

Typicality-based semantic treatment for anomia results in multiple levels of generalization

Gilmore, N., Meier, E.L., Johnson, J.P., & Kiran, S.

Natalie Gilmore *

Aphasia Research Laboratory

Speech, Language & Hearing Sciences, Boston University, Boston, MA

ngilmore@bu.edu

*Corresponding Author

Erin L. Meier

Aphasia Research Laboratory

Speech, Language & Hearing Sciences, Boston University, Boston, MA

emeier@bu.edu

Jeffrey P. Johnson

Aphasia Research Laboratory

Speech, Language & Hearing Sciences, Boston University, Boston, MA

johnsojp@bu.edu

Swathi Kiran

Aphasia Research Laboratory

Speech, Language & Hearing Sciences, Boston University, Boston, MA

kirans@bu.edu

Abstract

This study investigated the effects of typicality-based semantic feature analysis (SFA) treatment on generalization across three levels: untrained related items, semantic/phonological processing tasks, and measures of global language function. Using a single-subject design with group-level analyses, 27 persons with aphasia (PWA) received typicality-based SFA to improve their naming of atypical and/or typical exemplars. Progress on trained, untrained, and monitored items was measured weekly. Pre- and post-treatment assessments were administered to evaluate semantic/phonological processing and overall language ability. Ten PWA served as controls. For the treatment participants, the likelihood of naming trained items accurately was significantly higher than for monitored items over time. When features of atypical items were trained, the likelihood of naming untrained typical items accurately was significantly higher than for untrained atypical items over time. Significant gains were observed on semantic/phonological processing tasks and standardized assessments after therapy. Different patterns of near and far transfer were seen across treatment response groups. Performance was also compared between responders and controls. Responders demonstrated significantly more improvement on a semantic processing task than controls, but no other significant change score differences were found between groups. In addition to positive treatment effects, typicality-based SFA naming therapy resulted in generalization across multiple levels.

Keywords: aphasia, generalization, typicality, rehabilitation

Introduction

Generalization, the acquisition of a trained behavior influencing another related behavior or being performed in a new context (Webster, Whitworth, & Morris, 2015), is the ultimate goal of language rehabilitation and a sign of true treatment success. Without generalization, it would be necessary to train all items in all situations, which is not a viable option for clients and clinicians alike (Thompson, 1989). Anomia, the most pervasive symptom of aphasia, refers to difficulty retrieving the word for a concept that was accessible prior to brain injury (Goodglass & Wingfield, 1997). The multitude of words used to perform daily activities combined with the time pressure of most rehabilitation settings make generalization particularly important when developing naming treatments.

Cognitive psychological models propose that the naming process involves a series of steps starting with input processing (e.g., visual recognition of a pictured item) and ending with output processing (i.e., producing the name of the object aloud) (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, Roelofs, & Meyer, 1999; Plaut, 1995; Rapp & Goldrick, 2000). One example is the interactive two-step model (Dell et al., 1997; Schwartz, 2013), in which lexical access is a two-step process and word knowledge is represented in a network of three layers: semantic, lemma, and phoneme. The first step involves accessing the lemma, or a non-phonological representation of a word, which includes semantic and grammatical information. The second step involves linking the lemma to the phonological form of the word, or the sequence of phonemes associated with the word's meaning. Within this model, activation is transmitted between network layers in both a top-down and bottom-up manner.

Across lexical access models (Dell et al., 1997; Levelt et al., 1999), semantic and phonological processing occurs in stages. However, the degree to which these stages of

processing influence one another has been debated. Some models suggest that these stages are distinct, occurring in sequential order in one direction (Levelt et al., 1999). Other models suggest that these stages of processing are indeed interactive (Dell et al., 1997; Goldrick & Rapp, 2007; Plaut, 1995; Ralph, Moriarty, & Sage, 2002; Rapp & Goldrick, 2000), but the degree of interaction varies. The extent of interactivity can be largely restricted (Goldrick & Rapp, 2007; Rapp & Goldrick, 2000) (i.e., only lemma selection and lexical phonological processing influence one another; post-lexical phonological processing is independent), highly-interactive with some restriction (i.e., bi-directional activation with greater semantic influence during step one and greater phonological influence during step two) (Dell et al., 1997; Schwartz, 2013), or fully interactive with no restriction (Plaut, 1995; Ralph et al., 2002).

Models of lexical retrieval have been useful for testing hypotheses related to different aspects of language processing and for developing theoretically-driven treatment approaches (Goodglass & Wingfield, 1997). Semantic treatments, such as semantic feature analysis (SFA; Boyle, 2004; Boyle & Coelho, 1995; Coelho, McHugh, & Boyle, 2000; Davis & Stanton, 2005; DeLong, Nessler, Wright, & Wambaugh, 2015; Hashimoto & Frome, 2011; Kristensson, Behrns, & Saldert, 2015; Lowell, Beeson, & Holland, 1995), target the semantic system directly and result in both acquisition and generalization effects (Nickels, 2002; Wisenburn & Mahoney, 2009). Using the interactive activation model described above, this finding should not be surprising in that generalization to untrained items would be expected from both semantic, and phonologically-based treatments. However, although phonologically-based treatment approaches for anomia (i.e., phonologic, orthographic, indirect, guided and mixed cueing) have been shown to be effective for improving trained items, generalization to untrained items is less often seen as a function of these approaches with the exception of those incorporating phonomotor therapy

(Madden, Robinson, & Kendall, 2017; Wisenburn & Mahoney, 2009). Thus, a semantic-based approach, SFA, was chosen for this study, given its strong potential for promoting acquisition and generalization effects for PWA with anomia (Boyle, 2017).

SFA is an effective therapy for treating naming deficits for individuals with a range of aphasia types and severities (Boyle, 2010). The driving premise of this type of treatment is that when individuals generate semantic features of a target word (i.e., accessing their semantic network), they improve their ability to retrieve the target because they have strengthened access to its conceptual representation. The theoretical mechanism by which SFA promotes generalization comes from the spreading activation theory (Collins & Loftus, 1975), which posits that accessing/activating a particular lemma (or its features) results in activation of the lemmas of semantically-related concepts. Additionally, activation of the target lemma and the off-target semantically-related lemmas spreads downstream to the phoneme-level, thereby activating their unique phonemic representations. Activation of the target's semantic and phonological neighbors is thought to strengthen their representations, making them more accessible for retrieval despite not being explicitly trained (i.e., generalization to untrained semantically and/or phonologically-related items). However, generalization patterns from SFA-based treatment approaches have been mixed. Some studies have shown changes on untrained items (Boyle & Coelho, 1995; Coelho et al., 2000) and untrained tasks (Coelho et al., 2000; Hashimoto & Frome, 2011), yet others have not (Boyle & Coelho, 1995; Rider, Wright, Marshall, & Page, 2008; Wambaugh, Mauszycki, & Wright, 2014). Therefore, further investigation of generalization patterns to untrained items and untrained tasks associated with SFA-based treatment is warranted.

The treatment paradigm implemented in this study is further grounded in an influential language treatment principle: the complexity account of treatment efficacy (CATE) hypothesis

(Thompson, Shapiro, Kiran, & Sobecks, 2003; Thompson & Shapiro, 2007). CATE suggests that there is greater generalization from trained items to untrained items when more complex stimuli are trained than the reverse, as long as the untrained items/structures are a subset of the trained items. Kiran and colleagues have tested the CATE hypothesis in semantics with PWA (Kiran, 2007, 2008; Kiran & Johnson, 2008; Kiran, Sandberg, & Sebastian, 2011; Kiran & Thompson, 2003a). Their work has capitalized on previous research regarding the typicality effect, which infers special status for typical examples in a semantic category (i.e., accessed faster and more accurately than atypical examples; Rosch, 1975). An exemplar's typicality reflects how closely its semantic features match those of a prototypical exemplar of the category (e.g., prototypical bird: "robin"; typical bird: "sparrow"; atypical bird: "ostrich") based on an item's semantic features. The typicality effect has been demonstrated to influence semantic processing in a number of different experimental tasks with healthy controls (Hampton, 1995, 1979; Rosch, 1975; Rosch & Mervis, 1975) and PWA (Kiran, Ntourou, & Eubank, 2007; Kiran et al., 2011; Kiran & Thompson, 2003a; Meier, Lo, & Kiran, 2016; Rogers, Patterson, Jefferies, & Lambon Ralph, 2015; Rossiter & Best, 2013).

Typicality-based SFA treatment (e.g., (Kiran & Thompson, 2003b) likely results in generalization for two primary reasons. First, training features of atypical items strengthens access not only to those items but also to those of typical items whereas training features of typical items reinforces access to typical items only. This pattern occurs because atypical examples contain fewer prototypical features, possess more distinctive features, and share fewer features with other examples in the category. Thus, treatment focused on atypical exemplars inherently involves analysis of atypical features and typical features, exposing the patient to features representing the breadth of the category. In contrast, typical examples in the category

contain more core and prototypical features, possess fewer distinctive features, and have a higher number of shared features with other typical items. Therefore, there is less generalization to atypical items from training typical exemplars as features representing the spread of the category do not have to be trained to improve naming of typical items. In line with the CATE hypothesis, it would follow then that training more semantically-complex atypical exemplars of a category would result in gains on less semantically-complex typical exemplars of that same category without direct training. Second, given that the semantic and phonological levels are highly-interactive in Dell's model of lexical access (Schwartz, 2013), phonological representations of both atypical and typical items would also be strengthened when atypical items are trained (Kiran, 2007).

Using a typicality-based SFA treatment paradigm, Kiran and colleagues (Kiran, 2008; Kiran & Johnson, 2008; Kiran et al., 2011; Kiran & Thompson, 2003b) have demonstrated across different category types (i.e., animate, inanimate, well-defined) that PWA show gains in naming untrained typical items when features of atypical items are trained. However, these studies had relatively small sample sizes ($n=3-5$), were primarily conducted with participants with fluent and anomic aphasia, and found generalization patterns in line with the CATE hypothesis in some, but not all, participants. Stanczak and colleagues (Stanczak, Waters, & Caplan, 2006) also examined generalization patterns in two PWA from typicality-based SFA treatment and found mixed results with respect to expected typicality-based generalization. Specifically, one participant demonstrated greater generalization when atypical items were trained, but also marginally significant generalization when typical items were trained, which is inconsistent with the CATE hypothesis and previous work by Kiran and colleagues (Kiran & Thompson, 2003b). Additionally, Wambaugh et al. (Wambaugh, Mauszycki, Cameron, Wright,

& Nessler, 2013) used a typicality-based SFA treatment incorporating strategy training to target anomia in individuals with primarily Broca's aphasia and found limited expected within-category generalization overall (i.e., 4/9 participants showed no changes on untrained items) as well as some unexpected generalization patterns (i.e., three participants showed some generalization to atypical items from training on typical items). Results of this latter study are also inconsistent with generalization patterns suggested by spreading activation theory in the context of SFA treatment (Boyle & Coelho, 1995) and the CATE hypothesis with respect to typicality (Kiran, 2007). Taken together, these findings raise important questions about the consistency of generalization patterns that would be predicted by CATE, which could be addressed by studying a larger participant sample with a variety of aphasia types and severities.

Due to the nature of SFA treatment, generalization to other cognitive-linguistic abilities is possible. For example, while only the semantic system is directly targeted in therapy sessions, participants are also required to produce the names of items aloud. Thus, both semantic and phonological stages of lexical retrieval are rehearsed during therapy, and improvements on measures of semantic and phonological processing (i.e., "near transfer") would be likely. Gains on standardized assessments of cognitive-linguistic function (i.e., "far transfer") would then be expected in addition to the aforementioned treatment-related changes in language skills. At present, changes in phonological processing after typicality-based SFA treatment have not been consistently measured and effects have been variable on standardized semantic assessments (Kiran, 2008; Kiran & Johnson, 2008; Kiran et al., 2011; Kiran & Thompson, 2003b). Some PWA have also demonstrated improvements in broad language function after therapy, as indicated by their Aphasia Quotient on the Western Aphasia Battery-Revised (Kertesz, 2007), although the magnitude of those improvements varied by study and patient. Thus in addition to

examining treatment effects and generalization to untrained items, another aim of the current study was to investigate the near- and far-transfer effects of typicality-based SFA treatment.

To summarize, typicality-based SFA treatment has resulted in both the acquisition of trained items and generalization to untrained related items. However, the extent of generalization has varied across studies. Additionally, the effects of this treatment paradigm on untrained tasks of semantic and phonological processing have not been thoroughly examined and transfer of function to standardized tests has not been robust across all studies. Furthermore, studies conducted to examine the generalization effects of typicality-based treatments have historically implemented single-subject design methodology and examined gains at an individual subject-level, whereas the current study implemented group-level analyses to shed new light on the efficacy of this approach with a larger, more diverse participant sample. In light of these gaps, the following research questions were developed:

1. *Do PWA demonstrate greater improvement in their trained categories relative to their monitored categories after treatment?* We hypothesized that PWA would show greater gains in the trained categories than monitored categories (Kiran, 2008; Kiran & Thompson, 2003b) further confirming the efficacy of semantic-based treatment.
2. *Do PWA show greater generalization to untrained typical items than untrained atypical items after treatment?* Based on the CATE hypothesis, we hypothesized that PWA would show greater improvement in untrained typical items relative to untrained atypical items following training on atypical and typical category exemplars, respectively (Kiran, 2007).
3. *Do PWA demonstrate near transfer to untrained tasks of semantic and phonological processing after treatment?* We hypothesized that PWA would show improvements on tasks that tap into aspects of semantic and phonological processing that overlap with the

treatment tasks given the highly-interactive nature of semantic and phonological levels in the two-step interactive model of lexical access (Dell et al., 1997).

4. *Do PWA show far transfer to broad language skills after treatment?* We hypothesized that PWA would show gains on untrained language tasks and modalities. Participants should be able to rely on improved lexical retrieval skills during standardized assessments of naming and spoken language post-treatment. Also, typicality-based SFA treatment involves answering auditory and written feature questions; thus, improvements in auditory and reading comprehension may be expected post-therapy (Ellis & Young, 1988).

Methods

The present project was completed under the Center for the Neurobiology of Language Recovery (NIH/NIDCD 1P50DC012283; PI: Cynthia Thompson) (<http://cnlr.northwestern.edu/>).

Participants

Thirty-one individuals (20 male) with chronic aphasia following a left-hemisphere (LH) stroke were recruited from the greater Boston and Chicago areas. Ten of these individuals served as natural history controls in that they were tested at baseline and, again after a three- to six-month period without treatment. Twenty-seven of these individuals, five of whom started out in the natural history control group (BU08/BUc01, BU19/BUc02, BU16/BUc05, BU24/BUc06), participated in up to 12 weeks of therapy and thus comprised the treatment group in this study. See Table 1 for demographic information for both treatment and natural history control participants (i.e., age, months post onset (MPO), cognitive-linguistic severity, aphasia type, apraxia of speech (AOS) status, baseline performance on the confrontation naming screener, treatment/generalization effect sizes (ES), and average accuracy on trained and untrained

categories at pre- and post-treatment). All participants demonstrated adequate vision and hearing (i.e., passed screening at 40db HL bilaterally/or able to be corrected with increased volume); were English-proficient pre-morbidly (per self or family report); presented with stable neurological and medical status; and were not receiving concurrent individual speech and language therapy. None of the participants had neurodegenerative disease or active medical conditions that affected their ability to participate in the study. Written consent to participate in the study was obtained in accordance with the Boston University Institutional Review Board (IRB) protocols for 29 patients and per the Northwestern University IRB protocols for two participants.¹ Of note, data from the 27 individuals in the treatment group were also included in a separate study examining the influence of baseline language and cognitive skills on treatment success (Gilmore, Meier, Johnson, & Kiran, 2018, under revision).

The diagnosis of aphasia was determined through administration of the WAB-R. Participants were administered a 180-item confrontation naming screener consisting of items from five semantic categories (i.e., *birds*, *vegetables*, *fruit*, *clothing* and *furniture*) including 36 exemplars of each category, and further divided into half-categories by typicality (i.e., 18 typical; 18 atypical). During the screener, pictures were presented in random order and participants were instructed to name the items. In addition to production of the intended target, responses were considered correct if they were self-corrections, altered due to dialectal differences, distortions or substitutions of one phoneme, and/or produced correctly following a written self-cue. Participants were included in the study if they demonstrated stable performance of $\leq 65\%$ average accuracy in two different half-categories (e.g., *Atypical Birds*, *Typical Clothing*) across multiple baselines of the screener. Individuals with AOS were enrolled in the study as long as

¹Participants were recruited from the Chicago area as the Center for the Neurobiology of Language Recovery project was multi-site in nature. The same study procedures were followed for all participants.

they were stimuable to produce targets with a verbal model from a clinician. The AOS rating in Table 1 was based on the Screen for Dysarthria and Apraxia of Speech (S-DAOS) (Dabul, 2000) and clinical judgment. In cases when AOS severity per the S-DAOS and primary clinician's judgment did not align (e.g., patients with phonological errors judged as having AOS on S-DAOS), at least two trained speech-language pathologists (SLP) judged absence or presence of AOS based on pre-treatment speech samples.

Standardized Assessment

For the treatment group, standardized assessments were administered before and after therapy to assess far transfer to global language skills as a function of typicality-based SFA treatment, including the WAB-R (i.e., WAB Aphasia Quotient [AQ], Language Quotient [LQ], Cortical Quotient [CQ]), Cognitive Linguistic Test (Helm-Estabrooks, 2001), Boston Naming Test (BNT; Goodglass, Kaplan, & Weintraub, 2001), Pyramids and Palm Trees Test (PAPT; Howard & Patterson, 1992); subtests 1 (Same-Different Discrimination Using Nonword Minimal Pairs) and 51 (Word Semantic Association) of the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992), and the Northwestern Naming Battery Confrontation Naming Subtest (NNB-CN; Thompson, Lukic, King, Mesulam, & Weintraub, 2012). Of note, in this study, gains on the PAPT and PALPA subtests were considered far transfer as we hypothesized generalization to be less common to these measures given their minimal stimulus overlap with the treatment tasks.

Participants in the natural history control group were administered² the above assessments at the time of their enrollment in the study and, again after a three- to six-month

²Due to minor changes in study protocols over the course of the experiment, BUc07 did not receive Part Two of the WAB-R, which is necessary for computing the WAB-LQ and CQ; and BUc01, BUc02, and BUc08 were not given the BNT during the natural history phase. Thus, these patients were excluded from analyses using scores from those assessments.

period of no-treatment. Average scores for both the treatment and natural history control groups on standardized measures are shown in Table 2.

Semantic and Phonological Behavioral Tasks

A set of six tasks, developed and validated for use with PWA (Meier et al., 2016), were also administered³ before and after therapy to measure near transfer to semantic and phonological processing skills. There were three semantic tasks (i.e., Superordinate Category Verification [SCV], Category Coordinate Judgment [CCJ], Semantic Feature Verification [SFV]) and three phonological tasks (i.e., Syllable Judgment [SJ], Rhyme Judgment [RJ], Phoneme Verification [PV]) with two different versions (i.e., Phonological No-Name [PVNN], Phonological Name Provided [PVNP]). During the PVNN tasks, the name of the target was not provided and the participant had to access the word form to perform the task, thereby measuring processing at the level of the phonological output lexicon (POL). During the PVNP tasks, the name of the target was provided in order to measure processing at the level of the phonological output buffer. Both accuracy and response time data were collected. Table 2 contains treatment and natural history control group results for accuracy and reaction time, respectively, on both the semantic and phonological tasks.

Stimuli

This experiment included items and their relevant features from three animate categories (i.e., *birds, vegetables, fruit*) and two inanimate categories (i.e., *clothing and furniture*).

Items

³Of note, two treatment participants (i.e., BU06 and BU07) were excluded from analysis of the Syllable Judgment tasks because the response configuration (i.e., two- versus three-button response) differed at pre- and post-treatment time points for these participants. Additionally, one treatment participant was excluded from the Superordinate Category Verification analysis (i.e., BU11) and another participant from the Semantic Feature Verification analysis (i.e., BU12) because they did not complete the task at both time points. Lastly, BUc05, BUc07, BUc08, and BUc10 did not receive the semantic and phonological processing tasks before and after a no-treatment period and thus, were excluded from those analyses.

Picture naming stimuli were selected based on results of an MTurk pilot task (<https://www.mturk.com/mturk/>). Participants rated the typicality of items from semantic categories on a scale of one to five (1= most typical; 5 = most atypical), or as “not a member.” Items were assigned typicality rankings based on the average MTurk participant rating and organized from most to least typical. For use in the treatment tasks, eighteen items from the top and bottom of the list were selected as typical and atypical exemplars for each of the categories, respectively. A representative photo was chosen for each item for use during the naming screener, probes and treatment tasks. See Supplemental Material Table 1 for the average typicality rating for atypical and typical exemplars of each category used in the treatment and Supplemental Material Table 2 for a list of stimuli used in the tasks.

Semantic features

The semantic features used in the treatment tasks were also based on the results of an MTurk pilot task (<https://www.mturk.com/mturk/>). Participants in the pilot indicated whether a particular feature applied to the treatment items (i.e., yes/no). Core (i.e., relevant to all items in the category), prototypical (i.e., those relevant to typical items) and distinctive (i.e., those relevant to atypical items) features were assigned as applicable or not applicable to target treatment items, based on average of pilot participants’ responses. See Supplemental Material Table 3 for a sample of features used in the tasks.

Design

This study implemented a single-subject experimental design with group-level analyses. In line with single-subject design methodology, all participants completed three baseline naming assessments (i.e., before treatment or a period of no-treatment). Based on their performance on the naming screener, they were pseudo-randomly assigned two trained categories, a monitored

category, and an assessed category in a counter-balanced fashion. Treatment participants were only exposed to items in their monitored category during the weekly probe and to items in their assessed category during pre- and post-treatment assessments. They attended up to 24 two-hour treatment sessions either two or three times per week or until they reached criterion (i.e., $\geq 90\%$ accuracy on two consecutive weekly probes for both treatment categories). Each treatment participant was probed weekly on the full set of items in their two trained categories (i.e., 36 trained; 36 untrained) and their monitored category (i.e., 36 untrained). These data were utilized in group-level statistical analyses of treatment and generalization effects. The natural history control participants were not trained on any items at any point and performance across all of their assigned categories was only measured before and after a period of no intervention, not with weekly probes. See Supplemental Table 4 for treatment assignments for both groups.

Treatment

For each treatment participant, the two assigned half-categories were trained every week. Therapy tasks were administered through a computer with clinician assistance using E-prime (Schneider, Eschman, & Zuccolotto, 2002). Treatment proceeded according to a series of steps: 1) category sorting; 2) initial naming attempt; 3) written feature verification; 4) feature review; 5) auditory feature verification; and 6) second naming attempt. Of note, five participants who did not appear to be responding favorably to the treatment (BU16, BU20, BU21, BU22, BU24) were given a home exercise program to support carryover of gains to subsequent treatment sessions. These participants were interested in working on the stimuli outside of therapy sessions and demonstrated sufficient auditory comprehension and repetition to perform a home exercise program independently. See Supplemental Material Table 5 for full details of the treatment protocol, including home exercise programming. A 108-item probe consisting of a subset of

items from the full 180-item screener (i.e., 36 items each from three assigned categories: two trained and one monitored) was administered at the start of every other treatment session. To quantify treatment related-gains, treatment (T1, T2) and generalization (G1, G2) effect sizes (ES) for each half-category were calculated by subtracting the pre-treatment average score from the post-treatment average score and dividing that value by the pre-treatment standard deviation (Beeson & Robey, 2006), for both trained categories as shown in Table 1.

Data Analysis

Statistical analyses and plotting were completed in RStudio (R Studio Team, 2015) with the packages lme4 (Bates, Maechler, Bolker, & Walker, 2016), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2016), tidyverse (Wickham, 2017), and coin (Hothorn, Hornik, van de Wiel, Winell, & Zeileis, 2016). Analysis of treatment-related changes in participants' naming accuracy (i.e., item score) was completed using logistic mixed-effects regression models. Random effects included subjects and items with the maximal random effects structure permitted by the study design (Barr, Levy, Scheepers, & Tily, 2013). When convergence was not reached with the maximal model, random slopes that accounted for the least variance were removed in iterative fashion until convergence was reached. Treatment-induced changes in performance on the semantic and phonological processing tasks (i.e., accuracy and reaction time [RT]) and standardized assessments were analyzed using Wilcoxon signed-rank tests, as the data were non-normal. To account for leftward skew in the RT data, RTs more than three standard deviations from the overall mean reaction time (i.e., ≥ 2254.77 ms in the participant data; ≥ 12638.56 ms in the natural history control data) were considered outliers and removed from the dataset before statistical analysis.

In order to determine if the changes seen in the treatment group were due to the intervention as opposed to the passage of time, or repeated exposure to the tests/tasks administered as part of the study, statistical analyses were also performed comparing the treatment and natural history control groups.

Results

Experimental Patients: Treatment effects

To capture treatment effects at the group level, participants' naming accuracy (i.e., item score) on trained and monitored categories at all probe sessions (i.e., baseline, treatment, and post-treatment probes) was the dependent variable in a logistic mixed-effects model. Fixed effects included time (i.e., number of baseline, weekly probe and post-treatment assessments), a single variable for category exposure/typicality (i.e., trained/untrained, monitored, atypical/typical), and their interaction. To account for differences among participants' treatment response and across items, random effects included intercepts for subjects and items. Results demonstrated a significant time-by-training interaction effect, as depicted in Figure 1a. The likelihood of an accurate response on trained atypical and trained typical items was significantly greater than that of monitored items over time, $\beta = .14$, $SE = .01$, $t = 14.77$, $p < .001$, and $\beta = .08$, $SE = .01$, $t = 9.5$, $p < .001$, respectively. This finding provides evidence that typicality-based SFA treatment was effective for improving items that were directly trained.

Follow-up analyses of treatment effects

Despite robust treatment acquisition effects at the group-level, inspection of effect sizes for each participant revealed variability in the response to treatment (see Table 1). Therefore, Beeson and Robey's benchmarks (Beeson & Robey, 2006) were used to classify participants based on the effect sizes in their trained categories as responders (i.e., effect size greater \geq "small

effect,” or 4.0 in at least one trained category) and nonresponders (i.e., effect sizes < 4.0 in both trained categories). Then, follow-up analyses were conducted to assess treatment and generalization effects, changes on the semantic and phonological behavioral tasks, and standardized assessments separately for the responder (n = 17) and nonresponder (n = 10) groups (as described below). The same statistical procedures used for the primary analyses (i.e., with the full sample not split into subgroups) were applied with the responder and nonresponder data sets.

For responders, results were consistent with that of the full group, in that they demonstrated a significant time-by-training interaction effect. The likelihood of an accurate response on trained atypical and trained typical items was significantly greater than that of monitored items over time, $\beta = .19$, $SE = .01$, $t = 15.3$, $p < .001$, $\beta = .12$ $SE = .01$, $t = 11.5$, $p < .001$, respectively, as depicted in Figure 1b.

For nonresponders, the likelihood of an accurate response on trained atypical items was significantly greater than that of monitored items over time, $\beta = .09$, $SE = .02$, $t = 5.4$, $p < .001$, but the same was not true of trained typical items, $\beta = .03$, $SE = .02$, $t = 1.75$, $p = .08$. As illustrated in Figure 1c, these results reveal that participants in this group did indeed make gains on items on which they were directly trained, although to a lesser extent than those in the responder group. Furthermore, training features of atypical items resulted in greater acquisition than those of typical items. Typical items may have been more challenging for PWA to distinguish from other trained items and thus harder to learn as they possess more shared features than atypical items, which are made up of more distinctive and non-overlapping features, possibly making them easier to acquire and retain.

Experimental participants: Generalization effects

Within-category transfer

To directly compare the degree of generalization on untrained typical and atypical items, naming accuracy (i.e., item score) on untrained half-categories (i.e., untrained atypical, untrained typical) was entered as the dependent variable in a separate logistic mixed-effects model. Fixed effects included time, category exposure/typicality (i.e., untrained atypical, untrained typical), and their interaction. Random effects included intercepts for subjects and item. This analysis also demonstrated a significant time-by-typicality interaction effect. The likelihood of an accurate response on untrained typical items was significantly greater than the likelihood of an accurate response for untrained atypical items over time, $\beta = .03$, $SE = .01$, $t = 2.28$, $p = .02$. Participants were more likely to accurately name untrained typical items than atypical items over time, as illustrated in Figure 2a, which supports the CATE hypothesis (i.e., training features of atypical items results in significant improvements on untrained typical items, but training features of typical items does not result in significant improvements on untrained atypical items).

Follow-up Analyses of within-category transfer

As shown in Figure 2b, when generalization effects were examined separately for responders, results were again consistent with the full sample, in that they demonstrated a significant time-by-typicality interaction effect. Specifically, the likelihood of accuracy was significantly greater for untrained typical items than untrained atypical items over time, $\beta = .04$, $SE = .01$, $t = 3.25$, $p < .001$.

For the nonresponders group, there was no significant difference in the likelihood of an accurate response on untrained typical items and untrained atypical items over time, $\beta = .003$, $SE = .02$, $t = .127$, $p = .90$, as shown in Figure 2c. Given that the nonresponders did indeed demonstrate a treatment effect for atypical trained items, this finding suggests that a more robust response to trained items may be necessary for generalization to untrained typical items.

Near transfer

To assess for near transfer of treatment gains, pre- and post-treatment accuracy and RTs on the behavioral tasks were compared using Wilcoxon signed-rank tests as the data were non-normal. After treatment, participants' accuracy significantly improved on the Syllable Judgment Name Provided (SJNP) task ($W = 55, Z = -2.12, p = .04, r = -.42$), the Category Coordinate Judgment (CCJ) task ($W = 72.5, Z = -2.38, p = .02, r = -.46$), and the Superordinate Category Verification (SCV) task ($W = 60, Z = -2.20, p = .03, r = -.43$) after treatment as illustrated in Figure 3a.

Participants' RTs were significantly faster after than before treatment on a number of tasks: Category Coordinate Judgment (CCJ: $W = 296, Z = 2.57, p = .01, r = .49$), Superordinate Category Verification (SCV: $W = 236, Z = 2.01, p = .049, r = .39$) and Semantic Feature Verification (SFV: $W = 284, Z = 3.25, p = .001, r = .64$), as illustrated in Figure 4a.

Follow-up analyses of near transfer

Responders' accuracy significantly improved on the Syllable Judgment Name Provided (SJNP) ($W = 16.5, Z = -2.05, p = .046, r = -.53$), and Superordinate Category Verification (SCV) tasks ($W = 6, Z = -2.93, p = .004, r = -0.71$) after treatment. Nonresponders did not change significantly in their accuracy on any of the tasks. These results are depicted in Figures 3b and 3c, respectively

Responders' RTs were significantly faster on a number of tasks: Phoneme Verification No Name (PVNN: $W = 138, Z = 2.91, p = .002, r = .71$), Category Coordinate Judgment (CCJ: $W = 119, Z = 2.01, p = .045, r = .49$), Superordinate Category Verification (SCV: $W = 131, Z = 2.58, p = .008, r = .63$) and Semantic Feature Verification (SFV: $W = 123, Z = 2.79, p = .005, r = .68$), as shown in Figure 4b. Again, as depicted in Figure 4c, nonresponders did not change

significantly in their reaction time on any of the tasks post-treatment. These results suggest that there was transfer to semantic and phonological processing from treatment targeting the semantic system, consistent with the interactive activation model.

Far transfer

Pre- and post-treatment performance on standardized assessments was analyzed to determine if gains from treatment transferred to broad language abilities (e.g., auditory comprehension, naming) using Wilcoxon signed-rank tests as the data were non-normal. Participants improved significantly on the WAB-R and the BNT (WAB-LQ: $W = 94$, $Z = -2.10$, $p = .03$, $r = .40$; WAB-CQ: $W = 95$, $Z = -2.26$, $p = .02$, $r = .43$; WAB-AQ: $W = 87.5$, $Z = -2.44$, $p = .01$, $r = .47$; BNT: $W = 51.5$, $Z = -2.71$, $p = .005$, $r = .52$) as illustrated in Figure 5a. These findings demonstrate that gains in lexical retrieval from typicality-based SFA-treatment transferred to untrained tasks of global language function.

Follow-up analyses of far transfer

As illustrated in Figure 5b, responders improved significantly on the WAB-R and the BNT (WAB-LQ: $W = 33.5$, $Z = -2.03$, $p = .04$, $r = .49$; WAB-CQ: $W = 29$, $Z = -2.25$, $p = .02$, $r = .55$; WAB-AQ: $W = 29$, $Z = -2.25$, $p = .02$, $r = .55$; BNT: $W = 12.5$, $Z = -2.96$, $p = .002$, $r = .72$). These findings demonstrate that gains in lexical retrieval from typicality-based SFA-treatment transfer to untrained tasks of global language function. As revealed in Figure 5c, nonresponders did not change significantly in their accuracy on any of the standardized assessments post-treatment.

These findings further support that gains in lexical-semantic processing from this treatment approach resulted in gains in broad language function, as participants who responded

the most to treatment improved significantly on these measures post-therapy, whereas participants who did not respond as favorably to treatment did not show these gains.

Control Group Analyses

Only the responders were compared to the natural history control group for two reasons: 1) Nonresponders did not demonstrate within-category transfer, making it unlikely that they would demonstrate near/far transfer. Thus, including them with the nonresponders may have limited the possibility of demonstrating a difference between the treatment and control conditions, and 2) It allowed for more equally sized groups for comparison (i.e., Responders = 17; Natural History Controls = 10).

To assess for significant pre-existing differences between the treatment and natural history control group, Wilcoxon rank sum tests were conducted as the data were non-normal. No significant differences were found between the treatment and natural history control groups in terms of age, MPO, aphasia severity ([WAB-AQ], overall cognition ([CLQT-CS], and baseline performance on the naming screener (all $p > .05$), suggesting any differences found between these two groups' performance was not attributable to these factors.

Treatment and within-category transfer

Given that the data were non-normal, Wilcoxon rank sum tests were conducted to evaluate for significant differences between responders and natural history control groups in the magnitude of change in their naming of trained/pseudo-trained and untrained/pseudo-untrained items after a 12-week period with/without intervention.

Wilcoxon rank sum tests were used to tests for significant differences in post-treatment change scores for the groups while accounting for non-normal data. The average number of items responders could name in their trained categories following treatment increased by 35%, a

significantly greater amount than the 6% exhibited by the natural history control group, ($V=7.5$, $p < .001$). While the responder group also increased the average number of items they could name in their untrained categories after therapy to a greater degree than the natural history group (10% vs. 2%, respectively), the difference between change scores was not significant, ($V=54.5$, $p=.13$). It is important to note that performance on untrained typical and atypical items were combined for this analysis, which might have diluted potential differences between groups.

Near transfer

Additionally, Wilcoxon rank sum tests were used to assess for significant differences between groups in change scores on the semantic and phonological processing tasks (i.e., accuracy and RT). The responder group improved significantly more post-treatment on the Superordinate Category Verification task than the natural history control group (4% vs. -2%, respectively), $V = 11$, $p = .005$. Otherwise, no significant differences in change scores were found between groups in terms of their accuracy or reaction time on the semantic and phonological processing tasks.

Far transfer

Lastly, Wilcoxon rank sum tests were used to assess for significant differences between groups in change scores on standardized cognitive-linguistic measures. No significant differences in changes scores were found between groups on standardized assessments of cognitive-linguistic functioning.

Discussion

The results of this study provide valuable insight into acquisition and generalization effects of typicality-based SFA treatment for individuals with chronic aphasia. Treatment resulted in significant improvement on items that were directly targeted in therapy. With respect

to generalization, we also found evidence for the CATE hypothesis, such that the likelihood of naming untrained typical items was significantly higher than the likelihood of naming untrained atypical items over time. Transfer to closely related measures of semantic and phonological processing (i.e., near transfer) and standardized assessments of general language function (i.e., far transfer) was also observed. As close inspection of individuals' treatment response (i.e., ESs) revealed inconsistencies in the effect across participants, PWA were split into treatment responder and treatment nonresponder groups and then, follow-up analyses were conducted. Interestingly, different patterns of near and far transfer were seen across groups. Lastly, responders improved significantly more than natural history controls on the Superordinate Category Verification task after treatment/no-treatment. This result suggests that the significant gains treatment participants showed on this task may indeed represent near transfer, especially given they performed this activity every treatment session, whereas natural history controls had no such opportunity. Nonetheless, it must be acknowledged that no other significant differences were seen between the responder and natural history control groups' post-treatment change scores on the semantic/phonological processing tasks or cognitive-linguistic assessments. Each of these results is discussed in greater detail below.

The significant findings of our first analysis provide additional evidence supporting the efficacy for typicality-based SFA treatment in a diverse participant sample. Participants with a range of aphasia types responded favorably to treatment specifically, improvements were observed in 9/10 participants with anomic aphasia, 2/2 participants with conduction aphasia, 5/11 with participants with Broca's aphasia, and 1/1 participant with transcortical motor aphasia. However, out of 27 participants, ten individuals (i.e., BU02, BU05, BU11, BU12, BU16, BU21, BU23, BU24, BU25, BU26) did not show a robust effect of treatment based on effect sizes less

than 4.0 (i.e., the cut-off for a “small” effect), for both trained categories. One potential explanation for this individual variability is cognitive-linguistic severity. Indeed, this hypothesis was supported by the finding of significant baseline differences between responders and nonresponders on both the WAB-AQ ($p < .01$) and CLQT-Composite Severity ($p < .01$), suggesting that participants with stronger pre-treatment language and cognitive skills responded more favorably to treatment. However, language and overall cognitive severity do not fully explain the results. For example, BU20 had severe aphasia (WAB-AQ = 13) and moderate cognitive deficits (CLQT = 45%), but did improve in therapy, whereas BU11 who exhibited mild aphasia (WAB-AQ = 92.1) and mild cognitive deficits (CLQT = 75%) did not improve in either of his trained categories. Furthermore, previous research does suggest that individuals with more severe cognitive-linguistic deficits can make gains in therapy (Des Roches, Mitko, & Kiran, 2017; Meier, Johnson, Villard, & Kiran, 2017). As such, it is possible that those participants with more severe aphasia and concomitant cognitive deficits may have required a higher dosage or intensity of treatment to show gains comparable to the participants with milder aphasia (Brady, Kelly, Godwin, Enderby, & Campbell, 2016). Nevertheless, most participants in this study improved as a function of typicality-based SFA treatment.

Our remaining analyses were conducted to elucidate generalization effects of typicality-based SFA treatment. Participants improved in their naming of untrained typical items following training on semantic features of atypical items, a generalization pattern which builds upon the existing evidence for the CATE hypothesis with respect to typicality in rehabilitation of anomia for PWA (Kiran, 2007, 2008; Kiran & Johnson, 2008; Kiran et al., 2011; Kiran & Thompson, 2003b; Sandberg & Kiran, 2014; Stanczak et al., 2006). In general, the individual generalization patterns among our participants also align with CATE predictions. Note that 19 participants (i.e.,

BU01, BU03, BU07, BU08, BU09, BU10, BU11, BU12, BU13, BU14, BU15, BU17, BU19, BU20, BU21, BU22, BU24, BU25, BU26) were trained on atypical items for at least one of their two trained categories, and thus, had the opportunity to show a generalization pattern consistent with CATE. Of these 19 participants, five (i.e., BU01, BU10, BU13, BU14, BU17) demonstrated an effect size over 2.0 for untrained typical items. ESs of 2.0 or higher were classified as meaningful as they are half the size of the benchmark for a small effect for trained items in lexical retrieval therapy (Beeson & Robey, 2006). Furthermore, only three out of 22 participants (i.e., BU01, BU10, BU15) demonstrated generalization patterns inconsistent with CATE (i.e., unexpectedly robust effect sizes in their untrained atypical categories). In summary, these results demonstrate that when features of atypical items were trained there was *greater* generalization to untrained within-category typical exemplars. That being said, when features of typical items were trained, some PWA may have exhibited additional gains in untrained atypical items as a result of spreading activation and the top-down trickle effect of targeting the semantic system (i.e., improved semantic processing led to better overall word retrieval). Furthermore, it is also important to note that the generalization seen from training features of atypical items to gains in naming untrained typical items, while supported theoretically by research involving the typicality effect and CATE account, could also reflect that it is simpler to generalize to typical items, regardless of the typicality of items trained. Unfortunately, the experimental design used in this study did not allow for an examination of whether there is greater generalization to untrained atypical or typical within-category items when atypical items are targeted in therapy. Future work should investigate this valuable question regarding the underpinnings of generalization from typicality-based SFA.

We also investigated near transfer of gains from treatment to untrained semantic and phonological behavioral tasks. First, we found that participants' reaction times were significantly faster on all three semantic behavioral tasks after therapy, suggesting more efficient semantic processing. Secondly, regarding their accuracy on the tasks, participants showed significant gains post-treatment on two of three semantic tasks (i.e., Category Coordinate and Superordinate Category Verification), suggesting that strengthening the semantic system through typicality-based SFA treatment bolstered skills that could then be utilized by participants when performing other semantic tasks. No significant change was seen in post-therapy accuracy on the Semantic Feature Verification task, which may have been due to ceiling effects (i.e., Responders' Pre = 91%, Post = 91%). However, regarding the phonological tasks, participants did improve significantly in their accuracy on the Syllable Judgment task, providing some evidence for near transfer to phonological processing skills from typicality-based SFA treatment. Interestingly, we found that participants performed better at baseline on this task than either the Rhyme Judgment or Phoneme Verification tasks, suggesting that Syllable Judgment is the least challenging for PWA, which is consistent with findings of Meier et al. 2016. Unlike the Rhyme Judgment and Phoneme Verification tasks, which did not change with treatment, Syllable Judgment did not require the individual to maintain a given word or phoneme in working memory while retrieving the target lexical item (as in the no-name tasks) and comparing it to the target item (in both the no-name and name-provided tasks). Furthermore, a case has been made that syllables are actually lexically represented (Romani, Galluzzi, Bureca, & Olson, 2011) and thus, they would be closer to semantic levels than other phonological information, making it more probable for syllable-level generalization to occur. Overall, these results are consistent with previous work suggesting that semantic and phonological levels are indeed interactive and strengthening

semantic features of items results in better access to the phonological word form (Dell et al., 1997).

Interestingly, far transfer to standardized measures of semantic, phonological processing and lexical retrieval has not been consistently examined after SFA-based treatment. In previous studies, PWA have demonstrated gains on tests of semantic processing (e.g., PALPA spoken-word-to-picture matching; [Hashimoto & Frome, 2011]) and lexical retrieval (e.g., BNT; [Kiran & Thompson, 2003b]) following SFA-based treatment. Although examined with less frequency, phonological abilities (i.e., phoneme segmentation of words and nonwords [Hashimoto, 2012]; and repetition [Kiran, 2008; Kiran & Johnson, 2008]) have also improved post SFA-based therapy. In the present study, participants improved significantly on the BNT, but not on PALPA 1, PALPA 51, the NNB-CN or the PAPT. The lack of change on the latter set of tests may have been because these assessments required different skills than those trained in treatment (i.e. PALPA 1 – nonword same-different discrimination; PALPA 51 - semantic association relying on intact abstract word reading skills; NNB-CN – verb naming), or because participants were near ceiling on the measure at baseline (i.e., Responders' PAPT: Pre=90.33, Post=90.72).

Far transfer effects included improvement on standardized assessments of global language skills, which may have been due to the different components of the treatment protocol. Participants verified auditory and written features of targets in therapy as part of their weekly session, which provided an avenue for improved auditory and reading comprehension, as evidenced by gains on the WAB (i.e., AQ and LQ). Furthermore, they attended therapy twice weekly for two-hour sessions during which they were encouraged to sustain attention and distinguish items based on their features, which likely explains significant changes in cognitive-linguistic function as measured by the WAB-CQ. Additionally, improved lexical access abilities

as a function of treatment may have resulted in improved word retrieval and verbal expression on subtests of the WAB and the BNT. It should also be noted that participants did not improve significantly in their CLQT Composite Severity. One possibility is that typicality-based SFA treatment improved aspects of cognition (e.g., executive function, short-term memory; [Dignam et al., 2017]), but the CLQT composite severity score was not sensitive to subtle changes in cognitive subskills. Thus, future studies should investigate this relationship. Overall, these findings are consistent with results from other SFA-based treatment studies suggesting that gains from typicality-based SFA treatment may transfer to other communication contexts (Davis & Stanton, 2005; Edmonds & Kiran, 2006; Hashimoto, 2012; Kiran, 2008; Kiran & Johnson, 2008; Kiran et al., 2011; Kiran & Thompson, 2003b; Peach & Reuter, 2010; Rider et al., 2008; Wambaugh & Ferguson, 2007).

Of note, with the exception of the significant gains observed in post-treatment accuracy on the Superordinate Category Verification task, the significant improvements demonstrated by treatment participants in their accuracy (i.e., Syllable Judgment, Category Coordinate Judgment) and reaction time (i.e., Category Coordinate Judgment, Superordinate Category Verification, Semantic Feature Verification) on the semantic and phonological processing tasks and standardized assessments (i.e., WAB-AQ, WAB-LQ, WAB-CQ, BNT) must be considered in light of the fact that minimal significant differences in average change on these measures were found between the treatment and natural history control groups after a 12-week period of treatment/no-treatment. Furthermore, it is possible that the assessments used to assess for transfer from SFA-based treatment were not sensitive enough to capture this level of skill generalization. Future studies should investigate for cascading improvement on untrained related processes/tasks

as a function of semantic-based naming therapy using more fine-grained assessments and control tasks.

Our follow-up analyses investigating differences in treatment response demonstrate that the responders clearly drove the full sample results, showing robust training and generalization effects. Unfortunately, nonresponders did not exhibit the same magnitude of treatment effects and demonstrated no evidence of within-category transfer, near transfer, or far transfer. Nonetheless, these findings emphasize that while group-level analyses are possible with a large participant sample and serve to build the evidence base for aphasia treatment approaches broadly, it still remains essential to investigate individual profiles, given the persistent challenge of individual variability in aphasia rehabilitation (Lambon Ralph, Snell, Fillingham, Conroy, & Sage, 2010).

Another possible avenue for PWA who do not appear to be responding favorably to language therapy is to increase the intensity of practice through a home exercise program (Des Roches, Balachandran, Ascenso, Tripodis, & Kiran, 2015). While the primary aims of this study did not include investigating the efficacy of a home exercise program (HEP), they were implemented with a number of participants (BU16, BU20, BU21, BU22, BU24) and may have resulted in a better response to treatment for two of those participants (BU20, BU22). These initial results are promising, yet additional research examining the influence of factors such as candidacy, timing, and compliance home-based exercises is needed to determine the benefit of home exercise programs for individuals who are not responding favorably to SFA-based treatment.

Conclusion

Typicality-based SFA treatment was effective for improving lexical retrieval in twenty-seven individuals with aphasia of varying types and severities. Training features of atypical exemplars resulted in gains on those exemplars as well as untrained typical exemplars. PWA also showed significant gains on untrained tasks and standardized measures of language post-treatment, although follow-up analyses demonstrated that only participants who showed robust treatment effects (i.e., both trained atypical and typical categories) exhibited generalization across multiple levels. Typicality-based SFA treatment is an effective and efficient option for clinicians given its propensity to promote multi-level generalization in individuals with chronic anomia, although baseline individual language and cognitive severity should be considered before implementation.

Acknowledgments

This project is funded by NIH/NIDCD grants: P50DC012283 and T32DC013012. The authors thank Stefano Cardullo and Rachel Ryskin for their knowledge regarding mixed-effects models and R; Carrie Des Roches for her contributions to the project; and, many others for their assistance in data collection.

References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2016). lme4: Linear Mixed-Effects Models using “Eigen” and S4 (Version 1.1-12). Retrieved from <https://CRAN.R-project.org/package=lme4>
- Beeson, P. M., & Robey, R. R. (2006). Evaluating Single-Subject Treatment Research: Lessons Learned from the Aphasia Literature. *Neuropsychol Rev Neuropsychology Review*, 16(4), 161–169.
- Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*, 13(3), 236–249. [https://doi.org/10.1044/1058-0360\(2004/025\)](https://doi.org/10.1044/1058-0360(2004/025))
- Boyle, M. (2010). Semantic Feature Analysis Treatment for Aphasia Word Retrieval Impairments: What’s in a Name? *Topics in Stroke Rehabilitation*, 17(6), 411–422.
- Boyle, M. (2017). Semantic Treatments for Word and Sentence Production Deficits in Aphasia. *Seminars in Speech and Language*, 38(01), 052–061. <https://doi.org/10.1055/s-0036-1597256>
- Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, 4(4), 94–98.
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. In The Cochrane Collaboration (Ed.), *Cochrane*

- Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from <http://doi.wiley.com/10.1002/14651858.CD000425.pub4>
- Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology*, *14*(2), 133–142. <https://doi.org/10.1080/026870300401513>
- Collins, A. M., & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, *82*(6), 407–428.
- Dabul, B. (2000). *Apraxia Battery for Adults, Second Edition*. Austin, TX: Pro-Ed.
- Davis, L. A., & Stanton, S. T. (2005). Semantic feature analysis as a functional therapy tool. *Contemporary Issues in Communication Science & Disorders*, *32*, 85-92 8p.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review*, *104*(4), 801–838.
- DeLong, C., Nessler, C., Wright, S., & Wambaugh, J. (2015). Semantic Feature Analysis: Further Examination of Outcomes. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, *24*(4), 864–879.
- Des Roches, C. A., Balachandran, I., Ascenso, E. M., Tripodis, Y., & Kiran, S. (2015). Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. *Frontiers in Human Neuroscience*, *8*. <https://doi.org/10.3389/fnhum.2014.01015>
- Des Roches, C. A., Mitko, A., & Kiran, S. (2017). Relationship between Self-Administered Cues and Rehabilitation Outcomes in Individuals with Aphasia: Understanding Individual Responsiveness to a Technology-Based Rehabilitation Program. *Frontiers in Human Neuroscience*, *11*. <https://doi.org/10.3389/fnhum.2017.00007>

- Dignam, J., Copland, D., O'Brien, K., Burfein, P., Khan, A., & Rodriguez, A. D. (2017). Influence of Cognitive Ability on Therapy Outcomes for Anomia in Adults With Chronic Poststroke Aphasia. *Journal of Speech Language and Hearing Research*, *60*(2), 406. https://doi.org/10.1044/2016_JSLHR-L-15-0384
- Edmonds, L. A., & Kiran, S. (2006). Effect of semantic naming treatment on crosslinguistic generalization in bilingual aphasia. *Journal of Speech, Language, and Hearing Research*, *49*(4), 729–748. [https://doi.org/10.1044/1092-4388\(2006/053\)](https://doi.org/10.1044/1092-4388(2006/053))
- Ellis, A. W., & Young, A. W. (1988). *Human Cognitive Neuropsychology*. Hove, UK: Erlbaum.
- Gilmore, N., Meier, E. L., Johnson, J. P., & Kiran, S. (2018). *The multidimensional nature of aphasia: Using factors beyond language to predict treatment-induced recovery*.
- Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219–260. <https://doi.org/10.1016/j.cognition.2005.12.010>
- Goodglass, H., Kaplan, E., & Weintraub, S. (1983). *The Revised Boston Naming Test*. Philadelphia, PA: Lea & Febiger.
- Goodglass, H., & Wingfield, A. (1997). *Anomia: neuroanatomical and cognitive correlates*. San Diego, CA: Academic Press.
- Hampton. (1995). Testing the Prototype Theory of Concepts. *Journal of Memory and Language*, *34*(5), 686–708. <https://doi.org/10.1006/jmla.1995.1031>
- Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 441–461.

- Hashimoto, N. (2012). The use of semantic-and phonological-based feature approaches to treat naming deficits in aphasia. *Clinical Linguistics & Phonetics*, 26(6), 518–553. <https://doi.org/10.3109/02699206.2012.663051>
- Hashimoto, N., & Frome, A. (2011). The use of a modified semantic features analysis approach in aphasia. *Journal of Communication Disorders*, 44(4), 459–469. <https://doi.org/10.1016/j.jcomdis.2011.02.004>
- Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test: Examiner's Manual*. San Antonio, TX: The Psychological Corporation.
- Hothorn, T., Hornik, K., van de Wiel, M. A., Winell, H., & Zeileis, A. (2016). coin: Conditional Inference Procedures in a Permutation Test Framework (Version 1.1-3). Retrieved from <https://CRAN.R-project.org/package=coin>
- Howard, D., & Patterson, K. E. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Thames Valley Test Company.
- Kay, J., Lesser, R., & Coltheart, M. (1992). *The Psycholinguistic Assessment of Language Processing in Aphasia (PALPA)*. Erlbaum; Hove, U.K.
- Kertesz, A. (2007). *Western Aphasia Battery (Revised)*. San Antonio, TX: The Psychological Corporation.
- Kiran, S. (2007). Complexity in the treatment of naming deficits. *American Journal of Speech-Language Pathology*, 16(1), 18–29.
- Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research*, 51(6), 1550–1568. [https://doi.org/10.1044/1092-4388\(2008/07-0038\)](https://doi.org/10.1044/1092-4388(2008/07-0038))

- Kiran, S., & Johnson, L. (2008). Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories. *American Journal of Speech-Language Pathology, 17*(4), 389–400.
- Kiran, S., Ntourou, K., & Eubank, M. (2007). The effect of typicality on online category verification of inanimate category exemplars in aphasia. *Aphasiology, 21*(9), 844–866. <https://doi.org/10.1080/02687030600743564>
- Kiran, S., Sandberg, C., & Sebastian, R. (2011). Treatment of Category Generation and Retrieval in Aphasia: Effect of Typicality of Category Items. *Journal of Speech Language and Hearing Research, 54*(4), 1101. [https://doi.org/10.1044/1092-4388\(2010/10-0117\)](https://doi.org/10.1044/1092-4388(2010/10-0117))
- Kiran, S., & Thompson, C. K. (2003a). Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language, 85*(3), 441–450. [https://doi.org/10.1016/S0093-934X\(03\)00064-6](https://doi.org/10.1016/S0093-934X(03)00064-6)
- Kiran, S., & Thompson, C. K. (2003b). The Role of Semantic Complexity in Treatment of Naming Deficits: Training Semantic Categories in Fluent Aphasia by Controlling Exemplar Typicality. *Journal of Speech, Language, and Hearing Research, 46*(4), 773–787. [https://doi.org/10.1044/1092-4388\(2003/061\)](https://doi.org/10.1044/1092-4388(2003/061))
- Kristensson, J., Behrns, I., & Saldert, C. (2015). Effects on communication from intensive treatment with semantic feature analysis in aphasia. *Aphasiology, 29*(4), 466–487.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models (Version 2.0-33). Retrieved from <https://CRAN.R-project.org/package=lmerTest>
- Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: Both language and

- cognitive status are key predictors. *Neuropsychological Rehabilitation*, 20(2), 289–305.
<https://doi.org/10.1080/09602010903237875>
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(01), 1–38.
- Lowell, S., Beeson, P. M., & Holland, A. L. (1995). The Efficacy of a Semantic Cueing Procedure on Naming Performance of Adults with Aphasia. *American Journal of Speech-Language Pathology*, 4(4), 109–114.
- Madden, E., Robinson, R., & Kendall, D. (2017). Phonological Treatment Approaches for Spoken Word Production in Aphasia. *Seminars in Speech and Language*, 38(01), 062–074. <https://doi.org/10.1055/s-0036-1597258>
- Meier, E. L., Johnson, J. P., Villard, S., & Kiran, S. (2017). Does Naming Therapy Make Ordering in a Restaurant Easier? Dynamics of Co-Occurring Change in Cognitive-Linguistic and Functional Communication Skills in Aphasia. *American Journal of Speech-Language Pathology*, 26(2), 266. https://doi.org/10.1044/2016_AJSLP-16-0028
- Meier, E. L., Lo, M., & Kiran, S. (2016). Understanding semantic and phonological processing deficits in adults with aphasia: effects of category and typicality. *Aphasiology*, 30(6), 719–749. <https://doi.org/10.1080/02687038.2015.1081137>
- Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, 16(10/11), 935–979.
- Peach, R. K., & Reuter, K. A. (2010). A discourse-based approach to semantic feature analysis for the treatment of aphasic word retrieval failures. *Aphasiology*, 24(9), 971–990.
<https://doi.org/10.1080/02687030903058629>

- Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*(2), 291–321. <https://doi.org/10.1080/01688639508405124>
- R Studio Team. (2015). *RStudio: Integrated Development for R*. Boston, MA: R Studio, Inc. Retrieved from <http://www.rstudio.com/>
- Ralph, M. A. L., Moriarty, L., & Sage, K. (2002). Anomia is simply a reflection of semantic and phonological impairments: Evidence from a case-series study. *Aphasiology*, *16*(1–2), 56–82. <https://doi.org/10.1080/02687040143000448>
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460.
- Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L. (2008). Using semantic feature analysis to improve contextual discourse in adults with aphasia. *American Journal of Speech-Language Pathology*, *17*(2), 161–172. [https://doi.org/10.1044/1058-0360\(2008/016\)](https://doi.org/10.1044/1058-0360(2008/016))
- Rogers, T. T., Patterson, K., Jefferies, E., & Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: Effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220–239. <https://doi.org/10.1016/j.neuropsychologia.2015.04.015>
- Romani, C., Galluzzi, C., Bureca, I., & Olson, A. (2011). Effects of syllable structure in aphasic errors: Implications for a new model of speech production. *Cognitive Psychology*, *62*(2), 151–192. <https://doi.org/10.1016/j.cogpsych.2010.08.001>
- Rosch. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233.

- Rosch, E., & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573–605.
- Rossiter, C., & Best, W. (2013). “Penguins don’t fly”: An investigation into the effect of typicality on picture naming in people with aphasia. *Aphasiology*, 27(7), 784–798. <https://doi.org/10.1080/02687038.2012.751579>
- Sandberg, C., & Kiran, S. (2014). How justice can affect jury: Training abstract words promotes generalisation to concrete words in patients with aphasia. *Neuropsychological Rehabilitation*, 24(5), 738–769. <https://doi.org/10.1080/09602011.2014.899504>
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User’s Guide*. Pittsburgh: Psychology Software Tools Inc.
- Schwartz, M. F. (2013). Theoretical analysis of word production deficits in adult aphasia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120390–20120390. <https://doi.org/10.1098/rstb.2012.0390>
- Stanczak, L., Waters, G., & Caplan, D. (2006). Typicalitybased learning and generalisation in aphasia: Two case studies of anomia treatment. *Aphasiology*, 20(2–4), 374–383.
- Thompson, C. K. (1989). Generalization Research in Aphasia: A Review of the Literature (Vol. 18, pp. 195–222). Presented at the Clinical Aphasiology Conference, College-Hill Press.
- Thompson, C. K., Lukic, S., King, M. C., Mesulam, M. M., & Weintraub, S. (2012). Verb and noun deficits in stroke-induced and primary progressive aphasia: The Northwestern Naming Battery(). *Aphasiology*, 26(5), 632–655. <https://doi.org/10.1080/02687038.2012.676852>
- Thompson, C. K., Shapiro, L., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of

- treatment efficacy (CATE). *Journal of Speech, Language, and Hearing Research*, 46(3), 591–607. [https://doi.org/10.1044/1092-4388\(2003/047\)](https://doi.org/10.1044/1092-4388(2003/047))
- Thompson, C. K., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology*, 16(1), 30–42. [https://doi.org/10.1044/1058-0360\(2007/005\)](https://doi.org/10.1044/1058-0360(2007/005))
- Wambaugh, J. L., & Ferguson, M. (2007). Application of semantic feature analysis to retrieval of action names in aphasia. *Journal of Rehabilitation Research and Development*, 44(3), 381–394.
- Wambaugh, J. L., Mauszycki, S., Cameron, R., Wright, S., & Nessler, C. (2013). Semantic feature analysis: Incorporating typicality treatment and mediating strategy training to promote generalization. *American Journal of Speech-Language Pathology*, 22(2), S334–S369. [https://doi.org/10.1044/1058-0360\(2013/12-0070\)](https://doi.org/10.1044/1058-0360(2013/12-0070))
- Wambaugh, J. L., Mauszycki, S., & Wright, S. (2014). Semantic feature analysis: Application to confrontation naming of actions in aphasia. *Aphasiology*, 28(1), 1–24. <https://doi.org/10.1080/02687038.2013.845739>
- 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.
- Wickham, H. (2017). tidyverse: Easily Install and Load the “Tidyverse” (Version R package version 1.2.1.). Retrieved from <https://CRAN.R-project.org/package=tidyverse>
- Wisernburn, B., & Mahoney, K. (2009). A meta-analysis of word-finding treatments for aphasia. *Aphasiology*, 23(11), 1338–1352.

Table 1: Demographic information for the treatment and natural history control groups including age, time post onset in months (MPO); aphasia severity as demonstrated by baseline Western Aphasia Battery-Revised (WAB-R) Aphasia Quotient (AQ); overall cognitive severity as demonstrated by baseline Cognitive Linguistic Quick Test Composite Severity (CLQT-CS); aphasia type (Type); accuracy on the naming screener at baseline (BL Nam.) (Full = 180 items); and presence of apraxia of speech (AOS) based on the Screen for Dysarthria and Apraxia of Speech (S-DAOS) (Dabul, 2000). When participants did not produce sufficient verbal output to validly screen for AOS, it was rated as “undetermined.” Treatment (T1, T2) and generalization (G1, G2) effect sizes (ES) for each half-category were calculated by subtracting the pre-treatment average score from the post-treatment average score and dividing that value by the pre-treatment standard deviation (Beeson & Robey, 2006). Participants were classified as Responders (R) or Nonresponders (NR) post-hoc based on their treatment (Tx) response (i.e., ES in at least one trained category \geq “small,” or 4.0). Average accuracy at pre- and post-treatment on trained and untrained categories is also presented. Aphasia Type abbreviations: A=Anomic; B=Broca’s; C=Conduction; W=Wernicke’s; TCM=Transcortical Motor; G=Global; AOS abbreviations: AB=Absent; PR=Present; UN=Undetermined

TREATMENT GROUP																
ID	Age	MPO	WAB AQ	CLQT CS	Type	AOS	BL Nam.	T1 ES	T2 ES	Tx Resp.	G1 ES	G2 ES	Avg. Trained		Avg. Untrained	
													Pre	Post	Pre	Post
BU01	55	12	87	90	A	AB	58	3	20***	R	4	9	39	94	55	86
BU02	50	29	25	50	G	UN	1	0	0	NR	0	-1	4	4	0	1
BU03	63	62	52	55	C	PR	18	10**	9**	R	0	1	14	44	18	20
BU04	79	13	74	80	C	AB	68	4*	13***	R	1	1	50	100	77	83
BU05	67	10	31	50	W	AB	6	-1	-2	NR	-1	-1	10	4	7	1
BU06	49	113	67	80	B	PR	56	3	9**	R	1	0	56	90	72	73
BU07	55	137	48	60	B	AB	14	2	6*	R	1	1	25	44	13	19
BU08	49	57	83	100	A	PR	69	3	4*	R	1	1	72	100	65	70
BU09	71	37	95	95	A	AB	59	7**	11***	R	-1	1	41	87	51	45
BU10	53	12	80	80	A	AB	65	17***	3	R	7	4	60	93	62	82
BU11	78	22	92	75	A	AB	34	1	3	NR	1	1	38	57	22	29
BU12	68	104	40	60	B	PR	3	1	1	NR	1	1	4	11	6	8
BU13	42	18	93	95	A	AB	57	7*	15***	R	2	3	47	99	47	70
BU14	64	24	64	80	B	AB	41	8**	9**	R	6	10	35	85	40	71
BU15	71	74	87	60	A	AB	57	6*	1	R	2	1	44	60	38	52
BU16	50	71	34	70	B	UN	5	2	2	NR	1	1	1	8	2	5
BU17	61	152	74	90	A	AB	52	9**	15***	R	4	1	51	99	48	70
BU18	70	152	78	85	A	AB	48	6*	5*	R	2	0	54	92	49	52
BU19	80	22	29	70	B	PR	7	4	7*	R	0	0	13	30	6	6
BU20	48	14	13	45	B	PR	0	4	7*	R	0	0	0	42	0	0
BU21	65	16	12	45	B	PR	0	1	3	NR	0	-1	0	21	1	0
BU22	62	12	65	45	TCM	AB	7	5*	2	R	-1	-1	5	15	7	5
BU23	60	24	45	45	W	AB	5	0	1	NR	-1	2	6	9	5	7
BU24	69	169	40	70	B	PR	7	2	2	NR	-1	1	7	20	5	2
BU25	76	33	38	55	B	AB	2	2	-1	NR	0	1	2	7	1	2
BU26	64	115	58	30	B	AB	21	4	3	NR	1	0	27	41	14	17
BU27	65	17	84	75	A	AB	51	7*	2	R	1	-1	51	80	51	57
Mean	62	56	59	68			28	4	6		1	1	28	53	28	35
SD	10	52	26	19			26	4	6		2	3	23	37	26	32

Range	42-80	10-169	12-95	30-100			0-69	-1-17	-2-20		-1-7	-1-10	0-72	4-100	0-77	0-86
NATURAL HISTORY CONTROL GROUP																
BUc01	49	49	86	75	A	PR	69	-1	-1	NR	1	0	69	66	60	63
BUc02	79	10	32	70	B	PR	7	1	1	NR	1	0	11	15	4	5
BUc05	49	67	32	75	B	UN	5	0	-2	NR	1	-1	8	1	2	2
BUc06	69	164	39	75	B	PR	6	2	-2	NR	1	0	7	7	3	5
BUc07	39	18	71	70	C	PR	47	1	1	NR	-2	-1	41	50	58	49
BUc08	64	13	80	0	A	AB	46	0	1	NR	0	2	46	51	47	55
BUc09	62	21	92	75	A	AB	59	6	-2	R	3	1	26	55	69	82
BUc10	68	21	79	90	A	AB	42	-2	0	NR	1	-2	37	32	45	45
BUc11	58	23	62	50	B	PR	12	0	6	R	0	4	10	25	20	20
BUc12	53	467	91	100	A	AB	72	2	3	NR	2	0	54	66	74	78
Mean	59	85	66	68			36	1	1		1	0	31	37	38	40
SD	12	142	24	27			27	2	3		1	2	22	24	29	31
Range	39-79	10-467	32-92	0-100			5-72	-2-6	-2-6		-2-3	-2-4	7-69	.9-66	2-74	2-82

Note: Treatment effect sizes (ESs) were classified using benchmarks established Robey and Beeson (Beeson & Robey, 2006): small effect = 4.0-6.9 *; medium effect = 7.0-10.0 **; large effect = 10.1 + *** Generalization effect sizes greater than 2.0 (bold) were classified as “meaningful” as they are half the magnitude of a small effect for trained items.

Table 2: Treatment and natural history control group performance on Standardized Outcome Measures and Behavioral Tasks Administered before (Pre) and after (Post) treatment or no-treatment

	Accuracy						Reaction Time					
	Pre			Post			Pre			Post		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
TREATMENT GROUP												
WAB LQ	59.8	23.8	21.5-93.5	61.4*	24.3	18.3-97.0	n/a					
WAB CQ	63.9	21.4	25.6-94.4	65.6*	21.4	22.7-96.2						
WAB AQ	58.9	25.7	11.7-95.2	60.9*	25.4	9.7-96.0						
CLQT (%)	68.0	18.8	30.0-100.0	70.6	18.2	35.0-95.0						
BNT (%)	38.1	35.3	0.0-90.0	42.2*	36.3	0.0-93.3						
PAPT (%)	88.0	9.1	65.4-98.1	88.0	7.5	75.0-100.0						
PALPA 1 (%)	85.3	13.6	50.0-98.6	87.4	12.9	48.6-98.6						
PALPA 51 (%)	50.7	22.4	10.0-86.7	49.1	26.3	0.0-86.7						
NNB - CN (%)	52.5	39.3	0.0-99.0	52.1	39.8	0.0-100.0						
SCV	89.0	8.5	72.5-100.0	91.4*	10.2	60.0-100.0	3153.3	1304.2	874.5-6012.7	2867.9*	1179.7	978.9-5342.2
SFV	83.8	16.7	20.0-97.5	87.3	10.0	46.3-95.0	4185.1	1845.0	1022.7-9289.8	3428.0*	1337.8	1156.8- 7633.2
CCJ	87.6	9.7	62.5-100.0	90.5*	9.6	62.5-100.0	3710.3	1381.9	1583.1-6443.7	3126.5*	1026.0	1573.1-5064.3
SJNN	59.0	14.7	23.8-88.8	62.6	15.1	45.0-88.8	6190.1	2499.8	1568.8-10907.4	5552.4	1968.2	1162.3-9547.7
SJNP	68.1	21.1	26.3-98.8	73.8*	18.7	40.0-96.3	4949.4	2003.5	2671.7-11529.4	5020.3	2042.9	2390.8-9957.7
PVNN	57.5	9.9	46.3-90.0	58.4	12.3	40.0-90.0	6273.1	1995.8	2608.7-10870.6	6172.0	2121.8	3467.4-10602.9
PVNP	62.2	12.5	45.0-93.8	63.9	12.7	48.8-95.0	3982.5	2148.0	1097.8-9670.3	4141.2	2098.0	1610.3-10558.2
RJNN	56.6	11.1	47.5-88.8	57.5	14.3	41.3-91.3	5794.5	1907.1	3525.1-9894.9	5304.7	1819.6	2817.3-9046.8
RJNP	66.5	14.1	48.8-93.8	67.7	13.6	48.8-92.5	3907.4	2001.9	948.8-10111.1	3546.8	1554.2	1149.3-7616.7
NATURAL HISTORY CONTROL GROUP												
WAB LQ	68.0	24.9	36.3-93.6	67.4	24.2	36.1-95.6	n/a					

WAB CQ	72.2	21.2	44.0-93.7	71.6	21.0	45.5-95.7							
WAB AQ	66.4	23.7	32.3-92.0	68.8	25.7	28.9-94.0							
CLQT (%)	76.7	12.5	60-100	82	11.8	70.0-100.0							
BNT (%)	42.8	34.4	5.0-88.3	48.0	35.8	1.7-91.7							
PAPT (%)	94.1	4.0	85-100.0	91.3	6.2	78.9-98.1							
PALPA 1 (%)	90.5	6.4	81-100.0	90.3	6.3	76.4-97.2							
PALPA 51 (%)	51.4	19.5	7.0-73.0	49.7	24.3	10.0-86.7							
NNB - CN (%)	60.6	37.7	1.0-97.0	64.6	39.2	5.4-98.7							
SCV	96.5	2.9	92.5-100.0	94.2	3.0	88.8-96.3	2465.9	621.7	1599.0-3438.8	2006	607.4	874.5-2499.5	
SFV	91.3	3.8	86.3-95.0	91.5	2.3	88.8-95.0	3586	778.9	2832.1-5025.3	2695.9	949.3	1022.7-3525.6	
CCJ	91.9	4.2	85.0-96.3	92.3	2.3	88.8-95.0	3140.8	1217.5	1975.8-5083.7	2519.1	754.1	1538.1-2873.3	
SJNN	59.6	16.6	45.0-86.3	66.7	16.3	52.5-92.5	5800.1	1025.4	4362.9-6950.0	5091.6	1022.4	3893.3-6586.4	
SJNP	75.0	13.9	63.8-95.0	77.5	15.4	56.3-97.5	3365.1	655.2	2724.5-4391.5	3198.5	674.4	2497.6-4322.3	
PVNN	63.3	16.5	46.3-86.3	65.8	19.4	47.5-91.3	6064.2	1276.2	4272.0-7799.0	5158.2	1312.3	3309.9-7014.0	
PVNP	68.8	18.7	50.0-96.3	71.0	21.1	50.0-96.3	2821.9	689.9	1995.4-3940.5	2036.7	618	1398.2-2815.1	
RJNN	65.8	20.9	41.3-92.5	59.0	20.3	37.5-90.0	4904.2	722.8	4142.8-6020.7	4733.5	1066.4	3630.0-6437.4	
RJNP	71.9	18.5	51.3-93.8	69.6	18.4	51.3-97.4	2287.6	654.9	1734.4-3577.6	2213.8	613.2	1289.3-3017.6	

Note: Scores on tasks in which experimental participants improved significantly from the pre- to post-treatment time point are starred ($p < .05$). WAB LQ = Western Aphasia Battery (WAB) Language Quotient; CQ = Cortical Quotient; AQ = Aphasia Quotient; CLQT = Cognitive Linguistic Quick Test Composite Severity; BNT = Boston Naming Test; PAPT = Pyramid and Palm Trees Test; PALPA 1= Psycholinguistic Assessment of Language Processing in Aphasia Same-Different Discrimination Using Nonword Minimal Pairs; PALPA 51 = Word Semantic Association; NNB-CN = Northwestern Naming Battery Confrontation Naming Subtest; SCV = Superordinate Category Verification; SFV = Semantic Feature Verification; CCJ = Category Coordinate Judgment; SJNN = Syllable Judgment No Name; SJNP = Syllable Judgment Name Provided; PVNN= Phoneme Verification No Name; PVNP = Phoneme Verification Name Provided; RJNN= Rhyme Judgment No Name; RJNP = Rhyme Judgment Name Provided

Figures

Figure 1. Results of logistic mixed-effects regression examining treatment effects. The full sample (a) and the responders (b) showed a higher likelihood of responding accurately to atypical (thin dashed line) and typical trained items (thicker dashed line) than monitored items (solid black line) over time, demonstrating a direct effect of treatment. The nonresponders (c) demonstrated a higher likelihood of responding accurately to atypical trained items (thin dashed line) only.

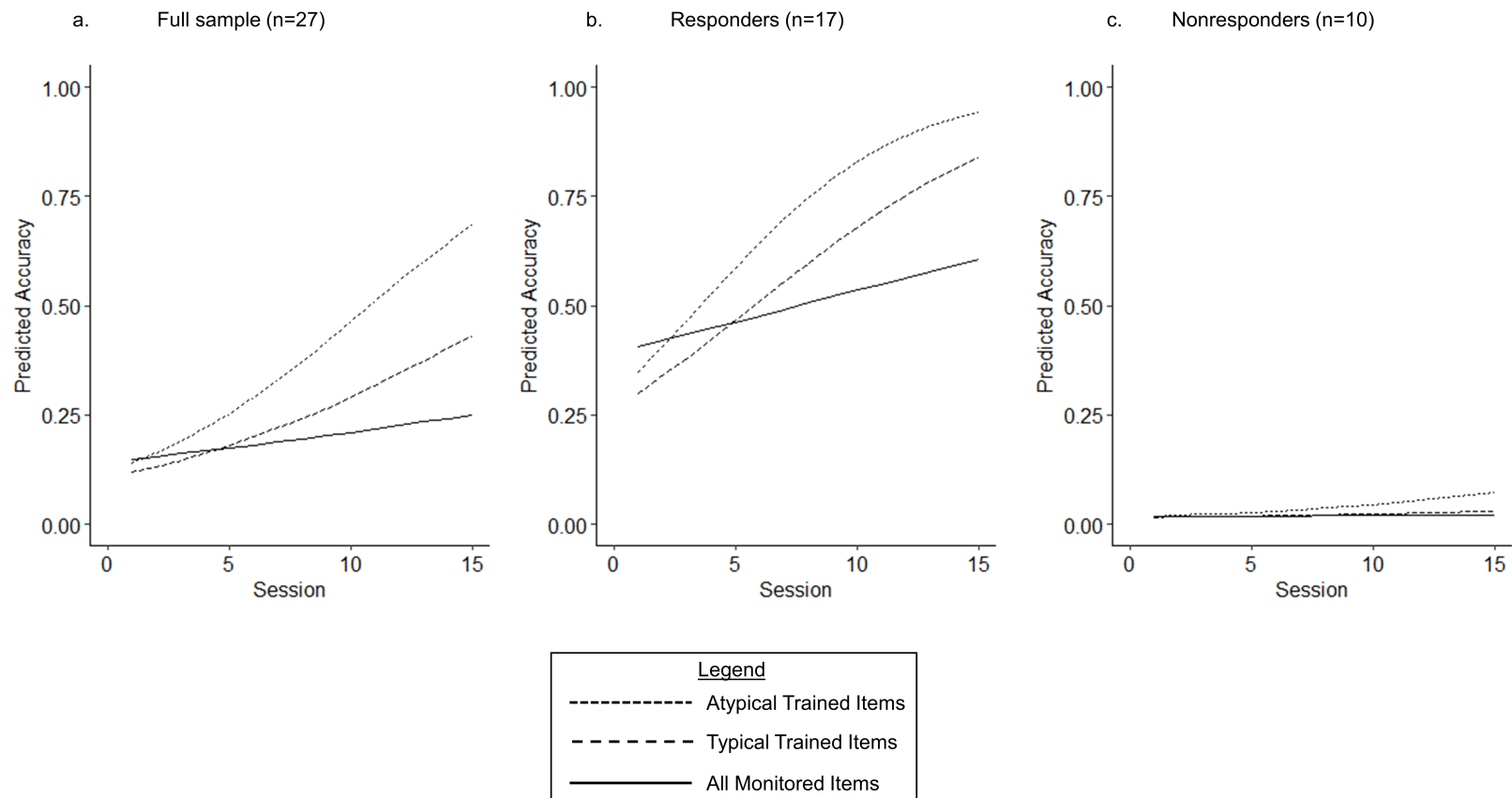


Figure 2. Results of logistic mixed-effects regression with only untrained items in the model to examine generalization. For the full sample (a) and the responders (b), the likelihood of naming untrained typical items accurately (thin dotted line) was significantly greater than the likelihood of naming untrained atypical items (thick black line), supporting the CATE hypothesis and typicality effect. For nonresponders, there was no significant difference in the likelihood of naming untrained typical items versus untrained atypical items.

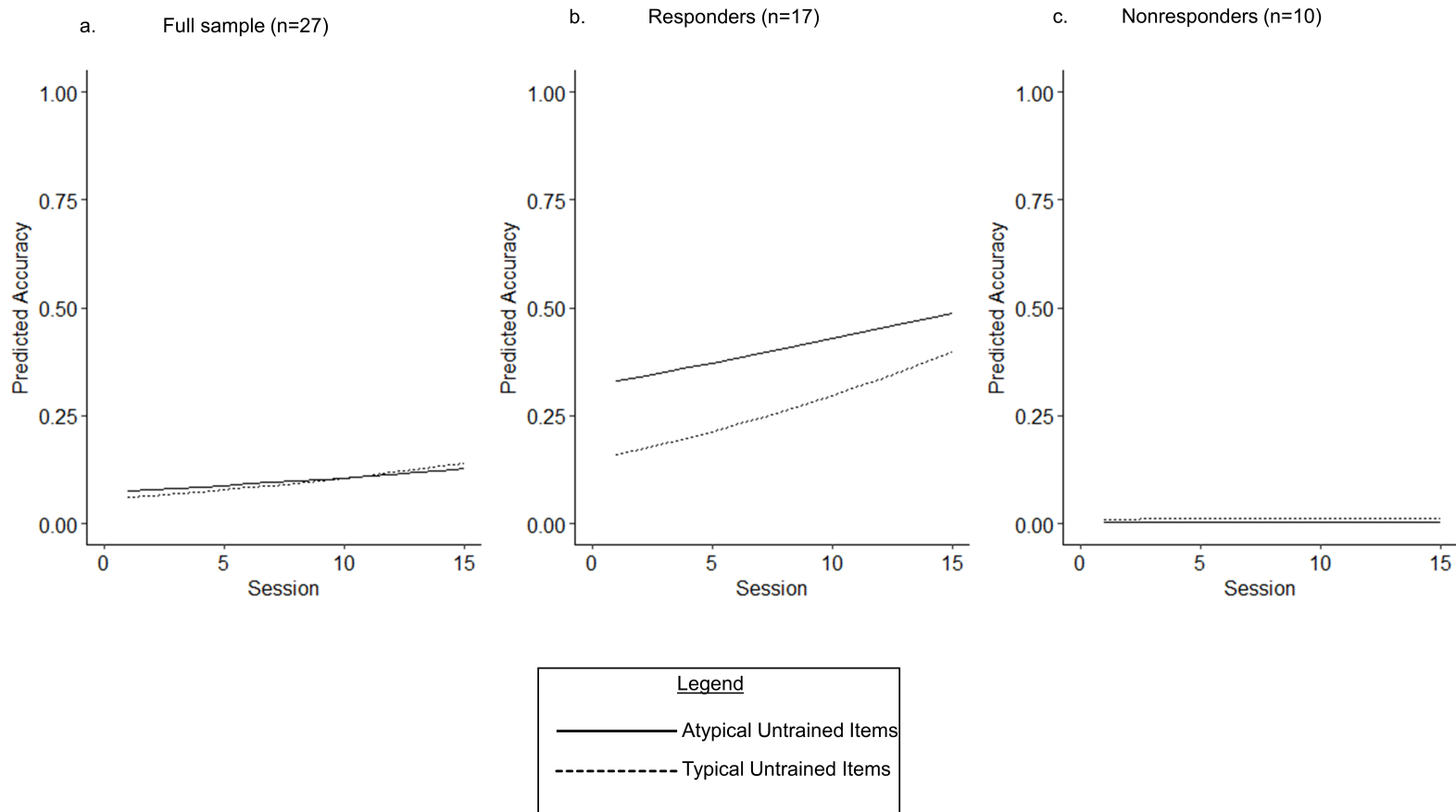
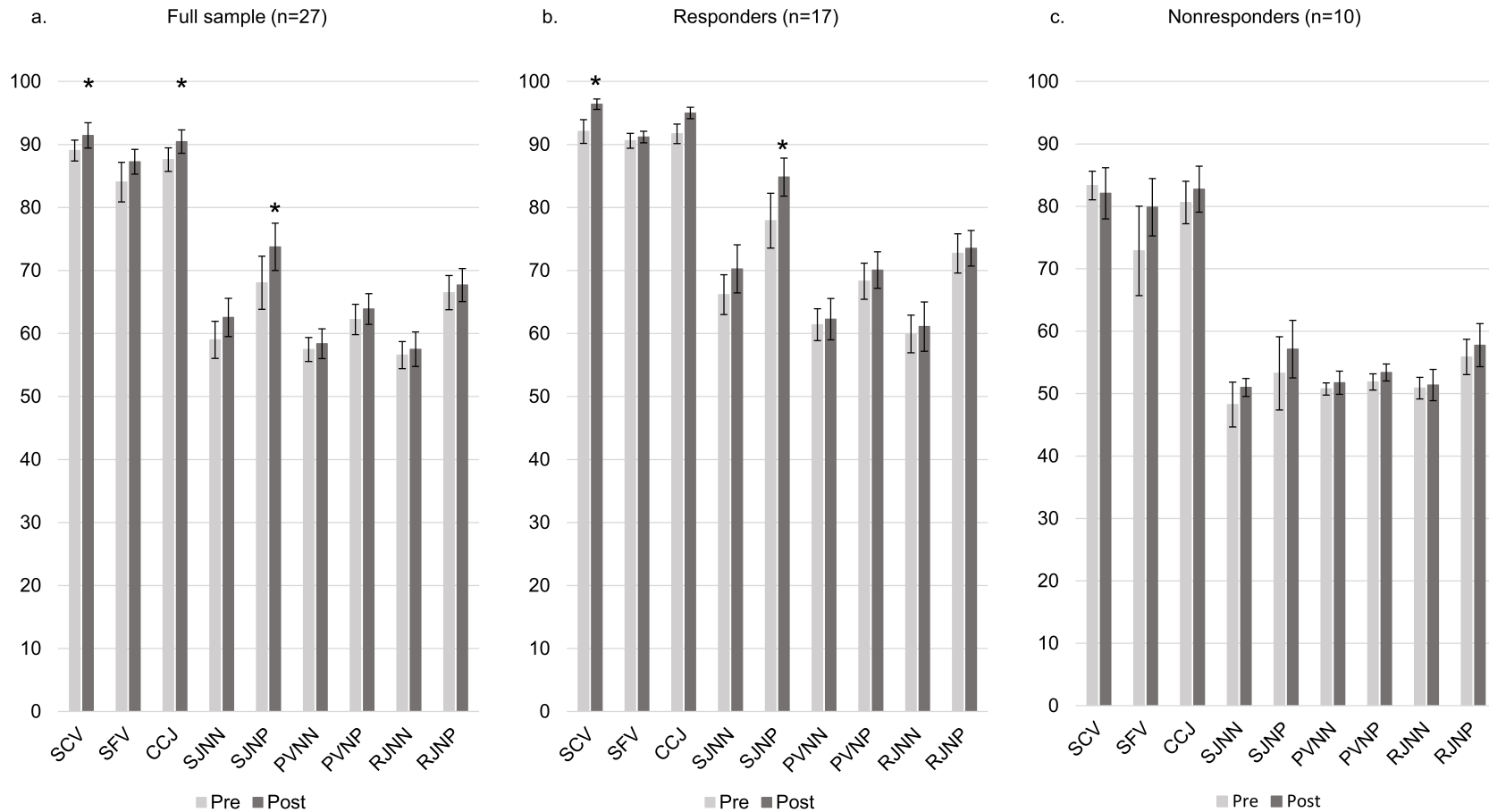
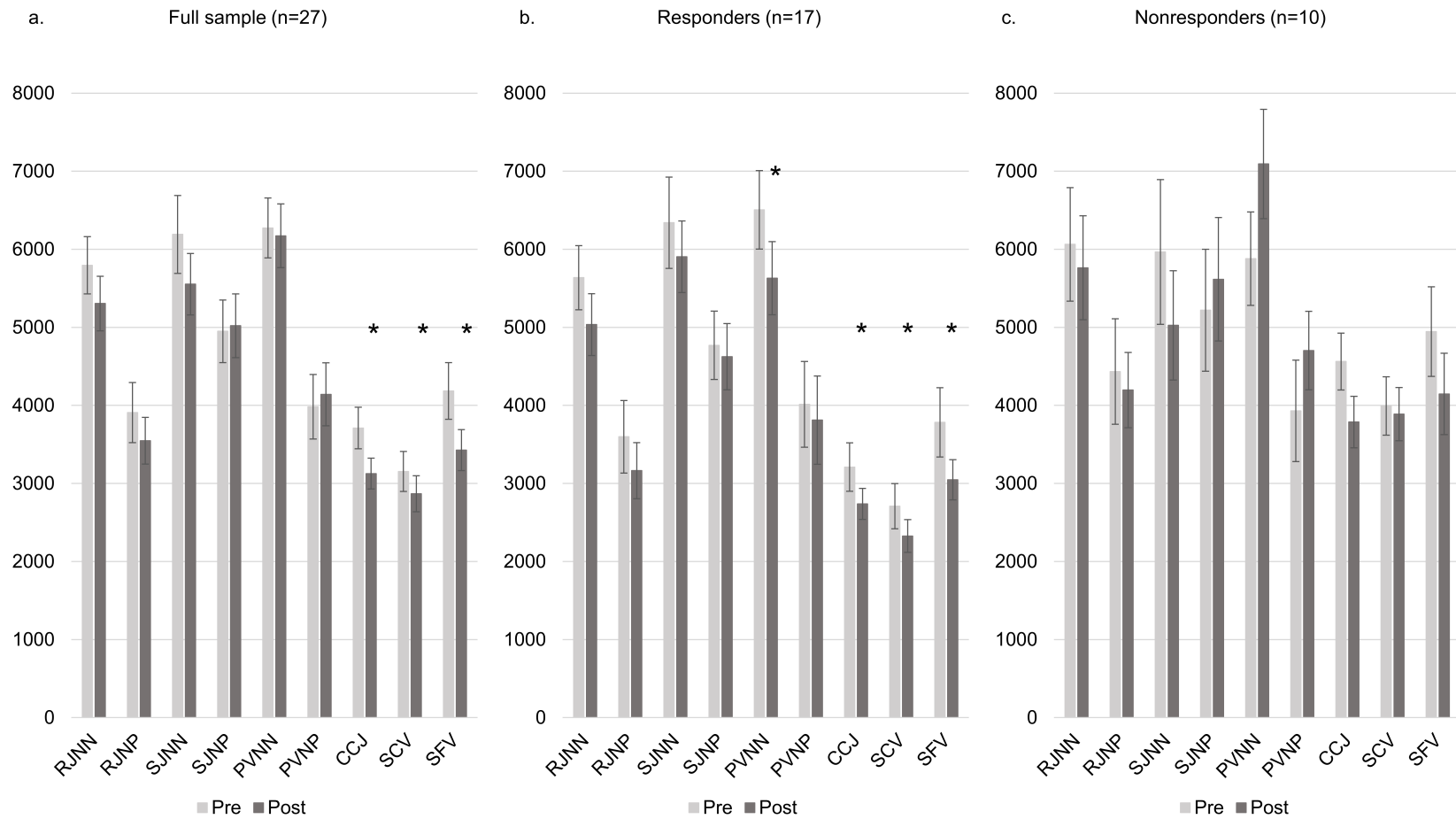


Figure 3: Changes pre- to post-treatment in average accuracy on semantic and phonological tasks for the full sample (a), responder (b) and nonresponder (c) groups. Significant gains were seen on tasks of semantic and phonological processing post-treatment for the full sample and responder groups (* = significant at $p < .05$ level).



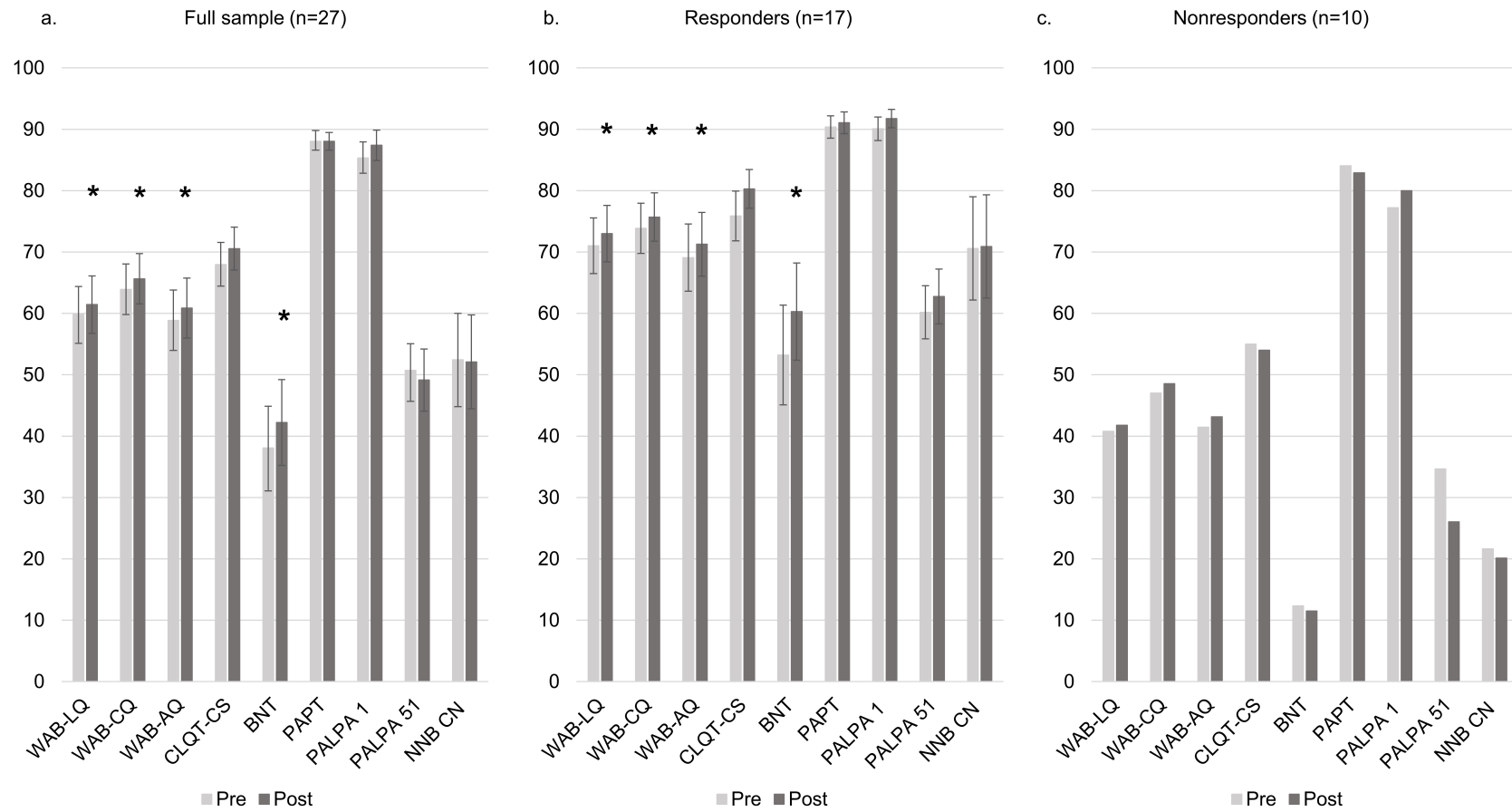
Note: NN = No-Name Condition; NP = Name-Provided Condition; SCV = Superordinate Category Verification; SFV = Semantic Feature Verification; CCJ = Category Coordinate Judgment; SJ = Syllable Judgment; PV = Phoneme Verification; RJ = Rhyme Judgment

Figure 4: Changes from pre- to post-treatment in average reaction time on the semantic and phonological tasks for the full sample (a), responder (b), and nonresponder (c) groups. Reaction times were faster on tasks of semantic processing post-therapy for the full-sample and responder analyses (* = significant at $p < .05$ level).



Note: NN = No-Name Condition; NP = Name-Provided Condition; RJ = Rhyme Judgment; SJ = Syllable Judgment; PV = Phoneme Verification; CCJ = Category Coordinate Judgment; SCV = Superordinate Category Verification; SFV = Semantic Feature Verification

Figure 5: Changes from pre- to post-treatment in average accuracy on standardized outcome measures of cognitive-linguistic function for the full sample (a), responder (b) and nonresponder (c) groups. Significant gains were seen on measures of global cognitive-linguistic functioning and naming ability post-treatment for the full-sample and responder group (* = significant at $p < .05$ level).



Note: WAB-LQ = Western Aphasia Battery-Language Quotient; WAB-CQ = WAB-Cortical Quotient; WAB-AQ = WAB-Aphasia Quotient; CLQT-CS = Cognitive Linguistic Quick Test-Composite Severity; BNT = Boston Naming Test; PAPT = Pyramids and Palm Trees Test; PALPA 1 = Same-Different Nonword Minimal Pair Task (auditory); PALPA 51 = Word Semantic Association (written); NNB CN = Northwestern Naming Battery-Confrontation Naming subtest