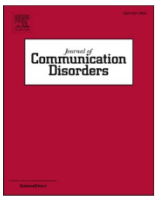
Journal of Communication Disorders 88 (2020) 106048

Contents lists available at ScienceDirect

Journal of Communication Disorders

journal homepage: www.elsevier.com/locate/jcomdis



Mechanisms underlying anomia treatment outcomes

Tijana Simic [a,b,c,d,*], Craig Chambers [d,e], Tali Bitan [a,f], Steven Stewart [d], Devora Goldberg [a], Laura Laird [a], Carol Leonard [a,c,g], Elizabeth Rochon [a,b,c,d]

[a] Department of Speech-Language Pathology, University of Toronto, 160-500 University Avenue, Toronto, ON, M5G 1V7, Canada
[b] Rehabilitation Sciences Institute, University of Toronto, 500 University Avenue, Toronto, ON, M5G 1V7, Canada
[c] Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, 600 Peter Morand Cres., Suite 206, Ottawa, ON, K1G 5Z3, Canada
[d] KITE-Toronto Rehabilitation Institute, University Health Network, 550 University Avenue, Toronto, ON, M5G 2A2, Canada
[e] Department of Psychology, University of Toronto, 3359 Mississauga Rd. N, Mississauga, ON, L5L 1C6, Canada
[f] Psychology Department, IIPDM, University of Haifa, Haifa, 3498838, Israel
[g] School of Rehabilitation Sciences, University of Ottawa, 451 Smyth Rd., Ottawa, ON, K1H 8M5, Canada

ARTICLE INFO

Keywords:
Aphasia
Anomia
Treatment
Responsiveness to cues
Repeated practice
Generalization

ABSTRACT

Treatments for anomia have demonstrated short- and long-term efficacy. However, individual outcomes can be variable, and evidence for treatment generalization is limited. We investigated whether treatment-related measures of access to- and learning of language, namely, a) responsiveness to cues, and b) during-treatment improvements in naming, are good predictors of treatment outcomes. In addition, we investigated mechanisms underlying treatment generalization. Ten adults with chronic, post-stroke aphasia received a phonological treatment for anomia three times a week for five weeks. Naming accuracy of treated and untreated words was assessed pre- and post-treatment and at four- and eight-week follow-ups. Generalization to an untrained naming task, which involved analyses of naming accuracy and speech errors, was also assessed; speech errors were analyzed according to the Interactive Activation (IA) model of word retrieval. Group analyses indicate significant improvements in naming treated compared to untreated words, at all timepoints after therapy. Additional analyses showed significant long-term improvements in naming untreated words. Initial responsiveness to cueing and early improvement emerged as significant predictors of overall pre- to post-treatment improvements in naming *treated* words; naming improvements made early-on in treatment were also predictive of improvements in naming of the *untreated* words at follow-up. Furthermore, our study is the first to demonstrate that generalization after a phonological treatment for anomia may be driven by a strengthening of lexical-phonological connections. This study provides novel insights regarding mechanisms driving anomia treatment outcomes. Understanding such mechanisms is critical to improving existing assessment practices, optimizing treatment selection and building treatment protocols that are more likely to generalize.

1. Introduction

One of the most pervasive and frustrating symptoms of aphasia is anomia, a difficulty finding the words for objects, people, or

* Corresponding author at: Speech-Language Pathology, Rehabilitation Sciences Institute, Rehabilitation Sciences Building, University of Toronto; 160-500 University Avenue, Toronto, ON, M5G 1V7, Canada.
E-mail address: tina.simic@mail.utoronto.ca (T. Simic).

https://doi.org/10.1016/j.jcomdis.2020.106048
Received 22 July 2019; Received in revised form 23 August 2020; Accepted 7 September 2020
Available online 29 September 2020
0021-9924/

actions (Nickels & Best, 1996). Anomia can be caused by impairments in accessing semantics or phonology, or in accessing and assembling phonemes (Laine & Martin, 2006). Along with other communication impairments, word-finding difficulty can have a profound negative impact on an individual's mood, vocational and social activities, independence, and overall quality of life (Gialanella, Bertolinelli, Lissi, & Prometti, 2011; Hilari, 2011; Kauhanen et al., 2000; Ross Graham, Pereira, & Teasell, 2011; Thomas & Lincoln, 2008; Tsouli, Kyritsis, Tsagalis, Virvidaki, & Vemmos, 2009).

Fortunately, many reviews of the literature have demonstrated the overall efficacy of anomia therapy, with various treatment approaches (e.g., semantic, phonological) yielding comparable improvements in naming ability (Best & Nickels, 2000; Doesborgh et al., 2004; Nickels, 2002; Wisenburn & Mahoney, 2009). However, individual outcomes can be variable; for some, treatment does not result in improved language performance (Wisenburn & Mahoney, 2009). Divergent outcomes have been reported even when the same treatment is administered to individuals with similar aphasia profiles (Helm-Estabrooks, 2002; Lazar & Antoniello, 2008; Laganaro, Di Pietro, & Schnider, 2006). In order to obtain critical insights into the underlying mechanisms responsible for variable treatment outcomes, it is important to investigate the predictors of recovery from aphasia and anomia (Laganaro, Di Pietro, & Schnider, 2006).

1.1. Predictors of recovery

Some studies have found that factors such as age, aphasia severity, lesion size and lesion location can be good predictors of aphasia recovery (e.g., Godecke, Rai, Ciccone, Armstrong, Granger & Hankey, 2013; Knoflach et al., 2012; Lazar et al., 2010; Plowman, Hentz, & Ellis, 2012; Price, Seghier, & Leff, 2010; Rijntjes, 2006). However, other studies have shown that language improvements are not predicted by such factors (e.g., Laska, Hellblom, Murray, Kahan, & Von Arbin, 2001; Pedersen, Stig Jørgensen, Nakayama, Raaschou, & Olsen, 1995), or that only a small proportion of the variance in recovery is explained by them (e.g., Lazar, Speizer, Festa, Krakauer, & Marshall, 2008). For example, although initial aphasia severity may be broadly informative, it may not be adequate to predict individual recovery patterns. Indeed, individuals with milder forms of aphasia do not always have positive clinical and/or functional improvements (Fridriksson, Nettles, Davis, Morrow, & Montgomery, 2006; Kiran, 2016; Persad, Wozniak, & Kostopoulos, 2013; Purdy, 2002; Ramsberger, 2005), and those with severe aphasia can make significant language gains (Godecke, Hird, Lalor, Rai, & Phillips, 2012; Persad et al., 2013; Robey, 1998).

Therefore, current evidence on the predictors of recovery from aphasia and anomia is mixed (Lazar & Antoniello, 2008), and the underlying factors that bring about differential changes in language as a result of treatment have yet to be delineated (Byng & Black, 1995; Cicerone et al., 2005, 2011; Nickels, 2002). This may be due, in part, to the fact that it is often unclear what is meant by recovery in prognostic studies. Namely, whether recovery is truly spontaneous, or a result of usual care, which involves some form of treatment and/or management, or otherwise due to a specific therapy approach. Mixed findings regarding the best predictors of aphasia recovery may also be due to the variability in language performance often observed in individuals with aphasia. For instance, even within a single session, an individual's trial-by-trial naming performance for a target stimulus can be inconsistent. This suggests that linguistic content is not altogether lost in aphasia, but *access* to such content may be disrupted and/or variable (Hula & McNeil, 2008; McNeil, Odell, & Tseng, 1991). Thus, treatment techniques aimed at facilitating access to linguistic content may hold important predictive value; in anomia therapy, one such technique is cueing.

1.1.1. Responsiveness to cues

Improved naming as a result of cueing has been widely documented and is considered integral to treatment (Marshall, Neuburger, & Phillips, 1992). Even a single cue can lead to correct naming of a target word both immediately and days later (Best, Herbert, Hickin, & Howard, 2002; Nickels, 2002). The facilitative effect of cueing has been ascribed to priming, whereby presentation of a related stimulus may subsequently boost the accessibility of a target word (Martin, Fink, Laine, & Ayala, 2004; Martin, Fink, & Laine, 2004). Though referred to as priming by some (e.g., Martin, Fink, Laine, Ayala et al., 2004; Martin, Fink, Laine et al., 2004), the effectiveness of a cue to facilitate naming has also been labelled *facilitation*. However, responsiveness to cues will be the term used in this paper.

In line with the Interactive Activation (IA) model of lexical retrieval, the facilitative effect of cues may be attributable to more targeted activation of network nodes (e.g., Meteyard & Bose, 2018). The IA model postulates that word retrieval occurs within a connectionist network, whereby nodes are linked across semantic, lexical and phonological levels of representation (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Schwartz, 2014). Feedforward and feedback connections allow for activation to spread across network nodes, such that stimulation at the phonological level can activate the semantic level, and vice versa. In the *Semantic-Phonological (SP)* IA model (Foygel & Dell, 2000), word retrieval is dependent on connection strength, or weights, between semantic and lexical nodes (i.e., *semantic s-weights*), or lexical and phonological nodes (i.e., *phonological p-weights*). Thus, cueing essentially serves to narrow the search space during lexical retrieval (Meteyard & Bose, 2018).

Phonological cues (e.g., initial phonemes) appear to be more effective than semantic cues (e.g., word-picture matching) in eliciting target words in anomia (Meteyard & Bose, 2018). However, the most effective cueing hierarchies may be those that activate semantic and phonological representations simultaneously (Howard, Hickin, Redmond, Clark, & Best, 2006; Meteyard & Bose, 2018). Importantly, there is evidence to suggest that responsiveness to cues may be a good predictor of treatment outcomes (Martin, Fink, Laine et al., 2004; Martin, Fink, Laine, Ayala et al., 2004; Hickin, Best, Herbert, Howard, & Osborne, 2002; Nardo, Holland, Leff, Price, & Crinion, 2017). For example, responsiveness to phonological cueing (i.e., word initial spoken and written letter cues) has been associated with short-term response to treatment (Hickin et al., 2002), and neuroimaging evidence supports this association (Nardo et al., 2017). In addition to boosting access to the mental lexicon (i.e., with cueing), treatment for anomia aims to stimulate and support learning (or relearning) of lexical items through repeated practice.

1.1.2. Learning and repeated practice

Studies have demonstrated that individuals with aphasia *can* learn novel vocabulary (Freedman & Martin, 2001; Breitenstein, Kamping, Jansen, Schomacher, & Knecht, 2004; Kelly & Armstrong, 2008; Marshall et al., 1992; Tuomiranta, Grönholm-Nyman, Rautakoski, Laine, & Martin, 2011). However, learning ability is generally impaired when compared to healthy controls (Gupta, Martin, Abbs, Schwartz, & Lipinski, 2006), and can vary across individuals (Dignam, Copland, Rawlings, O'Brien, Burfein & Rodriguez, 2016; Grossman & Carey, 1987; Tuomiranta, Rautakoski, Rinne, Martin, & Laine, 2012; Vallila-Rohter & Kiran, 2013). Importantly, one study found novel word learning ability to be a significant predictor of anomia treatment outcomes (Dignam et al., 2016), supporting the hypothesis that the mechanisms involved in novel word learning overlap with those involved in the relearning of words in anomia treatment. Therefore, (re)learning mechanisms may play a critical role in treatment success (Ferguson, 1999; Hopper & Holland, 2005; Lambon-Ralph & Fillingham, 2007).

A fundamental factor driving learning and memory consolidation in virtually all cognitive domains, and in rehabilitation, is repeated practice (Hauptmann & Karni, 2002; Hauptmann, Reinhart, Brandt, & Karni, 2005; Kleim & Jones, 2008; Raymer et al., 2008). Repeated practice is one of the principles of learning-induced neuroplasticity (Kleim & Jones, 2008), and has emerged as an essential element of successful novel word learning (Breitenstein et al., 2004) and language relearning (Raymer et al., 2008) in aphasia. For example, in anomia treatment, frequent naming attempts lead to greater improvements, even without feedback or explicit training (e.g., Fillingham, Sage, & Lambon Ralph, 2005; Off, Griffin, Spencer, & Rogers, 2016; Heath et al., 2015), and a greater number of sessions and/or exposures to stimuli leads to more enduring treatment effects (Friedman, Sullivan, Snider, Luta, & Jones, 2017; Godecke et al., 2013; Heath et al., 2015). Effortful retrieval during repeated practice further enhances the long-term retention of treated items (Middleton, Schwartz, Rawson, & Garvey, 2015; Roediger & Butler, 2011). Therefore, changes in naming performance as a result of repeated practice may represent treatment-induced learning; tracking such changes across individuals may thus hold important predictive value.

1.2. Generalization

Despite being a central goal of intervention, the evidence for treatment generalization is also inconsistent (Webster, Whitworth, & Morris, 2015). In general, treatment approaches that teach strategy use are more likely to transfer. Although the supporting mechanisms are poorly understood, semantic and phonological approaches for anomia have also demonstrated some transfer of treatment effects. In fact, individuals with impairments at the semantic level are less likely to benefit from anomia therapy in general (Best et al., 2013; Dignam et al., 2016; Martin, Fink, Renvall, & Laine, 2006; Martin, Fink, Laine et al., 2004). Regardless of the approach, generalization appears to be more common in individuals with impaired phonological encoding, but relatively intact semantics (Webster et al., 2015). However, identifying the locus of impairment in anomia is not straightforward. Individuals often present with deficits at multiple levels of word retrieval, and studies vary in their criteria for determining whether deficits are primarily semantic or primarily phonological.

In many cases, it is unclear whether generalization findings are truly an effect of treatment, or due to individual variability in responding, or multiple exposures to untreated stimuli. In addition, according to Webster et al. (2015), studies commonly consider post-treatment improvement in *any* language measure outside of the treatment task (e.g., various word-, sentence- or discourse-level assessments) to be generalization; instead, theoretically-driven measures may provide more valuable insights (Nickels & Best, 1996; Nickels, 2002; Webster et al., 2015; Wisenburn & Mahoney, 2009). Therefore, in order to gain a better understanding of generalization in anomia therapy, it may be necessary to measure it according to theoretical models of lexical retrieval.

As described above, the IA model postulates that word retrieval depends on spreading activation across network nodes. Disturbances at any level (i.e., semantic, lexical, phonological) can introduce noise into the network and impact lexical output, resulting in speech errors. Thus, speech errors, which are a common manifestation of anomia, can be important indicators of underlying word-retrieval mechanisms (Schwartz, Dell, Martin, Gahl, & Sobel, 2006; Schwartz, 2014). In contrast to broader (and more commonly used) measures, such as overall naming accuracy, changes in speech error production pre- to post-treatment may elucidate some of the factors driving treatment-related improvements (Abel, Willmes, & Huber, 2007; Jokel, Rochon, & Leonard, 2004; Kendall, Pompon, Brookshire, Minkina, & Bislick, 2013; Minkina et al., 2016), and may provide more nuanced insights into the mechanisms underlying generalization.

1.3. Summary and objectives

Although important developments have been made in the understanding and treatment of anomia (Laine & Martin, 2006), individual treatment outcomes can be variable; the critical factors driving this variability remain somewhat unclear. As a result, it is difficult to predict whether an individual will improve and/or maintain improvements after therapy. A review of the existing anomia treatment literature suggests that techniques which stimulate and support access to- and (re)learning of lexical items may be important indicators of treatment success. The mechanisms underlying treatment generalization are also poorly understood, and studies tend to lack theoretically driven measures of generalization (Webster et al., 2015).

Therefore, the goals of this study are twofold. First, in addition to measuring treatment outcomes at both the individual and group levels using a well-established treatment approach, we aim to investigate whether treatment effects generalize to an untrained naming task. To obtain more nuanced insights into the mechanisms underlying treatment generalization, we aim to investigate not only changes in naming accuracy but also speech error production, which will be quantified using the parameters of the SP computational model of lexical retrieval. Second, we aim to investigate whether two treatment-related measures, namely a) *responsiveness to cues* (i.e.,

Table 1
Participant characteristics .

Participant[1]	Sex	Age	Handedness[2]	Education (years)	MPO	Etiology	Stroke Type	Lesion Location	WAB AQ	Aphasia Severity[3]	Aphasia Type
P1	M	58	L	16	13	L MCA CVA	ischemic	fronto-parietal, temporal, insula	61.80	Moderate	Broca's
P2	M	35	L	17	8	L MCA CVA	ischemic + CP hemorrhage	large portion of MCA territory, basal ganglia	58.10	Moderate	Broca's
P3	M	75	R	20	12	L MCA CVA	ischemic	frontal, posterior parietal, insula	77.20	Mild	Anomic
P4	F	35	L	15	18	L MCA CVA	subarachnoid hemorrhage	sylvian fissure, temporal sulci, intrahemispheric fissure	39.60	Severe	Broca's
P6	M	56	L	17	12	L MCA CVA	ischemic	mass effect in frontal horn of lateral ventricle	64.90	Moderate	Broca's
P7	M	64	R	18	6	L MCA CVA	ischemic	temporo-parietal	78.60	Mild	Conduction
P8	M	55	L	14	10	L CVA	hemorrhagic	basal ganglia	66.80	Moderate	Broca's
P9	M	42	L	19	9	L MCA & ACA CVA	Ischemic + hemorrhagic transformation	frontal, temporal, insula	62.10	Moderate	Broca's
P10	M	79	R	13	74	L CVA	hemorrhagic	frontal	84.10	Mild	Anomic
P11	M	56	L	18	19	L MCA CVA	ischemic + CP hemorrhage	frontal, caudate, basal ganglia, internal and external capsules + occipital involvement	85.20	Mild	Anomic
Mean		**55.50**		**16.70**	**18.10**				**67.84**		
Median		**56.00**		**17.00**	**12.00**				**65.85**		
SD		*15.04*		*2.21*	*20.06*				*13.91*		

[1]One participant (P5) was screened but deemed ineligible for treatment; naming impairment was too mild.
[2]Refers to currently dominant hand; all participants were premorbidly right-hand dominant.
[3]Based on the WAB AQ, according to the WAB-R manual (Kertesz, 2006). N.B. All participants passed visuoperceptual and hearing screening at the time of enrollment. ACA - Anterior Cerebral Artery; CP - Cortical Petechial; CVA - Cerebrovascular Accident; L1 - First Language; L - Left; MCA - Middle Cerebral Artery; MPO - Months Post-Onset; R - Right; WAB AQ - Western Aphasia Battery Aphasia Quotient.

the effectiveness of cues to stimulate naming), and b) *during-treatment improvement* as a result of repeated practice, are predictive of treatment outcomes over time.

2. Methods

2.1. Participants

Participants were recruited from multiple referral sites in the Greater Toronto Area. Ethical approval for the study was granted by Research Ethics Boards at the University of Toronto, the Aphasia Institute and the March of Dimes Aphasia and Communication Disabilities Program. Participants were included in the study if they a) had a history of a single, left-hemisphere stroke, b) were in the chronic stage of stroke recovery (i.e., at least six months post-onset), c) were adults, 18 years of age and older, d) were pre-morbidly right-handed, and e) were educated in English, or primarily English-speaking. The primary inclusion criterion was the presence of moderate anomia, defined as 10–75 % naming accuracy on the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 2001). Participants were excluded from the study if they did not pass visual-perceptual screening (i.e., subtest 7 of the Birmingham Object Recognition Battery; Riddoch & Humphreys, 1993), and audiometric screening (i.e., 40 dB at 0.5, 1 and 2 kHz in at least one ear). In addition, individuals with a known history of drug or alcohol abuse, major psychiatric or neurological disorders, severe motor speech disorders (screened using *Tasks for Assessing Motor Speech Programming Capacity*; Duffy, 1995; Wertz, LaPointe, & Rosenbeck, 1984), or those receiving concurrent speech-language therapy were excluded from the study.

Ten participants (10 % female, 90 % male) were enrolled in the study, between 35 and 79 years of age (M = 55.50; SD = 15.04) and ranging from six to 74 months post-onset of stroke (M = 18.10; median = 12.00; SD = 20.06). According to the Revised Western Aphasia Battery (WAB-R; Kertesz, 2006), participants in our sample had Aphasia Quotients (AQs) between 39.60–85.20 (M = 67.84; SD = 13.91). Aphasia severity, based on the WAB-R AQ, ranged from mild (n = 4), moderate (n = 5), to severe (n = 1). Six participants presented with Broca's aphasia, three with Anomic aphasia, and one with Conduction aphasia (see Table 1). Data from the same sample of participants is reported in a recently published study (i.e., Simic et al., 2019), the focus of which was performance on a battery of executive control tasks. In contrast, the present study focuses on treatment outcomes, and participants' responsiveness to treatment.

2.2. Pre-treatment assessment

Participants' individual aphasia profiles were characterized using a large battery of language assessment measures administered prior to treatment, the raw scores of which are presented in Table 2. Language assessments which elicited verbal output from the participant were audio-recorded and scored offline by an independent rater blind to the purposes of the study.

To assess semantic processing, two versions of the Pyramids and Palm Trees Test (PPTT; Howard & Patterson, 1992) were administered: the three-picture and the one-spoken-word-two-picture versions. Subtests from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992) battery were also administered, namely, spoken word-picture matching (PALPA 47), written word-picture matching (PALPA 48), auditory synonym judgment (PALPA 49) and written synonym judgment (PALPA 50).

Additional subtests of the PALPA were administered to assess phonological input processing, including discrimination of nonword minimal pairs (PALPA 1), auditory lexical decision of real- and non-words (PALPA 5), rhyme judgment (PALPA 14), and auditory rhyme judgement (PALPA 15). The picture homophone matching subtest of the Psycholinguistic Assessment of Language (PAL; Caplan, 1993) was also administered. Repetition abilities were assessed with the syllable length (PALPA 7), auditory nonword (PALPA 8), and auditory word (PALPA 9) repetition subtests of the PALPA, whereas oral reading of single words was assessed with the PALPA 31. Finally, the BNT was used to assess naming accuracy, with scores ranging from 7/60 (12 %) to 34/60 (57 %; M = 18.60/60; SD = 8.60).

2.3. Treatment

Participants were treated using the Phonological Components Analysis (PCA) protocol (Leonard, Rochon, & Laird, 2008), which uses guided self-cueing to stimulate word-retrieval and production in individuals with aphasia (Madden, Robinson, & Kendall, 2017). PCA has been shown to be efficacious, significantly improving naming accuracy immediately following therapy and at one- and two-month follow-up periods (Leonard et al., 2008, 2015; Van Hees, Angwin, McMahon, & Copland, 2013), and demonstrating corresponding neural changes (Marcotte et al., 2018; Rochon et al., 2010; Van Hees, McMahon, Angwin, de Zubicaray, & Copland, 2014).

2.3.1. Baseline testing

Baseline naming performance was assessed on three separate occasions prior to treatment; participants were asked to name a battery of 198 coloured photographs of nouns at each baseline assessment. Presentation order of the baseline stimuli was randomized at each administration. On average, the three baselines were completed within 19.50 days (SD = 23.97; Median = 14.50 days). As indicated by the large standard deviation, one participant (P2) had a large lag between baselines one and two, due to health complications; with P2 removed, the average time to complete the three baselines was 12.00 days (SD = 3.71).

2.3.2. Stimulus Selection

Stimuli named incorrectly on at least two (out of three) baseline testing sessions were entered into a pool of potential words to be

Table 2
Pre-treatment language assessment scores across individuals .

			Participant (raw scores)											
Area	**Assessment Measure**	**Total**	**P1**	**P2**	**P3**	**P4**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**Mean**	*SD*
Aphasia Severity	Western Aphasia Battery - Aphasia Quotient	/100	61.80	58.10	77.20	39.60	64.90	78.60	66.80	62.10	84.10	85.20	**67.84**	*13.91*
Semantic Processing	PPTT - 3 Pictures	/52	46	44	40	47	48	50	44	51	48	46	**46.40**	*3.20*
	PPTT - 1 Spoken word/2 pictures	/52	47	43	31	44	48	46	39	46	42	47	**43.30**	*5.12*
	PALPA 47 Spoken Word-Picture Match	/39	38	35	30	35	36	38	33	37	38	36	**35.60**	*2.55*
	PALPA 48 Written Word-Picture Match	/39	39	36	29	30	35	39	34	38	37	37	**35.40**	*3.50*
	PALPA 49 Auditory Synonym Judgment	/60	55	29	39	49	55	24	52	46	38	*	**43.00**	*11.25*
	PALPA 50 Written Synonym Judgment	/60	49	53	48	36	33	57	56	53	47	56	**48.80**	*8.32*
Phonological Processing	PALPA 1 Nonword Minimal Pairs	/65	62	60	46	56	55	51	65	63	52	64	**57.40**	*6.40*
	PALPA 5 Auditory Lexical Decision - Real Words	/80	78	74	70	73	77	55	79	74	72	79	**73.10**	*7.06*
	PALPA 5 Auditory Lexical Decision - Nonwords	/80	73	76	77	69	72	59	67	52	55	73	**67.30**	*8.91*
	PALPA 7 Syllable Length Repetition	/24	21	23	18	16	21	14	24	23	19	21	**20.00**	*3.23*
	PALPA 8 Auditory Nonword Repetition	/30	24	23	12	13	16	0	19	19	9	24	**15.90**	*7.64*
	PALPA 9 Auditory Word Repetition	/80	69	78	59	54	61	28	75	75	56	74	**62.90**	*15.07*
	PALPA 14 Rhyming Judgment	/38	26	22	13	23	21	21	3	12	16	22	**17.90**	*6.90*
	PALPA 15 Auditory Rhyme Judgment	/58	52	57	34	41	57	54	55	53	40	54	**49.70**	*8.19*
	PALPA 31 Oral Reading	/80	54	43	76	24	38	57	67	40	71	73	**54.30**	*17.59*
	PAL 08 - Picture Homophone Matching	/32	23	21	18	16	20	29	18	21	26	23	**21.50**	*3.92*
Naming	Boston Naming Test (BNT)	/60	14	29	12	7	16	34	19	10	21	24	**18.60**	*8.57*

*Test not administered due to examiner error. PAL - Psycholinguistic Assessment of Language (Caplan, 1993); PALPA - Psycholinguistic Assessments of Language Processing in Aphasia (Kay et al., 1992); PPTT - Pyramids and Palm Trees Test (Howard & Patterson, 1992). *Adapted from "The role of executive control in post-stroke aphasia treatment" by* Simic et al., 2019, *Neuropsychological Rehabilitation, p.9. Copyright (2019) by Taylor & Francis Ltd. Adapted with permission.*

targeted in therapy. From this pool of words, two lists of approximately 30 words each[1] were created per participant and matched as closely as possible in terms of semantic category, word frequency and syllable length. One list was targeted in treatment, while the other served as a within-participant matched untreated (i.e., control) list. The treated words were further subdivided into two lists of roughly 15 words each.

2.3.3. PCA protocol

A target picture (e.g., *sweater*) is presented in the center of a whiteboard and the participant is given their first naming attempt (NA1). Regardless of their ability to name the picture, the participant is asked to identify five phonological components associated with the target word, in the following order: a rhyme word (e.g., *letter*), the first sound (i.e., */s/*), another word with the same first sound (e.g., *seat*), the last sound (i.e., */er/*), and the number of syllables (i.e., *2*). If the participant is unable to generate a component, they are provided with the correct answer. The therapist delivers specific verbal feedback on the accuracy of responses and writes each correct phonological component in turn on the whiteboard. The participant is subsequently given a second naming attempt (NA2); whether correct or incorrect, the therapist provides feedback and says the target word out loud for the first time during the trial. The components are reviewed, and the participant is given a third, and final naming attempt (NA3). Again, regardless of the participant's naming accuracy, the therapist provides specific feedback and says the target word a second and final time. Thus, on any given trial, the participant has three naming attempts, and hears the target word spoken by the therapist a total of two times. Participant responses are tracked in real-time (see also Leonard et al., 2008). Treatment was administered and/or supervised by a licensed speech-language pathologist.

2.3.4. Treatment schedule

Target words were divided into two lists of roughly 15 words each; both lists were treated during one round of treatment, consisting of two same-day sessions, with a five-minute break in between. Thus, each target word was trained once per round, and word order was randomized. PCA was administered three days a week for five weeks, and participants received a total of 15 rounds (i.e., 30 sessions) of treatment. An experienced research associate and/or licensed speech-language pathologist administered the therapy in a quiet room, either at the participants' homes or at the University of Toronto.[2] Participants received 21.05 h of treatment on average (SD = 3.35), and mean session length was 42.09 min (SD = 10.77). Session length and total time spent in treatment varied according to the pace of the participant.

2.3.5. Treatment fidelity

A random selection of six treatment sessions (i.e., 20 %) was video-recorded for each participant; two independent raters watched four trials of each recorded session, and answered a list of yes/no questions to evaluate whether each step of the treatment protocol (e.g., stimulus presentation, number of naming attempts, therapist feedback, phonological components) was administered accurately and reliably. Point-to-point inter-rater reliability for the assessment of treatment fidelity was 97.33 %. On average, the two raters determined that 98.54 % of the steps in the treatment protocol were administered correctly.

2.3.6. Outcome measures

2.3.6.1. Treated and untreated words. The primary outcome measure was naming accuracy on a set of coloured photographs of nouns, corresponding to each participant's treated and matched, untreated word lists. Outcomes were measured at four time points: pre- and post-treatment and at four- and eight-week follow-ups; presentation order of the stimuli was randomized at each time-point. The average time elapsed between administration of the third baseline and pre-treatment outcome assessment was 13.60 days (SD = 8.86), during which the language assessment battery described above was administered. Post-treatment assessment never occurred on the same day as therapy but was conducted within one week after the completion of treatment (M = 3.10 days, SD = 2.28). For individual outcome analyses, we extracted the naming accuracy of the selected treated and untreated words from each baseline testing session.

2.3.6.2. Response generalization. The secondary outcome measure was naming accuracy and speech error production on the Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996), which is comprised of 175 black-and-white line drawings of nouns. The PNT was administered pre- and post-treatment. PNT accuracy and error data were quantified using WebFit, an online tool (http://cogsci.uci.edu/~alns/webfit.html) that allows for computational modeling of speech errors in aphasia (Dell, Lawler, Harris, & Gordon, 2004; Walker & Hickok, n.d.). The number of correct naming responses, and semantic, formal, mixed, unrelated and nonword errors were entered into WebFit, which generated pre- and post-treatment *semantic s-weights* and *phonological p-weights* for each individual (corresponding to the *SP* model). Higher weight values indicate stronger (i.e. better) connections among levels of representation, while lower weight values indicate weaker connections (Dell et al., 2004; Foygel & Dell, 2000). For example, semantic

[1] Stimulus selection was based on baseline naming performance. Individuals who performed well at baseline had a smaller pool of potential treatment words to choose from. Therefore, although we aimed for each participant to have two lists of 30 words each, the lists were slightly shorter in some cases. The total number of words per list, for each individual, is indicated in Table 3.

[2] All participants received a total of 30 sessions of therapy, however three participants (P1, P2 & P3) received the PCA treatment protocol on alternate schedules: P2 and P3 (3 sessions per day, 4 days a week for 2.5 weeks); P1 (1 session per day, 3 days a week for 10 weeks). Importantly, all participants received 15 rounds of the PCA protocol per target word. No differences were noted in average session duration, total treatment hours, or in treatment performance as a function of treatment schedule.

connection weights decrease when an individual produces many unrelated speech errors, whereas phonological connection weights decrease with a greater number of phonologically related nonword errors. In this way, naming error profiles can be quantified using a single value to represent semantic and phonological connection weights, respectively. Impairments can occur at one or both levels. Thus, outcome measures derived from the PNT included naming accuracy, and *s*- and *p*- connection weights pre- and post-treatment.

2.3.6.3. Scoring. Primary and secondary outcome assessment sessions were audio recorded, transcribed and scored offline by an independent rater blind to assessment time, word list (where applicable), and study objectives. Naming of a target word was scored as either correct (1) or incorrect (0). Self-corrections were scored as correct, whereas paraphasias of all types were scored as incorrect, as were naming attempts uttered after a pause greater than ten seconds. Scoring and error coding followed the detailed procedures described by Roach et al. (1996), with corresponding materials obtained online (https://mrri.org/philadelphia-naming-test/); error analyses are only reported for the PNT data in the present study.

2.3.6.4. Reliability. A random subset of 20 % of the outcome data (i.e., both treated and untreated words, and the PNT) was scored by a second independent rater, blind to assessment time (e.g., pre-, post-treatment) and word list (i.e., treated or untreated), where applicable. First, interrater reliability was evaluated for the transcription of naming responses and was found to be excellent. Point-to-point agreement among the raters was 96.67 % for the transcription of the treated and untreated words, and 95.00 % for the transcription of the PNT. Subsequently, interrater reliability was evaluated for scoring of naming accuracy and coding of naming errors. Excellent inter-rater reliability was also found, with point-to-point agreement among the raters at 90.34 % for scoring and coding of the treated and untreated words (Fleiss' $\kappa = .92$; 95 % CI, .918–.921, $p < .01$), and 92.57 % for scoring and coding of the PNT data (Fleiss' $\kappa = .89$; 95 % CI, .892–.894, $p < .01$).

2.4. Statistical analysis

2.4.1. Individual outcomes

Individual participant naming accuracy (i.e., correct/incorrect) was analyzed by-item, using the WEighted STatistics (WEST) method outlined by Howard, Best and Nickels (2015). This method overcomes problems of autocorrelation inherent in repeated measures designs; a single weighted score representing repeated measurements of an item is obtained and analyzed using a one-sample *t*-test. Weighting factors are based on the null hypothesis of no change; namely, that the rate of change in the baseline phase does not differ from the rate of change in the treatment phase. As such, weighting factors always sum to zero, and are derived based on the number of baseline and follow-up tests in the study (the procedure for calculating weighting factors is detailed in Howard, Best, & Nickels, 2015). Separate analyses were conducted for each of the post-therapy assessment periods (i.e., immediately post-treatment, four- and eight-week follow-ups), thus weighting factors were derived based on three pre-treatment baselines and one post-test for each analysis.

2.4.1.1. WEST-Trend. The WEST-Trend method ensures that existing linear trends in the data are accounted for; as such, a significant WEST-Trend result indicates significant improvement, over and above existing trends in the data (Howard et al., 2015). By-item naming scores for the three baselines and each post-test were multiplied by factors of -3, -1, 1, and 3, respectively. WEST-Trend analyses were then conducted on the treated and untreated sets of words. A total of six WEST-Trend analyses were conducted per participant (i.e., two analyses each at post, four- and eight-week timepoints).

2.4.1.2. WEST-ROC. The WEST-ROC method confirms that the rate of change (ROC) post-treatment is significantly greater than the null rate of change expected at baseline, while also accounting for existing trends in the data (Howard et al., 2015). By-item naming scores for the three baselines and each post-test of interest were multiplied by factors of 2, -1, -4, and 3, respectively. Subsequently, WEST-ROC analyses were conducted on the treated and untreated sets of words. As above, a total of six analyses were conducted per participant.

Overall, a total of twelve one-sample t-tests (one-tailed) were conducted per participant. Using the Holm-Bonferroni procedure, alpha was initially set at $0.05/12 = .004$, and adjusted accordingly for all subsequent comparisons.

2.4.2. Group outcomes

Although not the primary aim of our study, we first wanted to establish treatment efficacy. As such, group data were analyzed by-item, using a mixed effects logistic regression, or generalized linear mixed model (GLMM). Fixed factors were assessment time (four levels: pre, post, 4-week and 8-week follow-ups), condition (two levels: treated and untreated), and the time by condition interaction. Participant, and items within participants were entered as random factors; only random intercepts were specified, and a variance components covariance structure was used. Assessment time was dummy coded, whereby *pre-treatment assessment* was the reference category. Alpha was set at 0.05 for all main effects.

2.4.2.1. Generalization. We used paired samples t-tests to compare pre- to post-treatment changes in naming accuracy, *semantic s*-weights, and *phonological p*-weights on the PNT (i.e., an untrained naming task).

2.4.3. Predictors of treatment outcomes

We investigated the predictive value of a) *initial responsiveness to cues* and b) *during-treatment improvement* as a result of repeated

practice on PCA treatment outcomes using GLMMs, where the dependent variable was naming accuracy (i.e., a score of 0 or 1). Given the larger variability in naming performance for treated words, we analyzed the treated and untreated sets of words separately. First, we conducted *time-only* models, where assessment time was entered as the sole predictor. Next, the predictors of interest were transformed to log-odds units in order to match the dependent variable. Predictors were found to be collinear (see Results below), thus separate analyses were conducted for each predictor of interest. As above, a variance components covariance structure was used, only random intercepts were specified, and assessment time was dummy coded. However, in these analyses, *post-treatment assessment* was used as the reference category, in order to distinguish the treatment and maintenance phases. Finally, we conducted likelihood ratio (LR) tests to evaluate whether addition of a predictor improved model fit, as compared to the *time-only* model. Alpha was set at 0.05.

2.4.3.1. Responsiveness to cues. To obtain a measure of responsiveness to cues, we calculated changes in naming accuracy as a result of cueing. In the context of PCA therapy, cueing entails the provision (or self-generation) of phonological components related to the target word. During PCA, participants receive one naming attempt prior to cueing (NA1), and one immediately after (NA2), without hearing the target word in the interim. Thus, responsiveness to cues was calculated as the difference in percent naming accuracy pre- (NA1) to post- (NA2) cueing. Based on previous research (e.g., Hickin et al., 2002), and in order to investigate the predictive value of participants' *initial* responsiveness to PCA cueing, we only used data from the first round of treatment (i.e., the first two sessions). Thus, the GLMMs evaluating this predictor specified assessment time, responsiveness to cues, and their interaction as fixed effects; random effects were participants and items within participants. Two models were fit for this predictor (i.e., for the treated and untreated words, respectively).

2.4.3.2. During-treatment improvement. In the present study, all participants received the same amount of practice (i.e., 15 repetitions of the PCA protocol for each target word). Thus, in contrast to previous studies that assessed how the *number* of task and/or stimulus repetitions impacted overall treatment outcome, we investigated *patterns of change*, or individual improvement as a result of repeated practice with each round of treatment. To do so, we tracked round-by-round accuracy of the first naming attempt (NA1) for each treated word. Subsequently, we calculated round-by-round changes in naming accuracy (i.e., difference scores from round one to two, two to three, etc.) for each participant, and identified notable patterns in the data. Based on these patterns, early and late improvements in treatment (described in Results), served as predictor variables in separate GLMMs, which specified assessment time, early (or late) improvement, and their interaction as fixed effects. As above, random effects were participants, and items within participants. Thus, we ran two models for early improvement (i.e., for the treated and untreated words), and two for late improvement. GLMM analyses were conducted in *Stata 16* StataCorp (2019); all other analyses were conducted in *IBM SPSS Statistics Software Version 25*.

3. Results

3.1. Individual treatment outcomes

Related samples Cochran's Q tests indicate that all but two participants (P1and P9) demonstrated stable baseline performance across the entire set of 198 baseline stimuli (range of $\chi^2(2) = .83$–4.57, ns). P9 showed increasing performance across baselines ($\chi^2(2) = 26.99, p < .01$), whereas P1 had significantly lower performance at baseline three compared to baselines one and two ($\chi^2(2) = 38.20, p < .01$). Baseline naming performance on the extracted treated and untreated items was stable for all but one participant (range of $\chi^2(2) =$

Table 3
Raw naming accuracy scores for treated and untreated words at baseline and at all assessment time points. Significant outcomes from individual WEST-Trend and WEST-ROC analyses are also shown.

		P1	P2	P3	P4	P6	P7	P8	P9	P10	P11	Mean	*SD*
Treated	**B1**	8	4	5	4	5	6	3	1	4	5	4.50	*1.84*
	B2	4	6	3	2	4	5	4	4	1	5	3.80	*1.48*
	B3	1	8	2	4	3	4	2	8	2	7	4.10	*2.64*
	PRE[1]	14	14	7	5	8	9	5	6	7	9	8.40	*3.27*
	POST	**19**	**22**	**17**	**15**	**15**	**28**	**27**	12	**17**	**18**	15.00	*3.00*
	4 W FU	**20**	**22**	**12**	9	**13**	**28**	**22**	12	**12**	16	12.33	*3.51*
	8 W FU	**21**	**22**	**10**	5	**19**	**26**	**21**	11	**11**	12	9.33	*3.79*
	Total words	/30	/30	/30	/30	/30	/29	/30	/30	/24	/29		
Untreated	**B1**	3	3	4	0	4	6	3	4	3	7	3.70	*1.89*
	B2	9	6	6	2	5	4	5	3	2	2	4.40	*2.27*
	B3	1	7	4	7	4	5	4	5	3	6	4.60	*1.84*
	PRE[1]	8	11	8	5	13	10	5	7	6	11	8.40	*2.76*
	POST	8	14	10	3	11	13	9	10	6	7	8.75	*3.37*
	4 W FU	**11**	14	14	7	9	**18**	**19**	6	7	11	9.00	*3.03*
	8 W FU	**11**	15	13	5	10	**15**	**17**	8	8	9	9.71	*3.35*
	Total words	/30	/30	/30	/30	/30	/29	/30	/30	/23	/28		

Underlined values indicate a significant WEST-Trend result (i.e., improvements in naming accuracy beyond existing linear trends in the data). Bolded values indicate a significant WEST-ROC result (i.e., greater rate of change during treatment, compared to baseline). See Appendix A for detailed WEST statistics.

[1]Pre-treatment outcome assessment scores were used in group GLMM analyses, whereas baseline scores were used in individual WEST analyses. Note: B - Baseline; 4 W FU - Four-week follow-up; 8 W FU - Eight-week follow-up.

.32–5.16, ns). As above, P1 demonstrated significantly decreased naming performance at baseline three ($\chi^2(2) = 7.92$, $p < .05$). Raw naming scores and statistically significant findings are presented in Table 3. Howard et al. (2015) indicate that steady baseline performance is not a requirement for the WEST analyses, given that existing linear trends in the data are accounted for by the weighting factors (described above). As such, despite unsteady baseline performance, P1 and P9 were both included in individual analyses of the data.

3.1.1. Treated words

According to WEST-Trend analyses, significant improvements in naming accuracy of the treated words, over and above existing linear trends in the data, were found for nine participants from baseline to post-treatment, and for seven participants from baseline to four- and eight-week follow-ups. In addition, WEST-ROC analyses reveal that eight participants demonstrated significantly greater rates of change post-treatment compared to baseline, and seven participants demonstrated significantly greater rates of change at four- and eight-week follow-ups.

3.1.2. Untreated words

WEST-Trend analyses indicate significant improvements in naming accuracy of the untreated words for one participant post-treatment, and for five participants at four- and eight-week follow-ups. According to WEST-ROC analyses, no participants showed a significant rate of change in naming the untreated words immediately post-treatment, however three individuals demonstrated significantly improved naming of the untreated words at four- and eight-week follow-ups. Detailed individual participant statistics are presented in Appendix A.

Although conservative, Howard et al., 2015 recommend that the effect of treatment be considered significant when both the WEST-Trend and WEST-ROC analyses yield significant results. Thus, a significant treatment effect for naming of the *treated* words was seen for eight participants immediately post-treatment (P1, P2, P3, P4, P7, P8, P10 and P11), for five participants at four-week follow-up (P1, P2, P7, P8 and P10), and for six participants at eight-week follow-up (P1, P2, P6, P7, P8 and P10). Significant improvements in naming of the untreated words were seen for two participants (P7 and P8) at four- and eight-week follow-ups.

3.2. Group treatment outcomes

Results from the GLMM analysis are presented in Table 4. Of note is the significant interaction of assessment time by condition, which indicates that compared to pre-treatment, naming accuracy was significantly better for the treated words than the untreated words at all timepoints following therapy (see Fig. 1).

3.2.1. Generalization

PNT raw scores and connection weights are shown in Table 5 (omission error data are presented in Appendix B). A paired-samples *t*-test comparing mean PNT naming accuracy pre- (M = 81.67/175, SD = 32.27) to post-treatment (M = 93.44/175, SD = 24.91) revealed a significant improvement (t(8) = -2.90, $p = .02$), suggesting that the effects of treatment generalized to naming on the PNT. Paired samples t-tests also indicate that average *phonological p-weights* pre- (M = .02, SD = .01) to post-treatment (M = .03, SD = .01) significantly improved (t(8) = -2.45, $p = .04$). In contrast, *semantic s-weights* did not change pre- (M = .02, SD = .01) to post-treatment (M = .02, SD = .01; t(8) = .13, ns)[3] . PNT data are based on a sample of nine individuals; P1 did not complete the PNT due to an administration error.

3.3. Predictors of treatment outcomes

The predictor variables of interest were a) *responsiveness to cues* (i.e., improvements in naming accuracy after cueing in round one), and b) *during treatment (i.e., early and late) improvements* secondary to repeated practice of the treatment protocol (described below). Please see Table 6 for individual participant scores for each predictor variable. As can be seen in the matrix plot in Fig. 2, predictors were collinear; responsiveness to cues and early improvement were significantly correlated ($r = .73$, $p = .02$). Thus, separate GLMMs were fit for each predictor below.

Prior to evaluating the predictors of interest, we conducted separate *time only* GLMMs (i.e., where assessment time was the only predictor) for the treated and untreated words. As mentioned in section 2.4.3 above, post-treatment was used as the reference category. Thus, although comparisons between pre- and post-treatment show negative beta coefficients, they nevertheless represent *improvements* in performance post-treatment. Results indicate significantly improved naming accuracy for the treated words pre- to post-treatment (β = -2.18; SE = .23, $p < .01$). Naming accuracy decreased post-treatment to four- (β = -0.50; SE = .21, $p = .01$) and eight-week (β = -0.66; SE = .21, $p < 0.01$) follow-ups. Results also indicate that the naming accuracy of untreated words significantly improved post-treatment to four- (β = .59; SE = 22, $p < .01$) and eight-week (β = .48; SE = .22, $p = .03$) follow-ups; untreated-word naming accuracy did not significantly improve pre- to post-treatment (β = -0.18; SE = .23, ns).

3.3.1. Responsiveness to cues

Based on GLMM results, responsiveness to cues emerged as a significant predictor of naming accuracy for the treated words immediately post-treatment (β =-0.97; SE = .25, $p < .01$). Again, the negative beta coefficient emerged as a result of post-treatment

[3] Some treated stimuli overlapped with items on the PNT; removal of these stimuli did not change the overall findings reported here.

Table 4
Results from the GLMM analysis of treatment effects.

	Coefficient	Std. Error	z	$p > \|z\|$	95 % CI		
Assessment Time							
Post	0.17	0.22	0.77	0.44	−0.26	–	0.60
4-week f/u	0.73	0.22	3.37	<.01	0.31	–	1.15
8-week f/u	0.62	0.22	2.87	<.01	0.20	–	1.04
Condition	−0.01	0.26	−0.02	0.98	−0.51	–	0.50
Assessment Time * Condition							
Post	2.12	0.32	6.71	<.01	1.50	–	2.74
4-week f/u	1.03	0.31	3.38	<.01	0.43	–	1.63
8-week f/u	0.97	0.31	3.19	<.01	0.38	–	1.57
Constant	−1.35	0.29	−4.68	<.01	−1.91	–	−0.78
Variance among participants	0.48	0.25			0.17	–	1.32
Variance among items within participants	2.40	0.35			1.82	–	3.18

Log likelihood = -1353.27; Wald $\chi^2(7) = 150.53$, $p < 0.01$ Notes: Number of observations = 2327. Reference categories are pre-treatment assessment (time), and untreated words (condition).

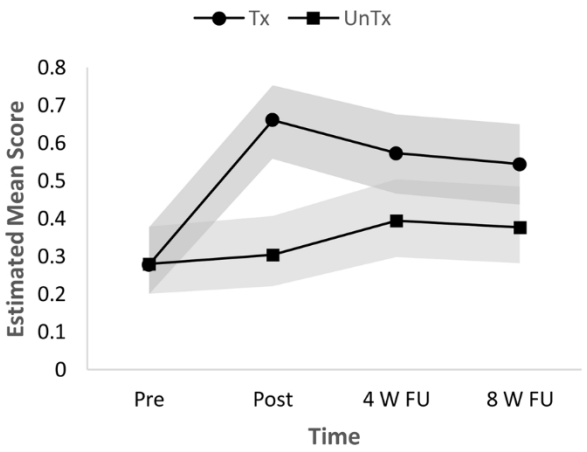


Fig. 1. GLMM estimated mean scores for naming accuracy of treated and untreated words, with 95 % confidence interval bands.

being the reference category. Responsiveness to cues did not emerge as a significant predictor of naming at four- (β = -0.30; SE = .24, ns) or eight-week (β = -0.22; SE = .24, ns) follow-ups, when compared to post-treatment. Responsiveness to cues also emerged as a significant predictor of naming accuracy of the untreated words at four-week follow-up (β = .50; SE = .24, p = .04), but not immediately post-therapy (β =-0.08; SE = .24, ns) nor at eight-week follow-up (β = .31; SE = .24, ns). Overall, results indicate that individuals who demonstrated the greatest responsiveness to cues in the first round of therapy also demonstrated the greatest gains in naming treated words immediately post-therapy, and in naming untreated words at four-week follow-up.

3.3.2. *During-treatment improvement*

Patterns of individual naming performance during treatment, according to round-by-round changes in NA1 naming accuracy (i.e., prior to cueing) for each participant, indicate that the greatest amount of improvement in naming occurred within the first four rounds of treatment (see Fig. 3). In fact, among the group, the total change in naming accuracy between rounds one and four (M = 24.85 %; SD = 17.21) was significantly higher than the total change in naming accuracy between rounds four and 15 (M = 9.30 %; SD = 6.74; $t(9) = 2.34$, p = .04, *two-tailed*).

However, individual variability exists. For example, while P7 and P8 demonstrate steep improvements from rounds one to four (see Table 6), P9 only begins to improve near the end of the therapy block (i.e., round 11). Given this individual variability in improvement, particularly within rounds one to four, we used summed change scores for a) rounds one to four, and b) rounds four to 15, as predictor variables in the analyses that follow. For ease of reading, we call these predictors *early* and *late improvement,* respectively.

3.3.2.1. Early improvement. Results from GLMM analyses indicate that early improvement in therapy is a significant predictor of naming accuracy of *treated* words immediately post-treatment (β = -1.37; SE = .31, p < .01). As above, the negative beta coefficient is a result of post-treatment being the reference category. Early improvement did not emerge as a predictor of treated-word naming performance at four- (β = -0.25, SE = 31, ns) or eight-week (β = -0.44, SE = .31, ns) follow-ups. However, early improvement did emerge as a significant predictor of naming accuracy of the *untreated* words at four-week follow-up (β = .86; SE = .30, p < .01). It was not a significant predictor of naming untreated words post-treatment (β = -0.21; SE = .29, ns) or at eight-week follow-up (β = .45; SE = .29, ns). Therefore, findings suggest that individuals who made the greatest gains early-on in therapy, demonstrated the greatest

Table 5
Individual participant PNT raw naming accuracy and error data pre- and post-treatment. Pre- and post-treatment *semantic s-* and *phonological p*-weights derived from computations of the *SP* model of lexical retrieval are also shown.

	Correct		**Semantic**		**Formal**		**Mixed**		**Unrelated**		**Nonword**[1]		**s-weight**		**p-weight**	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
P1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P2	110	121	7	13	4	2	5	10	0	2	7	9	0.026	0.023	0.024	0.026
P3	56	67	9	16	27	14	4	5	28	27	14	25	0.008	0.012	0.026	0.021
P4	51	56	11	23	2	8	12	9	2	5	26	9	0.019	0.014	0.016	0.023
P6	65	91	14	12	4	2	4	7	1	1	5	2	0.019	0.021	0.026	0.032
P7	123	123	2	5	8	8	0	0	1	0	39	34	0.040	0.036	0.013	0.014
P8	87	99	17	13	1	2	3	5	0	0	4	0	0.020	0.021	0.030	0.034
P9	31	67	12	11	4	4	4	8	2	8	33	22	0.016	0.019	0.013	0.019
P10	115	113	13	11	3	1	8	7	3	1	10	7	0.021	0.024	0.025	0.026
P11	97	104	9	9	4	3	7	5	1	5	7	3	0.023	0.021	0.026	0.031
Mean	**81.67**	**93.44**	**10.44**	**12.56**	**6.33**	**4.89**	**5.22**	**6.22**	**4.22**	**5.44**	**10.44**	**8.33**	**0.021**	**0.021**	**0.022**	**0.025**
SD	*32.27*	*24.91*	*4.36*	*4.95*	*7.98*	*4.29*	*3.42*	*2.95*	*8.97*	*8.53*	*10.68*	*10.21*	*0.009*	*0.007*	*0.006*	*0.007*

[1]Phonologically-related nonword errors.
Notes: Test not administered to P1 due to examiner error. The Philadelphia Naming Test has N = 175 stimuli. Webfit data were input on May 25th, 2019. Omission error data are presented in Appendix B.

Table 6
Percent naming accuracy of the first naming attempt (NA1, prior to cueing) in each round of treatment, across participants. Round one (R1) naming accuracy pre- (NA1) and post- (NA2) cueing is also shown. Responsiveness to cues and early improvement in treatment served as predictor variables in GLMM analyses.

	R1		R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	Responsiveness to cues (NA2-NA1)	Early Improvement[3]	Late Improvement[4]
Pt	NA 1[1]	NA 2[2]																	
P1	56.67	63.33	56.67	70.00	70.00	73.33	70.00	73.33	76.67	73.33	83.33	80.00	80.00	86.67	83.33	83.33	6.67	13.33	13.33
P2	60.00	76.67	73.33	90.00	86.67	76.67	83.33	86.67	80.00	76.67	93.33	83.33	83.33	96.67	80.00	86.67	16.67	26.67	0.00
P3	56.67	76.67	56.67	63.33	76.67	73.33	76.67	83.33	83.33	76.67	76.67	76.67	86.67	63.33	76.67	80.00	20.00	20.00	3.33
P4	16.67	50.00	16.67	46.67	43.33	40.00	36.67	43.33	36.67	46.67	53.33	40.00	50.00	56.67	60.00	46.67	33.33	26.67	3.33
P6	33.33	83.33	56.67	53.33	56.67	50.00	63.33	43.33	80.00	73.33	83.33	73.33	63.33	56.67	70.00	66.67	50.00	23.33	10.00
P7	27.59	73.33	51.72	68.97	82.76	82.76	79.31	93.10	82.76	93.10	86.21	89.66	89.66	93.10	96.55	89.66	45.75	55.17	6.90
P8	33.33	93.33	43.33	66.67	83.33	86.67	86.67	90.00	93.33	86.67	96.67	90.00	90.00	96.67	96.67	96.67	60.00	50.00	13.33
P9	53.33	60.00	33.33	50.00	50.00	56.67	60.00	56.67	60.00	53.33	53.33	70.00	76.67	70.00	80.00	73.33	6.67	−3.33	23.33
P10	41.67	66.67	45.83	50.00	54.17	62.50	50.00	70.83	62.50	66.67	58.33	62.50	70.83	58.33	79.17	66.67	25.00	12.50	12.50
P11	58.62	80.00	51.72	75.86	82.76	68.97	72.41	75.86	86.21	86.21	72.41	79.31	86.21	72.41	79.31	89.66	21.38	24.14	6.90
Mean	**43.79**	**72.33**	**48.59**	**63.48**	**68.64**	**67.09**	**67.84**	**71.65**	**74.15**	**73.26**	**75.70**	**74.48**	**77.67**	**75.05**	**80.17**	**77.93**	**28.55**	**24.85**	**9.30**
SD	*15.37*	*12.58*	*15.33*	*13.70*	*16.15*	*14.75*	*15.64*	*18.28*	*16.63*	*14.58*	*15.96*	*14.76*	*12.94*	*16.73*	*10.93*	*14.86*	*18.27*	*17.21*	*6.74*

[1]Percent naming accuracy pre-cueing, in round one of treatment. [2]Percent naming accuracy post-cueing, in round one of treatment [3]Total change in NA1 naming accuracy from rounds one to four. [4]Total change in NA1 naming accuracy from rounds four to 15.

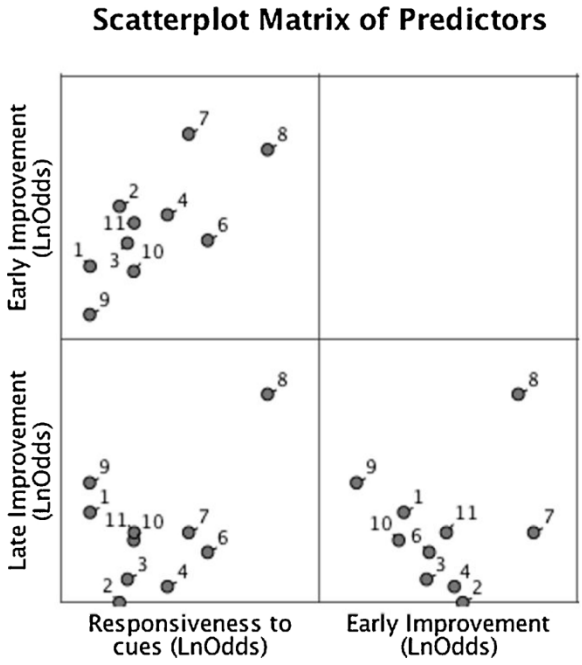


Fig. 2. Scatterplot matrix demonstrating correlations among the predictor variables; numbered data points represent each individual participant.

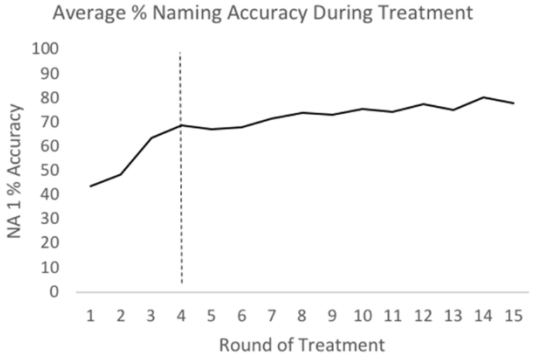


Fig. 3. Average percent naming accuracy of the first naming attempt (NA1, prior to cueing) across participants, for each round of treatment, corresponding to values presented in Table 6. The steepest improvements occurred within the first four rounds of treatment (i.e., left of the dashed line).

improvements in naming treated words immediately post-treatment, and in naming untreated words at four-week follow-up.

3.3.2.2. Late improvement. GLMM analyses indicate that late improvements made in therapy are a significant predictor of naming treated words immediately post-treatment (β = -1.09; SE = .46, p = .02), but not at four- (β = -0.07; SE = .44, ns) or eight-week (β = -0.04; SE = .44, ns) follow-ups. Late improvement did not predict naming accuracy of the untreated words post-treatment (β = -0.49; SE = .45, ns) or at four- (β = .66; SE = .44, ns) or eight-week (β = .57; SE = .44, ns) follow-ups. Therefore, it appears that individuals who made the greatest improvements in naming later-on in therapy also demonstrated the greatest improvements in naming treated words immediately post-therapy.

Findings from LR tests evaluating whether predictors improved model fit when compared to a time-only model (i.e., where time was the only predictor) are presented in Table 7. LR test results indicate that adding responsiveness to cues as a predictor (that could change across assessment timepoints), of treated-word naming accuracy, significantly improved model fit when compared to the time-only model. However, this predictor did not significantly improve model fit for naming of the untreated words. Early improvement also significantly improved model fit when added as a predictor to the time-only model, for both the treated and untreated words. Adding late improvement to the time-only model did not improve model fit for either the treated or untreated words. Overall, it appears that responsiveness to cues in session one of therapy and improvements made early-on in treatment are predictors of naming outcome for the *treated* words. In addition, improvements made early-on in treatment also appear to be a good indicator of naming outcome for the *untreated* words.

4. Discussion

Although treatment for anomia has been shown to be efficacious in both the short- and long-term, individual outcomes can be variable. Evidence regarding the best predictors of aphasia and anomia recovery is mixed. Thus, it remains difficult to predict whether an individual with anomia will improve after therapy, whether improvements will be long-lasting, or whether they will generalize (Nickels, 2002). A review of the literature suggests that mechanisms which stimulate access to language, and support learning (or relearning), may be valuable predictors of treatment outcomes. We investigated two such predictors (i.e., responsiveness to cues and during-treatment improvements) in the present study. Furthermore, we investigated the evolution of speech error production pre- to post-treatment, with the goal of obtaining more precise insights into the mechanisms supporting treatment-induced generalization.

Four important findings emerged from the present study: 1) PCA is efficacious in the short- and long-term, however individual treatment outcomes are variable; 2) participants demonstrated generalization to an untrained naming task (i.e., the PNT); these generalized improvements appear to be mediated by a strengthening of lexical-phonological connections (i.e., *phonological p*-weights), which is a novel application of IA computational models to treatment outcomes; 3) responsiveness to cues in the first round of treatment predicted improvements in naming accuracy of treated words; and 4) steep improvements early-on in therapy (i.e., within the first four rounds of PCA) predicted improved naming accuracy of both treated and untreated words.

4.1. Treatment outcomes

Consistent with previous work (e.g., Leonard et al., 2008, 2015; Van Hees et al., 2013), results from our group analyses lend further support for the short- and long-term efficacy of the PCA treatment approach. Despite a decrease in naming accuracy of the treated words from post-treatment to follow-up, naming accuracy remained significantly higher than at pre-treatment levels; participants demonstrated significantly improved naming of treated words pre- to post-treatment and pre-treatment to four- and eight-week follow-ups. Individual analyses of the data were able to provide a more nuanced breakdown of treatment outcomes. Eight participants overall demonstrated significant improvements in naming accuracy immediately post-treatment, and up to six participants maintained those improvements over the follow-up periods. Such immediate and maintained *item-specific* gains in naming accuracy are a common finding in the aphasia literature (e.g., Nickels, 2002).

In addition, group analyses show significantly improved naming of the untreated words at four- and eight-week follow-ups, corroborating existing reports of phonological treatments for anomia leading to broader improvements in naming, beyond the treated items (Best et al., 2013; Leonard et al., 2008; Nickels, 2002). According to individual analyses, naming accuracy of the untreated words significantly improved for two participants across follow-up periods; an additional four participants showed either a significant trend for improvement or a significant rate of change from baseline through to the follow-up periods. These improvements were primarily observed in the long-term (i.e., at follow-up stages).

Thus, it appears that for at least two and possibly six participants (as seen in Table 3), improvements in naming were not solely item-specific, but impacted a broader array of lexical items over time. Broadly stated, lexical items beyond the treated set appear to have been brought closer to selection thresholds. This may be a result of multiple exposures to the untreated stimuli over time, although these exposures were relatively infrequent (roughly once per month). An intriguing alternative is that the treatment itself led to improved performance in naming the untreated words; the generalized improvements noted on the PNT support the latter interpretation.

4.2. Generalization

Participants demonstrated significant pre- to post-treatment improvements in naming accuracy on an untrained naming task (i.e., the PNT). Participants also demonstrated significantly improved *phonological p*-weights pre- to post-treatment, which, according to the IA model of word retrieval, represents a strengthening of lexical-phonological connections. This is a particularly intriguing finding, as it suggests that the PCA treatment approach is indeed stimulating its intended target, namely the phonological level of representation.

Two mechanisms may be driving the strengthening of lexical-phonological connections. The first involves positive feedback loops between lexical and phonological levels of representation (as described by e.g., Foygel & Dell, 2000; Rapp & Goldrick, 2000), which allow for subsequent spreading activation among adjacent phonological nodes. In addition, according to Rapp & Goldrick (2000), spreading activation occurs primarily between the phonological and lexical levels, with minimal activation spreading to the semantic level. This may explain why lexical-semantic connections (i.e., *semantic s*-weights) remained unchanged post-treatment. The second mechanism supporting the observed increase in p-weights may involve boosted vertical access between semantic and phonological nodes, with subsequent activation feeding back and further activating the phonological level (as previously suggested by Leonard et al., 2015; Rochon et al., 2010). However, in this case an improvement in s-weights would be expected following therapy, which we did not observe; the timing and nature of outcome measurement may have impacted this finding.

Namely, although *generalization* to the PNT appears to be mediated by an increase in p-weights, preliminary data from our group (Mengad et al., 2019) suggest that improvements on *treated* words appear to be mediated by an increase in s-weights. Thus, it may be the case that the consistent and repeated practice of treated words over time eventually leads to improved access at the semantic level, whereas a single measure of generalization immediately post-treatment may only reflect changes at the phonological/lexical levels. Given that the target of our treatment approach is the phonological level of representation, it may take time for treatment effects to spread upstream to the semantic level.

Finally, it is important to highlight the possible limitations of using the IA computational model when evaluating treatment outcomes. For instance, an increase in semantic errors following therapy may in fact be a sign of improvement (especially if accompanied

Table 7
Results from GLMMs evaluating each predictor of interest and its interactions with each assessment time, as well as likelihood ratio tests comparing the fit of each predictor model against a time only model.

	TREATED			UNTREATED		
	Log likelihood	**Wald**	**p**	**Log likelihood**	**Wald**	**p**
Time only	−690.27	$\chi^2(3) = 97.65$	**<.01**	−662.747	$\chi^2(3) = 16.32$	**0.01**
Time x Responsiveness to cues	−680.81	$\chi^2(7) = 112.08$	**<.01**	−658.287	$\chi^2(7) = 24.20$	**0.01**
Time x Early Improvement	−674.91	$\chi^2(7) = 104.79$	**<.01**	−652.846	$\chi^2(7) = 33.86$	**<.01**
Time x Late Improvement	−685.85	$\chi^2(7) = 102.03$	**<.01**	−658.506	$\chi^2(7) = 23.30$	**<.01**
Likelihood Ratio Test						
Responsiveness to cues vs. Time only		$\chi^2(4) = 18.92$	**<.01**		$\chi^2(4) = 8.92$	0.06
Early Improvement vs. Time only		$\chi^2(4) = 30.72$	**<.01**		$\chi^2(4) = 19.80$	**<.01**
Late Improvement vs. Time only		$\chi^2(4) = 8.82$	0.07		$\chi^2(4) = 8.48$	0.08

by a decrease in omission errors; see Dell et al., 2004). However, this may lead to a decreased s-weight in the IA model. Although this did not appear to be an issue in the present study (i.e., decreased s-weights were not systematically linked to a tradeoff between semantic and omission errors), the assumptions of the IA model must be tested further in this regard[4] .

Broadly, our findings suggest that PCA therapy may induce a generalized "boost" in lexical retrieval, allowing for more precise lexical activation, and making words more available for use in daily interactions. More frequent use may incite further activation and strengthening of connections (akin to a ripple effect) beyond the treatment block, and induce a longer-lasting, positive impact. This may also explain why improvements in naming untreated words were not observed immediately in our data but developed over the follow-up periods. However, as mentioned above, it is possible that multiple (albeit infrequent) exposures to the untreated stimuli may also explain our results. It is important for future research to employ longitudinal study designs in order to distinguish whether (and to what degree) generalized improvements are a result of the treatment itself, compared to multiple exposures to stimuli. Regardless, the present findings have important clinical implications, which are discussed below.

4.3. Responsiveness to cues

Previous behavioral and neuroimaging studies have demonstrated the predictive value of responsiveness to cues (i.e., facilitation) in other phonological treatment approaches (Hickin et al., 2002; Nardo et al., 2017); to our knowledge, our findings are the first to show this relationship in PCA therapy. Individuals who were most responsive to cueing in the first round of treatment demonstrated the greatest acquisition of treated words. As described above, according to the IA model, the facilitative effect of cues may be ascribed to more targeted activation and therefore, a more precise search space during lexical retrieval (Meteyard & Bose, 2018).

Responsiveness to cueing is likely a measure of stimulability. In this study, the measurement of responsiveness was based on the effectiveness of a single administration of the PCA cueing technique (i.e., in round one), suggesting that stimulability is an inherent state, which varies at baseline across individuals. This may be why responsiveness to cues emerged as a predictor of naming accuracy for both the treated and untreated words. However, in the latter case, responsiveness to cueing did not add significant predictive value as compared to a time only model. Therefore, responsiveness to cues is perhaps a better indicator of *item-specific* improvement immediately post-treatment. This is consistent with the supposition that *treated* words are most salient and most easily accessible at this stage. However, it remains unclear to what extent the treatment technique used impacts an individual's level of stimulability, and whether stimulability can be supported and/or heightened in order to optimize treatment outcomes. This is an intriguing avenue for future research.

It is important to note that responsiveness to cues was not a predictor of treatment maintenance; findings from a recently published study suggest instead that the maintenance of treatment effects is dependent on baseline executive control ability, which is implicated in learning and memory consolidation (Simic et al., 2019). Therefore, immediately post-treatment, naming performance may primarily reflect the current salience and/or activation levels of treated words. However, at follow-up stages, naming performance may be reflective primarily of longer-term learning and memory consolidation, once activation levels of the treated words have subsided.

4.4. Early improvement during treatment

In the present sample, the greatest during-treatment improvements in naming accuracy occurred between rounds one and four of therapy, with much smaller improvements thereafter. This is most evident in the round-by-round naming performance of participants P7 and P8 (see Table 6). In fact, this type of performance trajectory is a common finding for cognitive skill acquisition in healthy adults and is referred to as the *power law of practice* (Anderson, 1982; Logan, 1988). The steep early learning of P7 and P8 may have supported their superior outcomes in treatment, as they went on to make significant improvements in naming both the treated and untreated words following therapy. On the other hand, P9, who did not respond to therapy, demonstrated some improvement in performance later on in treatment, but made inconsistent and/or minimal gains in the initial rounds of treatment. Indeed, late improvement did not improve model fit when compared to a time only model. Taken together, these findings suggest that the early stages of (re)learning may hold important predictive value.

[4] The authors would like to thank the anonymous reviewers for this important contribution.

Building upon previous work, we found that individuals who made the greatest improvements early-on in PCA treatment (i.e., within the first four rounds) demonstrated the greatest short-term improvements in naming treated words. Thus, early improvement may be a good measure of item-specific (re)learning. Steep improvements (or relearning) early-on also may have allowed for a greater number of correct practice trials during treatment. In turn, this may have increased the accessibility of treated stimuli, especially immediately post therapy. For example, P2, P7 and P8 achieved roughly 80 % (and above) naming accuracy by round four, and P1 did so by round ten (see Table 6). Thus, sessions following attainment of this arbitrary criterion of 80 % accuracy, may be comparable to an overtraining period (i.e., repeated practice even after high levels of accuracy have been achieved).

Furthermore, previous studies have linked overtraining to better maintenance of learned items (e.g., Friedman et al., 2017). Given that the number of practice repetitions was held constant in our study (i.e., all participants received a total of 15 repetitions of the PCA protocol, for each treated word), we were unable to replicate such findings; early improvement did not emerge as a significant predictor of treatment maintenance. However, in line with these studies, all participants who achieved consistently high naming accuracy early-on during treatment and had longer subsequent periods of "overtraining" (e.g., P2, P7, P8), also demonstrated good overall maintenance of treatment effects.

Early item-specific improvements in naming during therapy also emerged as a good overall predictor of improved naming of the *untreated* words. This is in line with studies showing the positive impact of repeated retrieval practice for one set of items, on learning transfer to another set of items (Roediger & Butler, 2011). Furthermore, it supports the hypothesis that the treatment itself played a role in the improved naming accuracy of the untreated words. However, as above, we cannot discount the potential impact of multiple exposures to the untreated stimuli; it may be that early improvement is simply a measure of how an individual responds to repeated practice in general. This is an empirical question for future studies. Nevertheless, our findings add to the existing literature by demonstrating that repeated practice of a treatment protocol does not benefit individuals equally despite general group trends, and that early improvements in particular are important indicators of both item-specific, and broader treatment outcomes.

4.5. Clinical implications

Together, these findings have important clinical implications. First, individual outcome analyses demonstrated roughly three patterns of response to treatment. Namely, those who did not improve after treatment, those who made short- and long-term item-specific improvements, and those who also demonstrated broader improvements in naming untreated words. Clinically, it may be important to distinguish among responders in this way. Although more research is needed in this area, it may eventually be possible to distinguish predictors of item-specific improvement, from predictors of broader, more generalized improvement. If borne out, measuring initial responsiveness to cueing or early improvements in therapy may be implemented as a relatively simple assessment practice, which could assist with treatment selection by informing on an individual's potential for improvement with a variety of treatment techniques.

For example, if steep improvements are noted early-on for a given treatment (e.g., within the first four sessions), patients may be more likely to show generalized improvements. As such, treatment plans for these individuals may employ larger stimulus sets (e.g., see Laganaro, Di Pietro, & Schnider, 2006), which are rotated every fourth repetition or round of treatment. Furthermore, given the present findings, four iterations of a given treatment may be a critical point at which to assess the effectiveness of a treatment approach and make changes as needed. However, these suggestions are speculative, and must be empirically tested.

In the present study, we found significant improvements in the naming accuracy of untreated words. Although we are of the opinion that these improvements were influenced, at least in part, by the treatment itself, it is nevertheless a possibility that they were simply a result of multiple exposures to the untreated stimuli over time. However, if the latter is indeed the case, repeated exposure to items could be incorporated into treatment, especially given its potential to lead to significant naming improvements. For example, a given anomia treatment protocol which trains personally relevant and specific items, may be interspersed with periods of retrieval practice and/or exposure to a broader set of items. Such a combination approach to treatment may ultimately maximize language improvements.

Further research is needed to determine the most efficient method (e.g., the minimum number of trials necessary) to obtain valid measures of early improvement and responsiveness, in order to implement them clinically. Finally, techniques and/or environmental modifications that heighten an individual's stimulability to cues or support steeper learning in initial therapy sessions may also maximize treatment-related gains.

4.6. Limitations

The primary limitation of this study is the small sample size; although the findings provide a stepping stone for future research on this topic, results must be interpreted with caution. In addition, our sample was comprised mostly of males. As such, replication in a larger and more balanced group of participants would be required, to determine whether the current findings are applicable to the broader aphasia population. In the present study, multicollinearity among predictors prevented us from making direct comparisons of the predictor variables; variations in study design, and a larger sample size may allow for such comparisons in future studies. Furthermore, treated and untreated stimuli were balanced within participants, but not between participants. This may be an important factor to manipulate in future studies, in order to characterize the types of words that are (re)learned compared to those that are not. Finally, the present study differs from traditional facilitation studies, by providing five phonological components to stimulate naming, as opposed to the one or two cues that are typically provided. Nevertheless, our results offer converging evidence for the predictive value of responsiveness to cues, corroborating findings from previous studies.

5. Conclusions

Our study contributes to research which aims to understand the mechanisms driving anomia (and aphasia) treatment. Notably, it underscores the importance of considering the fundamental components of treatment itself; responsiveness to treatment, perhaps not surprisingly, is important to treatment outcome. Individual responsiveness may vary depending on the therapy approach taken and as such, careful treatment selection is imperative. In addition, the nature and evolution of individual responses (i.e., error types) during and after therapy may prove particularly informative. Based on current and previous findings, responsiveness to cueing appears to be a good indicator of item-specific acquisition immediately following phonological treatments for anomia, and early improvement in therapy appears to be important to both item-specific, and more generalized treatment gains. Future research must identify whether the measures investigated in this study are applicable across many therapy approaches (i.e., treatment-general), or whether they are treatment-specific indicators of success. Importantly, our findings suggest that longitudinal treatment outcomes are critical to broadening our understanding of the mechanisms contributing to treatment success, as immediate treatment acquisition, long-term maintenance, short- and long-term generalization may be mediated by different factors. Ultimately, identifying such factors may assist with optimizing stimulus selection and individualizing therapy, in addition to building treatment programs that are more likely to generalize.

Funding

This work was supported by the Heart and Stroke Foundation [grant#7308,#7015; awarded to E. Rochon and C. Leonard], and by theCanadian Partnership for Stroke Recovery [scholarship awarded to T. Simic]. Funding sponsors had no further involvement in the study.

Author note

This research was approved under the Research Ethics Board of the University of Toronto (ethics protocol #32663). Ethics approval was also obtained from the referring sites, the Aphasia Institute and the March of Dimes Aphasia and Communication Disabilities Program. Informed consent was obtained from all participants.

CRediT authorship contribution statement

Tijana Simic: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Craig Chambers:** Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision. **Tali Bitan:** Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision. **Steven Stewart:** Formal analysis, Writing - review & editing. **Devora Goldberg:** Validation, Formal analysis. **Laura Laird:** Methodology, Investigation, Data curation, Writing - review & editing. **Carol Leonard:** Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision, Funding acquisition. **Elizabeth Rochon:** Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors report no conflicts of interest.

Acknowledgments

The authors would like to thank the participants and their families, as well as our referral sites across the Greater Toronto Area (the Aphasia Institute, and the March of Dimes Aphasia and Communication Disabilities Program). The authors would also like to acknowledge the following members of the Language Sciences Laboratory for assistance with data collection and scoring: Fiona Hobler, So Yeun Kim, and Claire Dreyfuss.

Appendix A. Detailed statistics for individual WEST-Trend and WEST-ROC treatment outcome analyses

*Significant; based on one-sample *t* -test (one-tailed) after Holm-Bonferroni procedure (i.e., where alpha was initially set at .05/12 = .004 and adjusted accordingly thereafter).

Appendix B. Number of omission errors on the Philadelphia Naming Test (PNT) pre- and post-treatment by participant

		Baseline to Post						Baseline to 4 week follow-up						Baseline to 8 week follow-up					
WEST-Trend		**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**	**Mean**	**SD**	**t-value**	**df**	**95% CI**	**p-value**	**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**
	P1	1.00	1.89	2.89	29	[0.29–1.71]	**.004***	1.10	1.97	3.06	29	[0.36–1.84]	**.003***	1.20	1.99	3.30	29	[0.46–1.94]	**.002***
	P2	1.87	1.81	5.64	29	[1.19–2.54]	**<.001***	1.87	1.93	5.31	29	[1.15–2.59]	**<.001***	1.87	1.81	5.64	29	[1.19–2.54]	**<.001***
	P3	1.17	1.88	3.40	29	[0.47–1.87]	**.001***	0.67	1.65	2.22	29	[0.05–1.28]	.018	0.47	2.00	1.28	29	[-0.28 to 1.21]	.105
	P4	1.17	1.86	3.44	29	[0.47–1.86]	**.001***	0.57	1.76	1.77	29	[-0.09 to 1.22]	.044	0.17	1.53	0.60	29	[-0.41 to 0.74]	.278
(Treated words only)	**P6**	0.97	2.34	2.26	29	[0.09–1.84]	.016	0.77	2.19	1.92	29	[-0.05 to 1.59]	.033	1.37	1.56	4.79	29	[0.78–1.95]	**<.001***
	P7	2.24	1.41	8.59	28	[1.71–2.78]	**<.001***	2.24	1.41	8.59	28	[1.71–2.78]	**<.001***	2.03	1.50	7.31	28	[1.46–2.60]	**<.001***
	P8	2.33	1.27	10.08	29	[1.86–2.81]	**<.001***	1.83	1.68	5.97	29	[1.21–2.46]	**<.001***	1.73	1.57	6.03	29	[1.15–2.32]	**<.001***
	P9	1.23	1.76	3.85	29	[0.58–1.89]	**.001***	1.23	1.76	3.85	29	[0.58–1.89]	**.001***	1.13	1.61	3.85	29	[0.53–1.74]	**.001***
	P10	1.67	1.52	5.36	23	[1.02–2.31]	**<.001***	1.04	1.46	3.50	23	[0.43–1.66]	**.001***	0.92	1.67	2.70	23	[0.21–1.62]	**.007***
	P11	1.41	2.01	3.79	28	[0.65–2.18]	**.001***	1.21	2.14	3.03	28	[0.39–2.02]	**.003***	0.79	1.76	2.43	28	[0.12–1.46]	.011
WEST-Trend		**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**	**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**	**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**
	P1	0.23	1.36	0.94	29	[-0.27 to 0.74]	.177	0.53	1.53	1.92	29	[-0.04 to 1.10]	.033	0.53	1.53	1.92	29	[-0.04 to 1.10]	.033
	P2	1.13	1.72	3.62	29	[0.49–1.77]	**.001***	1.13	1.72	3.62	29	[0.49–1.77]	**.001***	1.23	1.74	3.89	29	[0.59–1.88]	**.001***
	P3	0.53	1.83	1.59	29	[-0.15 to 1.22]	.061	0.93	1.78	2.87	29	[0.27–1.60]	**.004***	0.83	1.68	2.71	29	[0.21–1.46]	**.006***
	P4	0.47	1.11	2.31	29	[0.05 to 0.88]	.014	0.87	1.38	3.43	29	[0.35–1.38]	**.001***	0.67	1.27	2.88	29	[0.19–1.14]	**.004***
	P6	0.67	1.81	2.02	29	[-0.01–1.34]	.027	0.47	1.70	1.51	29	[-0.17 to 1.10]	.072	0.57	1.70	1.83	29	[-0.07 to 1.20]	.039
(Untreated words only)	**P7**	0.76	2.18	1.87	28	[-0.07 to 1.59]	.036	1.28	1.94	3.54	28	[0.54–2.02]	**.001***	0.97	2.24	2.32	28	[0.11–1.82]	**.014***
	P8	0.57	1.70	1.83	29	[-0.07 to 1.20]	.039	1.57	1.94	4.42	29	[0.84–2.29]	**<.001***	1.37	1.94	3.86	29	[0.64–2.09]	**.001***
	P9	0.67	1.71	2.14	29	[0.03–1.31]	.021	0.27	1.44	1.02	29	[-0.27 to 0.80]	.159	0.47	1.72	1.49	29	[-0.17 to 1.11]	.074
	P10	0.43	1.53	1.36	22	[-0.23 to 1.10]	.094	0.57	1.70	1.59	22	[-0.17 to 1.30]	.063	0.70	1.92	1.74	22	[-0.13 to 1.53]	.048
	P11	0.14	1.18	0.64	27	[-0.31 to 0.60]	.263	0.57	2.10	1.44	27	[-0.24 to 1.39]	.081	0.36	1.62	1.17	27	[-0.27 to 0.98]	.126
WEST-ROC		**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**	**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**	**Mean**	**SD**	**t-value**	**df**	**95 % CI**	**p-value**
	P1	2.17	2.02	5.88	29	[1.41–2.92]	**<.001***	2.27	1.76	7.05	29	[1.61–2.92]	**<.001***	2.37	1.71	7.57	29	[1.73–3.01]	**<.001***
	P2	1.20	2.61	2.52	29	[.23–2.17]	**.009***	1.20	2.27	2.90	29	[.35–2.05]	**.004***	1.20	2.61	2.52	29	[.23–2.17]	**.009***
	P3	1.67	2.07	4.40	29	[.89–2.44]	**<.001***	1.17	2.00	3.19	29	[.42–1.91]	**.002***	0.97	1.54	3.43	29	[.39–1.54]	**.001***
	P4	1.17	2.20	2.91	29	[.35–1.99]	**.004***	0.57	1.85	1.68	29	[-.13 to 1.26]	.052	0.17	1.93	0.47	29	[-.56 to .89]	.320
(Treated words only)	**P6**	1.30	1.51	4.71	29	[.74–1.87]	**<.001***	1.10	1.67	3.61	29	[.48–1.72]	**<.001***	1.70	2.28	4.09	29	[.85–2.55]	**<.000***
	P7	2.59	1.86	7.48	28	[1.88–3.29]	**<.001***	2.59	1.86	7.48	28	[1.88–3.29]	**<.001***	2.38	1.97	6.50	28	[1.63–3.13]	**<.001***
	P8	2.50	1.53	8.98	29	[1.93–3.07)	**<.001***	2.00	1.64	6.68	29	[1.39–2.61]	**<.001***	1.90	1.84	5.64	29	[1.21–2.59]	**<.001***
	P9	0.07	1.86	0.20	29	[-.63 to .76]	.423	0.07	2.12	0.17	29	[-.72 to .86]	.432	−0.03	2.19	−0.08	29	[-.85 to .78]	.467
	P10	2.08	2.24	4.55	23	[1.14–3.03]	**<.001***	1.46	2.32	3.08	23	[.48–2.44]	**.003***	1.33	2.20	2.97	23	[.40–2.26]	**.004***
	P11	1.07	2.59	2.22	28	[.08–2.05]	.015	0.86	2.23	2.08	28	[.01–1.71]	.023	0.45	2.92	0.83	28	[-.66 to 1.56]	.208
WEST-ROC		**Mean**	**SD**		**df**	**95 % CI**		**Mean**	**SD**		**df**	**95 % CI**		**Mean**	**SD**		**df**	**95 % CI**	

(*continued on next page*)

(*continued*)

		Baseline to Post						Baseline to 4 week follow-up						Baseline to 8 week follow-up					
				t-value			**p-value**			**t-value**			**p-value**			**t-value**			**p-value**
(Untreated words only)	**P1**	0.57	1.61	1.93	29	[-.04 to 1.17]	.032	0.87	1.70	2.80	29	[.23–1.50]	**.005***	0.87	1.70	2.80	29	[.23–1.50]	**.005***
	P2	0.47	2.47	1.03	29	[-.46 to 1.39]	.155	0.47	2.26	1.13	29	[-.38 to 1.31]	.133	0.57	2.51	1.23	29	[-.37 to 1.51]	.114
	P3	0.53	2.18	1.34	29	[-.28 to 1.36]	.095	0.93	2.13	2.40	29	[.14–1.73]	.012	0.83	2.29	1.99	29	[-.02 to 1.69]	.028
	P4	−0.70	1.88	−2.04	29	[-1.40 to .00]	.025	−0.30	2.04	−0.81	29	[-1.06 to .46]	.213	−0.50	1.83	−1.49	29	[-1.19 to .19]	.073
	P6	0.67	2.15	1.69	29	[-.14 to 1.47]	.051	0.47	2.30	1.11	29	[-.39 to 1.33]	.138	0.57	2.30	1.35	29	[-.29 to 1.43]	.094
	P7	0.93	2.23	2.24	28	[.08–1.78]	.017	1.45	2.44	3.19	28	[.52–2.38]	**.002***	1.14	2.03	3.02	28	[.37–1.91]	**.003***
	P8	0.40	2.14	1.02	29	[-.40 to 1.20]	.158	1.40	1.65	4.64	29	[.78–2.02]	**<.001***	1.20	1.92	3.43	29	[.48–1.92]	**.001***
	P9	0.50	2.24	1.22	29	[-.34 to 1.34]	.116	0.10	2.25	0.24	29	[-.74 to .94]	.405	0.30	2.23	0.74	29	[-.53 to 1.13]	.234
	P10	0.43	2.25	0.93	22	[-.54 to 1.41]	.183	0.57	2.06	1.31	22	[-.33 to 1.46]	.101	0.70	1.92	1.74	22	[-.13 to 1.53]	.048
	P11	0.32	2.98	0.57	27	[-.84 to 1.48]	.287	0.75	2.49	1.60	27	[-.22 to 1.72]	.061	0.54	2.59	1.10	27	[-.47 to 1.54]	.142

	Omission Errors	
	Pre	Post
P1	n/a	n/a
P2	33.00	14.00
P3	32.00	13.00
P4	30.00	39.00
P6	78.00	57.00
P7	0.00	4.00
P8	54.00	53.00
P9	35.00	31.00
P10	7.00	7.00
P11	39.00	41.00
Mean	**34.22**	**28.78**
SD	*23.10*	*19.99*

References

Abel, S., Willmes, K., & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology, 21*(5), 411–447.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Best, W., & Nickels, L. (2000). From theory to therapy in aphasia: Where are we now and where to next? *Neuropsychological Rehabilitation, 10*(3), 231–247. https://doi.org/10.1080/096020100389147

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalization from anomia therapy occur and is it predictable? A case series study. *Cortex, 3*, 1–13.

Best, W., Herbert, R., Hickin, J., Osborne, F., & Howard, D. (2002). Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. *Aphasiology, 16*(1/2), 151–168.

Breitenstein, C., Kamping, S., Jansen, A., Schomacher, M., & Knecht, S. (2004). Word learning can be achieved without feedback: Implications for aphasia therapy. *Restorative Neurology and Neuroscience, 22*(6), 445–458.

Byng, S., & Black, M. (1995). What makes a therapy? Some parameters of therapeutic intervention in aphasia. *International Journal of Language & Communication Disorders, 30*(3), 303–316. https://doi.org/10.3109/13682829509021444

Caplan, D. (1993). Toward a psycholinguistic approach to acquired neurogenic language disorders. *American Journal of Speech-language Pathology, 2*(1), 59–83. https://doi.org/10.1044/1058-0360.0201.59

Cicerone, K. D., Dahlberg, C., Malec, J. F., Langenbahn, D. M., Felicetti, T., Kneipp, S., … Catanese, J. (2005). Evidence-based cognitive rehabilitation: Updated review of the literature from 1998 through 2002. *Archives of Physical Medicine and Rehabilitation, 86*(8), 1681–1692. https://doi.org/10.1016/j.apmr.2005.03.024

Cicerone, K. D., Langenbahn, D. M., Braden, C., Malec, J. F., Kalmar, K., Fraas, M., … Ashman, T. (2011). Evidence-based cognitive rehabilitation: Updated review of the literature from 2003 through 2008. *Archives of Physical Medicine and Rehabilitation, 92*(4), 519–530.

Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology, 21*(2-4), 125–145.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and non-aphasic speakers. *Psychological Review, 104*, 811–838.

Dignam, J., Copland, D., Rawlings, A., O'Brien, K., Burfein, P., & Rodriguez, A. D. (2016). The relationship between novel word learning and anomia treatment success in adults with chronic aphasia. *Neuropsychologia, 81*, 186–197. https://doi.org/10.1016/j.neuropsychologia.2015.12.026

Doesborgh, S. J. C., van de Sandt-Koenderman, M. W. M. E., Dippel, D. W. J., van Harskamp, F., Koudstaal, P. J., & Visch-Brink, E. G. (2004). Cues on request: The efficacy of multicue, a computer program for wordfinding therapy. *Aphasiology, 18*(3), 213–222.

Duffy, J. R. (1995). *Motor speech disorders: Substrates, differential diagnosis, and management (1st ed)*. St.Louis: Mosby.

Ferguson, A. (1999). Learning in aphasia therapy: It's not so much what you do, but how you do it! *Aphasiology, 13*(2), 125–150. https://doi.org/10.1080/026870399402244

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

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*, 182–216.

Freedman, M. L., & Martin, R. C. (2001). Dissociable components of short-term memory and their relation to long-term learning. *Cognitive Neuropsychology, 18*(3), 193–226. https://doi.org/10.1080/02643290126002

Fridriksson, J., Nettles, C., Davis, M., Morrow, L., & Montgomery, A. (2006). Functional communication and executive function in aphasia. *Clinical Linguistics & Phonetics, 20*(6), 401–410.

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. https://doi.org/10.1037/neu0000318

Gialanella, B., Bertolinelli, M., Lissi, M., & Prometti, P. (2011). Predicting outcome after stroke: The role of aphasia. *Disability and Rehabilitation, 33*(2), 122–129.

Godecke, E., Rai, T., Ciccone, N., Armstrong, E., Granger, A., & Hankey, G. (2013). Amount of therapy matters in very early aphasia rehabilitation after stroke: A clinical prognostic model. *Seminars in Speech and Language, 34*(3), 129–141. https://doi.org/10.1055/s-0033-1358369

Godecke, E., Hird, K., Lalor, E. E., Rai, T., & Phillips, M. R. (2012). Very early poststroke aphasia therapy: A pilot randomized controlled efficacy trial. *International Journal of Stroke, 7*(8), 635–644. https://doi.org/10.1111/j.1747-4949.2011.00631.x

Grossman, M., & Carey, S. (1987). Selective word-learning deficits in aphasia. *Brain and Language, 32*, 306–324.

Gupta, P., Martin, N., Abbs, B., Schwartz, M., & Lipinski, J. (2006). New word learning in aphasic patients: Dissociating phonological and semantic components. *Brain and Language, 99*, 8–9.

Hauptmann, B., & Karni, A. (2002). From primed to learn: The saturation of repetition priming and the induction of long-term memory. *Cognitive Brain Research, 13*(3), 313–322. https://doi.org/10.1016/S0926-6410(01)00124-0

Hauptmann, B., Reinhart, E., Brandt, S. A., & Karni, A. (2005). The predictive value of the leveling off of within-session performance for procedural memory consolidation. *Cognitive Brain Research, 24*(2), 181–189. https://doi.org/10.1016/j.cogbrainres.2005.01.012

Heath, S., McMahon, K. L., Nickels, L. A., Angwin, A., MacDonald, A. D., Van Hees, S., … Copland, D. A. (2015). An fMRI investigation of the effects of attempted naming on word retrieval in aphasia. *Frontiers in Human Neuroscience, 9*(May). https://doi.org/10.3389/fnhum.2015.00291

Helm-Estabrooks, N. (2002). Cognition and aphasia: A discussion and a study. *Journal of Communication Disorders, 35*(2), 171–186.

Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology, 16*(10-11), 981–999. https://doi.org/10.1080/02687030244000509

Hilari, K. (2011). The impact of stroke: Are people with aphasia different to those without? *Disability and Rehabilitation, 33*(3), 211–218.

Hopper, T., & Holland, A. L. (2005). Aphasia and learning in adults: Key concepts and clinical considerations. *Topics in Geriatric Rehabilitation, 21*(4), 315–322. https://doi.org/10.1097/00013614-200510000-00008

Howard, D., & Patterson, K. (1992). *The pyramids and palm trees test*. Suffolk, UK: Thames Valley Test Company.
Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology, 29*(5), 526–562. https://doi.org/10.1080/02687038.2014.985884
Howard, D., Hickin, J., Redmond, T., Clark, P., & Best, W. (2006). Re-visiting "semantic facilitation" of word retrieval for people with aphasia: Facilitation yes but semantic no. *Cortex, 42*(6), 946–962. https://doi.org/10.1016/S0010-9452(08)70439-8
Hula, W. D., & McNeil, M. R. (2008). Models of attention and dual-task performance as explanatory constructs in aphasia. *Seminars in Speech and Language, 29*(3), 169–187. https://doi.org/10.1055/s-0028-1082882
Jokel, R., Rochon, E., & Leonard, C. (2004). Testing predictions of the interactive activation model in recovery from aphasia after treatment. *Brain and Cognition, 54*, 251–253.
Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test (2nd ed.)*. Philadelphia: Lippincott, Williams & Wilkins.
Kauhanen, M., Korpelainen, J. T., Hiltunen, P., Määttä, R., Mononen, H., Brusin, E., et al. (2000). Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. *Cerebrovascular Diseases, 10*(6), 455–461. https://doi.org/10.1159/000016107
Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessments of language processing in aphasia (PALPA)*. Hove, UK: Lawrence Erlbaum Associates Ltd.
Kelly, H., & Armstrong, L. (2008). New word learning in people with aphasia. *Aphasiology, 23*, 1398–1417.
Kendall, D. L., Pompon, R. H., Brookshire, C. E., Minkina, I., & Bislick, L. (2013). An analysis of aphasic naming errors as an indicator of improved linguistic processing following phonomotor treatment. *American Journal of Speech-language Pathology, 22*(2), S240–S249. https://doi.org/10.1044/1058-0360(2012/12-0078)
Kertesz, A. (2006). *Western aphasia battery-revised*. Texas: Harcourt Assessments.
Kiran, S. (2016). How does severity of aphasia influence individual responsiveness to rehabilitation? Using big data to understand theories of aphasia rehabilitation. *Seminars in Speech and Language, 37*(1), 48–59. https://doi.org/10.1055/s-0036-1571358
Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech Language and Hearing Research, 51*(1), S225–S239.
Knoflach, M., Matosevic, B., Rücker, M., Furtner, M., Mair, A., Wille, G., et al. (2012). Functional recovery after ischemic stroke - A matter of age: Data from the austrian stroke unit registry. *Neurology, 78*(4), 279–285. https://doi.org/10.1212/WNL.0b013e31824367ab
Laganaro, M., Di Pietro, M., & Schnider, A. (2006b). What does recovery from anomia tell us about the underlying impairment: The case of similar anomic patterns and different recovery. *Neuropsychologia, 44*(4), 534–545. https://doi.org/10.1016/j.neuropsychologia.2005.07.005
Laganaro, M., Di Pietro, M., & Schnider, A. (2006a). Computerised treatment of anomia in acute aphasia: Treatment intensity and training size. *Neuropsychological Rehabilitation, 16*(6), 630–640. https://doi.org/10.1080/09602010543000064
Laine, M., & Martin, N. (2006). *Anomia: Theoretical and clinical aspects*. Hove, UK: Psychology Press.
Lambon-Ralph, M. A., & Fillingham, J. K. (2007). Chapter 9: The importance of memory and executive function in aphasia: Evidence from the treatment of anomia using errorless and errorful learning. In A. S. Meyer, L. R. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 193–213). Hove: Psychology Press.
Laska, A. C., Hellblom, A., Murray, V., Kahan, T., & Von Arbin, M. (2001). Aphasia in acute stroke and relation to outcome. *Journal of Internal Medicine, 249*(5), 413–422. https://doi.org/10.1046/j.13652796.2001.00812.x
Lazar, R. M., & Antoniello, D. (2008). Variability in recovery from aphasia. *Current Neurology and Neuroscience Reports, 8*(6), 497–502. https://doi.org/10.1007/s11910-008-0079-x
Lazar, R. M., Minzer, B., Antoniello, D., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2010). Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke, 41*(7), 1485–1488. https://doi.org/10.1161/STROKEAHA.109.577338
Lazar, R. M., Speizer, A. E., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2008). Variability in language recovery after first-time stroke. *Journal of Neurology, Neurosurgery, and Psychiatry, 79*(5), 530–534. https://doi.org/10.1136/jnnp.2007.122457
Leonard, C., Laird, L., Burianová, H., Graham, S., Grady, C., Simic, T., et al. (2015). Behavioural and neural changes after a "choice" therapy for naming deficits in aphasia: Preliminary findings. *Aphasiology, 29*(4), 506–525.
Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology, 22*(9), 923–947.
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492
Madden, E. B., Robinson, R. M., & Kendall, D. L. (2017). Phonological treatment approaches for spoken word production in aphasia. *Seminars in Speech and Language, 38*(1), 62–73. https://doi.org/10.1055/s-0036-1597258
Marcotte, K., Laird, L., Bitan, T., Meltzer, J., Graham, S., Leonard, C., et al. (2018). Therapy-induced neuroplasticity in chronic aphasia after Phonological Component Analysis: A matter of intensity. *Frontiers in Neurology, 9*, 1–7. https://doi.org/10.3389/fneur.2018.00225
Marshall, R. C., Neuburger, S. I., & Phillips, D. S. (1992). Effects of facilitation and cueing on labelling of 'novel' stimuli by aphasic subjects. *Aphasiology, 6*(6), 567–583. https://doi.org/10.1080/02687039208249492
Martin, N., Fink, R. B., Renvall, K., & Laine, M. (2006). Effectiveness of contextual repetition priming treatments for anomia depends on intact access to semantics. *Journal of the International Neuropsychological Society, 12*(6), 853–866. https://doi.org/10.1017/S1355617706061030
Martin, N., Fink, R., & Laine, M. (2004). Treatment of word retrieval deficits with contextual priming. *Aphasiology, 18*(5/6/7), 457–471.
Martin, N., Fink, R., Laine, M., & Ayala, J. (2004). Immediate and short-term effects of contextual priming on word retrieval in aphasia. *Aphasiology, 18*(10), 867–898. https://doi.org/10.1080/02687030444000390
McNeil, M. R., Odell, K., & Tseng, C. H. (1991). Toward the integration of resource allocation into a general theory of aphasia. *Clinical Aphasiology, 20*, 21–39.
Mengad, S., Mosu, R., Laird, L., Gordon, J. K., Rochon, E., & Leonard, C. (2019). The influence of phonological components analysis treatment on lexical access in individuals with aphasia. *Conference Abstract: Academy of Aphasia 55th Annual Meeting*. https://doi.org/10.3389/conf.fnhum.2017.223.00031
Meteyard, L., & Bose, A. (2018). What does a cue do? Comparing phonological and semantic cues for picture naming in aphasia. *Journal of Speech Language and Hearing Research, 61*(3), 658–674. https://doi.org/10.1044/2017_JSLHR-L-17-0214
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. https://doi.org/10.1037/xlm0000091
Minkina, I., Oelke, M., Bislick, L. P., Brookshire, C. E., Hunting Pompon, R., Silkes, J. P., et al. (2016). An investigation of aphasic naming error evolution following phonomotor treatment. *Aphasiology, 30*(8), 962–980. https://doi.org/10.1080/02687038.2015.1081139
Nardo, D., Holland, R., Leff, A. P., Price, C. J., & Crinion, J. T. (2017). Less is more: Neural mechanisms underlying anomia treatment in chronic aphasic patients. *Brain, 140*(11), 3039–3054. https://doi.org/10.1093/brain/awx234
Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology, 16*(10-11), 935–979.
Nickels, L., & Best, W. (1996). Therapy for naming disorders (part I): Principles, puzzles and progress. *Aphasiology, 10*(1), 21–47. https://doi.org/10.1080/02687039608248397
Off, C. A., Griffin, J. R., Spencer, K. A., & Rogers, M. A. (2016). The impact of dose on naming accuracy with persons with aphasia. *Aphasiology, 30*(9), 983–1011. https://doi.org/10.1080/02687038.2015.1100705
Pedersen, P. M., Stig Jørgensen, H., Nakayama, H., Raaschou, H. O., & Olsen, T. S. (1995). Aphasia in acute stroke: Incidence, determinants, and recovery. *Annals of Neurology, 38*(4), 659–666. https://doi.org/10.1002/ana.410380416
Persad, C., Wozniak, L., & Kostopoulos, E. (2013). Retrospective analysis of outcomes from two intensive comprehensive aphasia programs. *Topics in Stroke Rehabilitation, 20*(5), 388–397. https://doi.org/10.1310/tscir2001-38
Plowman, E., Hentz, B., & Ellis, C. (2012). Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors. *Journal of Evaluation in Clinical Practice, 18*(3), 689–694. https://doi.org/10.1111/j.1365-2753.2011.01650.x
Price, C. J., Seghier, M. L., & Leff, A. P. (2010). Predicting language outcome and recovery after stroke: The PLORAS system. *Nature Reviews Neurology, 6*(4), 202–210. https://doi.org/10.1038/nrneurol.2010.15

Purdy, M. (2002). Executive function ability in persons with aphasia. *Aphasiology, 16*(4-6), 549–557.
Ramsberger, G. (2005). Achieving conversational success in aphasia by focusing on non-linguistic cognitive skills: A potentially promising new approach. *Aphasiology, 19*(10-11), 1066–1073.
Rapp, Brenda, & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460–499. https://doi.org/10.1037/0033-295X.107.3.460
Raymer, A. M., Beeson, P., Holland, A., Kendall, D., Maher, L. M., Martin, N., … Gonzalez Rothi, L. J. (2008). Translational research in aphasia: From neuroscience to neurorehabilitation. *Journal of Speech Language and Hearing Research, 51*(1), S259–S275.
Riddoch, M. J., & Humphreys, G. W. (1993). *Birmingham object recognition battery*. Hove, UK: Lawrence Erlbaum Associates Ltd., Publishers.
Rijntjes, M. (2006). Mechanisms of recovery in stroke patients with hemiparesis or aphasia: New insights, old questions and the meaning of therapies. *Current Opinion in Neurology, 19*(1), 76–83. https://doi.org/10.1097/01.wco.0000203886.28068.38
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.
Robey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech Language and Hearing Research, 41*(1), 172–187. https://doi.org/10.1044/jslhr.4101.172
Rochon, E., Leonard, C., Burianova, H., Laird, L., Soros, P., Graham, S., et al. (2010). Neural changes after phonological treatment for anomia: An fMRI study. *Brain and Language, 114*, 164–179.
Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003
Ross Graham, J., Pereira, S., & Teasell, R. (2011). Aphasia and return to work in younger stroke survivors. *Aphasiology, 25*(8), 952–960. https://doi.org/10.1080/02687038.2011.563861
Schwartz, M. (2014). Theoretical analysis of word production deficits in adult aphasia. *Philosophical Transactions Biological Sciences, 369*, 1–10.
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*, 228–264.
Simic, T., Bitan, T., Turner, G., Chambers, C., Goldberg, D., Leonard, C., et al. (2019). The role of executive control in post-stroke aphasia treatment. *Neuropsychological Rehabilitation*, 1–40. https://doi.org/10.1080/09602011.2019.1611607
Thomas, S. A., & Lincoln, N. B. (2008). Predictors of emotional distress after stroke. *Stroke, 39*(4), 1240–1245. https://doi.org/10.1161/STROKEAHA.107.498279
Tsouli, S., Kyritsis, A. P., Tsagalis, G., Virvidaki, E., & Vemmos, K. N. (2009). Significance of aphasia after first-ever acute stroke: Impact on early and late outcomes. *Neuroepidemiology, 33*(2), 96–102.
Tuomiranta, L., Grönholm-Nyman, P., Kohen, F., Rautakoski, P., Laine, M., & Martin, N. (2011). Learning and maintaining new vocabulary in persons with aphasia: Two controlled case studies. *Aphasiology, 25*(9), 1030–1052. https://doi.org/10.1080/02687038.2011.571384
Tuomiranta, L., Rautakoski, P., Rinne, J. O., Martin, N., & Laine, M. (2012). Long-term maintenance of novel vocabulary in persons with chronic aphasia. *Aphasiology, 26*(8), 1053–1073. https://doi.org/10.1080/02687038.2012.693583
Vallila-Rohter, S., & Kiran, S. (2013). Nonlinguistic learning in individuals with aphasia: Effects of training method and stimulus characteristics. *American Journal of Speech-language Pathology, 22*(2), S426–S437. https://doi.org/10.1044/1058-0360(2013/12-0087)
Van Hees, S., Angwin, A., McMahon, K., & Copland, D. (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia. *Neuropsychological Rehabilitation, 23*(1), 102–132.
Van Hees, S., McMahon, K., Angwin, A., de Zubicaray, G., & Copland, D. A. (2014). Neural activity associated with semantic versus phonological anomia treatments in aphasia. *Brain and Language, 129*(1), 47–57. https://doi.org/10.1016/j.bandl.2013.12.004
StataCorp. (2019). *Stata Statistical Software: Release 16*. College Station, TX: StataCorp: LLC.
Walker & Hickok (n.d.). Computational Models of Aphasia. Retreived data on May 25th, 2019 from, http://cogsci.uci.edu/~alns/webfit.html.
Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop "fishing"? A review of generalisation following aphasia intervention. *Aphasiology, 29*(11), 1240–1264. https://doi.org/10.1080/02687038.2015.1027169
Wertz, R. T., LaPointe, L. L., & Rosenbeck, J. C. (1984). *Apraxia of speech in adults: The disorder and its management*. Orlando, FL: Grune & Stratton.
Wisenburn, B., & Mahoney, K. (2009). A meta-analysis of word-finding treatments for aphasia. *Aphasiology, 23*(11), 1338-135.