Research Article

The Roles of Retrieval Practice Versus Errorless Learning in Strengthening Lexical Access in Aphasia

Julia Schuchard^a and Erica L. Middleton^a

Purpose: The purpose of this study was to determine how 2 methods known to improve naming impairment in aphasia (i.e., retrieval practice and errorless learning) affect lexical access. We hypothesized that instances of naming during retrieval practice use and strengthen item-specific connections in each of 2 stages of lexical access: Stage 1, meaning-to-word connections, and Stage 2, word-to-phonology connections. In contrast, errorless learning prioritizes opportunities for repeating words, which we expect to primarily strengthen item-specific connections in Stage 2 because repetition circumvents the need for semantically driven word retrieval.

Method: We tested the outcomes of retrieval practice versus errorless learning training for items that were selected because the naming errors they elicited suggested weakened connections at Stage 1 or at Stage 2 of lexical access for each of 10 individuals with chronic aphasia. Each participant's Stage 1 items and Stage 2 items were divided evenly between the 2 training conditions. Naming tests were administered 1 day and 1 week after training to assess retention of training

gains. We also examined whether the participants' pretraining naming error profiles were associated with the relative efficacy of retrieval practice versus errorless learning.

Results: The posttraining naming tests showed an advantage of retrieval practice over errorless learning for Stage 1 items and an advantage of errorless learning over retrieval practice for Stage 2 items. In addition, greater percentages of phonological error naming responses prior to training were associated with greater posttraining accuracy in the errorless learning condition relative to the retrieval practice condition.

Conclusions: The findings suggest that the advantage of retrieval practice for naming impairment in aphasia largely results from greater strengthening of practiced semantic–lexical connections compared with errorless learning, which prioritizes repetition and, therefore, mainly confers strengthening of practiced lexical–phonological connections. Understanding how specific training conditions improve naming can help predict the relative efficacy of each method for individuals with aphasia.

Individuals with aphasia vary across a large number of neurological and cognitive–linguistic variables that may affect the outcomes of language treatment. Moreover, these variables may have quantitatively or qualitatively different effects on the impact of one treatment method compared with another. These factors exacerbate the difficulty of predicting the optimal course of treatment for a person with aphasia. To address the challenge of optimizing language rehabilitation, it is important to understand a treatment's mechanisms of action, that is, how specific training methods and learning experiences exert change on a damaged process (Baddeley, 1993; Gonzalez-Rothi,

2006; Nickels, 2002a; Whyte, 2006). The focus of this study is the treatment of naming impairment in aphasia (i.e., difficulty accurately retrieving the names for known entities, such as common, everyday objects).

The purpose of this study was to determine how two methods known to improve naming in aphasia (i.e., retrieval practice and errorless learning) affect specific stages of processing in the course of naming that involve lexical access (i.e., semantically driven retrieval of a word and its phonology). A wealth of studies in the psychology of learning and memory show that retrieval practice, or practice retrieving to-be-learned (i.e., target) information from long-term memory, confers superior learning of target information compared with nonretrieval-based learning experiences, such as reexposure to target information (i.e., restudy; for reviews, see Delaney, Verkoeijen, & Spiguel, 2010; Karpicke, Lehman, & Aue, 2014; Roediger & Karpicke, 2006; Rowland, 2014). In the context of this study, *retrieval practice* refers to

^aMoss Rehabilitation Research Institute, Elkins Park, PA
Correspondence to Julia Schuchard: jrschuchard@gmail.com
Editor-in-Chief: Sean Redmond
Editor: Charles Ellis

Received September 17, 2017

Revision received January 31, 2018

Accepted March 15, 2018

https://doi.org/10.1044/2018_JSLHR-L-17-0352

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

familiarization of a depicted object and its name, followed by attempts to name the picture without any additional cues present. We compare the effects of retrieval practice to errorless learning, which, in the context of naming training, refers to oral repetition of names with no opportunities for self-generated naming attempts. Here, we present evidence that (a) retrieval practice and errorless learning differentially strengthen specific stages of lexical access and (b) understanding these mechanisms of action can help predict the efficacy of these training methods for individuals with aphasia.

Models of Lexical Access

Of the series of cognitive and speech motor processes that comprise the act of naming, this study focuses on lexical access. Lexical access is generally agreed to involve at least partially independent stages of (a) mapping from meaning to an intermediary lexical representation, followed by (b) mapping from the lexical representation to its constituent phonemes (G.S. Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000; Walker & Hickok, 2016). It is debated whether the intermediary representation is modality-specific or amodal and how it stores or interfaces with syntactic information (see Rapp & Goldrick, 2000). In this article, we do not take a position on these issues; we will refer to the intermediary lexical representation using the neutral term *word* (see Figure 1).

The interactive two-step model of lexical access depicted in Figure 1 has been successful in explicating naming error patterns in aphasia and neurotypical speakers (G.S. Dell et al., 1997; Foygel & Dell, 2000; Schwartz, Dell, Martin, Gahl, & Sobel, 2006). In this model, Stage 1 begins with the activation of the semantic features of the target word. Activation spreads in the network to the target word and to words that share semantic features with the target, and the most active word is selected. Stage 2 then begins by activating the selected word. Activation spreads to the phoneme layer, and the most active phonemes in each position (e.g., onset, vowel, and coda) are selected to form the final response. Naming accuracy and error types for individuals with aphasia can be simulated by altering the strength of connections at Stage 1 (i.e., s-weights) and at Stage 2 (i.e., p-weights; Schwartz et al., 2006).

This model has also been extended to account for word repetition (i.e., hearing a word and orally producing it) in aphasia. The model can predict word repetition performance on the basis of patients' naming performance when assuming that (a) the auditory word recognition aspect of repetition is intact and (b) naming and repetition both use Stage 2 of lexical access (G.S. Dell, Martin, & Schwartz, 2007). That is, naming requires both semantically driven word retrieval (Stage 1) and phonological retrieval (Stage 2). The lexical route for repetition bypasses Stage 1 because the word is activated by auditory input, followed by Stage 2 phonological retrieval. Alternatively, the nonlexical route for repetition bypasses lexical access entirely by mapping

directly from input to output phonology, such as when repeating nonwords.

The interactive two-stage model and similar spreading activation theories of lexical access do not currently account for long-lasting changes that result from typical language production and from language treatment. However, they are not incompatible with models that posit that attempting to retrieve a word exerts persistent changes in the connections between representational levels (e.g., between semantic features and words), thereby changing the future accessibility of that word and related words (Breining, Nozari, & Rapp, 2016; Damian & Als, 2005; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010). Such models are able to account for the finding that retrieving a word leads to greater difficulty in subsequent retrieval of a semantically related word (i.e., semantic interference effect; e.g., Belke, Meyer, & Damian, 2005; Damian, Vigliocco, & Levelt, 2001; Kroll & Stewart, 1994). These semantic interference effects are relatively insensitive to the retrieval of intervening unrelated words or the passage of time (Damian & Als, 2005; Howard et al., 2006; Hsiao, Schwartz, Schnur, & Dell, 2009; Schnur, 2014), suggesting that the use of a particular semantic-lexical pathway exerts persistent changes on the underlying connections, even in the adult language system.

Retrieval Practice Versus Errorless Learning

The inclusion of self-generated naming attempts in naming treatment is in line with robust retrieval practice principles demonstrated in basic learning and memory research. Specifically, learning studies show that the act of retrieving information from long-term memory enhances posttraining performance (i.e., retention) for that information compared with restudy (i.e., reexposure to the information), even though the restudy condition provides more exposures to the correct information (see Delaney et al., 2010; Karpicke et al., 2014; Roediger & Karpicke, 2006; Rowland, 2014). Similarly, practice retrieving the names of objects or actions for written or oral production ameliorates naming impairment in aphasia (e.g., Hillis, 1989; Howard, Patterson, Franklin, Orchard-Lisle, & Morton, 1985; Nickels, 2002a) and can exert beneficial effects even in the absence of clinician-provided feedback or error correction (Fillingham, Sage, & Lambon Ralph, 2005a, 2005b; Nickels, 2002b).

In contrast, errorless learning interventions for cognitive rehabilitation prioritize exposure to correct information, thereby resembling the restudy conditions in the learning literature. Errorless methods are motivated by evidence that making errors can strengthen incorrect stimulus-response associations and reduce treatment efficacy, particularly for patients with impaired explicit memory (Baddeley & Wilson, 1994; Evans et al., 2000; Wilson & Evans, 1996). Across rehabilitation domains, errorless methods rarely promote completely errorless performance during training but, rather, reduce the frequency of errors compared with alternative methods. In naming treatment, errorless approaches reduce

Figure 1. An example of spreading activation in the interactive two-step model of lexical access (G.S. Dell et al., 1997; Foygel & Dell, 2000; Schwartz et al., 2006). Naming “cat” involves mapping from semantic features to the word (Stage 1, top) and from the word to its phonological form (Stage 2, bottom).

the frequency of errors by eliminating self-generated naming attempts, instead providing repeated exposure to the training stimuli and oral repetition of the names by the patients to bolster associations between depicted objects and their names (for reviews, see Fillingham, Hodgson, Sage, & Lambon Ralph, 2003; Middleton & Schwartz, 2012). Naming treatment studies have shown that an errorless learning approach can be at least equally beneficial as “errorful” learning (i.e., picture naming attempts) for people with aphasia (Conroy, Sage, & Lambon Ralph, 2009; Fillingham et al., 2005a, 2005b; Fillingham, Sage, & Lambon Ralph, 2006; McKissock & Ward, 2007). However, the errorful learning conditions in these studies did not provide initial familiarization of the names and, in some cases, omitted feedback, thereby limiting opportunities for successful name retrieval during training. These conditions are likely to reduce the benefits of a retrieval-based approach (Jang, Wixted, Pecher, Zeelenberg, & Huber, 2012; Kornell, Bjork, & Garcia, 2011; Rowland, 2014; Rowland & DeLosh, 2015). Hence, errorful conditions in these studies share some characteristics with retrieval practice-based learning methods but do not provide optimal learning conditions.

In contrast, recent research compared the effects of retrieval practice and errorless learning on naming impairment in aphasia using methods that included initial familiarization for each to-be-learned name and provided correct answer feedback. Reduced error rates were observed during errorless learning relative to retrieval practice, but subsequent naming tests showed higher accuracy for the items in the retrieval practice conditions (Friedman, Sullivan, Snider, Luta, & Jones, 2017; Middleton, Schwartz, Rawson, & Garvey, 2015; Middleton, Schwartz, Rawson, Traut, & Verkuilen, 2016). These results suggest that the benefits of retrieval practice typically outweigh the benefits of error reduction for the retention of naming gains. Nevertheless, findings in the errorless learning literature suggest that some individuals with aphasia respond better to errorless learning than to retrieval-based methods (Conroy et al., 2009; Fillingham et al., 2005a, 2005b, 2006; McKissock & Ward, 2007). In the following section, we explicate how retrieval practice and errorless learning may differentially affect lexical access, which, in turn, can help account for how individuals differ in their responses to these two training methods.

The Present Study

In this study, we draw on the hypothesis that using a pathway alters the connections underlying that pathway. For the purposes of this study, the critical difference between retrieval practice and errorless learning is that the former provides opportunities for self-generated naming and for repeating words (in the initial familiarization trials and during feedback after each naming attempt), whereas errorless learning only provides opportunities for repeating words. In the previous section, we reviewed evidence that naming requires the use of both Stage 1 and Stage 2 of lexical access, whereas word repetition uses Stage 2 but relies less on Stage 1. Thus, we hypothesized that instances of naming during retrieval practice use and strengthen item-specific connections in both Stage 1 and Stage 2 of lexical access, whereas instances of name repetition during errorless learning primarily use and strengthen item-specific connections in Stage 2. Here, we assume that people with aphasia use Stage 2 connections to repeat words, rather than relying solely on a nonlexical route for repetition. This assumption is supported by Nozari, Kittredge, Dell, and Schwartz (2010), who compared the effect of word frequency (an index of lexical–phonological retrieval) on naming and repetition in aphasia and found that people with aphasia use the lexical route (alone or in combination with a nonlexical route) when repeating words. It is unlikely that repeating words during errorless learning benefits only Stage 2 processing, considering that the errorless learning studies discussed above found that individuals with a variety of linguistic symptom profiles showed some degree of naming improvement from this approach. In the Discussion section, we return to possible mechanisms by which repetition may affect Stage 1 connections. Nevertheless, we contend that retrieval practice results in stronger changes at Stage 1 than errorless learning because semantically driven word retrieval is required for naming but not repetition.

These predicted effects of errorless learning versus retrieval practice could be tested at the level of individuals by comparing across participants with different linguistic profiles or at the level of specific items by comparing performance for different items within individuals. For the former method, one would identify individuals with aphasia whose linguistic profiles indicated weakened connections at Stage 1 but not at Stage 2 or vice versa and examine the outcomes of errorless learning versus retrieval practice for these individuals. However, it can be difficult to identify individuals with naming impairment isolated to one stage of naming and matched across other potentially relevant factors, such as severity of impairment. In this study, we used the item-level method, which allows for within-participant comparisons. We examined the outcomes of each training method for selected items that exhibited evidence of weakened connections at Stage 1 or at Stage 2 of lexical access for each of 10 individuals with aphasia. These items were identified based on the types of naming errors they elicited for an individual (primarily semantic errors indicating Stage 1 impairment and phonological errors indicating Stage 2 impairment).

Because the same type of error may arise from different underlying speech-language impairments, we selected a relatively homogeneous group of participants who primarily exhibited lexical access impairment (as detailed in the Method section).

Following from our hypotheses, we predicted that retrieval practice should result in higher naming accuracy than errorless learning for the Stage 1 type of items at posttraining naming tests that assess retention of gains, whereas the two types of training should have similar effects on the Stage 2 type of items. However, effects of accuracy during training may complicate these predicted patterns. Previous evidence suggests that successful retrieval attempts during training confer substantially greater benefit than failed retrieval attempts, even when both kinds of trials are followed by correct answer feedback (Middleton et al., 2015). Hence, in addition to analyzing accuracy at post-training naming tests, we included analyses of the difference between training accuracy and test accuracy to examine the persistence of training gains over time.

Although this study focuses on item-level effects, similar predictions would apply to individual-level differences between persons with aphasia. That is, if an individual exhibits relatively more difficulty at one stage of lexical access than another, then the training method that provides more opportunities to use that stage should be more efficacious for that person. This study tests our item-level predictions, and a secondary analysis investigates these individual-level predictions as well.

Method

Participants

Participants were 10 native English speakers (three female) with chronic aphasia resulting from stroke. These individuals were selected because difficulty in naming was a primary language deficit; all participants scored lower than 90% correct on the Philadelphia Naming Test (Roach, Schwartz, Martin, Grewal, & Brecher, 1996). Participants had completed at least 12 years of education ($M = 14$ years) and had no indications of developmental disorder, neurodegenerative disorder, or premorbid major psychiatric illness. We did not collect detailed information regarding prior or concurrent language treatment, but participants were past the period in which formal language treatment is typically administered (nine out of 10 were more than 3 years post-stroke). Table 1 provides demographic and language testing information for each participant and lesion information for participants whose brain scans were available for examination. Participants gave informed consent under a protocol approved by the Institutional Review Board of Einstein Healthcare Network and were paid at a rate of \$15 per hour.

As detailed below, participants' naming errors were used as evidence of weakened connections at Stage 1 or at Stage 2 of lexical access. However, we cannot completely rule out other causes of participants' naming errors. For example, semantic errors can result from impairment to

Table 1. Participant information.

Test	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Age (years)	71	52	52	53	60	58	56	68	66	72	61
Years post-onset	4	4	11	7	8	1	16	35	5	9	10
Primary site(s) of lesion (all left hemisphere)	inferior frontal; temporal; parietal	inferior frontal; temporal; parietal	frontal; superior temporal anomic	inferior frontal anomic	temporal; parietal anomic		inferior frontal; temporal; parietal anomic			temporal; parietal anomic	
Western Aphasia Battery (WAB) classification	Broca's	Broca's				anomic		anomic	anomic		
Apraxia of speech	moderate	mild	mild	none	mild	none	none	none	none	none	
WAB Aphasia Quotient	58.8	77.0	81.5	83.2	85.6	86.4	90.0	90.1	90.4	91.2	83.4
WAB Auditory Comprehension	7.1	9.7	8.6	8.0	9.2	9.2	10.0	9.5	9.1	9.2	8.9
WAB Repetition	6.1	7.8	9.0	8.8	8.1	8.0	7.4	8.3	8.4	9.5	8.1
Philadelphia Repetition Test	89	95	97	95	84	97	85	93	89	88	91
Nonword Repetition Test	23	50	65	77	35	47	35	32	37	45	45
Pyramids and Palms Test	94	94	83	92	98	96	96	88	85	94	92
Synonymy Triplets	93	83	80	70	87	90	93	80	73	87	84
Philadelphia Naming Test	57	85	74	85	81	69	79	82	88	85	79
Common Object Naming Corpus											
Correct	53	65	60	69	61	57	57	52	69	60	60
Semantic error	9	14	5	11	6	18	12	15	11	17	12
Phonological error	21	15	7	12	27	12	22	12	8	11	15
Description	1	2	1	1	1	4	3	4	3	4	2
Omission	13	2	26	5	2	1	2	11	4	2	7
Other errors	3	2	1	2	3	8	4	6	5	6	4

Note. WAB subtests have a maximum score of 10. All other test scores represent percentages with maximum of 100%. P = participant.

the semantic representations themselves or access to those representations (Gainotti, Miceli, Caltagirone, Silveri, & Masullo, 1981; Hillis, Rapp, Romani, & Caramazza, 1990; Jefferies & Lambon Ralph, 2006). To limit the influence of these additional causes of naming errors, we excluded individuals who scored poorly on tests of word comprehension or nonverbal semantic associations (see Table 1 for relevant scores on the Western Aphasia Battery; Kertesz, 1982, the Pyramids and Palm Trees Test; Howard & Patterson, 1992, and the Synonymy Triplets Test; Martin, Schwartz, & Kohn, 2006). In addition, phonological errors can result from impairments in postlexical phonemic buffering/sequencing or articulatory processes (Goldrick & Rapp, 2007; Romani & Galluzzi, 2005; Romani, Galluzzi, & Olson, 2011; Shallice, Rumiat, & Zadini, 2000), but participants in this study were unlikely to have an exclusively postlexical origin for phonological errors because we excluded individuals with severe apraxia of speech (as evaluated by the Apraxia Battery for Adults–Second Edition; Dabul, 2000) and those with worse than mild word repetition performance (see Table 1 for scores on repetition tests). One individual with moderate apraxia of speech (P1 in Table 1) was included because the research team judged that lexical access impairment, not only speech motor difficulty, contributed to his naming errors. For example, this participant produced a variety of error types on the Common Object Corpus, not only phonetic errors.

Stimuli

Stimuli consisted of items from a Common Object Corpus, which includes pictures of 660 common objects obtained from published picture corpora (Szekely et al., 2004) and Internet sources. All target names were nouns, ranging in length from one to five syllables, one to 13 phonemes, and two to 15 letters. For the purposes of a separate experiment, 88% of the 660 items were divided between 19 experimenter-determined semantic categories (e.g., clothing, musical instruments, vehicles) and were rated by healthy adults for their “goodness of fit” with other corpus items in the same category. Table 2 provides details about the stimuli selected for training in this study, including average goodness of fit ratings for items that belonged to one of the 19 categories. Audio recordings of the names were created by a female native English speaker. All stimuli were presented on a PC computer using E-Prime software.

Response Coding

Participant naming and repetition responses were recorded, transcribed, and coded for accuracy and error type using a modified version of the coding system established for the Philadelphia Naming Test. The research assistant who conducted transcription and coding for this study had been trained in these tasks using speech samples from people with various types and severity of aphasia and apraxia of speech. The research assistant practiced until she independently and consistently matched the transcriptions

and coding established by the speech pathologist on the research team (with at least 90% agreement). She was then considered fully trained and reliable for this study.

The first complete utterance in each trial was coded for accuracy and error type. Subsequent utterances within the response interval or responses produced after the response interval were not coded. A response was coded as correct if the first complete utterance matched the target response, with all the correct phonemes produced in the correct order. A *semantic error* was defined as a noun related to the target by association or as a category coordinate, superordinate, or subordinate (e.g., “water” for “raincoat”; “horn” for “clarinet”). Noun responses that were semantically and phonologically related to a target (i.e., mixed errors) were classified as semantic errors because the lexical status of the error and its semantic relationship to the target suggests origination from retrieval of the wrong word rather than faulty phonological mapping. A *description* was defined as a nonnoun or a multiple-word response that describes the item (e.g., “I use it to cook” for “spatula”). A *phonological error* was defined as a semantically unrelated word or nonword that shares more than one phoneme with the target or a single phoneme in the same position compared with the target (e.g., “turkey” for “trophy”). An omission was defined as no response or a personal comment (e.g., “I like it”). Other error types included morpheme omissions in multiple-morpheme names (e.g., “rope” for “jump rope”), picture parts (e.g., “bread” for “sandwich”), and unrelated errors (e.g., “outside” for “peacock”).

Item Selection

In the initial item selection phase of the study, participants were instructed to name each picture in the Common Object Corpus. The pictures were presented in a random order two times, with the two administrations occurring in separate weeks, typically requiring a total of four sessions. Each naming session lasted between one and 2 hr, including as many breaks as the participant needed. The experimenter ended each naming trial when the participant pointed to a “thumbs up” picture to indicate that she had given her final answer or after 20 s if a final answer was not provided. For the purposes of a separate study, auditory and visual presentation of the name was provided for the participant to repeat after a randomly selected 50% of trials in the first administration. This single presentation of feedback was not expected to influence training outcomes in this study because at least 3 weeks elapsed between the first administration and training. Nevertheless, the number of items that received feedback was matched across conditions during item assignment (see Table 2).

To select items with weakened connections at Stage 1 of lexical access (hereafter, “Stage 1 items”), we identified those that elicited a semantic error or a description in at least one administration of the Common Object Corpus. Semantic errors are associated with incorrect mapping from semantics to words in Stage 1 of the two-step interactive model of lexical access (Schwartz et al., 2006), and evidence

Table 2. Stimuli means and standard deviations.

Category	Name agreement	Frequency	Phonemes	Syllables	Letters	Visual complexity	Semantic category	Feedback
Errorless learning (all)	0.92 (0.06)	0.75 (0.46)	6.3 (2.0)	2.3 (0.9)	7.3 (2.2)	2.7 (0.7)	1.9 (0.7)	48%
Stage 1 items	0.91 (0.06)	0.81 (0.49)	5.9 (1.8)	2.1 (0.8)	6.8 (2.2)	2.7 (0.7)	1.9 (0.7)	47%
Stage 2 items	0.93 (0.06)	0.71 (0.41)	6.9 (2.2)	2.5 (0.9)	7.9 (2.3)	2.7 (0.7)	2.0 (0.7)	50%
Filler items	0.91 (0.05)	0.53 (0.31)	6.7 (1.9)	2.3 (0.8)	7.7 (2.0)	2.6 (0.6)	2.0 (0.8)	43%
Retrieval practice (all)	0.91 (0.06)	0.75 (0.46)	6.3 (2.0)	2.2 (0.9)	7.3 (2.3)	2.7 (0.8)	2.0 (0.7)	47%
Stage 1 items	0.90 (0.06)	0.80 (0.49)	5.9 (1.9)	2.1 (0.8)	6.9 (2.2)	2.7 (0.7)	2.0 (0.8)	46%
Stage 2 items	0.93 (0.06)	0.71 (0.43)	6.8 (2.0)	2.4 (1.0)	7.8 (2.2)	2.7 (0.8)	1.9 (0.7)	48%
Filler items	0.93 (0.04)	0.56 (0.32)	6.6 (2.3)	2.5 (1.0)	7.8 (2.5)	2.7 (0.8)	2.0 (0.7)	43%

Note. Frequency values represent the log of frequency per million, obtained from SUBTLEX_{US} (Brysbaert & New, 2009). Other values were obtained from published corpora or norming experiments with healthy adult participants. Name agreement values represent the proportion of participants who responded to the picture with the target name when asked to name the picture. Visual complexity values represent ratings of the visual complexity of the picture from 1 (*very simple*) to 5 (*very complex*). Semantic category values represent ratings of the item's goodness of fit with other corpus items in the same semantic category from 1 (*fits very well*) to 5 (*does not fit very well*). For the purposes of a separate study, auditory and visual presentation of the name was provided for the participant to repeat after a randomly selected 50% of trials in the first administration of the Common Object Corpus; "feedback" denotes the percentage of items that received this feedback.

from lesion analyses in aphasia suggests that descriptions also commonly result from difficulty in mapping from semantics to words (Antonucci, Beeson, Labiner, & Rapcsak, 2008; Walker et al., 2011). Although we prioritized items that elicited a semantic error in both pretraining administrations, obtaining a sufficient number of Stage 1 items per person occasionally required including items that elicited a different type of error or a correct response in one of the two administrations. For example, an item that was named correctly on the first administration but elicited a semantic error on the second administration would meet the criteria for a Stage 1 item. Collapsing across the two administrations, the majority of pretraining naming responses for the selected Stage 1 items were semantic errors or descriptions: 66% semantic errors, 8% descriptions, 4% omissions, 2% phonological errors, 8% other error types, and 12% correct. Appendix A displays the error types produced by each participant in each of the two administrations separately.

To select items with weakened connections at Stage 2 (hereafter, "Stage 2 items"), we identified those that elicited a phonological error in at least one administration, prioritizing those that elicited a phonological error in both administrations. Formals (i.e., phonologically related word errors) can arise in Stage 1, with feedback from the phonological level resulting in the selection of a word phonologically related to the target, or in Stage 2, with incorrect phoneme retrieval forming a word rather than a nonword by chance (Schwartz et al., 2006). On an administration of the Common Object Corpus in this study, however, the majority of each participant's phonological errors were nonwords ($M = 81\%$), which are associated with correct Stage 1 word selection followed by incorrect mapping to output phonology in Stage 2 (Schwartz et al., 2006). Collapsing across the two administrations of the Common Object Corpus, pretraining naming responses for the selected Stage 2 items in this study were 77% phonological errors, 5% omissions, 1% other error types, and 17% correct. (See Appendix A for individual participant error types.)

Each participant was assigned 88 items for training. This number of items was determined by the number of training trials that could be administered in approximately 1 hr of training. Six participants were assigned equal numbers of Stage 1 items and Stage 2 items. The other four participants did not produce a sufficient number of phonological errors to form a full set of Stage 2 items. To equate the length of training across participants, these four participants were assigned additional Stage 1 items (analyzed with the other Stage 1 items) or filler items (excluded from the primary analyses) that had elicited other error types in the pretraining administrations of the Common Object Corpus. Exact numbers of Stage 1 and Stage 2 items for each participant are listed in Appendix A. Across the 10 participants, the final sample of training items included 494 Stage 1 items and 330 Stage 2 items. For each participant, each item type was divided evenly between the errorless learning and retrieval practice training conditions, matching as closely as possible for variables known to impact naming performance (see Table 2).

Training and Test Procedures

The training session consisted of one block of errorless learning and one block of retrieval practice, including 44 items per block. The order of the two blocks was counter-balanced across participants. In the errorless learning block, each item was presented for three repetition trials. On a repetition trial, the picture of the object was presented simultaneously with the written and auditory forms of the name. The participant was asked to repeat the name once. The picture and written name remained on the computer monitor for 8 s. After that time, correct answer feedback was provided (i.e., the auditory and written forms of the name were presented for the participant to repeat). In the retrieval practice block, the first trial for an item was a repetition trial. The subsequent two trials were naming trials. On a naming trial, the picture was presented, and the participant attempted to name the picture. After 8 s, correct answer

feedback was provided. Thus, all trials in both training conditions ended in correct answer feedback, and the timing of the trials was matched across conditions. In both conditions, 10 intervening items were administered between each of the three trials for a given item.

A naming test for the participant's trained items was administered 1 day after training (hereafter, "Test 1") and 1 week after training (hereafter, "Test 2"). The procedures for Test 1 and Test 2 were identical to the item selection phase (i.e., the participant was given 20 s to try to name each picture), with no feedback administered.

Data Analyses

Data analyses were performed with Stata 14.2 using mixed-effects regression models. These models are used to test main effects and interaction effects, as an analysis of variance can. However, mixed-effects regression models can also account for by-participant and by-item variability in the random-effects structures and have documented advantages compared with analyses of variance (e.g., Baayen, Davidson, & Bates, 2008; Jaeger, 2008). Although there is no standard method to determine the optimal sample size for a mixed-effects model, research suggests that reasonable estimates are obtained when at least 10 participants are included in the random-effects structure (McNeish & Stapleton, 2016). The random-effects structure of each model in this study included by-participant random intercepts. Random slopes were also included if they significantly improved model fit in a model comparison procedure, which was evaluated using the change in the deviance statistic (-2 times the log-likelihood). Item-level random intercepts were included in models analyzing repeated measures of the items (e.g., at Test 1 and at Test 2); otherwise, they were not included because the majority of items occurred no more than twice in the dataset.

In addition to the experimental conditions, psycholinguistic characteristics of the stimuli may affect participants' naming accuracy. By examining data from similar studies previously conducted in the lab, we identified word length (i.e., number of phonemes), lexical frequency, and name agreement as the variables most likely to affect participants' performance in this study. Using the model comparison procedure described above, we found that the inclusion of the target words' length as a covariate significantly improved the model fits in the analyses of this study, whereas frequency and name agreement values did not. Although psycholinguistic variables were matched as closely as possible across the experimental conditions, including a significant covariate is useful to statistically control for variability in participant performance that does not arise from the experimental conditions of interest. Hence, we included word length as a covariate in each model.

The primary set of analyses included the items that had been assigned to the four cells of the experimental design (training type, i.e., errorless learning vs. retrieval practice, crossed with item type, i.e., Stage 1 vs. Stage 2 items). The dependent variable in each model was response accuracy

(correct or incorrect). We first constructed models to test the predicted Training Type $\times$ Item Type interaction on naming accuracy at Test 1 and at Test 2 (see Table 3, Models 1 and 2). Data from both tests were included in subsequent models that examined (a) posttraining differences between the two training conditions for each item type separately (see Table 3, Models 3a and 3b) and (b) the Training Type $\times$ Training versus Test Phases interaction of the study for each item type (see Table 3, Models 4a and 4b).

In a secondary analysis, we constructed a mixed-effects regression model to examine whether the participants' pretraining naming error profiles were associated with the relative efficacy of errorless learning versus retrieval practice (see Table 3, Model 5). Using the data from the first administration of the Common Object Corpus, the critical predictors entered in the model were the participants' percentage of pretraining naming responses classified as semantic errors (serving as a measure of Stage 1 impairment) and percentage of pretraining naming responses classified as phonological errors (serving as a measure of Stage 2 impairment). The dependent variable was calculated by applying the following formula to all items for each participant at Test 1 and, again, at Test 2: percent correct in the retrieval practice condition $-$ percent correct in the errorless learning condition. Filler items were included in this measure because item type was not a factor of interest in this analysis.

Results

Figure 2 displays accuracy in the errorless learning and the retrieval practice conditions across the training and test sessions for the Stage 1 items (Figure 2a) and for the Stage 2 items (Figure 2b). Table 3 summarizes the parameter estimates for the critical terms in each statistical model, along with their standard errors and p values estimated using the normal approximation for the z value ($z = \text{coefficient} / \text{standard error}$). Additional details and results for each model are reported in Appendix B.

Posttraining naming test results showed that the effect of retrieval practice versus errorless learning depended on the type of item. The Training Type $\times$ Item Type interaction effect on naming accuracy was significant at Test 1, coefficient = -0.62 , $SE = 0.31$, $p < .05$ (see Table 3, Model 1), and at Test 2, coefficient = -0.85 , $SE = 0.31$, $p < .01$ (see Table 3, Model 2). Because the pattern of results was similar at the two test times, data from both tests were included together in subsequent analyses. Across Test 1 and Test 2, accuracy was significantly higher in the retrieval practice condition (61%) relative to the errorless learning condition (54%) for the Stage 1 items, coefficient = 0.51 , $SE = 0.25$, $p < .05$ (see Table 3, Model 3a). In contrast, accuracy was significantly lower in the retrieval practice condition (37%) relative to the errorless learning condition (45%) for the Stage 2 items, coefficient = -0.55 , $SE = 0.26$, $p < .05$ (see Table 3, Model 3b). These results reflect the effects of training type and item type controlling for word length. As expected, word length also had a significant effect on naming accuracy (see Appendix B).

Table 3. Key statistical analysis results.

Research question	Model	Fixed effect	Coefficient	SEM	p value
Do the effects of retrieval practice versus errorless learning on Test 1 accuracy differ between Stage 1 items and Stage 2 items?	Model 1	Training Type × Item Type	−0.62	0.31	.046*
Do the effects of retrieval practice versus errorless learning on Test 2 accuracy differ between Stage 1 items and Stage 2 items?	Model 2	Training Type × Item Type	−0.85	0.31	.01*
Across Tests 1 and 2, does accuracy in the retrieval practice condition differ from the errorless learning condition for Stage 1 items?	Model 3a.	Training Type	0.51	0.25	.04*
Across Tests 1 and 2, does accuracy in the retrieval practice condition differ from the errorless learning condition for Stage 2 items?	Model 3b.	Training Type	−0.55	0.26	.04*
For Stage 1 items, do the effects of retrieval practice versus errorless learning differ between the training and test phases?	Model 4a.	Training Type × Time	1.64	0.25	< .001*
For Stage 2 items, do the effects of retrieval practice versus errorless learning differ between the training and test phases?	Model 4b.	Training Type × Time	−0.06	0.25	.82
Do the participants' rates of semantic or phonological naming errors prior to training predict the difference between retrieval practice and errorless learning effects on naming after training?	Model 5	Percent semantic errors	0.02	0.52	.97
		Percent phonological errors	−1.06	0.34	< .01*

* $p < .05$.

Figure 2. Average accuracy for Stage 1 items (a) and Stage 2 items (b) assigned to the errorless learning and retrieval practice conditions. The data indicate repetition response accuracy for training Trials 1–3 in the errorless learning condition and for training Trial 1 in the retrieval practice condition; all other points indicate naming response accuracy. Pretraining accuracy is averaged across two administrations of the items in the item selection phase.

Analyses of accuracy across the training and test sessions showed that the relative advantage of errorless learning compared with retrieval practice during training persisted at test for the Stage 2 items only. For the Stage 1 items (Figure 2a), higher accuracy in the errorless learning condition relative to retrieval practice during training reversed to lower accuracy at subsequent test sessions, resulting in a significant Training Type $\times$ Time (training vs. tests) interaction, coefficient = 1.64, $SE = 0.25$, $p < .001$ (see Table 3, Model 4a). For the Stage 2 items (Figure 2b), accuracy remained higher in the errorless learning condition compared with retrieval practice at the test sessions, resulting in no Training Type $\times$ Time interaction, coefficient = -0.06 , $SE = 0.25$, $p = .82$ (see Table 3, Model 4b).

In addition to accuracy, the distribution of error types produced during and after training is informative. Most types of errors were relatively rare during training, presumably due to the fact that, in both training conditions, participants saw and heard the correct names in the initial familiarization (i.e., repetition) trial for each item and as feedback in each trial. However, the exposure to the correct names during training did not always preempt participants' phonological difficulty when attempting to repeat or name an item. Hence, phonological errors constituted 98% of errors during errorless learning and 88% of errors during retrieval practice, indicating that any differences in training accuracy primarily resulted from differences in phonological difficulty. For example, training performance was generally lower for Stage 2 items versus Stage 1 items (see Figure 2). In the subsequent naming tests, the two item types elicited numerically equal or fewer phonological errors following errorless learning relative to retrieval practice (see Figure 3). The advantage of errorless learning for reducing the frequency of phonological errors is particularly apparent for the Stage 2 items, which rarely elicited other error types (Figure 3b). In contrast, the advantage of retrieval practice for Stage 1 items resulted from a lower frequency of semantic and other nonphonological error types at the posttraining tests (Figure 3a).

The results of the secondary analysis identified participants' pretraining phonological error rate as a significant predictor of the relative efficacy of the two types of training. Greater percentages of phonological error naming responses prior to training were associated with greater posttraining accuracy in the errorless learning condition relative to the retrieval practice condition, coefficient = -1.06 , $SE = 0.34$, $p < .01$ (see Table 3, Model 5). Participants' pretraining semantic error rate was not a significant predictor of posttraining differences between the two training conditions, coefficient = 0.02 , $SE = 0.52$, $p = .97$ (see Table 3, Model 5). Figure 4 depicts the relationship between individuals' pretraining naming error percentages and their posttraining differences in naming accuracy.

Discussion

In this study, the effect of retrieval practice versus errorless learning on subsequent naming accuracy depended on whether the target names had showed evidence of

weakened connections at Stage 1 or at Stage 2 of lexical access. Errorless learning boosted word production accuracy during training compared with retrieval practice for both types of items, but this advantage persisted at subsequent tests for the Stage 2 items only. Naming tests administered 1 day and 1 week after training showed an advantage of retrieval practice for Stage 1 items and an advantage of errorless learning for Stage 2 items. These findings suggest that the advantage of retrieval practice for naming impairment in aphasia largely results from greater strengthening of practiced semantic-lexical (Stage 1) connections, whereas the prioritization of repetition during errorless learning mainly confers strengthening of practiced lexical-phonological (Stage 2) connections.

These results should motivate the extension of models of use-dependent, incremental learning in semantic-lexical connections (Damian & Als, 2005; Howard et al., 2006; Oppenheim et al., 2010) to a two-stage framework. When naming difficulty arises due to weakened connections at Stage 1 of lexical access, greater change should occur via training methods that use those connections. This is consistent with the observed advantage of retrieval practice for Stage 1 items because instances of naming in retrieval practice require Stage 1 mapping, whereas instances of name repetition in errorless learning do not. When naming difficulty arises due to weakened connections at Stage 2, at least equally strong changes should occur via errorless learning compared with retrieval practice because both training methods use Stage 2 connections. In this study, naming accuracy was significantly higher for the Stage 2 items following errorless learning. One interpretation of this effect is that successful Stage 2 processing confers greater benefit than failed Stage 2 processing. Hence, errorless learning may confer greater benefit than retrieval practice at Stage 2 because errorless learning provides more opportunities to correctly practice that stage via repetition.

There are several possible ways in which repetition could affect Stage 1 connections as well. When repeating a word, semantics may serve as input to the mapping from the auditory word form to the word unit. The informativeness of such input to target selection could provide a signal for learning. In addition, outcomes at a lexical-phonological level may influence mappings at an earlier semantic-lexical level, for example, via backpropagation mechanisms implemented in certain models. In this study, many items classified as Stage 1 likely had some degree of weakening at Stage 2 as well (and vice versa); hence, repetition during errorless learning could benefit Stage 1 items via strengthening of phonological connections. These possibilities are an important topic for future investigations of naming and repetition effects on lexical access.

Another essential research direction will be to investigate how opportunities for naming versus repetition interact with other training techniques in multiple-step treatment protocols like semantic feature analysis and Verb Network Strengthening Treatment, which have been shown to promote generalized improvement in naming for some individuals with aphasia (Boyle, 2010; Boyle & Coelho, 1995;

Figure 3. Types of naming responses at pretraining (collapsed across two separate administrations), Test 1, and Test 2 for Stage 1 items (a) and Stage 2 items (b) assigned to the errorless learning and retrieval practice conditions.

Figure 4. Pretraining naming error profiles in relation to posttraining results for each of the 10 participants. Retrieval practice effect values on the vertical axis represent the percent correct in the retrieval practice condition minus the percent correct in the errorless learning condition at Test 1 (circles) and at Test 2 (squares). Solid shapes indicate positive values, that is, better performance in the retrieval practice condition. Unfilled shapes represent 0 or negative values, that is, equal or better performance in the errorless learning condition.

Edmonds, Mammino, & Ojeda, 2014; Edmonds, Nadeau, & Kiran, 2009). In particular, specifying processes that confer benefits on untrained words is important to increase the impact of naming treatment. According to Oppenheim et al.'s (2010) model of incremental learning, retrieval of a word facilitates future access to that word and weakens future access to semantically related words. In the learning literature, however, retrieval of target information can either weaken or strengthen memory for related information (Anderson, 2003; Chan, 2009, 2010; Chan, McDermott, & Roediger, 2006; Rowland & DeLosh, 2014). The conditions that lead to these two different retrieval practice effects are a topic of ongoing investigation, and similar studies of naming training will be important to reveal conditions in which retrieval practice generalizes to untrained words.

Although this study focused on item-level effects, the findings also have implications for individual-level differences in naming treatment outcomes. An individual who primarily exhibits difficulty in mapping from meaning to words should strongly benefit from opportunities to practice semantically driven word retrieval, whereas an individual who primarily exhibits difficulty in retrieving the correct phonological form of the word may benefit equally or more from a repetition-based approach that reduces phonological errors during treatment. These predictions were supported by a case comparison study in which a significant advantage of retrieval practice relative to errorless learning was observed for an individual with Stage 1 impairment but not for an individual with Stage 2 impairment (Schuchard & Middleton, *in press*). Similarly, in this study individuals who showed higher percentages of phonological error naming responses prior to training were associated with greater benefit from errorless learning relative to retrieval practice. The percentage of pre-training semantic errors, however, was not a significant predictor. Semantic-lexical difficulty may be better measured with the inclusion of other error types, such as omissions (Chen, Middleton, & Mirman, 2018; G. Dell, Lawler, Harris, & Gordon, 2004), but testing this possibility will require further research.

Limitations of this study include relatively short retention intervals and confounding variables related to the distinction between Stage 1 and Stage 2 items. Regarding the former, we chose 1-day 1-week retention intervals because similar previous studies found that the effects of a single training session can persist 1 week later (Middleton et al., 2015; 2016). For findings from this study to be translated to a clinical setting, further studies with multiple training sessions and longer retention intervals will be required. With regard to confounding variables, words that elicit difficulty at Stage 2 tend to have greater length and lower lexical frequency than those that elicit difficulty at Stage 1 (see Table 2). These differences likely contributed to overall lower accuracy for Stage 2 items relative to Stage 1 items during and after training (see Figure 2). In fact, the analyses clearly showed that word length had a significant effect on participants' performance in the study. However, we can be confident that the observed effects do not entirely rely on differences in psycholinguistic variables of the stimuli because

the two training conditions were closely matched on these variables, and we statistically controlled for word length in the analyses.

In addition, this study's relatively small, homogeneous sample does not allow for the investigation of how impairment at each stage of lexical access may interact with other cognitive-linguistic deficits in their effects on treatment efficacy. Previous research focused on errorless learning in aphasia-included participants with greater variability in their cognitive-linguistic profiles and showed that better performance on tests of cognitive skills like memory and problem solving were associated with better outcomes across errorless and errorful training conditions (Fillingham et al., 2005a, 2005b, 2006). In one of these studies, Fillingham et al. (2006) also noted that the participants who showed an advantage of errorful learning over errorless learning were the participants with the best scores on tests of memory and attention (Fillingham et al., 2006).

However, the focus of this study on a group of participants selected for lexical access impairment resulted in the first identification of a significant linguistic predictor of the relative efficacy of errorless learning versus retrieval practice for naming impairment (i.e., pretraining phonological error rates). In addition, the findings from this study suggest testable predictions for future research with more heterogeneous groups of people with aphasia. If the benefit of errorless learning is primarily derived from increased instances of successful word-to-phonology mapping when repeating words, then severely impaired repetition abilities should reduce or eliminate the advantage of this approach relative to others. Similarly, if the benefit of retrieval practice is primarily derived from increased use of meaning-to-word mapping, then impairment to semantic representations themselves is likely to reduce the retrieval practice advantage.

This study illustrates the benefits of determining how well-specified treatment methods and learning experiences in rehabilitation change the underlying system, as others have proposed (Baddeley, 1993; Gonzalez-Rothi, 2006; Nickels, 2002a; Whyte, 2006). The resulting knowledge can help identify individual characteristics that predict the efficacy of a training method and the advantage of that method relative to others. In this case, the effects of errorless learning and retrieval practice on lexical access suggest that naming error profiles constitute a significant predictor of the relative advantages of these two training methods. Knowledge about a behavioral treatment's "active ingredients" can also be applied to adjust training protocols to be maximally beneficial for each individual. For example, an individual's locus of lexical access impairment may suggest a need for increasing word retrieval opportunities or word repetition opportunities during the course of an established naming treatment protocol. Hence, the growing body of research examining not only what works for people with aphasia but also how treatment works will be essential for addressing the challenge of what works best for whom.

Acknowledgments

This work was supported by the National Institutes of Health under Award Numbers R01 DC015516-01A1 (awarded to Erica Middleton), R03 DC012426 (Erica Middleton), and T32 HD071844 (John Whyte) and by the Albert Einstein Society, Einstein Healthcare Network, Philadelphia, PA, under Research Grant 17-06 (Julia Schuchard). The authors would like to acknowledge Mackenzie Stabile and Taylor Echevarria for assistance in data processing.

References

- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49(4), 415–445.
- Antonucci, S. M., Beeson, P. M., Labiner, D. M., & Rapsack, S. Z. (2008). Lexical retrieval and semantic knowledge in patients with left inferior temporal lobe lesions. *Aphasiology*, 22(3), 281–304.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baddeley, A. (1993). A theory of rehabilitation without a model of learning is a vehicle without an engine: A comment on Caramazza and Hillis. *Neuropsychological Rehabilitation*, 3(3), 235–244.
- Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*, 32(1), 53–68.
- Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology*, 58(4), 667–692.
- Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation*, 17(6), 411–422.
- Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, 4(4), 94–98.
- Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, 23(2), 500–506.
- Brysbaert, M., & New, B. (2009). 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, Instruments, & Computers*, 41, 977–990.
- Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61(2), 153–170.
- Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, 18(1), 49–57.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135(4), 553–571.
- Chen, Q., Middleton, E., & Mirman, D. (2018). Words fail: Lesion-symptom mapping of errors of omission in post-stroke aphasia. *Journal of Neuropsychology*. Advance online publication. <https://doi.org/10.1111/jnp.12148>
- Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology*, 23(11), 1311–1337.
- Dabul, B. (2000). *Apraxia Battery for Adults—Second Edition*. Austin, TX: Pro-Ed.
- Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1372–1384.
- Damian, M. F., Vigliocco, G., & Levelt, W. J. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, 81(3), B77–B86.
- Delaney, P. F., Verkoeijen, P. P. J. L., & Spiguel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *The Psychology of Learning and Motivation*, 53, 63–147.
- Dell, G., Lawler, E., Harris, H., & Gordon, J. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, 21(2), 125–145.
- Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56(4), 490–520.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4), 801–838.
- Edmonds, L. A., Mammino, K., & Ojeda, J. (2014). Effect of Verb Network Strengthening Treatment (VNeST) in persons with aphasia: Extension and replication of previous findings. *American Journal of Speech-Language Pathology*, 23(2), S312–S329.
- Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, 23(3), 402–424.
- Evans, J. J., Wilson, B. A., Schuri, U., Andrade, J., Baddeley, A., Bruna, O., . . . Taussik, I. (2000). A comparison of “errorless” and “trial-and-error” learning methods for teaching individuals with acquired memory deficits. *Neuropsychological Rehabilitation*, 10(1), 67–101.
- Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation*, 13(3), 337–363.
- Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2005a). Treatment of anomia using errorless versus errorful learning: Are frontal executive skills and feedback important? *International Journal of Language & Communication Disorders*, 40(4), 505–523.
- Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (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, J. K., Sage, K., & Lambon Ralph, M. A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation*, 16(2), 129–154.
- Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, 43(2), 182–216.
- Friedman, R. B., Sullivan, K. L., Snider, S. F., Luta, G., & Jones, K. T. (2017). Leveraging the test effect to improve maintenance of the gains achieved through cognitive rehabilitation. *Neuropsychology*, 31(2), 220–228.
- Gainotti, G., Miceli, G., Caltagirone, C., Silveri, M. C., & Masullo, C. (1981). The relationship between type of naming error and semantic-lexical discrimination in aphasic patients. *Cortex*, 17(3), 401–409.

- Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, 102(2), 219–260.
- Gonzalez-Rothi, L. J. (2006). Cognitive rehabilitation: The role of theoretical rationales and respect for the maturational process needed for our evidence. *The Journal of Head Trauma Rehabilitation*, 21(2), 194–197.
- Hillis, A. E. (1989). Efficacy and generalization of treatment for aphasic naming errors. *Archives of Physical Medicine and Rehabilitation*, 70(8), 632–636.
- Hillis, A. E., Rapp, B., Romani, C., & Caramazza, A. (1990). Selective impairment of semantics in lexical processing. *Cognitive Neuropsychology*, 7(3), 191–243.
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, 100(3), 464–482.
- 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.
- Howard, D., Patterson, K., Franklin, S., Orchard-Lisle, V., & Morton, J. (1985). Treatment of word retrieval deficits in aphasia: A comparison of two therapy methods. *Brain*, 108(4), 817–829.
- Hsiao, E. Y., Schwartz, M. F., Schnur, T. T., & Dell, G. S. (2009). Temporal characteristics of semantic perseverations induced by blocked-cyclic picture naming. *Brain and Language*, 108(3), 133–144.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Jang, Y., Wixted, J. T., Pecher, D., Zeelenberg, R., & Huber, D. E. (2012). Decomposing the interaction between retention interval and study/test practice: The role of retrievability. *The Quarterly Journal of Experimental Psychology*, 65(5), 962–975.
- Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, 129(8), 2132–2147.
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation*, 61, 237–284.
- Kertesz, A. (1982). *Western Aphasia Battery*. New York, NY: Grune & Stratton.
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85–97.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33(2), 149–174.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38.
- Martin, N., Schwartz, M. F., & Kohn, F. P. (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.
- McKissock, S., & Ward, J. (2007). Do errors matter? Errorless and errorful learning in anomic picture naming. *Neuropsychological Rehabilitation*, 17(3), 355–373.
- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28(2), 295–314.
- Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation*, 22(2), 138–168.
- Middleton, E. L., Schwartz, M. F., Rawson, K. A., & Garvey, K. (2015). Test-enhanced learning versus errorless learning in aphasia rehabilitation: Testing competing psychological principles. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(4), 1253–1261.
- Middleton, E. L., Schwartz, M. F., Rawson, K. A., 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.
- Nickels, L. (2002a). Improving word finding: Practice makes (closer to) perfect? *Aphasiology*, 16(10–11), 1047–1060.
- Nickels, L. (2002b). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, 16(10–11), 935–979.
- Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, 63(4), 541–559.
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, 114(2), 227–252.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107(3), 460–499.
- Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, 24, 121–133.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.
- Romani, C., & Galluzzi, C. (2005). Effects of syllabic complexity in predicting accuracy of repetition and direction of errors in patients with articulatory and phonological difficulties. *Cognitive Neuropsychology*, 22(7), 817–850.
- Romani, C., Galluzzi, C., & Olson, A. (2011). Phonological-lexical activation: A lexical component or an output buffer? Evidence from aphasic errors. *Cortex*, 47(2), 217–235.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463.
- Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: Retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin & Review*, 21(6), 1516–1523.
- Rowland, C. A., & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory*, 23(3), 403–419.
- Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, 75, 27–44.
- Schuchard, J., & Middleton, E. L. (in press). Word repetition and retrieval practice effects in aphasia: Evidence for use-dependent learning in lexical access. *Cognitive Neuropsychology*.
- Schwartz, M., Dell, G., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, 54(2), 228–264.
- Shallice, T., Rumiat, R. I., & Zadini, A. (2000). The selective impairment of the phonological output buffer. *Cognitive Neuropsychology*, 17(6), 517–546.
- Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., ... Federmeier, K. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, 51(2), 247–250.
- Walker, G. M., & Hickok, G. (2016). Bridging computational approaches to speech production: The semantic-lexical-auditory-

motor model (SLAM). *Psychonomic Bulletin & Review*, 23(2), 339–352.

Walker, G. M., Schwartz, M. F., Kimberg, D. Y., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2011). Support for anterior temporal involvement in semantic error production in aphasia: New evidence from VLSM. *Brain and Language*, 117(3), 110–122.

Whyte, J. (2006). Using treatment theories to refine the designs of brain injury rehabilitation treatment effectiveness studies. *The Journal of Head Trauma Rehabilitation*, 21(2), 99–106.

Wilson, B. A., & Evans, J. J. (1996). Error-free learning in the rehabilitation of people with memory impairments. *The Journal of Head Trauma Rehabilitation*, 11(2), 54–64.

Appendix A

Pretraining Response Types

Items	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Stage 1 items											
Total number	44	44	44	54	44	44	44	44	44	88	
Pretraining 1											
Correct	13.6%	34.1%	36.4%	24.1%	29.5%	20.5%	6.8%	11.4%	31.8%	27.3%	23.5%
Semantic error	56.8%	52.3%	43.2%	61.1%	52.3%	59.1%	61.4%	68.2%	50.0%	42.0%	54.6%
Description	0.0%	2.3%	2.3%	1.9%	2.3%	9.1%	22.7%	20.5%	6.8%	13.6%	8.1%
Omission	20.5%	0.0%	15.9%	1.9%	2.3%	0.0%	6.8%	0.0%	0.0%	3.4%	5.1%
Phonological error	0.0%	9.1%	2.3%	7.4%	0.0%	2.3%	2.3%	0.0%	2.3%	5.7%	3.1%
Other error	9.1%	2.3%	0.0%	3.7%	13.6%	9.1%	0.0%	0.0%	9.1%	8.0%	5.5%
Pretraining 2											
Correct	0.0%	0.0%	0.0%	0.0%	0.0%	2.3%	0.0%	0.0%	0.0%	1.1%	0.3%
Semantic error	77.3%	84.1%	75.0%	88.9%	84.1%	90.9%	56.8%	79.5%	90.9%	67.0%	79.5%
Description	2.3%	6.8%	4.5%	0.0%	0.0%	0.0%	27.3%	15.9%	0.0%	18.2%	7.5%
Omission	6.8%	0.0%	18.2%	0.0%	0.0%	0.0%	6.8%	0.0%	0.0%	1.1%	3.3%
Phonological error	4.5%	0.0%	0.0%	1.9%	0.0%	0.0%	6.8%	0.0%	2.3%	0.0%	1.5%
Other error	9.1%	9.1%	2.3%	9.3%	15.9%	6.8%	2.3%	4.5%	6.8%	12.5%	7.9%
Stage 2 items											
Total number	44	44	44	22	44	44	44	22	22	0	
Pretraining 1											
Correct	2.3%	40.9%	25.0%	27.3%	0.0%	52.3%	2.3%	40.9%	59.1%		27.8%
Phonological error	97.7%	54.5%	43.2%	72.7%	100.0%	43.2%	93.2%	40.9%	31.8%		64.1%
Omission	0.0%	4.5%	29.5%	0.0%	0.0%	4.5%	0.0%	18.2%	9.1%		7.3%
Other error	0.0%	0.0%	2.3%	0.0%	0.0%	0.0%	4.5%	0.0%	0.0%		0.8%
Pretraining 2											
Correct	11.4%	13.6%	18.2%	0.0%	0.0%	6.8%	0.0%	22.7%	13.6%		9.6%
Phonological error	88.6%	86.4%	68.2%	100.0%	100.0%	88.6%	95.5%	63.6%	86.4%		86.4%
Omission	0.0%	0.0%	11.4%	0.0%	0.0%	0.0%	0.0%	13.6%	0.0%		2.8%
Other error	0.0%	0.0%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		0.3%
Filler items											
Total number	0	0	0	12	0	0	0	22	22	0	
Pretraining 1											
Correct				16.7%				0.0%	13.6%		10.1%
Omission				50.0%				95.5%	45.5%		63.6%
Other error				33.3%				4.5%	40.9%		26.3%
Pretraining 2											
Correct				0.0%				0.0%	0.0%		0.0%
Omission				83.3%				95.5%	90.9%		89.9%
Other error				16.7%				4.5%	9.1%		10.1%

Note. P = participant.

Appendix B (p. 1 of 3)Data Analyses

Model 1 (Test 1) and Model 2 (Test 2):**Dependent variable:**

Accuracy (correct or incorrect)

Fixed effects:

Two-way full factorial: Training Type (errorless learning vs. retrieval practice) × Item Type (Stage 1 vs. Stage 2)

Target word length

Random effects:

By-participant random intercepts and random slopes (for item type)

Stata command: `meglm Accuracy TrainingType##ItemType Length, || Participant: ItemType, covariance(unstructured)family(binomial) link(logit)`**Model 1 (Test 1)**

Fixed effect	Coefficient	SEM	<i>p</i>
Intercept	2.18	0.33	< .001
Training Type	0.33	0.20	.09
Item Type	−0.28	0.32	.39
Training Type × Item Type	−0.62	0.31	.046
Length	−0.34	0.04	< .001
Random effect			
Participant			
(Intercept)	1.04	0.80	
(Item Type)	0.54	0.37	

Model 2 (Test 2)

Fixed effect	Coefficient	SEM	<i>p</i>
Intercept	2.31	0.32	< .001
Training Type	0.33	0.20	.09
Item Type	0.15	0.30	.63
Training Type × Item Type	−0.85	0.31	.01
Length	−0.37	0.04	< .001
Random effect			
Participant			
(Intercept)	0.83	0.66	
(Item Type)	0.42	0.29	

Appendix B (p. 2 of 3)

Data Analyses

Model 3 (Test 1 and Test 2):

Dependent variable:

Accuracy (correct or incorrect)

Fixed effects:

Training Type (errorless learning vs. retrieval practice)

Test (Test 1 vs. Test 2)

Target word length

Random effects:

By-participant and by-item random intercepts

Stata command: `meglm Accuracy TrainingType Test Length, || Participant:, covariance(unstructured) || Item:, covariance(unstructured)family(binomial) link(logit)`

3a. Stage 1 items only	Fixed effect	Coefficient	SEM	p
	Intercept	4.17	0.59	< .001
	Training Type	0.51	0.25	.04
	Test	−0.15	0.17	.39
	Length	−0.66	0.08	< .001
	Random effect			
	Participant	0.45	0.28	
	Participant: Item	3.20	0.83	
3b. Stage 2 items only	Fixed effect	Coefficient	SEM	p
	Intercept	2.13	0.52	< .001
	Training Type	−0.55	0.26	.04
	Test	0.27	0.20	.17
	Length	−0.37	0.07	< .001
	Random effect			
	Participant	0.12	0.13	
	Participant: Item	2.22	0.74	

Model 4 (Training, Test 1, and Test 2):

Dependent variable:

Accuracy (correct or incorrect)

Fixed effects:

Two-way full factorial: Training Type (errorless learning vs. retrieval practice) × Time (training versus tests)

Target word length

Random effects:

By-participant and by-item random intercepts

Stata command: `meglm Accuracy TrainingType##Time Length, || Participant:, covariance(unstructured) || Item:, covariance(unstructured)family(binomial) link(logit)`

4a. Stage 1 items only	Fixed effect	Coefficient	SEM	p
	Intercept	6.12	0.49	< .001
	Training Type	−1.14	0.24	< .001
	Time	−2.62	0.20	< .001
	Training Type × Time	1.64	0.25	< .001
	Length	−0.56	0.06	< .001
	Random effect			
	Participant	0.35	0.20	
	Participant: Item	2.85	0.43	
4b. Stage 2 items only	Fixed effect	Coefficient	SEM	p
	Intercept	2.54	0.44	< .001
	Training Type	−0.53	0.24	.03
	Time	−0.56	0.18	< .01
	Training Type × Time	−0.06	0.25	.82
	Length	−0.33	0.05	< .001
	Random effect			
	Participant	0.23	0.16	
	Participant: Item	2.45	0.41	

Appendix B (p. 3 of 3)Data Analyses

Model 5: Error Profiles**Dependent variable:**

Retrieval practice effect (% correct in retrieval practice condition – % correct in errorless learning condition), calculated for each participant at Test 1 and at Test 2

Fixed effects:

Percentage of pretraining naming responses classified as semantic errors

Percentage of pretraining naming responses classified as phonological errors

Test (Test 1 vs. Test 2)

Random effects:

By-participant random intercepts

Stata command: mixed RetrievalPracticeEffect PercentSemanticErrors PercentPhonologicalErrors i.Test, ll Participant:, covariance(identity)

Fixed effect	Coefficient	SEM	p
Intercept	16.88	9.26	.07
Percent semantic errors	0.02	0.52	.97
Percent phonological errors	–1.06	0.34	< .01
Test	–1.36	2.65	.61
Random effect			
Participant			
(Intercept)	25.57	20.86	
(Residual)	35.23	15.75	