Test-Enhanced Learning Versus Errorless Learning in Aphasia Rehabilitation: Testing Competing Psychological Principles

Erica L. Middleton,
Research Department, Moss Rehabilitation Research Institute

Myrna F. Schwartz,
Research Department, Moss Rehabilitation Research Institute

Katherine A. Rawson, and
Psychology Department, Kent State University

Kelly Garvey
Research Department, Moss Rehabilitation Research Institute.

Department of Speech, Language and Hearing Sciences, Boston University.

Abstract

Because individuals with acquired language disorders are frequently unable to reliably access the names of common everyday objects (i.e., naming impairment), rehabilitation efforts often focus on improving naming. The present study compared two rehabilitation strategies for naming impairment, reflecting contradictory prescriptions derived from different theoretical principles. The prescription derived from psychological research on test-enhanced learning advocates providing patients opportunities to retrieve target names from long-term memory (i.e., *retrieval practice*) in the course of treatment. In contrast, the errorless learning approach derived from cognitive rehabilitation research eschews retrieval practice in favor of methods that minimize naming errors, and thus the potential for error learning, in the course of treatment. The present study directly compared these approaches and showed that, despite superior (and errorless) performance during errorless treatment, treatment that prioritized retrieval practice produced greater retention 1-day and 1-week following treatment. These findings have implications for clinical practice, as well as theoretical accounts of lexical access and test-enhanced learning.

Keywords

naming; aphasia; test-enhanced learning; errorless learning; lexical access

Fast and accurate word retrieval is essential for efficient speech. However, naming impairment (difficulty producing names for everyday objects) is ubiquitous in acquired language disorders, such as in aphasia from stroke and traumatic brain injury. What methods

should a clinician use to rehabilitate naming impairment? Two literatures are relevant: cognitive rehabilitation research on errorless learning, and psychological research on test-enhanced learning. Unfortunately, these literatures have minimal contact with one another and advocate contradictory practices based on different theoretical foundations. The current study directly compares the prescriptions derived from these two literatures for treating spoken naming impairment in aphasia. This research provides an important first step for adjudicating between the two research traditions as applied to rehabilitation of naming impairments in acquired language disorders. More broadly, the current research provides foundational empirical and methodological groundwork to bridge the two disparate literatures.

Prescriptions from Cognitive Rehabilitation Research

Cognitive rehabilitation research shows mounting interest in *errorless learning* treatments, motivated by the hypothesis that errors committed during treatment (e.g., misnaming an object presented for naming; sequencing errors in a manual task) may be learned and deleteriously impact treatment efficacy. To evaluate this hypothesis, studies compare 'errorless' treatments—where therapists scaffold errorless performance by providing guidance or modeling correct responses on all trials to avoid errors—to 'errorful' treatments. Because the interest in such studies concerns how errors impact efficacy, errorful procedures are often designed to promote errors, e.g., by having participants complete tasks with self-generated solutions without prior or sufficient familiarization with the treatment targets for mastery (for review, Middleton & Schwartz, 2012).

To date, errorless treatments have been most comprehensively investigated in memory-impaired populations, with errorless methods generally showing superior benefits over errorful methods, particularly for severe explicit memory impairment (Clare & Jones, 2008; Middleton & Schwartz, 2012). Interest in errorless learning treatments is growing, with numerous recent investigations in other populations (e.g., schizophrenia; Leshner, Tom, & Kern, 2013; executive dysfunction, Bertens, Fasotti, Boelen, & Kessels, 2013; mild cognitive impairment, Lubinsky, Rich, & Anderson, 2009; dementia including Alzheimer's; Li & Liu, 2012).

Outside the amnesia literature, the errorless learning approach has been most extensively studied in aphasia, with emphasis on treating naming impairment (Fillingham, Hodgson, Sage, & Lambon Ralph, 2003; Middleton & Schwartz, 2012). The impetus for this work is the possibility that errorful *speech* is self-reinforcing due to Hebbian learning (e.g., Fillingham et al., 2003). Hebbian binding of a stimulus and an elicited response may increase "the likelihood of making the same response in the future, whether correct or incorrect" (Fillingham et al., 2003, p. 341), given the same stimulus. Hence, the standard errorless naming treatment, in which the experimenter provides the target name with the picture and the patient repeats it (typically without error), aims to strengthen only the association between the stimulus and the correct response, based on Hebbian learning principles.

Several studies have evaluated application of errorless learning principles to naming impairment in aphasia by comparing treatment outcomes for items trained with errorless versus errorful methods. In a series of single-subject controlled comparisons (Fillingham, Sage, & Lambon Ralph, 2005a; 2005b; 2006), an errorful treatment (naming with a phonological cue) and errorless treatment both improved naming for most participants. However, group-level analyses were not reported to compare their relative efficacy. Available group studies have shown either a trend (Conroy, Sage, & Lambon Ralph, 2009) or a reliable advantage (McKissock & Ward, 2007) for errorless over errorful methods. However, these studies were designed to produce marked differences in error rates between the errorless and errorful conditions, which motivated design features that deviate from clinical practice and may have limited efficacy of the errorful method. For example, in McKissock and Ward's errorful condition, no prior familiarization with correct names was provided, and responding on each trial was strongly encouraged, forcing participants to produce errors when otherwise they may have refrained from responding. Below, we consider training methods that (like errorful treatment) encourage self-generated responding but (unlike errorful treatment) emphasize the retrieval of correct information from long-term memory over production of errors. To foreshadow, under these conditions powerful learning is the norm.

Prescriptions from Test-Enhanced Learning Research

Substantial research on *test-enhanced learning* establishes that tests can bolster learning and retention of content as varied as foreign vocabulary, word lists, facts, picture-word associations, and text (Rawson & Dunlosky, 2011). The typical paradigm for demonstrating test-enhanced learning begins with initial study (e.g., participant studies Inuit-English word pair *angyak—boat*). This is followed by either *restudy* of the same information (*angyak—boat*) or a *test* (*angyak--_____*) where the participant is given a comparable amount of time as restudy to attempt to retrieve the target (*boat*) from long-term memory. A "testing effect" refers to the typical advantage for the test condition over restudy on measures of retention. Testing effects typically become more pronounced at longer retention intervals and when feedback (e.g., provision of the target for further study) is presented after the retrieval attempt. However, the testing effect is robust even in the absence of feedback (for reviews, see Rawson & Dunlosky, 2011; Roediger & Karpicke, 2006).

A good deal is known about how testing benefits are maximized. First, retrieving correct information from long-term memory (*retrieval practice*) significantly enhances its retention (termed *direct benefits of testing*; Roediger & Karpicke, 2006). Second, retrieval of correct information confers more learning when it requires more effort. For example, in Pyc and Rawson (2009) greater effort (as reflected in increased time to retrieve a target) predicted better long-term retention. Though conditions that make retrieval effortful (e.g., extending the interval between repeated practice trials for an item; Carpenter & DeLosh, 2005) typically increase error rates during learning when compared to restudy or conditions that make retrieval practice easy (e.g., shorter intervals between trials), effortful retrieval practice proves superior for subsequent retention (Carpenter & DeLosh, 2005; Pashler, Zarow, & Triplett, 2003). Lastly, even when retrieval fails, tests can potentiate learning by promoting

deeper encoding of correct information presented as feedback (termed *test-potentiated learning*; Arnold & McDermott, 2013; Kornell, 2014).

A naming treatment designed in accordance with these principles for maximizing testing benefits would look quite different from the typical errorless naming treatment. Not unlike the restudy condition of test-enhanced learning designs, errorless naming treatments eschew or minimize retrieval of information from long-term memory. For example, repetition of the named target, a popular errorless approach, can be achieved in whole or part by mapping from activated phonological representations in short-term memory.

Research suggests that testing is particularly effective for forging new (or reinforcing existing) associative pathways between units of information (e.g., the words *cucumber-frog* in paired-associate learning; Carpenter, 2009; Pyc & Rawson, 2012). Emerging accounts of this process include *elaborative retrieval* (Carpenter, 2009) and *mediator-effectiveness* theories (Pyc & Rawson, 2012). According to the elaborative retrieval hypothesis, 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*). In contrast, restudy does not require a memory search (i.e., the target is presented with the cue), so elaborative information is less likely to be activated or encoded with the cue-target pair. Similarly, mediator-effectiveness theory proposes that *failed* retrieval can trigger elaborative encoding of feedback, thereby strengthening cue-target associations (Pyc & Rawson, 2010; 2012).

If testing is particularly beneficial for enhancing associations between units of information, testing should ameliorate naming impairment in aphasia, as this impairment is commonly attributable to disrupted *access* to representations in the mental lexicon. The access deficit is thought to operate at two levels: one level associates semantic features with lexical units, or words; the other associates words with their phonological syllables and segments. Thus, testing could benefit naming in aphasia by bolstering associations at either or both levels.

The present study compared standard errorless naming treatment to two retrieval practice methods. First, each pictured object and its name was presented for one exposure trial to prime the association between the object and its name (similar to initial study in the test-enhanced learning paradigm). Five minutes later, each item underwent one training trial of either errorless treatment (i.e., *repetition),* retrieval practice with cued naming, or retrieval practice with uncued naming. Feedback was administered in all three conditions. The dependent measure of primary interest was performance on a naming test 1-day later. The errorless account predicts better retention in the repetition condition, insofar as that condition, relative to the retrieval practice conditions, elicits fewer errors during training and, consequently, less opportunity for Hebbian-based error learning. A secondary prediction of the errorless account is that retention will be poorest for uncued naming, which is expected to elicit the most errors, and most error-learning, during training. In contrast, the

test-enhanced learning account predicts superior retention in both retrieval practice conditions compared to repetition.

Method

A challenge to experimental investigations with patients is the frequent heterogeneity of cognitive-linguistic deficits even among those with the same diagnosis. Our design addressed this in the following manner: (1) We recruited participants with aphasia who were relatively homogeneous in severity and type of naming impairment. (2) We maximized the number of observations per condition per participant by starting with a very large picture corpus of common everyday objects (615 items). (3) From this corpus, in an *item selection phase* prior to the experiment, oral naming responses were elicited and items that had resulted in a naming error were reserved for use in the experiment for that participant (i.e., a participant's 'personalized' set of items). Maximizing number of observations increased power to detect differences within and across participants, and using error-prone items increased experiment sensitivity.[1] Because of the substantial resources required for data collection and processing for each patient, we enrolled eight participants.

Coding Considerations

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. In other contexts, it is desirable to instead code naming responses using a continuous measure of similarity between a response and a target. For example, such a measure may more precisely capture incremental differences in the difficulty individual items pose for naming, a variable that may be important for stimulus control.

In this work, all responses were coded in two ways. For each trial the first complete response was coded for *phonological overlap* (Lecours & Lhermitte, 1969), which provides a continuous measure of phonological similarity to a target that is standardized across different word lengths (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 (e.g., /kakt/ for cat was not considered correct). Semantic errors and descriptions (including all non-noun responses) received an overlap score of zero, because these were clear instances where an incorrect name was produced, and we did not want to reward coincidental phonological similarity to a target.

---

[1]The selection of error-prone items for use in the experiment introduces the possibility that changes in performance from item-selection to retention test may partially be due to regression to the mean. However, the main research question concerns differences in performance at delayed test as a function of training condition. Because items were matched into the conditions while controlling for phonological overlap, there is no reason to expect regression to the mean would be greater in any one condition.

$$\text{phonological overlap} = \frac{\#\ \text{shared phonemes in target and response} \times 2}{\Sigma \text{phonemes in target and response}} \quad (1)$$

Next, a binary variable of *response accuracy* was derived from the phonological overlap measure. A clear majority of target phonemes was taken to indicate the correct name was produced (phonological overlap .75). Less than .75 was coded as an incorrect response.

This binary response was the primary outcome measure in this study.

Participants

Participants gave informed consent under a protocol approved by the Institutional Review Board of Einstein Healthcare Network. Participants were reimbursed $15 per hour of participation. The sample included eight participants (three female) who were right-handed with chronic aphasia secondary to left-hemisphere stroke. They represented both fluent and nonfluent subtypes of aphasia, but at time of testing, the predominant deficit was in naming (see Tables A1 and A2 in the Appendix). The naming impairment in our participants was principally attributable to failure to reliably and fluently retrieve known words. Alternative sources of difficulty were ruled out by background tests and other data. For example, from their generally good performance on tests of nonverbal semantics and word comprehension (Table A1), one can be confident that the participants did not suffer from a central semantic deficit, which can compromise the semantic input to word retrieval. Neither were their naming failures caused by dysfunction in the phonological-phonetic encoding that follows word retrieval. This type of dysfunction typically manifests in errors where a target word is retrieved but produced with minor deviations in word form. These were rare (Column 3 in Table A2) compared to errors where the wrong word or no word was produced (Incorrect Category, Table A2). Their generally good word repetition (Table A1) further weighs against phonological-phonetic dysfunction as a major contributor to our participants' naming impairment.

Materials and Procedure

Pictures of 615 common objects were collected from published picture corpora (Szekely et al., 2004) and Internet sources. Visual complexity and name agreement values were taken from published corpora when available. Otherwise, these values were obtained in normative studies, with a minimum of 40 responses per item. Frequency values for all names were taken from the SUBTLEX$_{US}$ project (Brysbaert & New, 2009).

In the item selection phase, the entire picture set was administered in random order for naming to each participant twice over two weeks. Any item that yielded a response that was not an exact match to the name (all the target phonemes in the correct order) at least once was tagged for that participant's personalized item set. We allocated these items into the three training conditions, matching as closely as possible for phonological overlap and for several variables known to impact pathological and normal naming performance (Table 1). Differences in the severity of naming impairment produced a range (54-116; *M*=86) of observations per condition per participant.

Each participant completed the main experiment in multiple sessions spanning several weeks. Training and testing for one condition was completed prior to initiation of training for the next condition, with the order of conditions counterbalanced across participants. At the beginning of a training session, items were presented in random order for one exposure trial in which the picture and name were presented together (the name was seen and heard) and the participant repeated the name. After a five-minute break in which participants watched a video, the pictures were re-presented in random order for one training trial involving either retrieval practice (cueing or no-cueing) or repetition, depending on the condition. In the cueing condition, when the picture appeared, the word onset (consonant plus a shortened schwa or vowel, e.g., /t™¶/ for 'chicken'; /æ/ for 'apple') was played once and the onset's corresponding letter/letters (e.g., ch_____; a_____) was shown. In the no-cueing condition, only the picture was presented. On repetition trials, at picture onset the written name appeared and the auditory form of the word was played once. All trials ended after eight seconds, and the picture and orthography (when present) was shown for the duration of the trial. Participants were instructed to name as best they could (cueing and no-cueing conditions) or to repeat the word once (repetition condition) and then quietly wait for the end of the trial. All training trials were followed immediately by feedback, where the picture was shown with the orthographic/auditory form of the word and the participant repeated the name once. Immediately following this, the trial was advanced.

At the end of training, participants watched a 10-minute video before administration of the first retention test (i.e., *same-session retention test*). A naming test was also administered the next day (hereafter, *delayed test*) and seven days after training (hereafter, *follow-up test*).[2] In each retention test, pictures were presented in random order and the participant had up to 20 seconds to name the picture. To advance to the next trial, the participant was instructed to indicate when they had given their 'final answer,' a procedure adopted to eliminate experimenter feedback regarding accuracy of response. The entire experimental protocol (including item-selection testing) required between 11-14 sessions per participant (*M*=12 sessions). Participants with more items per condition required more sessions in the main experiment. Each session lasted between 10-60 minutes (*M*=25 minutes).

Analysis

Due to experimenter error, it was necessary to exclude data for one item from analyses. Response accuracy (correct/incorrect response) was modeled with mixed logistic regression using the lme4 package in R version 2.15.3 (R Development Core Team, 2012; for an introduction to mixed-effects models, see Baayen, Davidson, and Bates, 2008). All models described below included random intercepts for participants and items to capture the correlation among observations that can arise from multiple participants giving responses to

---

[2]Half of the items trained in a session were presented during the same-session retention test, and all items trained in a session were presented on the delayed and follow-up tests. We included a same-session test to examine retention shortly after training (comparable to the same-session retention tests administered in prior testing research). However, a same-session retention test may also function as an additional practice trial that influences performance after a longer delay. Thus, only including half of the items on the same-session test afforded comparison of performance on the delayed test as a function of whether items were versus were not included on the same-session test. In fact, performance on the delayed test did not differ systematically for items that had versus had not been included on the same-session test. Thus, we collapse across this factor in the main analyses.

overlapping sets of items.[3] For the main analyses, the fixed effect included in the models was a three-level factor of condition (cueing/no-cueing/repetition).

Results

Table 2 reports mean response accuracy during training and on retention tests for each condition. Table 3 reports the main analyses results.[4] As expected, errorless treatment (repetition condition) was associated with near perfect performance during training. Relative to repetition, the two retrieval practice conditions showed lower accuracy during training, with no-cueing showing the worst performance (p<.001 for all pairwise differences; see Table 3). Because retrieval practice training was more errorful than repetition, this meets the conditions for testing the contrasting predictions of errorless versus test-enhanced learning principles on the retention tests.

Differences between conditions were minimal at the same-session test (ps>.12 for all pairwise comparisons), aligning with prior work showing the advantage for retrieval practice over restudy is weak or reversed at short retention intervals (Roediger & Karpicke, 2006). Results at longer retention intervals strongly supported the test-enhanced learning approach. At the delayed test, performance in both the cueing (p=.003) and no-cueing condition (p-=. 038) exceeded that of the repetition condition. At the follow-up test, the advantage over repetition persisted for the cueing condition (p=.002), but not no-cueing (p=.154). The cueing and no-cueing conditions did not differ statistically at either the delayed test or at follow-up (all ps>.10). For individual participant results per condition, see Table A3 in the Appendix.

Consistent with expectations from research on test-enhanced learning, retrieval practice yielded superior delayed test performance compared to repetition. We next investigated whether this advantage was related to whether the correct name was retrieved on training trials for cued and non-cued naming training. Because retrieval outcome in training is likely to be confounded by variations in intrinsic item difficulty (i.e., easier items are more likely to be successfully retrieved), we derived a dependent variable that controlled for item difficulty, as measured by performance in the item selection phase.[5] In two linear mixed models, we modeled this dependent variable as a function of a three-level fixed factor corresponding to type of event during training (successful retrieval, failed retrieval, repetition), with one model focusing on the cueing and repetition data, and the other model focusing on the no-cueing and repetition data. Random intercepts for items and participants and a by-participant random slope for the fixed effect were included in each model. Results

---

[3]In all mixed-model analyses, we tested a by-participants random slope for the fixed effect(s), a model term that captures individual differences between participants in their response to a fixed effect (e.g., to the effect of condition). A random slope was included in a model if it improved model fit by chi-square deviance in model log likelihoods (Baayen, Davidson, and Bates, 2008). In the models for the main analyses, a by-participants random slope of condition did not improve model fit and was not included (Table 3). However, inclusion of the random slope in these models never impacted the significance of the fixed-effects coefficients.
[4]Parallel analyses as those reported in Table 3 were conducted using linear mixed modeling of the continuous measure of phonological overlap. There was no substantive change in any of the results using either phonological overlap or response accuracy.
[5]To control for effects of item difficulty, we used linear regression to derive a residualized delayed test performance measure. This involved modeling phonological overlap at delayed test performance as a function of average phonological overlap per item at item selection. We used phonological overlap rather than response accuracy to more effectively partial out incremental variations in the difficulty items pose for naming.

indicated that the advantage of retrieval practice over repetition at delayed test was rooted in the trials involving successful retrieval during training (in the cueing condition, coefficient = 0.11, *SE* = 0.02, *t*=5.02; in the no-cueing condition, coefficient = 0.13, *SE* = 0.03, *t*=4.31). Failed retrieval items did not significantly differ from items that underwent repetition (cueing, *t*=-0.43; no-cueing, *t*=-1.70).[6] These results indicate that successful retrieval, but not unsuccessful retrieval, confers an advantage over repetition in retention performance. With that said, it is noteworthy that items treated in the repetition condition were not retained better than failed retrieval items. This casts doubt on the guiding principle behind the errorless naming treatment approach, that retrieval errors in the course of treatment limit efficacy.

Discussion

We compared treatment methods for naming impairment based on errorless learning versus test-enhanced learning principles. Whereas errorless treatment was associated with superior rates of target production at training, retrieval practice methods conferred superior benefits at a retention test after one day, with the advantage persisting for the cueing condition after one week (for similar results in foreign language learning, see Kang, Gollan & Pashler, 2013). The advantage for the retrieval practice conditions originated from training trials involving successful retrieval of target names. This outcome is consistent with the elaborative retrieval hypothesis (Carpenter, 2009), which states that successful retrieval elaborates the retrieval structure associated with a target, improving subsequent accessibility. In contrast, the finding that training trials involving failed retrieval yielded similar benefits at delayed test as repetition training fails to support the mediator effectiveness hypothesis (Pyc & Rawson, 2012), which states that failed retrieval attempts enhance encoding of feedback relative to conditions where retrieval is not attempted (here, repetition). However, the present design was not ideal for evaluating the mediator effectiveness hypothesis because presenting the target name for a full eight seconds on each trial may have conferred an unfair advantage to the repetition condition. In any case, making retrieval errors did not produce significant decrements, and successful retrieval yielded substantial benefit. Taken together, the results demonstrate superior performance for treatment methods that prioritized retrieval practice over an errorless condition that minimized the opportunity to retrieve target names from long-term memory.

In aphasia treatment, as in other domains of neurorehabilitation, there is widely perceived need for research that establishes the effectiveness of available treatments. The present study makes progress towards addressing this need by establishing retrieval practice as a robust mechanism for ameliorating naming impairment. Given the extensive literature on ways to maximize the benefits of retrieval practice (Rawson & Dunlosky, 2011; Roediger & Karpicke, 2006), there are many promising avenues of future research that could build on this work. For instance, future studies could provide more than one retrieval practice opportunity per item (which typically enhances testing effects) and evaluate how to schedule such opportunities over time to maximize the benefits from effortful—yet successful—

---

[6]The function for generating p-values in languageR package is not implemented for models with random correlation parameters; for large data sets such as the current one, significance (α = .05) can be assumed for t-values 2, (Baayen et al., 2008).

retrieval of target information. It will also be important to understand how the relative effectiveness of retrieval practice methods versus errorless learning treatment relate to an individual's profile of cognitive-linguistic deficits. For example, people with aphasia may vary in their ability to filter out errors and incorporate feedback (e.g., from impaired explicit memory or executive dysfunction). In some cases, a hybrid approach may be desirable, e.g., coupling initial errorless learning training sessions with later sessions dedicated to retrieval practice.[7]

In the current study, retrieval practice ameliorated naming impairment where neuropsychological evidence suggested the deficit originated from difficulty in reliably *accessing* existing lexical representations. This extends beyond prior test-enhanced learning research, which has primarily focused on the effects of retrieval practice for *acquisition* of new knowledge, associations, or skills. The results suggest that retrieval from long-term memory is a pervasive learning mechanism spanning diverse domains of cognition including (minimally) episodic and semantic memory, and language processing.

The present work also bears on theories of lexical access by demonstrating that retrieval from long-term memory affects the persistent accessibility of lexical representations. Whereas such a mechanism has not been explicitly acknowledged in lexical access theories, some phenomena are consistent with it. For example, repetition priming studies show that naming can facilitate later retrieval of the same word even after long delays (e.g., 6 weeks; Mitchell & Brown, 1988). While repetition priming studies have not delineated a privileged role for retrieval practice in affecting long-term changes in accessibility (i.e., beyond the degree of priming expected from encoding that does not involve retrieval, e.g., reading or repetition), some long-term repetition priming effects may implicate retrieval practice, in alignment with the present results (Wheeldon & Monsell, 1992).

Long-term repetition priming has been an impetus for the view that each act of naming constitutes a learning event (incremental lexical learning; Damian & Als, 2005), modeled by weight-strengthening (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010). Within this framework, the present work provides evidence that retrieval from long-term lexical memory affects greater weight changes to a retrieved target compared to production that does not involve retrieval, an unacknowledged yet important constraint for incremental models of lexical learning.

In closing, the present work evaluated conflicting prescriptions for how to treat naming impairment in aphasia derived from the heretofore disconnected literatures on errorless learning and test-enhanced learning. By using rigorous experimentation to establish that retrieval practice ameliorates naming impairment in aphasia, this work provides an important foundation for future work exploring the application of test-enhanced learning principles to the treatment of language disorders.

---

[7]We thank an anonymous reviewer for this suggestion.

Acknowledgments

Appendix

Appendix. Table A1

Neuropsychological characteristics of study participants.

Participant	Type	Nonverbal Comp	Word Comp	Word Rep
S1	A	88	98	94
S2	A	83	89	98
S3	TCM	92	94	95
S4	B	85	91	96
S5	A	96	86	91
S6	C	94	98	94
S7	A	81	95	95
S8	A	96	92	87
Average		89	92.9	93.8

Note. *Type* = Aphasia type, where A = anomic, B = Broca's, C = conduction, TCM = Transcortical motor. *Nonverbal Comp* = A picture-picture association test for nonverbal comprehension, in percentages (Howard & Patterson, 1992). *Word Comp* = Spoken word to picture matching test of comprehension in percentages (Mirman et al., 2010). *Word Rep* = A test of immediate word repetition, in percentages (Mirman et al., 2010).

Table A2

Breakdown of Correct and Incorrect Responses On the Item Selection Test.

	Correct			Incorrect		
	Fully Correct	Minor Deviations	Phonol Err	Sem Err	NR/D	Other
Phonological Overlap:	1	.75-.99	0-.74	na	na	na
Participant						
S1	0.67	0.05	.06	.11	.08	.03
S2	0.59	0.03	.04	.13	.17	.03
S3	0.51	0.07	.06	.13	.19	.05
S4	0.64	0.07	.06	.13	.08	.02
S5	0.63	0.03	.05	.11	.12	.05
S6	0.60	0.07	.07	.10	.12	.04
S7	0.72	0.03	.04	.12	.06	.03
S8	0.57	0.06	.08	.14	.10	.04
Average	0.62	0.05	.06	.12	.12	.04

Note. *Fully correct,* Phonological overlap score=1.0; *Minor Deviations,* Overlap score between .75-.99; *Phonol Err,* phonologically related word or nonword response, with overlap score between 0 - .74; *Sem Err,* semantically-related word substitution. *NR/D* = null response or description involving either a multi-word phrase or single non-noun response. *Other* = unrelated response; named picture part.

Table A3

Mean Response Accuracy at Training, Delayed Test, and Follow-Up Test per Condition.

Participant	Condition	Training	Delayed Test	Retrieval Practice Advantage	Follow-up Test	Retrieval Practice Advantage
S1	Repetition	1.00	.778		.741	
	No-cueing	.796	.870	.093	.889	.148
	Cueing	.852	.889	.111	.944	.204
S2	Repetition	1.00	.552		.542	
	No-cueing	.396	.469	−.083	.385	−.156
	Cueing	.656	.531	−.021	.500	−.042
S3	Repetition	.957	.621		.543	
	No-cueing	.672	.784	.164	.724	.181
	Cueing	.733	.698	.078	.733	.190
S4	Repetition	.988	.845		.786	
	No-cueing	.854	.829	−.016	.720	−.066
	Cueing	.869	.845	.000	.798	.012
S5	Repetition	.989	.648		.648	
	No-cueing	.614	.670	.023	.739	.091
	Cueing	.682	.773	.125	.693	.045
S6	Repetition	1.00	.717		.707	
	No-cueing	.790	.860	.143	.740	.033
	Cueing	.840	.850	.133	.740	.033
S7	Repetition	1.00	.877		.860	
	No-cueing	.828	.897	.019	.897	.037
	Cueing	.897	.931	.054	.914	.054
S8	Repetition	.977	.625		.648	
	No-cueing	.614	.614	−.011	.693	.045
	Cueing	.750	.625	.000	.727	.080

Note. *Retrieval practice advantage* corresponds to the difference between each retrieval practice condition and the repetition condition. Positive values correspond to greater accuracy in the retrieval practice condition being compared.

References

Arnold KM, McDermott KB. Test-potentiated learning: Distinguishing between direct and indirect effects of tests. Journal of Experimental Psychology-Learning Memory and Cognition. 2013; 39(3): 940–945. doi: 10.1037/a0029199.

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59(4):390–412.

Bertens D, Fasotti L, Boelen DHE, Kessels RPC. A randomized controlled trial on errorless learning in goal management training: Study rationale and protocol. BMC Neurology. 2013; 13:64. [PubMed: 23786651]

Brysbaert M, New B. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods. 2009; 41(4):977–990. doi: 10.3758/brm.41.4.977. [PubMed: 19897807]

Carpenter SK. Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. Journal of Experimental Psychology-Learning Memory and Cognition. 2009; 35(6):1563–1569. doi: 10.1037/a0017021.

Carpenter SK, DeLosh EL. Application of the testing and spacing effects to name learning. Applied Cognitive Psychology. 2005; 19(5):619–636.

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

Conroy P, Sage K, Lambon Ralph MA. Errorless and errorful therapy for verb and noun naming in aphasia. Aphasiology. 2009; 23(11):1311–1337.

Damian MF, Als LC. Long-lasting semantic context effects in the spoken production of object names. Journal of Experimental Psychology-Learning Memory and Cognition. 2005; 31(6):1372–1384. doi: 10.1037/0278-7393.31.6.1372.

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

Fillingham JK, Sage K, Lambon Ralph MA. 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. 2005a; 19(7):597–614.

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

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

Howard D, Nickels L, Coltheart M, Cole-Virtue J. Cumulative semantic inhibition in picture naming: experimental and computational studies. Cognition. 2006; 100(3):464–482. [PubMed: 16413014]

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

Kang SHK, Gollan TH, Pashler H. Don't just repeat after me: Retrieval practice is more effective than imitation for foreign language learning. Psychonomic Bulletin and Review. 2013; 20:1259–1265. [PubMed: 23681928]

Kornell N. Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. Journal of Experimental Psychology-Learning Memory and Cognition. 2014; 40(1):106–114. doi: 10.1037/a0033699.

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

Leshner AF, Tom SR, Kern RS. Errorless learning and social problem solving ability in schizophrenia: An examination of the compensatory effects of training. Psychiatry Research. 2013; 206:1–7. [PubMed: 23158835]

Li R, Liu KPY. The use of errorless learning strategies for patients with Alzheimer's disease: A literature review. International Journal of Rehabilitation Research. 2012; 35(4):292–298. [PubMed: 23044898]

Lubinsky T, Rich JB, Anderson ND. Errorless learning and elaborative self-generation in healthy older adults and individuals with amnestic mild cognitive impairment: Mnemonic benefits and mechanisms. Journal of International Neuropsychological Society. 2009; 15:704–716.

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

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

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

Mitchell DB, Brown AS. Persistent repetition priming in picture naming and its dissociation from recognition memory. Journal of Experimental Psychology-Learning, Memory, and Cognition. 1988; 14(2):213–222.

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

Pashler H, Zarow G, Triplett B. Is temporal spacing of tests helpful even when it inflates error rates? Journal of Experimental Psychology-Learning Memory and Cognition. 2003; 29(6):1051–1057.

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

Pyc MA, Rawson KA. Why testing improves memory: Mediator effectiveness hypothesis. Science. 2010; 330:335. [PubMed: 20947756]

Pyc MA, Rawson KA. Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. Journal of Experimental Psychology-Learning, Memory, and Cognition. 2012; 38(3):737–746.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna: 2012. URL http://www.r-project.org/

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

Roediger HL, Karpicke JD. The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science. 2006; 1(3):181–210.

Szekely A, Jacobsen T, D'Amico S, Devescovi A, Andonova E, Herron D, Bates E. A new on-line resource for psycholinguistic studies. Journal of Memory and Language. 2004; 51(2):247–250. [PubMed: 23002322]

Wheeldon LR, Monsell S. The locus of repetition priming of spoken word production. Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology. 1992; 44(4): 723–761.

Table 1

Mean Item Characteristics per Condition, across Individual Participants' Stimuli Sets.

	Phonological Overlap[†]	Name Agreement	Log Frequency (per million)	Visual Complexity	Number of Phonemes
Condition	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Repetition	.39 (0.05)	.91 (0.01)	0.92 (0.04)	2.76 (0.06)	5.63 (0.30)
No-cueing	.39 (0.06)	.91 (0.01)	0.91 (0.06)	2.73 (0.07)	5.62 (0.31)
Cueing	.38 (0.06)	.91 (0.01)	0.92 (0.04)	2.70 (0.03)	5.62 (0.36)

Note.

[†]Average phonological overlap from the item-selection phase. SD = standard deviation.

Table 2

Mean (Standard Error) Response Accuracy at Training and Retention Naming Tests per Condition.

Condition	Training	Same-session Test	Delayed Test	Follow-up Test
Repetition	.99 (.01)	.78 (.03)	.71 (.04)	.68 (.04)
No-cueing	.70 (.05)	.77 (.05)	.75 (.05)	.72 (.06)
Cueing	.78 (.03)	.82 (.04)	.77 (.05)	.76 (.05)

Table 3

Mixed Logistic Model Coefficients and Associated Test Statistics

Response Accuracy at Training				
Fixed Effects	**Coef.**	**SE**	**Z**	**p**
Intercept	4.74	0.42	11.40	<.001
Cueing[a]	−3.23	0.37	−8.78	<.001
No-cueing[a]	−3.76	0.37	−10.31	<.001
Cueing[b]	0.54	0.13	4.09	<.001
Random Effects	s^2			
Participants	0.36			
Items	0.35			

Response Accuracy at Delayed Retention Test				
Fixed Effects	**Coef.**	**SE**	**Z**	**p**
Intercept	1.15	0.28	4.11	<.001
Cueing[a]	0.41	0.14	3.01	.003
No-cueing[a]	0.28	0.13	2.08	.038
Cueing[b]	0.13	0.14	0.93	.353
Random Effects	s^2			
Participants	0.54			
Items	0.73			

Response Accuracy at Follow-up Retention Test				
Fixed Effects	**Coef.**	**SE**	**Z**	**p**
Intercept	0.99	0.26	3.82	<.001
Cueing[a]	0.39	0.13	3.04	.002
No-cueing[a]	0.18	0.13	1.43	.154
Cueing[b]	0.21	0.13	1.61	.107
Random Effects	s^2			
Participants	0.46			
Items	0.49			

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; s^2 = Random effect variance.

[a] Reference is repetition condition

[b] Reference is no-cueing condition.