravalloti, & DeLuca, 2010; Sumowski et al., 2013). In contrast, research on retrieval practice and spacing effects in speech-language rehabilitation is sparse (Friedman, Sullivan, Snider, Luta, & Jones, 2017; Middleton et al., 2015, 2016).

Middleton et al. (2015) first investigated the impact of retrieval practice on disrupted lexical access in a group study of eight PWA with naming impairment. Lexical access deficit is a common cause of naming impairment across the subtypes of aphasia that manifests as difficulty reliably and fluently retrieving lexical representations (i.e., words and their forms) for production (for review, see Mirman & Brit, 2014). Middleton et al. compared how forms of naming treatment that paralleled conditions in the standard retrieval practice paradigm

affected later naming performance. In a within-participants design, sets of pictures of common, everyday objects that a PWA experienced difficulty naming were assigned into different training conditions. For all items selected for training for a PWA, each picture and its name were first displayed to the participant, analogous to initial familiarization in the standard retrieval practice paradigm. After the initial familiarization phase, each item then appeared for one trial in one of three training conditions: (1) cued retrieval practice, in which the picture was displayed and accompanied by name onset to facilitate naming, (2) non-cued retrieval practice, in which the picture alone was displayed for naming, or (3) errorless learning naming treatment, in which the name was presented at picture onset and the participant repeated the name (i.e., repetition training). Naming treatment based on principles of errorless learning attempts to maximize accurate production of target names for pictures. Accuracy is prioritized because of the possibility that failed naming (e.g., production of the wrong name) promotes future errorful responding on those items, decreasing treatment efficacy (for reviews, see Fillingham et al., 2003; Middleton & Schwartz, 2012). The most common form of errorless learning in the naming treatment literature employs word repetition training (e.g., Fillingham, Sage, & Lambon Ralph, 2005a, 2005b, 2006; McKissock & Ward, 2007). However, repetition training is akin to the restudy condition in the standard retrieval practice paradigm. In repetition training, given that the target name is activated from input phonology (Nozari, Kittredge, Dell, & Schwartz, 2010), production is achieved without requiring retrieval of the target name from semantic memory. Because of this, we expected less improvement in naming from repetition training than from the retrieval practice conditions. Outcomes confirmed this expectation. Both retrieval practice conditions outperformed repetition training on an initial naming test administered the next day (i.e., 1-day test). This finding provided the first demonstration of *retrieval practice effects* in naming treatment. Furthermore, the advantage of cued retrieval practice over repetition persisted on a follow-up test administered after one week (i.e., 1-week test).

Following up on Middleton et al. (2015), Middleton et al. (2016) further examined the effects of retrieval practice in naming treatment but additionally manipulated the spacing of trials. The study involved a group of four PWA. The stimuli used were entities eliciting proper nouns. Prior to the experiment, each PWA underwent extensive testing to identify a set of proper noun entities in which the participant indicated knowledge of each entity and its name but experienced difficulty naming. Sets of these items were then presented for non-cued retrieval practice or for repetition training according to either massed or spaced schedules. On tests one day and one week later, retrieval practice outperformed repetition training, replicating the retrieval practice effect observed in Middleton et al. (2015). Furthermore, spacing outperformed massing at both test intervals. This finding provided the first demonstration of a *spacing effect* in naming treatment.

In a recent study, Friedman et al. (2017) compared retrieval practice and repetition training as techniques for treating naming impairment in three PWA. In that study, errorful items for each PWA were first trained via alternating blocks of repetition training and retrieval practice administered in each of multiple sessions. After the PWA had mastered an item to a criterion in which the item was accurately named at the beginning of each of two consecutive sessions, the item was administered in multiple additional sessions for 'overlearning' involving either further repetition training or retrieval practice. On naming

tests administered 1-month and 4-months after the end of training, performance was greater when overlearning involved retrieval practice versus repetition. These results suggest that retrieval practice is more beneficial than repetition for prolonging the maintenance of mastery. However, they do not inform whether retrieval practice or repetition should be prioritized to achieve initial mastery from multi-session training.

In sum, only three prior studies have examined retrieval practice effects in speech-language rehabilitation, only one prior study examined spacing effects, and none have examined both factors in a clinically-relevant, multiple-session training regimen.

1.3   Overview of Current Research

Four PWA completed the current study, which employed a within-participant design. The primary focus was on results across the group. Neuropsychological characterization of the PWA revealed naming impairment with lexical access deficit as a major contributor to their naming difficulties. As in Middleton et al. (2016), the stimuli involved proper noun entities. Prior to the experiment, each PWA underwent extensive testing involving a large picture corpus of proper noun entities to identify items in which the PWA indicated they knew the entities and their names but experienced difficulty naming. The items identified for each participant were then divided equally into four conditions formed by crossing a two-level factor of type of training (retrieval practice versus repetition training) with a two-level factor of spacing (spaced versus massed schedule). Items were trained in their assigned condition in each of four sessions administered over two weeks. The effects of the training factor and spacing factor were assessed at a test administered one week following the final training session (hereafter, *retention test*). We also examined the persistence of such effects on another test one month following the final training session (hereafter, *follow-up test*).

As a brief aside, untreated items can be included in naming treatment studies to measure spontaneous recovery, generalization from treatment, or regression to the mean that can result from selecting error-prone items for treatment. Our main research question concerned differences in performance at the retention test and follow-up as a function of the training and spacing manipulations. Because items assigned to the conditions were equally errorful, there is no reason to expect improvements that are unrelated to training to be greater in any one condition. Thus, in the current study, we did not administer or assess change in untreated items.

An acknowledged shortcoming of the retrieval practice and spacing literatures is the almost exclusive focus on the impact of these factors when items are trained in a single session (Bahrick, 1979; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; for discussion, see Rawson & Dunlosky, 2011, 2013). In psychological research, the few studies that have compared retrieval practice to restudy when items are trained in multiple sessions have generally reported retrieval practice effects (Cull, 2000; Metcalf, Kornell, & Son, 2007; Morris & Fritz, 2002; Rawson, Dunlosky, Sciartelli, 2013) with one exception (Goossens et al., 2016). Hence, we tentatively expected to observe a retrieval practice effect in the current multi-session design. In the aphasia literature, to the degree that similar principles apply to overlearned items as to items that have yet to be mastered, the results reported by Friedman et al. (2017) provide additional

support for predicting a retrieval practice effect at the post-training tests in the present design.

As regards the spacing manipulation, to our knowledge only one prior study has examined spacing effects when items are administered in massed versus spaced fashion in each of multiple sessions (Rawson & Dunlosky, 2013, Experiment 3). In that study, items were presented for retrieval practice in a massed versus spaced schedule in an initial session, after which the items underwent *relearning* in multiple, subsequent sessions. During relearning, each item was presented for retrieval practice in either a massed schedule or a spaced schedule followed by correct-answer feedback until the item was successfully retrieved once during the session. In each relearning session, performance on the first test trial was consistently greater for spaced items versus massed items. These findings indicate that the spacing effect can persist across multiple sessions. Similarly, the current study investigated whether the spacing effect persists across training sessions. To do so, we examined performance on the first trial per item in the retrieval practice condition for items in the spaced versus massed schedule in training sessions 2 through 4. One possibility is that both spaced and massed items benefit from each long (multi-day) lag between training sessions, and thus the spacing effect may diminish across training sessions. Attenuation of the spacing effect across sessions is of clinical relevance because it could indicate that spacing effects observed early in training may ultimately disappear, given enough training sessions. Thus, we assess whether the magnitude of the spacing effect in the retrieval practice condition changes from training sessions 2–4. However, the outcome of greater interest concerns whether we observe a spacing effect across both training conditions on the post-training retention test and follow-up test.

Lastly, we will inspect a possible interaction of the training and spacing factors. Some studies have reported an interaction of these factors at final test, with an enhanced advantage for retrieval practice over restudy in spaced compared to massed schedules (e.g., Carpenter & DeLosh, 2005; Cull, 2000). However, such investigations have been limited to single-session studies. Thus, it is an open question whether an interaction will obtain in the current design.

2. Method

Even among those diagnosed with the same subtype of aphasia (e.g., Broca's), PWA exhibit substantial variability in the form and severity of deficits in cognitive and language-based processes. It is reasonable to expect that such differences in deficits might increase between-participants variability in response to experimental manipulations. Likewise, neurological damage can increase variability in how an individual responds to trials within the same task (MacDonald, Nyberg, & Bäckman, 2006). To increase experimental power, the strategy adopted in the present study as well as in our prior work (Middleton et al., 2015, 2016) involved (1) selecting individuals for study that are relatively homogeneous in their profile of cognitive-linguistic deficits, and (2) administering large numbers of items per condition per participant to produce stable results within and across individuals. In Middleton et al. (2016), 50 observations per condition for each of four PWA provided sufficient power for detecting learning effects of interest. Thus, our current recruitment target was four PWA,

with 48 observations per condition per participant. Capping recruitment at four individuals was also necessitated by the substantial resources required for data collection and processing for each PWA, including a minimum of 17 sessions per PWA to complete the protocol.

2.1 Participants

Participants gave informed consent under a protocol approved by the institutional review board of Einstein Healthcare Network. They were reimbursed $15 for each hour of participation. The four participants (1 male) were right-handed with chronic aphasia of the anomic subtype secondary to left-hemisphere stroke. Hereafter, individual participants will be denoted by P1, P2, P3 and P4. See Table 1 for demographics and test battery scores for the four participants. The participants were recruited from a large pool (N = 133) of potential research volunteers with chronic aphasia secondary to stroke who had completed an extensive language test battery. The goal of recruitment was to select participants with oral naming impairment attributable to difficulty retrieving known words and/or their forms (i.e., lexical access deficit). The group showed very good performance on a word comprehension task (Mirman et al., 2010) and good performance on a nonverbal semantic comprehension task (Howard & Patterson, 1992). Slightly worse performance on a task of semantic relatedness of words (synonymy triplets; Martin, Schwartz, & Kohen, 2006) indicated subtle semantic deficits, but overall their comprehension performance suggests generally intact semantic processing and access to the meanings of words. In light of this, the preponderance of semantic substitution errors (e.g., *caterpillar* for *butterfly*) and omissions (i.e., failure to produce a naming attempt) on a test of oral picture naming (Philadelphia Naming Test; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) likely reflect problems mapping from semantics to words (Chen, Middleton, & Mirman, in press; Schwartz et al., 2009). Phonological errors in naming were the second most common error type. These errors can reflect lexical-phonological access problems or deficits in stages of production after lexical-phonological retrieval (i.e., deficit in post-lexical phonological encoding, Goldrick & Rapp, 2007, or in articulation). However, post-lexical disruption as a major contributor to the group's naming impairment is unlikely, given that they tended to show very good word repetition ability. Compared to word repetition, nonword repetition ability was more impaired. However, nonword repetition is additionally sensitive to disorders in phonological input processing and to working-memory or short-term memory deficits. Verbal short-term memory deficits were present in our sample in degrees of severity that appeared to track nonword performance (see STM span; Martin, Shelton, & Yaffee, 1994). Overall, background testing pointed to lexical access deficit as a major (though not necessarily exclusive) source of naming impairment in our group.

2.2 Materials

To develop the 700-item proper noun picture corpus, images for 1073 proper noun entities were first collected from Internet sources. From a sample of 10 neurotypical older adults (mean age = 65.18, *SD* = 7.44; mean education = 13.71, *SD* = 1.90), between 7 to 10 entity recognition and name recognition judgments were collected for each of the 1073 images. For a description of the rating scales, see section 2.4.1. Based on these ratings, we selected 700 entities that were rated the most recognizable (*M* = 1.30, *SD* = .23) with the most recognizable names (*M* = 1.28, *SD* = .25) to comprise the 700-item proper noun picture

corpus in the current study. Thus, entities and their names in the 700-item corpus were generally highly recognizable to people with demographic characteristics similar to our sample of PWA. The 700-item proper noun corpus comprised 81% famous people (e.g., movie stars, politicians, historical figures), 9% fictional characters (e.g., Pink Panther), and 11% films with iconic movie posters (e.g., Back to the Future). See Middleton et al. (2016) for additional information regarding corpus development. Lastly, from the remaining 373 of the 1073 original set of images, we selected 42 entities that were rated the most recognizable with the most recognizable names to fulfill other purposes described in section 2.4.1. Materials are available upon request.

2.3 Response Accuracy

Trained research assistants transcribed responses using the International Phonetic Alphabet. Extensive details regarding response coding are reported in Middleton et al. (2016). To summarize here, response coding proceeded in two stages. First, we calculated a continuous measure of phonological similarity (Lecours & Lhermitte, 1969) between the naming response and target name. Second, responses with a phonological similarity value of .75 or greater were coded as 'correct,' and responses with phonological similarity values less than .75 were coded as 'incorrect.' The rationale for coding productions that included most of the target name as correct was to give credit for successful name retrieval while ignoring minor deviations in word form that can occur during post-lexical phonological-phonetic encoding. Across the four participants, about a third of correct responses on the post-training tests contained minor deviations in form ($M = .35$, $SD = .15$). Although the participants had very good word repetition ability indicating generally intact post-lexical processing (see section 2.1), this rate of distortion is not unexpected due to the phonologically complex nature of the proper noun target names (i.e., multi-syllabic and multi-morphemic; see Table 2).

2.4 Procedure

2.4.1 Item selection phase.—The goal of the present study was to investigate the impact of retrieval practice and spacing on the accessibility of existing lexical representations rather than on acquisition of new vocabulary. Using procedures from Middleton et al. (2016), each PWA completed extensive testing on the 700-item proper noun picture corpus prior to the experiment to select items for training. The outcome of item selection was identification of 192 *personalized* items that the PWA consistently experienced difficulty naming, but in which each entity and its name were highly familiar to the participant. The item selection phase started with a picture-to-name matching task in which the 700-item picture corpus was presented one at a time in random order. Participants were instructed 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 Grace Kelly, foils were Ingrid Bergman, Vivian Leigh, and Greta Garbo), and the fourth option was "None of the above." Using other pictures, we presented an additional 42 items among the 700 items in which the correct answer was "None of the above" to prevent participants from ignoring this option. The picture-to-name matching task required between 1 to 3 sessions for completion.

In a week following administration of the picture-to-name matching task, the 700-item corpus was presented twice in random order for oral naming and recognition judgments,

with 2–4 sessions of testing required for each administration. On each naming trial, the participant saw a picture and had up to 20 seconds to attempt to name it. Participants were instructed to indicate when they were finished attempting to name the picture, at which point the experimenter advanced to the next trial. This procedure was designed to remove experimenter feedback of any kind. If the participant did not give their final answer within 20 seconds, the trial advanced automatically. Each naming trial was immediately followed by a query asking whether the PWA recognized the entity ("Do you recognize this person or thing? 1=yes, 2=not sure, 3=no") and a query whether the PWA recognized 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"). To summarize, each participant provided two naming responses, two entity recognition judgments, and two name recognition judgments for each item in the 700-item corpus.

From this procedure, we selected a set of 192 personalized items per participant. Each participant's personalized item set was divided into the four conditions, resulting in 48 items per condition per participant that were matched for the variables reported in Table 2. As the purpose of item selection was to identify items for training for each participant in which the entity and name were known but the PWA experienced difficulty naming, items selected for a participant were those that most closely met all of the following four criteria: (1) the item was associated with a correct response on the picture-to-name matching task; (2) the item received a '1' response on both trials of the entity recognition task; (3) the item received a '1' response on both trials of the name recognition task; (4) responses on both naming trials were incorrect. Of the 768 items selected for training across the four participants, 100% of items met criterion 1, 94% met criterion 2, 81% met criterion 3, and 98% met criterion 4 (i.e., 98% of items elicited naming error on both trials during item selection). Thus, at least 94% and 81% of items for entity recognition and name recognition, respectively, were consistently given a '1' indicating full recognition across two independent trials per item for a participant. In addition, for both the entity recognition and name recognition tasks, rating averages in the conditions were very close to 1 on the 1–3 scale (see Table 2). We take these results as strong evidence that the participants had pre-existing conceptual and linguistic representations for the proper noun entities in their personalized item sets, and that naming problems on the items primarily reflected word retrieval failure.

2.4.2   Main experiment.—The protocol included two cycles. Each cycle included six sessions. The first four sessions in a cycle were dedicated to training and were administered two per week, with 1 to 4 days between consecutive sessions. The fifth session was held approximately 7 days following the final training session. Session 5 involved the retention test of naming for the items trained in that cycle. The sixth session was held one month (28 days) after the final training session. Session 6 involved the follow-up test of naming for the same items. For all participants, Session 6 for the first cycle was administered prior Session 1 of the second cycle. In each cycle, a participant received training and testing on 24 items per each of the four conditions, which corresponded to half (96 items) of their entire personalized set (192 items). Thus, because items from the four conditions were trained in each session in a cycle, the timing of sessions and other aspects of administration were equated across the conditions.

Each training session included two blocks of 159 trials. Each block included 144 experimental trials and 15 filler trials. The items used for filler trials were other items from the 700-item corpus that were not in a participant's personalized item set. These were items in which the participant indicated recognition of the entity and its name at item selection, but the participant experienced less difficulty with naming compared to items in their personalized item set. Easier items were selected as fillers with the expectation that they could help ameliorate frustration during training.

Each block involved training for 12 items per each of the four conditions. Each item was presented for 3 trials according to its assigned schedule of learning. In the spaced schedule, an item's trials were presented with an intervening 24 trials (Lag 24). In the massed schedule, an item's trials were separated by one intervening trial (Lag 1). Each block began and ended with 5 filler trials to avoid privileging memory for items appearing at the beginning or end of the block. Five additional filler trials were used to help fill out the lag sequence for the experimental items in the block. The average ordinal position within a block was equated for the items in the four conditions. Participants were encouraged to rest between the two blocks.

At the beginning of the first training session only, the first block was preceded by ten practice trials composed of a mix of retrieval practice and repetition training trials on filler items to help participants become accustomed to completing the two kinds of training trials in interleaved fashion.

The procedure for repetition and retrieval practice trials was adopted from Middleton et al. (2016). In the repetition condition, an item was presented for three repetition trials in each training session. On a repetition trial, a picture and its written name appeared on the screen and remained on the display. At picture onset, the name was also presented once auditorally. The participant was instructed to repeat the name once. After 8 seconds from picture onset, a *feedback procedure* concluded the trial in which the auditory form of the name was played again and the participant repeated the name. In the retrieval practice condition, in the first training session in a cycle, the first trial for each item was a repetition trial in analogy to initial study in the typical retrieval practice paradigm. All subsequent trials for items in this condition were retrieval practice trials, with two retrieval practice trials per item in the first session and three retrieval practice trials per item in each of the three subsequent training sessions. On a retrieval practice trial, the picture was presented and the participant was provided 8 seconds to attempt to name the picture. After 8 seconds, the trial concluded with the feedback procedure described above. In sum, each item appeared in one block of training trials in each of 4 sessions, receiving a total of 12 training trials according to its assigned condition.

For the retention test and follow-up test, each item was presented for one naming trial using the same procedure as during item selection (i.e., 20 seconds for naming with no feedback). Each item selection session and training session required approximately 2 hours. Retention test and follow-up test sessions required approximately 30 minutes.

2.5 Analyses

No part of the study procedures or analyses was preregistered prior to the research being undertaken. Data and analysis code is available at https://osf.io/cwq75/. We conducted separate analyses on retention test and follow-up test performance. In each analysis, response accuracy (correct/incorrect) was modeled with mixed logistic regression using Stata 14.2 (StataCorp, 2015) with items included as a random effect in the regression models. It was not appropriate to treat participants as a random effect in the regression models because of the small sample of PWA, and because they were non-randomly selected from a larger pool of PWA volunteers based on their cognitive-linguistic profile. However, it is possible to calculate unbiased estimates of the effects of the experimental factors in a group analysis if the participants respond similarly to the experimental factors. In all models, we entered participants as a categorical variable and inspected potential interactions between the participants factor and the experimental factors to assess possible differences in how the participants responded to the experimental factors (for similar approaches, see Middleton et al., 2016; Park, Goral, Verkuilen, & Kempler, 2013).

With the fixed effects of training (retrieval practice versus repetition training), spacing (spaced versus massed schedule), and participants, we modeled all possible combinations of main effects and interactions of the fixed effects. As an initial criterion for model selection, we selected as best the model with the lowest Bayesian information criterion (BIC) and Akaike information criterion (AIC). BIC and AIC are indices of model fit that penalize for complexity (i.e., the number of terms in a model) to mitigate overfitting. As a second criterion for model selection, if the difference between nested models in terms of BIC and AIC was negligible, we report the more complex model when warranted by a chi-square test of deviance in model log likelihood (alpha = .05). To illustrate, in the analysis on retention test performance, the best model identified by AIC and BIC included terms for a main effect of spacing and an interaction of participants and training, reported in Table 4. Furthermore, this model provided a better fit to the data [χ^2 (3) = 20.70, p<.0001] compared to the simpler nested model containing main effect terms for spacing, training, and participants with no interactions. In other words, the model selection procedure revealed that not all participants responded the same way to the training factor at the retention test. Thus, for this analysis we also conducted participant-specific ordinary logistic regression models and report the effects of the experimental factors for each participant (see Table 5). In the analysis on follow-up performance, the best model identified by BIC and AIC included main effect terms for spacing, training, and participants with no interactions (see Table 6). Compared to this model, more complex models were not warranted by chi-square tests of deviance in log likelihood. In other words, at follow-up, the model selection procedure revealed the participants responded similarly to the spacing and training factors.

To summarize, we used mixed logistic regression to evaluate whether a retrieval practice effect and a spacing effect were apparent at the retention test (1-week interval) and whether the effects persisted at the follow-up test (1-month interval). The model selection procedure enabled statistical evaluation of whether the experimental factors interacted at either retention interval, and whether the effects were consistent across participants. Additional

models were compared to assess possible change in magnitude of the spacing effect in the retrieval practice condition across training sessions, described in greater detail below.

3. Results

Figure 1 displays mean response accuracy across participants as a function of the training factor (left panel) and spacing factor (right panel) at retention test and follow-up. Table 3 reports mean response accuracy in each of the four conditions for each of the four participants, along with retrieval practice and spacing effects for each participant at retention test and follow-up. Tables 4, 5, and 6 report regression output for the group model at retention test, individual participant models at retention test, and the group model at follow-up, respectively.

3.1   Interaction of Training and Spacing

In both analyses of retention test performance and follow-up performance, the model selection procedure did not identify as best a model that included the interaction of training and spacing. Nested model comparisons revealed no improvement in model fit by chi-square deviance in log likelihood with addition of the interaction term of training and spacing (all $ps > .62$). Thus, we next report the effects of training when collapsing over the spacing factor, followed by the effects of spacing when collapsing over the training factor.

3.2   Retrieval Practice Effects

To revisit, at retention test the model selection procedure identified as best the model with a main effect of spacing and an interaction between training and participants (see Table 4). Reported in Table 5, the individual participant models applied to retention test performance revealed a significant advantage for retrieval practice training over repetition for three of the four participants: P1 (*Coef.*=−1.11, *SE*=0.32, *Z*=−3.44, *p*=.001), P2 (*Coef.*=−0.63, *SE*=0.29, *Z*=−2.15, *p*=.031), and P4 (*Coef.*=−1.56, *SE*=0.32, *Z*=−4.93, *p*<.001). The differences between training conditions were quite large for these three participants—retention test performance was 24%, 16%, and 36% higher after retrieval practice compared to repetition for P1, P2, and P4, respectively (see Table 3). In terms of raw number of items, of the 96 items trained in each condition per participant, 23, 15, and 35 more items were successfully named at retention test after retrieval practice compared to repetition for P1, P2, and P4, respectively. The remaining participant, P3, showed a nonsignificant but numerical advantage of 8% favoring repetition over retrieval practice (*Coef.*=0.42, *SE*=0.33, *Z*=1.28, *p*=.199; Table 5). However, in the follow-up test analysis, the best model included main effects of training, spacing, and participants but no interactions. The main effect of training was significant (*Coef.*=−0.42, *SE*=0.17, *Z*=−2.54, *p*=.01; see Table 6) with group means showing a 9% advantage for retrieval practice over repetition (Figure 1). At follow-up, all participants showed numerically higher performance in the retrieval practice condition compared to repetition (Table 3). Specifically, P1, P2, P3, and P4 showed a 15%, 9%, 3%, and 7% advantage for retrieval practice over repetition at follow-up, respectively, translating to 14, 9, 3, and 7 more items successfully named with retrieval practice.

3.3 Spacing Effects

The group analysis of retention test performance revealed a significant advantage for the spaced over massed schedule of learning (*Coef.*=0.37, *SE*=0.18, *Z*=2.08, *p*=.038; Table 4). On the follow-up test, the advantage for spaced over massed was marginal (*Coef.*=0.30, *SE*=0.17, *Z*=1.76, *p*=.078; Table 6). The group means advantage for spacing over massing was 7% and 6% at retention test and follow-up, respectively (Figure 1). P1, P2, P3, and P4 showed a 1%, 5%, 13%, and 9% advantage for spacing over massing at retention test (see Table 3), respectively, translating to 1, 5, 12, and 9 more items successfully named with spacing. At follow-up, P1, P2, P3, and P4 showed a 6%, 3%, 7%, and 7% advantage for spacing over massing, respectively, translating to 6, 3, 7, and 7 more items successfully named with spacing.

In the Introduction, we considered the possibility that the multiple-session regimen may progressively diminish the magnitude of the spacing effect across the training sessions. We raised this possibility because the long lags between training sessions could cumulatively benefit all items regardless of the spacing condition to which they are assigned. To assess possible diminishment of the effect of spacing across training sessions, we applied mixed logistic regression to response accuracy on the first trial per item in the retrieval practice condition at sessions 2, 3, and 4. With training session coded as a continuous variable, a model with main effects of participants, session, and spacing was compared to a model with a main effect of participants and an interaction of spacing and session. The main effects-only model was preferred because it was associated with superior AIC and BIC model fit indices, and addition of the interaction of spacing and session did not improve model fit by a chi-square test of deviance in log likelihoods (*p*= .23). In other words, the size of the spacing effect did not significantly grow or diminish in the retrieval practice condition across training sessions. However, the main effect of spacing was highly robust (*Coef.*=1.19, *SE*=0.23, *Z*=5.13, *p*<.001). In sum, the spaced schedule conferred an advantage over the massed schedule in the retrieval practice condition across sessions 2–4, an effect that did not wax or wane across training sessions (see Figure 2).

4. Discussion

In the initial studies upon which the current work builds, selected vocabulary for PWA was presented for a small number of training trials in a single session (Middleton et al., 2015, 2016). In those studies, performance on naming tests administered one day and one week following the training session was greater following retrieval practice versus repetition training and greater following spacing versus massing. The current work provides the critical next step to determining the clinical relevance of retrieval practice and spaced learning for naming treatment by investigating their impact when training is administered in multiple sessions and when performance is assessed after longer delays. Multiple session regimens are common in conventional naming treatment, and they aim to foster cumulative benefits across treatment sessions with durable retention of treatment gains. Thus, in the present study, vocabulary selected for treatment was trained in each of multiple sessions. The difference between training conditions was first measured at one week (retention test), with persistence of such differences inspected at one month following training (follow-up). At the

retention test, performance was significantly higher in the retrieval practice condition compared to the repetition condition for three of the four participants. At follow-up, a group analysis showed a persisting, significant advantage for retrieval practice over repetition. Concerning spacing, performance was significantly higher in the spacing condition compared to the massed condition on the retention test, with a similar marginal trend at follow-up.

The spacing manipulation also resulted in superior performance for spaced over massed items during training in the retrieval practice condition. To revisit, we initially considered the possibility that both spaced and massed items would benefit from each long (multi-day) lag between training sessions and thus that the spacing effect might diminish across training sessions. Although the spacing effect in the retrieval practice condition was present and did not diminish across the training sessions, it is possible that the between-session lag factor may have ultimately contributed to the marginal spacing effect at follow-up. Because the current study is one of only two studies (Rawson & Dunlosky, 2013, Experiment 3) to evaluate the effects of within-session spacing when items are trained in several sessions, much additional research is required to investigate whether prescriptions derived from spacing studies involving single-session manipulations are applicable to knowledge acquisition or cognitive rehabilitation goals that require multiple sessions.

For various reasons, we are cautious in deriving recommendations from the current study results alone for how to improve the naming of errorful vocabulary in PWA. Obvious reasons for caution include the small, non-randomly selected sample of participants, and the individual differences observed in response to the training factor at retention test. With that said, the current findings converge with those in our prior work (Middleton et al., 2015; 2016). Taken together, they arguably constitute significant progress towards an evidence base for eventual clinical recommendation. In the present study, retrieval practice lead to statistically significant and robust advantages over repetition for three of the four PWA at retention test. Furthermore, a persisting retrieval practice effect was confirmed at the group level one-month following training. Each of two group studies of PWA (Middleton et al. 2015; 2016) found superior benefits from retrieval practice compared to repetition at one-day and one-week retention intervals. Thus, aggregating across the studies, we have found consistent retrieval practice effects over a range of schedules of treatment (i.e., one or multiple sessions), dosages (i.e., one to several trials per item), and retention intervals (i.e., one day, one week, one month). Likewise, in the current work, spacing outperformed massing, significantly so at the retention test and marginally at follow-up. Although the spacing effect was weaker at follow-up, there certainly was no hint of an advantage for the massed condition. Along with Middleton et al. (2016), who reported a spacing effect at one-day and one-week retention intervals in group analyses, we have confirmed spacing effects in designs involving one or multiple training sessions with a few to several training trials per item. Overall, the consistency in results across studies points to retrieval practice and spacing as promising principles of learning for informing practice given similar treatment goals and patient characteristics.

The preponderance of research on retrieval practice and spacing effects has focused on the application of these learning factors for promoting the acquisition and retention of new

knowledge representations, memories, or skills in relatively young, neurotypical populations. However, as discussed in the Introduction, various domains of neurorehabilitation include attempts to explore the applicability of these learning factors in teaching neuropsychological patients new functional information or skills. Investigations of their application to language treatment are more scant, but the current work joins a small number of studies that collectively suggest these learning factors are relevant for language therapy (Friedmann et al., 2017; Middleton et al., 2015, 2016). In the current study, the targeted dysfunctional process was lexical access deficit (i.e., difficulty reliably retrieving known words for production from the mental lexicon). In addition to aspects of the sample's neuropsychological characterization, we contend that the learning effects reflected improvements in lexical access. This is because in each individual, training focused on items for which the participant could choose the name for the entity from among several options, and the participant deemed the entities and their names recognizable. This aspect of the design was important for establishing the relevance of retrieval practice and spacing for altering the persistent accessibility to existing lexical representations, which demonstrates the applicability of these learning factors to language therapy. Furthermore, this observation is of theoretical significance because it suggests that just as these factors are important for the acquisition of new knowledge, they are important for exerting persistent changes in lexical accessibility, pointing to the domain generality of these learning principles.

At present, much future work is needed to elucidate the learning mechanisms that underlie the effects of retrieval practice and spacing on lexical access. Schuchard and Middleton (2018a; 2018b) made progress towards explicating retrieval practice effects in lexical access by testing a theoretical extension of a prominent computational model of use-dependent learning in lexical access (Oppenheim, Dell, & Schwartz, 2010). According to Oppenheim et al. (2010), each act of naming is a learning event wherein the retrieval connections (i.e., weights) between semantics and the word that is retrieved are strengthened. In that framework, learning is use-dependent because the process of mapping between representations provides important information to the learning algorithm that determines the manner and degree of weight change. Schuchard and Middleton (2018a; 2018b) proposed that use-dependent learning impacts each of the two major stages of lexical access, which involve mapping from semantics to words (Stage-1), followed by mapping from the selected word to its phonological constituents (Stage-2). Schuchard and Middleton hypothesized that retrieval practice can be more efficacious than repetition training because production of the target word during retrieval practice requires and thus can strengthen Stage-1 mapping. In support of this hypothesis, in a group study of PWA, retrieval practice was more effective than repetition training for improving performance on items that elicited naming errors attributable to Stage-1 (e.g., semantic errors) and not Stage-2 (e.g., phonological errors) mapping failure (Schuchard & Middleton, 2018a). Likewise, retrieval practice was more effective than repetition training for improving naming in a PWA with Stage-1 mapping impairment, a pattern that did not hold for a second PWA whose naming impairment was localizable to Stage-2 (Schuchard & Middleton, 2018b). In light of these results, retrieval practice and spacing effects in lexical access may be underpinned by use-dependent learning mechanisms, which may be implicit. A full theoretical account of such effects may also require explicit memory components, such as episodic memory for experiences during

training. Adjudicating between these theoretical possibilities is an important topic for future research.

Although the results from the present study are promising for the ultimate clinical applicability of retrieval practice and spacing effects to naming rehabilitation, a great deal of additional research is needed to explore possible application to other types of language treatment, as well as individual differences in response to these factors. The retrieval practice literature suggests that successful retrieval of target representations is important for learning and may confer more beneficial change than failed retrieval (e.g., Hinze & Wiley, 2011; Kang, Pashler, Cepeda, Rohrer, Carpenter, & Mozer, 2011; Kornell, Bjork, & Garcia, 2011; Pashler, Cepeda, Wixted, & Rohrer, 2005). In prominent theoretical models of spacing effects (e.g., Benjamin & Tullis, 2010; Hintzman, 2004, 2010; Thios& D'Agostino, 1976), learning is predicated upon successful linking between the target information presented on the current trial and previously encoded representations of that information. Without such linking, the power of spacing over massing may be undermined. In cases of severe naming impairment, a person with aphasia may require items receive interleaved repetition and retrieval practice (to enhance rates of successful retrieval practice) and shorter intervals of spacing (to scaffold linking to representations in long-term memory) to derive maximal benefit from retrieval practice and spacing.

In the present work, we did observe a statistically significant interaction between the participant factor and type of training at retention test because P3 showed a numerical advantage for repetition training over retrieval practice, in contrast to the strong advantage for retrieval practice over repetition for the other three participants. Two details of P3's performance may illuminate why that participant showed a pattern different from the group. First, P3 was the most impaired of the sample, performing on average 27% and 12% lower on the retention and follow-up tests, respectively, compared to the other participants. During training in the retrieval practice condition, P3's rate of successful retrieval practice was on average 25% lower than the other three participants, lending support to the possibility that lower rates of successful retrieval practice can undermine its advantage over repetition training. Thus, this participant may have benefitted from more repetition prior to attempting retrieval practice, or perhaps could have benefitted from shorter lags between retrieval practice trials.

When considering the more general clinical applicability of retrieval practice and spacing effects to language treatment, it is important to consider the goals of treatment. Aphasia interventions typically take one of two general approaches, compensatory or 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, the current work exemplifies a *restorative* approach, in which the goal is to promote access to affected vocabulary through (re-)learning techniques. Here, the focus is on item-specific benefits, with more tentative a priori expectations regarding generalization to untreated items than in compensatory approaches. 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. However, it is reasonable to expect that the learning factors studied here may have

implications for generalization. For example, as described above, evidence suggests retrieval practice is more effective at strengthening Stage-1 mapping in the course of lexical access than repetition training (Schuchard & Middleton, 2018a, 2018b). In cases of aphasia where the naming impairment is at least in part attributable to degradation or dysregulation of semantic representations (e.g., Jefferies & Lambon Ralph, 2006), assuming a distributed semantic representational system and backpropagation of learning, the targeted strengthening of weights between semantics and words from retrieval practice for a core set of vocabulary could promote refinement of semantic distinctions more generally. Such semantic refinement may promote enhanced generalization over repetition, an important topic for future research. Lastly, when anomia is severe, the value of expanding a PWA's functional vocabulary—particularly relating to personally relevant topics--should not be underestimated.

5. Conclusion

Devising an effective treatment plan for a person with aphasia requires adoption and implementation of treatment methods that maximize and sustain improvement in the aspects of language function that are impaired. In speech-language pathology as in other domains of neurorehabilitation, treatment research strives to provide a base of evidence for the efficacy of existing and new treatments to inform clinical decision-making (Robey & Schultz, 1998). We subscribe to the view that an important phase of efficacy research involves characterization of how the damaged system responds to different kinds of learning experiences (Baddeley, 1993). Because different learning experiences are variably provided by common naming treatments, studies such as the current work are an important building block to informing how to design interventions that maximize persistent improvement.

Acknowledgements

References

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(3), 235–244

Bahrick HP (1979). Maintenance of knowledge: Questions about memory we forgot to ask. Journal of Experimental Psychology: General, 108, 296–308.

Barcroft J (2007). Effects of opportunities for word retrieval during second language vocabulary learning. Language Learning, 57(1), 35–56.

Benjamin AS, & Tullis J (2010). What makes distributed practice effective? Cognitive Psychology, 61, 228–247. [PubMed: 20580350]

Butler AC, & Roediger HL, (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Memory and Cognition, 36(3), 604–616. [PubMed: 18491500]

Carpenter SK, Cepeda NJ, Rohrer D, Kang SHK, & Pashler H (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. Educational Psychology Review, 24(3), 369–378.

Carpenter SK, & DeLosh EL (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 and Cognition, 20(6), 633–642. [PubMed: 1435266]

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

Cepeda NJ, Vul E, Rohrer D, Wixted JT, & Pashler H (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. Psychological Science, 19(11), 1095–1102. [PubMed: 19076480]

Chen Q, Middleton EL, & Mirman D (in press). Words fail: Lesion-symptom mapping of errors of omission in post-stroke aphasia. Journal of Neuropsychology.

Clare L, & Jones RSP (2008). Errorless learning in the rehabilitation of memory impairment: A critical review. Neuropsychology Review, 18, 1–23. [PubMed: 18247118]

Coyne JH, Borg JM, DeLuca J, Glass L, & Sumowski JF (2015). Retrieval practice as an effective memory strategy in children and adolescents with traumatic brain injury. Archives of Physical Medicine and Rehabilitation, 96(4), 742–745. [PubMed: 25312580]

Cull WL (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. Applied Cognitive Psychology, 14(3), 215–235.

Delaney PF, Verkoeijen PPJL, & Spirgel A (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. Psychology of Learning and Motivation, 53, 63–147.

Dignam JK, Rodriquez AD, & Copland D (2016). Evidence for intensive aphasia therapy: Consideration of theories from neuroscience and cognitive psychology. PM&R, 8(3), 254–267. [PubMed: 26107539]

Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, & Willingham DT (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. Psychological Science in the Public Interest, 13(1), 4–58.

Fillingham JK, Hodgson C, Sage K, & Lambon Ralph MA (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. Neuropsychological Rehabilitation, 13(3), 337–363. [PubMed: 21854317]

Fillingham JK, Sage K, & Lambon Ralph MA (2005a). Treatment of anomia using errorless versus errorful learning: Are frontal executive skills and feedback important? International Journal of Language and Communication Disorders, 40(4), 505–523. [PubMed: 16195203]

Fillingham JK, Sage K, & Lambon Ralph MA (2005b). Further explorations and an overview of errorless and errorful therapy for aphasic word-finding difficulties: The number of naming attempts during therapy affects outcome. Aphasiology, 19(7), 597–614.

Fillingham JK, Sage K, & Lambon Ralph MA (2006). The treatment of anomia using errorless learning. Neuropsychological Rehabilitation, 16(2), 129–154. [PubMed: 16565031]

Friedman RB, Sullivan KL, Snider SF, Luta G, & Jones KT (2017). Leveraging the test effect to improve maintenance of the gains achieved through cognitive rehabilitation. Neuropsychology, 31(2), 220–228. [PubMed: 27732041]

Goldrick M, & Rapp B (2007). Lexical and post-lexical phonological representations in spoken production. Cognition, 102, 219–260. [PubMed: 16483561]

Goossens NAMC, Camp G, Verkoeijen PPJL, Tabbers HK, Bouwmeester S, & Zwaan RA (2016). Distributed practice and retrieval practice in primary school vocabulary learning: A multi-classroom study. Applied Cognitive Psychology, 30, 700–712.

Goverover Y, Arango-Lasprilla JC, Hillary FG, Chiaravalloti N, & DeLuca J (2009). Application of the spacing effect to improve learning and memory for functional tasks in traumatic brain injury: A pilot study. American Journal of Occupational Therapy, 63, 543–548. [PubMed: 19785253]

Goverover Y, Basso M, Wood H, Chiaravalloti N, & DeLuca J (2011). Examining the benefits of combining two learning strategies on recall of functional information in persons with multiple sclerosis. Multiple Sclerosis Journal, 17(12), 1488–1497. [PubMed: 21586486]

Goverover Y, Hillary FG, Chiaravalloti N, Arango-Lasprilla JC, & DeLuca J (2009). A functional application of the spacing effect to improve learning and memory in persons with multiple sclerosis. Journal of Clinical and Experimental Neuropsychology, 31(5), 513–522. [PubMed: 18720184]

Hintzman DL (2004). Judgment of frequency vs. recognition confidence: Repetition and recursive reminding. Memory and Cognition, 32, 336–350. [PubMed: 15190724]

Hintzman DL (2010). How does repetition affect memory? Evidence from judgments of recency. Memory and Cognition, 38, 102–115. [PubMed: 19966243]

Hinze SR, & Wiley J (2011). Testing the limits of testing effects using completion tests. Memory, 19, 290–304. [PubMed: 21500089]

Howard D, & Patterson K (1992). Pyramids and Palm Trees: A test of semantic access from pictures and words. Bury St. Edmunds, Suffolk: Thames Valley Test Company.

Jefferies E, & Lambon Ralph MA (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case–series comparison. Brain, 129, 2132–2147. [PubMed: 16815878]

Kang SHK, Pashler H, Cepeda NJ, Rohrer D, Carpenter SK, & Mozer MC (2011). Does incorrect guessing impair fact learning? Journal of Educational Psychology, 103, 48–59.

Kertesz A (1982). The western aphasia battery. San Antonio, TX: The Psychological Corporation (Harcourt Brace Jovanovich) Publishers.

Kornell N (2009). Optimizing learning using flashcards: Spacing is more effective than cramming. Applied Cognitive Psychology, 23, 1297–1317.

Kornell N, Bjork RA, & Garcia MA (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. Journal of Memory and Language, 65, 85–97.

Kornell N, & Vaughn KE (2016). How retrieval attempts affect learning: A review and synthesis. Psychology of Learning and Motivation, 65, 183–215.

Kromann CB, Jensen ML, & Ringsted C (2009). The effects of testing on skills learning. Medical Education, 43, 21–27. [PubMed: 19140995]

Lecours AR, & Lhermitte F (1969). Phonemic paraphasias: Linguistic structures and tentative hypotheses. Cortex, 5, 193–228. [PubMed: 5391877]

Lyle KB & Crawford NA (2011). Retrieving essential material at the end of lectures improves performances on statistics exams. Teaching of Psychology, 38(2), 94–97.

MacDonald SWS, Nyberg L, & Bäckman L (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. Trends in Neuroscience, 29(8), 474–480.

Martin N, Schwartz MF, Kohen FP (2006).Assessment of the ability to process semantic and phonological aspects of words in aphasia: A multi-measurement approach. Aphasiology, 20(2–4), 154–166.

Martin RC, Shelton JR, Yaffee LS (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. Journal of Memory and Language, 33(1), 83–111.

McKissock S, & Ward J (2007). Do errors matter? Errorless and errorful learning in anomic picture naming. Neuropsychological Rehabilitation, 17(3), 355–373. [PubMed: 17474061]

Metcalf J, Kornell N, & Son LK (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. European Journal of Cognitive Psychology, 19(4/5), 743–768. [PubMed: 19148303]

Middleton EL (2019, 2 26). Retrieval practice and spacing effects in multi-session treatment of naming impairment in aphasia. Retrieved from osf.io/cwq75

Middleton EL, & Schwartz MF (2012). Errorless learning in cognitive rehabilitation: A critical review. Neuropsychological Rehabilitation, 22(2), 138–168. [PubMed: 22247957]

Middleton EL, Schwartz MF, Rawson KA, & 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(4), 1253–1261.

Middleton EL, Schwartz MF, Rawson KA, Traut H, & Verkuilen J (2016). Towards a theory of learning for naming rehabilitation: Retrieval practice and spacing effects. Journal of Speech, Language, and Hearing Research, 59(5), 1111–1122.

Mirman D, & Brit A (2014). What we talk about when we talk about access deficits. Philosophical Transactions of the Royal Society B, 369, 20120388.

Mirman D, Strauss TJ, Brecher A, Walker GM, Sobel P, Dell GS, & Schwartz MF (2010). A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive functions. Cognitive Neuropsychology, 27(6), 495–504. [PubMed: 21714742]

Morris PE, & Fritz CO (2002). The improved name game: Better use of expanding retrieval practice. Memory, 10(4), 259–266. [PubMed: 12097210]

Nozari N, Kittredge AK, Dell GS, Schwartz MF (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. Journal of Memory and Language, 63(4), 541–559. [PubMed: 21076661]

Oppenheim GM, Dell GS, & Schwartz MF (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. Cognition, 114(2), 227–252. [PubMed: 19854436]

Oren S, Willerton C, & Small J (2014). Effects of spaced retrieval training on semantic memory in Alzheimer's disease: A systematic review. Journal of Speech, Language, and Hearing Research, 57, 247–270.

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(7), 811–827. [PubMed: 23914001]

Pashler H, Cepeda NJ, Wixted JT, & Rohrer D (2005). When does feedback facilitate learning of words? Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 3–8.

Pastötter B, Weber J, & Bäuml KT (2013). Using testing to improve learning after severe traumatic brain injury. Neuropsychology, 27(2), 280–285. [PubMed: 23527656]

Pyc MA, & Rawson KA (2007). Examining the efficiency of schedules of distributed retrieval practice. Memory and Cognition, 35, 1917–1927. [PubMed: 18265608]

Pyc MA, & Rawson KA (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? Journal of Memory and Language, 60(4), 437–447.

Pyc MA, & Rawson KA (2012a). Are judgments of learning made after correct responses during retrieval practice sensitive to lag and criterion level effects? Memory and Cognition, 40, 976–988. [PubMed: 22399224]

Pyc MA, & Rawson KA (2012b). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38, 737–746.

Rawson KA, & Dunlosky J (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? Journal of Experimental Psychology: General, 140(3), 283–302. [PubMed: 21707204]

Rawson KA, & Dunlosky J (2013). Relearning attenuates the benefits and costs of spacing. Journal of Experimental Psychology: General, 142(4), 1113–1129. [PubMed: 23088488]

Rawson KA, Dunlosky J, & Sciartelli SM (2013). The power of successive relearning: Improving performance on course exams and long-term retention. Educational Psychological Review, 25, 523–548.

Rawson KA, Vaughn KE, Walsh M, & Dunlosky J (2018). Investigating and explaining the effects of successive relearning on long-term retention. Journal of Experimental Psychology: Applied, 24, 57–71. [PubMed: 29431462]

Roach A, Schwartz MF, Martin N, Grewal RS, and Brecher A (1996). The Philadelphia Naming Test: scoring and rationale. Clinical Aphasiology, 24, 121–133.

Robey RR, & Schultz MC (1998). A model for conducting clinical-outcome research: An adaptation of the standard protocol for use in aphasiology. Aphasiology, 12(9), 787–810.

Roediger HL, & Butler AC (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences, 15, 20–27. [PubMed: 20951630]

Roediger HL, & Karpicke JD (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17(3), 249–255. [PubMed: 16507066]

Roediger HL, Putnam AL, & Smith MA (2011). Ten benefits of testing and their applications to educational practice. Psychology of Learning and Motivation, 44, 1–36.

Rowland CA (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. Psychological Bulletin, 140(6), 1432–1463. [PubMed: 25150680]

Schwartz MF, Kimberg DY, Walker GM, Faseyitan OK, Brecher AR, Dell GS, & Coslett HB (2009). Anterior temporal involvement in semantic word retrieval: Voxel-based lesion symptom mapping evidence from aphasia. Brain, 132, 3411–3427. [PubMed: 19942676]

Schuchard J, & Middleton EL (2018a). The roles of retrieval practice versus errorless learning in strengthening lexical access in aphasia. Journal of Speech, Language, and Hearing Research, 61, 1700–1717.

Schuchard J, & Middleton EL (2018b). Word repetition and retrieval practice effects in aphasia: Evidence for use-dependent learning in lexical access. Cognitive Neuropsychology, 35(5–6), 271–287. [PubMed: 29724159]

StataCorp (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Sumowski JF, Chiaravalloti N, & DeLuca J (2010). Retrieval practice improves memory in multiple sclerosis: Clinical application of the testing effect. Neuropsychology, 24(2), 267–272. [PubMed: 20230121]

Sumowski JF, Coyne JH, Cohen A, & Deluca J (2014). Retrieval practice improves memory in survivors of severe traumatic brain injury. Archives of Physical Medicine and Rehabilitation, 95, 397–400. [PubMed: 24231401]

Sumowski JF, Leavitt VM, Cohen A, Paxton J, Chiaravalloti ND, Deluca J (2013). Retrieval practice is a robust memory aid for memory-impaired patients with MS. Multiple Sclerosis Journal, 19, 1943–1946. [PubMed: 23674564]

Sumowski JF, Wood HG, Chiaravalloti N, Wylie GR, Lengenfelder J, DeLuca J (2010). Retrieval practice: A simple strategy for improving memory after traumatic brain injury. Journal of the International Neuropsychological Society, 16, 1147–1150. [PubMed: 20946709]

Thios SJ, & D'Agostino PR (1976). Effects of repetition as a function of study-phase retrieval. Journal of Verbal Learning and Verbal Behavior, 15(5), 529–536.

Toppino TC, & Gerbier E (2014). About practice: Repetition, spacing, and abstraction. Psychology of Learning and Motivation, 60, 113–189.

Vaughn KE, Dunlosky J, & Rawson KA (2016). Effects of successive relearning on recall: Does relearning override effects of initial learning criterion? Memory & Cognition, 44, 897–909. [PubMed: 27027887]

Vaughn KE, & Rawson KA (2011). Diagnosing criterion level effects on memory: What aspects of memory are enhanced by repeated retrieval? Psychological Science, 22, 1127–1131. [PubMed: 21813798]

Whyte J, Gordon W, Rothi LJG (2009). A phased developmental approach to neurorehabilitation research: The science of knowledge building. Archives of Physical Medicine and Rehabilitation, 90(11 Suppl 1):S3–10. [PubMed: 19892072]
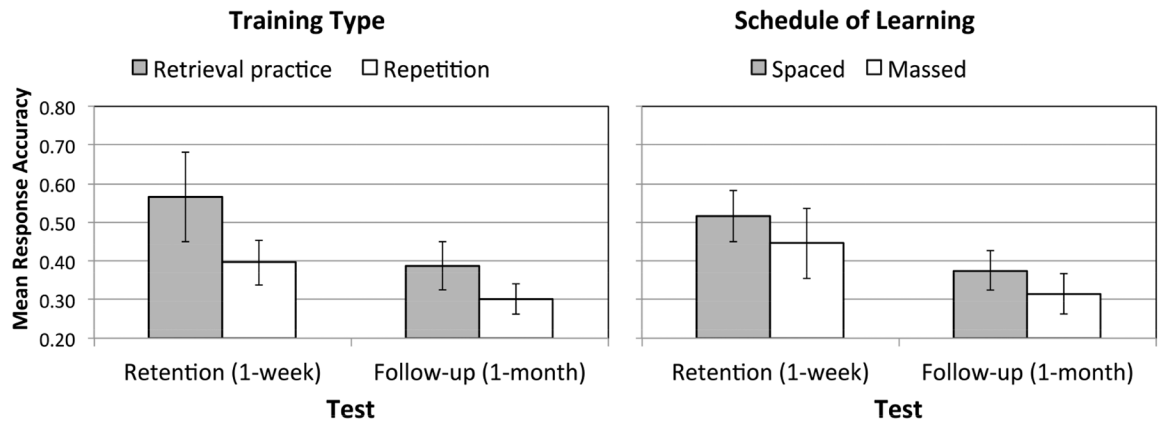
Figure 1.
Mean response accuracy (with standard errors) across participants as a function of training condition (left panel) and spacing condition (right panel) at retention test and follow-up.

Schedule of Learning

☐ Spaced ☐ Massed
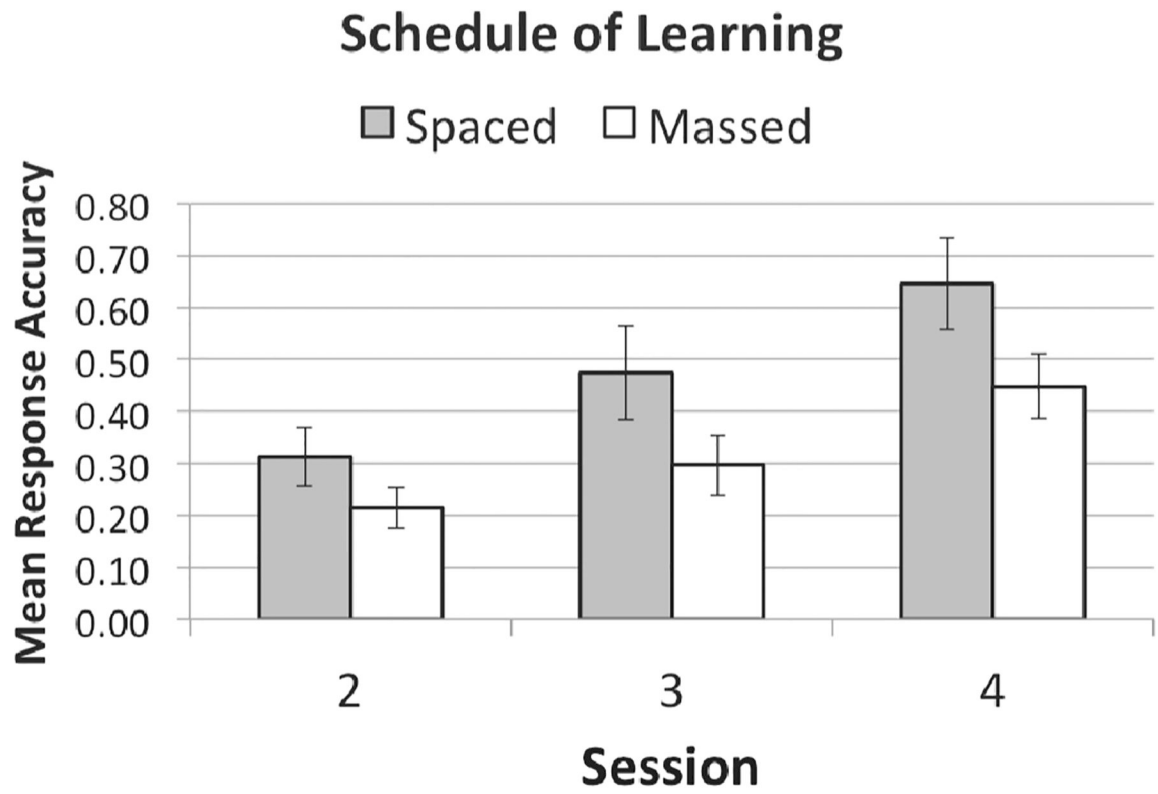
Figure 2.
Mean response accuracy (with standard errors) across participants for the first trial per session in the retrieval practice condition as a function of spacing in sessions 2–4.

Table 1.

Participant Battery Test Scores and Demographics.

Demographic	P1	P2	P3	P4	
Age at current testing	51	64	50	77	
Years of education	13	13	13	12	
Months postonset	44	41	121	79	
Test					**Average**
AQ	81.1	90.4	81.5	90.2	85.8
Subtype	anomic	anomic	anomic	anomic	
Picture naming					
Correct	74	88	74	82	80
Semantic error	4	1	6	4	4
Omission	9	2	9	9	7
Phonological error	9	9	7	3	7
Other	5	1	4	2	3
Word repetition	93	89	97	95	94
Nonword repetition	43	37	65	73	55
STM span	2.6	3	3.2	4.4	3.3
Nonverbal comprehension	92	85	83	90	88
Word comprehension	96	89	94	94	93
Synonymy triplets	77	73	80	83	78
Nouns	73	87	87	80	82
Verbs	80	60	73	87	75

Note. *AQ*=Western Aphasia Battery Aphasia Quotient score out of 100 (Kertesz, 1982). *Subtype* = aphasia subtype as determined by the Western Aphasia Battery. *Picture naming* = Percentage of correct responses and various error types on oral picture naming (*Philadelphia Naming Test*; Roach et al., 1996). *Word repetition* = A test of immediate word repetition, in percentages (Mirman et al., 2010). *Nonword repetition* = A test of immediate repetition of nonwords, in percentages (Mirman et al., 2010). STM span = test of verbal short term memory where participants repeat lists of words of increasing lengths; max score = 5 (Martin et al., 1994). *Nonverbal comprehension* = A picture-picture association test for nonverbal semantic comprehension, in percentages (Howard & Patterson, 1992). *Word comprehension* = Spoken word to picture matching test of comprehension in percentages (Mirman et al., 2010). Synonymy triplets = Test of verbal semantic relatedness involving choosing two of three words most similar in meaning, in percentages, with noun and verb subscores (Martin et al., 2006).

Table 2

Mean (Standard Deviation) per Variable Across Participants' Personalized Item Sets as a Function of Condition

Training	Spacing	Response Accuracy[a] *M (SD)*	# Syllables *M (SD)*	# Phonemes *M (SD)*	Entity Recognition[b] *M (SD)*	Name Recognition[b] *M (SD)*
Repetition	Massed	.02 (0.02)	3.70 (0.02)	9.81 (0.12)	1.04 (0.08)	1.13 (0.17)
	Spaced	.01 (0.02)	3.69 (0.09)	9.78 (0.09)	1.03 (0.06)	1.13 (0.17)
Retrieval Practice	Massed	.01 (0.02)	3.73 (0.07)	9.76 (0.13)	1.05 (0.10)	1.11 (0.15)
	Spaced	.01 (0.01)	3.64 (0.05)	9.76 (0.14)	1.04 (0.07)	1.12 (0.16)

Note.

[a]Mean response accuracy from the item selection phase.

[b]Scores can range from 1–3, with 1 = yes to recognition prompt on both trials for an item for a participant, and 3 = no to recognition prompt on both trials. See section 2.4.1 for details.

Table 3

Mean Response Accuracy by Condition and Retrieval Practice and Spacing Effects Per Participant at Retention Test and Follow-Up

			Retention Test	Follow-up Test
Participant	**Training**	**Spacing**	**_M_**	**_M_**
P1	Retrieval Practice	Spaced	.81	.60
		Massed	.75	.52
	Repetition	Spaced	.52	.44
		Massed	.56	.40
	Retrieval Practice Effect		.24	.15
	Spacing Effect		.01	.06
P2	Retrieval Practice	Spaced	.63	.44
		Massed	.56	.33
	Repetition	Spaced	.46	.27
		Massed	.42	.31
	Retrieval Practice Effect		.16	.09
	Spacing Effect		.05	.03
P3	Retrieval Practice	Spaced	.25	.31
		Massed	.23	.23
	Repetition	Spaced	.44	.27
		Massed	.21	.21
	Retrieval Practice Effect		−.08	.03
	Spacing Effect		.13	.07
P4	Retrieval Practice	Spaced	.73	.35
		Massed	.56	.31
	Repetition	Spaced	.29	.31
		Massed	.27	.21
	Retrieval Practice Effect		.36	.07
	Spacing Effect		.09	.07

Note. *Retrieval practice effect* corresponds to the magnitude of the advantage for retrieval practice compared to repetition collapsed over the spacing factor (negative value indicates higher performance in the repetition condition). *Spacing effect* corresponds to the magnitude of the advantage for the spaced over massed schedule collapsed over the training factor.

Table 4

Group Model Results: Retrieval Practice and Spacing Effects at Retention Test

Fixed Effects	Coef.	SE	Z	p
Intercept	1.23	0.29		
Retrieval Practice Effect (P1)[a]				
Repetition[b]	−1.19	0.36	−3.30	.001
Spacing Effect				
Spaced[c]	0.37	0.18	2.08	.038
Participant Effect				
P2[d]	−0.96	0.36	−2.70	.007
P3[d]	−2.71	0.41	−6.64	<.001
P4[d]	−0.70	0.36	−1.93	.053
Interaction of Training and Participant				
Repetition[b] × P2[d]	0.48	0.49	0.99	.32
Repetition[b] × P3[d]	1.60	0.51	3.15	.002
Repetition[b] × P4[d]	−0.57	0.50	−1.13	.26
Random Effect	s^2			
Items	.62			

Note. Excluding the intercepts, *Coef.* = 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, two-tailed; s^2 = Random effect variance.

[a] Because the model includes an interaction term for participants and training, estimate corresponds to the training effect for the reference level of the participants factor (P1).

[b] Reference is retrieval practice condition.

[c] Reference is massed condition.

[d] Reference is participant 1 (P1).

Table 5

Participant-Specific Model Results: Retrieval Practice and Spacing Effects at Retention Test

Participant 1				
Fixed Effects	*Coef.*	*SE*	*Z*	*p*
Intercept	1.25	0.30		
Retrieval Practice Effect				
Repetition[a]	−1.11	0.32	−3.44	.001
Spacing Effect				
Spaced[b]	0.05	0.32	0.16	.875
Participant 2				
Fixed Effects	*Coef.*	*SE*	*Z*	*p*
Intercept	0.27	0.25		
Retrieval Practice Effect				
Repetition[a]	−0.63	0.29	−2.15	.031
Spacing Effect				
Spaced[b]	0.21	0.29	0.73	.466
Participant 3				
Fixed Effects	*Coef.*	*SE*	*Z*	*p*
Intercept	−1.50	0.33		
Retrieval Practice Effect				
Repetition[a]	0.42	0.33	1.28	.199
Spacing Effect				
Spaced[b]	0.63	0.33	1.91	.056
Participant 4				
Fixed Effects	*Coef.*	*SE*	*Z*	*p*
Intercept	0.39	0.26		
Retrieval Practice Effect				
Repetition[a]	−1.56	0.32	−4.93	<.001
Spacing Effect				
Spaced[b]	0.44	0.31	1.39	.165

Note. Excluding the intercepts, *Coef.* = 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, two-tailed.

[a] Reference is retrieval practice condition.

[b] Reference is massed condition.

Table 6

Group Model Results: Retrieval Practice and Spacing Effects at Follow-up Test

Fixed Effects	Coef.	SE	Z	p
Intercept	0.05	0.20		
Retrieval Practice Effect				
Repetition[a]	−0.42	0.17	−2.54	.011
Spacing Effect				
Spaced[b]	0.30	0.17	1.76	.078
Participant Effect				
P2[c]	−0.72	0.24	−2.96	.003
P3[c]	−1.18	0.26	−4.56	<.001
P4[c]	−0.92	0.24	−3.85	<.001
Random Effect	s^2			
Items	.49			

Note. Excluding the intercepts, *Coef.* = 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, two-tailed; s^2 = Random effect variance.

[a] Reference is retrieval practice condition.

[b] Reference is massed condition.

[c] Reference is participant 1 (P1).