AJSLP

Review Article

Acquisition and Generalization Responses in Aphasia Naming Treatment: A Meta-Analysis of Semantic Feature Analysis Outcomes

Yina M. Quique,[a,b] William S. Evans,[a,b] and Michael Walsh Dickey[a,b]

Purpose: This meta-analysis synthesizes results from published studies that used semantic feature analysis (SFA) treatment to improve naming for people with aphasia. It examines how both person- and treatment-related variables affected the likelihood of correct naming responses in individual probe sessions for both acquisition (treated) and generalization (untreated) stimuli.
Method: The meta-analysis compiled data from 12 studies analyzing a total of 35 participants with aphasia. It used mixed-effects models as a novel statistical tool to examine the effects of 2 sets of variables on naming performance: treatment-related variables, including treatment phase (baseline vs. treatment), dosage (number of treatment sessions), and stimulus type (treated vs. untreated, semantically related vs. unrelated items), and person-specific variables, including degree of language impairment and demographic variables (age, time poststroke).
Results: Results of the meta-analysis revealed that SFA intervention promoted increased naming accuracy during naming probes when comparing baseline and treatment phases. In addition, increased dosages of SFA were associated with increased naming accuracy, and treatment-related gains were larger for acquisition (treated) than generalization (untreated) stimuli, likewise for related versus unrelated generalization stimuli. Furthermore, a subset of person-specific variables was predictive of SFA-related gains: Language impairment variables were related to treatment-related changes in naming performance, but demographic variables were not.
Conclusion: These results provide group-level evidence for the efficacy of SFA as well as preliminary estimates of how much naming performance benefit is engendered by varying dosages of SFA. The results also provide promising and previously unobserved evidence of potential person-level predictors of SFA treatment response.

Semantic feature analysis (SFA) is commonly used in the treatment of word retrieval deficits in people with aphasia (PWA). This treatment involves a chart with which a patient is cued to retrieve object names by describing critical features of those objects in five predefined categories including physical properties, typical location, personal associations, category, use, and actions one might perform with/on the object. Repeated, structured practice in retrieving semantic features using these methods is intended to facilitate lexical retrieval and help ameliorate word-finding deficits. Massaro and Tompkins (1994) described the first systematic protocol for SFA's application to remediate word-finding deficits (originally for individuals with traumatic brain injury). Since then, SFA has been applied in multiple treatment studies with PWA.

The theoretical foundation of SFA rests on long-established spreading activation models of semantic processing (Collins & Loftus, 1975), in which activating one concept spreads activation to related concepts. It is also consistent with recent models of lexical access, in which lexical representations of a concept are activated first and then phonologically encoded (Foygel & Dell, 2000). These models hold that, in early stages of word production, the semantic properties of a target word are active in a semantic network and that triggering or activating concepts or features around the target should enhance the activation of the word. Ultimately, the most strongly activated representation will be selected for phonological encoding. Because SFA uses cueing of semantic properties of a set of objects,

[a]Department of Communication Science and Disorders, University of Pittsburgh, PA
[b]Geriatric Research Education and Clinical Center, VA Pittsburgh Healthcare System, PA
Correspondence to Yina M. Quique: yinaquique@pitt.edu

Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Lisa Edmonds
Received September 15, 2017
Revision received February 16, 2018
Accepted April 22, 2018
https://doi.org/10.1044/2018_AJSLP-17-0155
Publisher Note: This article is part of the Special Issue: Select Papers From the 47th Clinical Aphasiology Conference.

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

the treatment should increase the activation of the target object by activating its properties as well as increasing activation of related words and concepts. This should promote improved production of target words as well as better access to related but untreated words that share semantic features with treated words.

The vast majority of studies that have examined the effects of SFA in improving naming function have used single-subject experimental designs. These experimental designs describe effects of treatments in single individuals or small groups, using multiple baselines across either participants or behaviors to establish experimental control (e.g., Thompson, 2006). By examining the course of intervention across potentially variable baseline and treatment intervals, these studies take into account the individual variability present in stroke survivors with aphasia. As noted by Beeson and Robey (2006), single-subject experimental designs play a foundational role in aphasia treatments and, in general, are important in the development and implementation of evidence-based practice in the field of communication sciences and disorders (Byiers, Reichle, & Symons, 2012).

However, because the evidence for the efficacy of SFA to date comes almost exclusively from small-sample single-subject experimental designs, this evidence is limited in important ways. For example, examining treatment response in only single individuals or small groups will obscure person-level differences in treatment response that may only be observed in larger and potentially more representative samples. Similarly, studies differ in the amount of dosage (e.g., Gravier et al., 2018) or intensity of SFA treatment administered, and the effects of these differences may only be observed by aggregating and comparing data across multiple studies. Determining the effects of such person-level and treatment-related factors thus requires large participant samples and/or aggregating data from multiple studies. Establishing the effects of such factors on SFA treatment response is critical for improving identification of appropriate SFA treatment candidates and for determining SFA's efficacy across a broad sample of individuals.

In addition to identifying factors that may contribute to better or worse response to SFA, it is also critical to examine different types of treatment response: acquisition versus generalization. Impairment-focused treatments like SFA in general aim to establish (or improve) a target communicative behavior and to ensure the generalization of that behavior, both to novel-related examples (response generalization) and to novel situations or contexts (stimulus generalization; see Thompson, 1989). As noted above, because the focus of SFA is on stimulating semantic representations through structured practice, this should result in improved retrieval of not only treated words (acquisition) but also untreated words that share semantic features with the treated words (response generalization). Given that generalization is the ultimate goal of effective aphasia treatment (Thompson, 1989), it is critical that group-level studies of SFA and other impairment-focused treatments address not only acquisition but also generalization responses to intervention.

Examining acquisition versus generalization responses to SFA intersects with the importance of identifying person-level and treatment-related factors relevant to characterizing SFA response, described above. For example, it is currently unknown whether the person-level factors that promote naming treatment response are the same for acquisition and generalization (see Abel, Weiller, Huber, & Willmes, 2014; Meinzer et al., 2010). Similarly, it is unclear how treatment-related factors (such as stimulus choice or dosage) may influence the two responses to intervention. Questions of how treatment intensity and dosage may affect response to aphasia treatments have received increasing attention in recent work (Bhogal, Teasell, & Speechley, 2003; Cherney, 2012; Cherney, Patterson, & Raymer, 2011; Hinckley & Craig, 1998; see discussion in Gravier et al., 2018) and are critical to optimizing delivery of SFA and other aphasia treatments. For example, the optimal dosage for promoting generalization might differ from the optimal dosage for promoting acquisition. However, answering these questions again requires aggregating single-subject data in some way, so that population-level inferences can be drawn about treatment for this disorder population in general.

Several articles to date have reviewed the single-subject SFA treatment literature with the aim of providing group-level and more generalizable evidence regarding the efficacy of SFA. However, these studies have notable limitations, particularly with regard to identifying and quantifying larger patterns of SFA treatment response. Boyle (2010) conducted a systematic review of seven published SFA treatment studies, examining whether patient generation of semantic features or verification of clinician-provided features resulted in better response to SFA intervention. Boyle reported that 16 of 17 individuals showed improvement in naming of treated items, and 13 of those people showed some evidence of generalization or improved naming of untreated items. Boyle also concluded that studies involving SFA variants that required feature generation elicited better treatment response than studies involving feature verification. Evidence from this systematic review does indicate that SFA may be efficacious in improving naming performance for people with chronic aphasia. However, the review did not attempt to examine individual variability in generalization after SFA or identify factors that may have distinguished the 13 individuals who showed evidence of generalization from the four who did not. Furthermore, the review did not attempt to quantify the magnitude of treatment-related gains across studies or to measure the effects of person-level variables (such as demographic or language impairment variables) or treatment-related factors (such as dosage) on SFA treatment response.

A subsequent systematic review of SFA treatment studies (Maddy, Capilouto, & McComas, 2014) examined a larger sample of SFA treatment studies. It also used quantitative analysis methods to provide more systematic evidence for the efficacy of SFA. Maddy and colleagues identified 11 studies that were experimental in nature, addressed SFA as a treatment, included adults with neurological damage, and did not combine SFA with other treatment

methods. Of these, 10 were found to be well-designed single-subject experimental design studies (on the basis of criteria from Tate et al., 2008). Treatment efficacy was measured via effect sizes comparing the pretreatment and posttreatment phases. When there were not enough data to calculate a *d* statistic, the authors used the percentage of nonoverlapping data (PND), which is the percentage of nonoverlap between the treatment phase and the baseline. From the 24 participants, the systematic review reported a large effect size for two participants, a medium effect size for six participants, and a small effect size for eight participants. PND was calculated for seven participants; six of them showed a large treatment effect (PND > 90%), and one showed a moderate treatment effect (PND = 85%).

As Maddy and colleagues suggest, these effect size and PND measures indicate that there was a medium to large treatment effect for at least some of the participants treated with SFA. However, these analyses did not examine the variability in SFA response across participants or try to identify factors that may have contributed to it, as the authors themselves note. Furthermore, the analyses did not address generalization to untrained items or the role of treatment-level factors such as dosage, stimulus selection, or intensity in the observed SFA treatment response. It is worth noting that the choice to limit analyses to comparing means from pretreatment versus posttreatment data (as for effect size measures) or to measuring overlap between the entire baseline and intervention phases (as in PND analyses) may make it more difficult to examine the effects of varying levels of dosage in particular. Such analyses either ignore (effect size measures) or abstract away from (PND analyses) session-level variability during the intervention phase and, more importantly, fail to account for the fact that dosage increases cumulatively over the course of treatment. Given this observation, analyses that directly measure or model performance of session-by-session treatment probes during the treatment interval may be better suited to addressing the effects of dosage and other treatment-related variables.

A more recent meta-analysis (Oh, Eom, Park, & Sung, 2016) aimed to provide a systematic review of the treatment and generalization effects of SFA at both the item and discourse levels. Oh and colleagues identified 11 studies that met their inclusion criteria, four of which overlapped with the systematic review from Maddy et al. (2014). The 11 studies included 35 participants. They calculated effects sizes using Cohen's *d*, finding medium to large effects for trained items and small effect sizes for untrained items. Their analysis of effect sizes on discourse production, using correct information units (CIUs), varied from small to large depending on the type of outcome measures (CIU's average, CIUs per minute, percentage of CIUs). Because the choice of discourse production measure varied widely across studies, it is difficult to determine the effects of SFA on generalization to discourse (a potential measure of stimulus generalization or broader treatment-related gains). In a complementary set of analyses examining SFA-related changes on impairment-level assessments, Oh and colleagues examined performance on standardized tests using the standardized mean difference (SMD) measure. They found that, across studies, participants showed significant changes on the Western Aphasia Battery–Aphasia Quotient (WAB-AQ) score (SMD = 0.188, $p < .01$). However, changes on the Boston Naming Test (BNT) were not significant (SMD = 0.098, $p > .05$).

Oh et al.'s meta-analysis examined a significantly larger sample of participants than previous studies, using quantitative methods that directly examine the magnitude of SFA-related changes. It again found group-level evidence for the efficacy of SFA, not only for improved naming performance (treated and untreated items) but also for improvements on impairment-level assessments (such as the WAB-AQ). It also examined both acquisition and generalization treatment responses, confirming clinical intuition and informal observation that gains for treated items (acquisition) are greater than gains for untreated items (response generalization). However, like the two systematic reviews described above, this meta-analysis does not systematically examine variability across participants in either the acquisition or generalization responses to SFA, nor does it attempt to identify person-specific factors that might contribute to this variability. It also does not examine the contribution of treatment-related factors to either response to SFA. For example, stimulus selection for untreated items varies across SFA studies: Some studies (e.g., Wambaugh, Mauszycki, Cameron, Wright, & Nessler, 2013) probe naming performance on semantically related and unrelated items, as a measure of response generalization. Other studies (e.g., Boyle & Coelho, 1995) do not. Given that the spreading activation mechanism hypothesized to underlie SFA's effects should promote generalization to untreated words that share semantic features with treated items, the stimulus selection variable should affect the magnitude of SFA response generalization: Related items would be expected to show greater naming improvement than unrelated items.

Given the importance of understanding dosage and individual differences in SFA treatment effects and the lack of existing meta-analyses addressing these considerations, the current study aggregated data from across the SFA single-subject literature to address the following questions: First, what is the contribution of treatment-related variables (treatment phase: baseline vs. treatment, dosage, and stimulus type: treated vs. untreated/generalization, related vs. unrelated generalization stimuli) to response to SFA? Second, how do person-related variables (language impairment and demographic measures) contribute to response to SFA? The current study used generalized linear mixed-effects modeling (Jaeger, 2008), a novel application of an approach that is becoming standard in the psycholinguistic literature. This approach was used to combine individual session-level data from participants across treatment studies into a single repeated-measures analysis, which provided a statistically rigorous and generalizable evaluation of SFA treatment efficacy (described in further detail below).

This analysis approach has not been used to date for meta-analysis of existing treatment study data (although see Kronmüller & Barr, 2015, and Nicenboim & Vasishth, 2018, for applications of related mixed-model methods for

meta-analysis of experimental results from the psycholinguistic literature). However, it has several advantages over other meta-analytic methods for answering questions that are the focus of the current study. First, because this approach makes predictions at the trial level (e.g., the chance of naming a given treatment probe item correctly) instead of averaging across performance for an entire study phase, it is possible to look directly at the effects of dosage on probe naming performance. Standard effect size measures calculated on single-subject designs such as Cohen's *d* compare baseline with performance in the posttreatment maintenance phase and completely ignore session-by-session changes occurring during the treatment phase (Beeson & Robey, 2006). In contrast, mixed-effects models can be used to look at the cumulatively increasing effects of dosage session by modeling the change in performance over time.

A second advantage of using a mixed-effects modeling approach to examine single-subject data is that these models permit the use of multiple regression techniques. These techniques can be used to examine the simultaneous influence of multiple factors on performance (in this case, treatment probe performance). They may therefore be used to look directly at the relationships between different factors of interest, such as the interaction between treatment dosage and aphasia severity. Furthermore, by accurately specifying sources of subject- or item-specific variability using appropriate random-effects structures (Barr, Levy, Scheepers, & Tily, 2013), this approach also increases the statistical power to detect robust effects of such variables of interest, when compared with traditional effect size calculations based on mean performance.

Finally, these mixed-effects models allow the impact of both fixed effects (e.g., main variables of interest like dosage) and random effects (e.g., subject-specific variability) to be accounted for and interpreted. When the model estimates of fixed effects are standardized, they can be interpreted as measures of effect size (e.g., as a measure of the effect of one additional unit of treatment of treatment probe performance). The random effects can be used to make participant- or item-specific adjustments to these fixed-effect model predictions, thereby accounting for the error variance attributable to these sources. This analysis approach therefore permits drawing stronger population-level inferences about what SFA really does for participants in general and not only for the specific individuals studied.

Method

This study used a novel meta-analytical method on the basis of mixed-effects modeling to examine how SFA treatment-related and person-specific variables affect the likelihood of a correct response for both acquisition (improvement on treated stimuli) and generalization (improvement for untreated stimuli). Published SFA treatment data from 35 PWA taken from 12 published single-subject controlled studies were analyzed. Naming accuracy data for individual probe sessions were extracted from published figures and were modeled using mixed-effects models.

Eligibility Criteria

The starting point for this meta-analysis was Boyle's (2010) systematic review, which surveyed seven published SFA treatment studies. All seven studies from the Boyle review were included in this meta-analysis. PubMed and Google Scholar were searched to identify further published treatment studies that used SFA treatment in the context of a single-subject controlled experimental design. The inclusionary criteria for incorporating studies in this meta-analysis involved four main categories, related to research design, protocol, type of stimuli treated, and participants. In relation to research design, we included studies with single-subject designs that reported probe data per session, specifically the number of correct and incorrect responses per session. Being able to identify the number of correct responses out of the number of naming attempts per naming probe was critical, because this is the dependent variable in the mixed-model analyses reported below; this will be referred to as *session-level probe accuracy data* below. Group designs and qualitative research were excluded. In terms of the protocol, we included studies using both feature generation and feature verification variants of SFA using well-defined feature categories (see Boyle, 2010) and excluded studies where SFA was combined with other treatment approaches. The choice to require that the studies use a well-defined set of categories was because Boyle (2010), following Massaro and Tompkins (1994), identifies this property as a critical feature of SFA treatment. Related to the type of stimuli treated, we only incorporated studies that targeted nouns and excluded verb treatment studies with SFA. Finally, in relation to participants, we included studies enrolling people with diagnosed aphasia due to either stroke or closed head injury. Following Boyle (2010), we also included monolingual and bilingual speakers and participants who were 6 or more months postonset of aphasia.

The boundary for the additional search was up to the year 2015. Using the above search strategy, we found 19 additional articles. Only five of those articles fulfilled the inclusionary and exclusionary criteria above. Five articles were excluded because of the design, as they focused on discourse or did not report item-level probe data per session. One of these studies was a neuroimaging study that also did not report item-level probe data. An additional study was excluded because we were not able to determine the number of correct responses per probe session on the basis of the description of naming probe methods provided in the article. Five additional articles were excluded because of the protocol, either the treatment that they used did not include well-defined feature categories or the treatment was not SFA. The articles by Wambaugh and Ferguson (2007) and Kristensson, Behrns, and Saldert (2015) were excluded because of the type of stimuli treated, because for treatment, they used action names instead of object names. Finally, Massaro and Tompkins's (1994) article was excluded because of the participant criterion, as it did not characterize their participants as having aphasia. See Table 1 for a summary of excluded articles and additional information about the

Table 1. Excluded articles and overlap with published meta-analysis and systematic review.

Reference	Source: meta-analysis[a]/ systematic review[b]/additional search[c]	Category of exclusion: explanation
Antonucci (2009)	Meta-analysis	Design: focused on discourse; did not report session-level probe data
Falconer & Antonucci (2012)	Meta-analysis	Design: focused on discourse; did not report session-level probe data
Peach & Reuter (2010)	Systematic review	Design: did not report session-level probe data; used action names
Kiran et al. (2013)	Additional search	Design: did not report session-level probe data
Marcotte & Ansaldo (2010)	Systematic review	Design: did not report session-level probe data
Davis & Stanton (2005)	Systematic review	Design: unable to determine the number of correct responses per probe session
Kiran (2008)	Meta-analysis	Protocol: did not use well-defined feature categories following SFA treatment chart
Kiran & Johnson (2008)	Additional search	Protocol: did not use well-defined feature categories following SFA treatment chart
Kiran et al. (2011)	Additional search	Protocol: did not use well-defined feature categories following SFA treatment chart
Kiran & Thompson (2003)	Meta-analysis	Protocol: did not use well-defined feature categories following SFA treatment chart
Stanczak et al. (2006)	Additional search	Protocol: did not use well-defined feature categories following SFA treatment chart
Wambaugh & Ferguson (2007)	Systematic review	Type of stimuli: treated action names instead of objects
Kristensson et al. (2015)	Meta-analysis	Kind of stimuli: treated action names
Massaro & Tompkins (1994)	Systematic review	Participants: Participants were not reported to have aphasia

Note. None of these articles was included in the review by Boyle (2010).
[a]Meta-analysis by Oh et al. (2016). [b]Systematic review by Maddy et al. (2014). [c]Additional search from the current meta-analysis authors.

overlap between the studies included in the meta-analysis from Oh et al. (2016) and the systematic review from Maddy et al. (2014). Having applied the inclusionary criteria, we selected a set of 12 studies to be analyzed. Refer to Table 2 for details of the selected studies for the current meta-analysis and additional information about the overlap between the studies included in the meta-analysis from Oh et al. (2016) and the systematic review from Maddy et al. (2014).

Data Variables of Interest

The effects of both treatment-related and person-specific variables on performance were measured. Treatment-related variables were associated with properties and parameters of the SFA treatment, as opposed to being associated with the participants. There were three such variables: treatment phase, dosage, and stimulus type. (a) Treatment phase was whether the naming probe was in the baseline or the treatment phase. We hypothesized that people would have a higher likelihood of a correct response in the treatment phase compared with the baseline phase; this would be an evidence of a positive effect of SFA treatment. Data for the baseline phase were drawn exclusively from the baseline phases of the studies analyzed. The treatment phase data were drawn from the first treatment/intervention phase of each of the studies. Stadie and colleagues (2008) argue that later treatment phases in single-subject studies reflect the cumulative effects of treatment in previous phases, targeting different stimulus types, and may also reflect general learning or adaptation to the treatment task, making interpretation of data from later phases difficult. Analyzing only data from the first treatment phase avoids these issues. Each observation in the treatment phase represents a presession probe, administered immediately before the treatment session; this probe procedure is standard in the single-subject literature (e.g., Thompson, 2006). (b) Dosage was operationalized as the number of SFA treatment sessions that people received during the intervention phase. Performance in each presession probe reflects the effect of the number of treatment sessions completed before that probe: For example, the fifth presession probe would reflect the impact of receiving four sessions of SFA treatment. More fine-grained measures of dosage, such as the number of minutes of treatment received or the number of treatment trials completed, could not be reconstructed on the basis of the information available from the published articles (see Gravier et al., 2018, for a discussion of how different measures of dosage may be related to SFA response). We hypothesized that people would have a higher likelihood of a correct response with a greater number of SFA treatment sessions. (c) Stimulus type was defined in two ways. The first one was acquisition (treated items) versus generalization (untreated items). This comparison allowed us to examine and measure the two treatment responses directly. The second one was semantically related items versus semantically unrelated items among the generalization stimuli. We hypothesized that participants would be more likely to have a correct naming response for treated items compared with untreated items, consistent with clinical impression and findings by Oh et al. (2016). Because

Table 2. Treatment studies included in the present analysis.

Primary study	Meta-analysis participant number	Reported bilingual by study	Task	Number of treatment sessions per week	Reported semantic relatedness in the generalization items
Boyle (2004)[a,b,c]	1	No	SFG	3	No
	2	No	SFG	3	No
Boyle & Coelho (1995)[b,c]	3	No	SFG	3	No
Lowell et al. (1995)[c]	4	No	SFG + SFR	3	Yes
	5	No	SFG + SFR	3	Yes
	6	No	SFG + SFR	3	Yes
Rider et al. (2008)[a,b,c]	7	No	SFG	2–3	No
	8	No	SFG	2–3	No
	9	No	SFG	2–3	No
Coelho et al. (2000)[b,c]	10	No	SFG	3	No
Wallace & Kimelman (2013)[d]	11	No	SFG + SFR		Yes
	12	No	SFG + SFR		Yes
	13	No	SFG + SFR		Yes
Wambaugh et al. (2013)[a,b]	14	No	SFG	3	Yes
	15	No	SFG	3	Yes
	16	No	SFG	3	Yes
	17	No	SFG	3	Yes
	18	No	SFG	3	Yes
	19	No	SFG	3	Yes
	20	No	SFG	3	Yes
	21	No	SFG	3	Yes
	22	No	SFG	3	Yes
Edmonds & Kiran (2006)[c]	23	Yes	SFR	2	Yes
	24	Yes	SFR	2	Yes
	25	Yes	SFR	2	Yes
Kiran & Roberts (2010)[c]	26	Yes	SFR		Yes
	27	Yes	SFR		Yes
	28	Yes	SFR		Yes
	29	Yes	SFR		Yes
Conley & Coelho (2003)[a]	30	No	SFG	3	No
Hashimoto & Frome (2011)[a,b]	31	No	SFG + SFR	2	No
DeLong et al. (2015)[a]	32	No	SFG	3	Yes
	33	No	SFG	3	Yes
	34	No	SFG	3	Yes
	35	No	SFG	3	Yes
	36	No	SFG	3	Yes

Note. SFG = semantic feature generation; SFR = semantic feature review.

[a]Meta-analysis by Oh et al. (2016). [b]Systematic review by Maddy et al. (2014). [c]Review by Boyle (2010). [d]Additional search from the current meta-analysis authors.

SFA targets the semantic network, we also hypothesized that related items would show greater likelihood of a correct response compared with unrelated stimuli for the generalization probes.

In addition, we analyzed person-specific variables, as results from these analyses could help identify which particular individuals are more likely to show greater acquisition or generalization in response to SFA. Two sets of person-specific variables were included. (a) The first set examined language impairment variables and included overall aphasia severity, as measured by WAB-AQ (Kertesz, 1982), and naming impairment severity, as measured by the BNT raw score (Kaplan, Goodglass, & Weintraub, 2001). Outcomes on both these variables were examined in Oh et al.'s (2016) SFA meta-analysis. We hypothesized that less impaired performance on these measures would be associated with better responses to SFA treatment. Previous work has found that people with milder aphasia poststroke (e.g., as measured by higher WAB-AQ scores) tend to show better recovery (e.g., Watila & Balarabe, 2015). The question of whether pretreatment naming ability is predictive of response to naming treatments has not been systematically addressed in the literature. However, recent findings from an ongoing SFA treatment study (Gravier et al., 2018) suggest that pretreatment naming ability (as measured by naming performance on the Comprehensive Aphasia Test; Swinburn, Porter, & Howard, 2004) was not predictive of treatment-related naming improvements. (b) The second set of person-specific variables were demographic variables, including age and months postonset of aphasia.

We included these particular person-specific variables because these variables were most consistently reported across the studies. Some studies provided additional testing data, like performance in the Test of Adolescent/Adult Word Finding, the Reading Comprehension Battery for

Aphasia–Second Edition, and the Aphasia Diagnostic Profiles. However, these were not included because only some of the articles reported them. Likewise, we included the stimulus types described above (treated vs. untreated, semantically related vs. semantically unrelated) because they were reported for most studies. Some studies had additional stimulus-type manipulations, like typicality of the stimulus items (Wambaugh et al., 2013) and treatment of few versus many exemplars (Boyle & Coelho, 1995). However, we collapsed across these additional distinctions in the current analyses. See Table 2 for details regarding the selected studies.

Data Extraction Process

Three different coders performed data extraction for 11 of the studies in the current meta-analysis. Two coders independently extracted the variables from the same article to provide a reliability measure. The third coder reconciled the data points for which there was disagreement between the two initial coders.

The first step that coders followed was to extract information related to the study participants: (a) number of participants; (b) demographic variables for those participants, including age, time postonset in months, and gender; and (c) language impairment variables for the participants, including aphasia severity as measured by the WAB-AQ and naming impairment severity as measured by raw score on the BNT. The next step involved extracting information pertinent to treatment-related variables, including the number of treatment phases, number of exemplars/items in each treatment or baseline list, duration (in hours/minutes) of each treatment session (where available), number of treatment sessions per week (where available), and whether participants produced features as part of SFA treatment procedures or verified them. The following variables were consistently reported across studies and were extracted and coded: (a) treatment phase, corresponding to baseline versus treatment phases; (b) stimulus type, corresponding to treatment versus generalization, as well as generalization stimulus type, corresponding to related versus unrelated; and (c) dosage, measured in the number of sessions in each treatment phase. These variables served as predictor variables in the statistical models described below.

Once these variables were extracted, the coders then extracted the data for the dependent variable, session-level probe accuracy data. To do this, two pieces of information were extracted. First, the total number of items that were probed in each naming probe was identified, on the basis of the methods described in each published article. Second, the raw naming accuracy scores for each naming probe were extracted, measured in terms of the number of probe items that were correctly named out of the total number of items that were probed in each naming probe. If the article reported percentages, those were converted into raw naming accuracy scores. Once the two coders had completed the information, a third rater compared the two coders' data-entry variables, one participant at a time. If there were disagreements or mismatches, the rater went back to the article to verify the correct information. For the 12th study in this meta-analysis, Wambaugh et al. (2013) shared raw study data in SigmaPlot 13.0 (Systat Software, 2018) format. These data were extracted using SigmaPlot 13.0 and combined with the information obtained from the other 11 articles.

Statistical and Modeling Techniques

The likelihood of a correct response in a naming probe (item-level naming probe accuracy) was modeled using generalized mixed-effects logistic regression models (Jaeger, 2008) and included random intercepts and random slopes for subjects (Baayen, Davidson, & Bates, 2008). The analyses were conducted in the R statistical software (R Development Core Team, 2008) using RStudio Version 1.0.136 with a logistic link function in the "lme4" package (Bates, Mächler, Bolker, & Walker, 2015). Figures were made with the "ggplot2" package, Version 2.2.1 (Wickham & Chang, 2008). Our data set and our analysis scripts are freely available for download at http://osf.io/h6mqj/

Fixed Effects (Predictor Variables)

The effects of the predictor variables of interest on item-level probe data (raw naming accuracy) were examined in a series of logistic mixed-effects regression models. (a) The models that aimed to identify the effect of treatment phase included the fixed effects of treatment phase (baseline and treatment phases only), stimulus type (treated vs. untreated generalization items), and their interaction. (b) The models that aimed to identify the effects of dosage included the fixed effects of dosage (number of SFA treatment sessions that people received during the intervention phase), stimulus type (treatment vs. generalization), and their interaction. (c) The models that aimed to identify the effects of generalization stimulus type included the fixed effects of semantic relatedness (related vs. unrelated generalization items), dosage, and their interaction. (d) The models that examined the effects of language impairment variables included fixed effects of aphasia severity and naming impairment severity (both converted into z scores), treatment phase, stimulus type (treated vs. generalization), and their two- and three-way interactions. Separate models were also fit for aphasia and naming impairment severity variables to look at their overall effects on task performance. (e) Finally, the models examining demographic variables included age or months postonset (centered around the mean), treatment phase, stimulus type (treated vs. generalization), and their interaction, all as fixed effects. Separate models were fit for age and months postonset.

Random Effects

The random-effects structure included both random intercepts and random slopes nested within participants for the fixed effects of interest described above. Because of convergence issues, the random slopes and random intercepts were uncorrelated. Random intercepts were included in the random effects to capture the fact that participants

could vary in their average accuracy; for example, some participants may exhibit higher response accuracy than others, independent of other factors (such as treatment-related variables like stimulus type or dosage) that might affect response accuracy. Inclusion of random slopes in the models captured the fact that participants could vary in the magnitude of their fixed effects; for example, some participants might show larger or smaller dosage effects. Accounting for this variability via the random-effects structure increases the specificity of the population level inferences when interpreting the fixed effects, as noted above. Because item-specific data were not provided in the published articles (e.g., specific items included in the naming probes), items were not included in the random-effects structure. It has been argued that random-effects structures affect the overall generalizability of an analysis (Barr et al., 2013); for that reason, models were kept as maximal as the data permitted. See Table 3 for fixed- and random-effects structures and variable coding used in each of the models reported below.

Outcome Variable (Dependent Variable)

The outcome variable was session-level probe accuracy, converted to log-odds. In other words, the models were designed to predict accuracy of naming responses in each naming probe, measure in the log-odds likelihood of a correct response to a naming probe. Using this approach, we can observe the relation between the fixed effects (described above) and the predicted variable (naming accuracy) in a direct manner: The regression coefficients for each fixed effect (predictor variable) represent the amount that one additional unit of that fixed effect increases or decreases in the likelihood of a correct response. For example, the regression coefficient for the dosage variable may be interpreted as the effect of one additional session of SFA treatment on the log likelihood of correct naming probe response.

Variation Across Studies

It should be noted that there is a great deal of variability across the set of studies included in this meta-analysis and across individual participants within those studies. For example, the exact number of treatment trials per session differed between studies, and some studies included treatment for bilingual participants in other languages (e.g., Edmonds & Kiran, 2006). However, the inclusion of well-specified random effects for subjects should largely control for these differences, allowing model fixed effects to be interpreted as accurate inferences about SFA treatment response in the population with aphasia in general. This is because, if a study-specific source of variability significantly affected an individual's performance, this would be reflected in their specific slope and intercept random-effect adjustments and modulate the overall contribution of their data to the final fixed-effect estimates.

Results

We analyzed 12 studies that included 35 participants: 15 female and 21 male (see Table 4 for participant information). Participants' age ranged from 30 to 87 years (M = 60.13 years, SD = 10.46 years), and their time postonset of aphasia ranged from 6 to 384 months (M = 53.9 months, SD = 72.53 months). Two of the studies had bilingual participants (Edmonds & Kiran, 2006; Kiran & Roberts, 2010). In regard to the protocol, three of the 12 studies used a combination of semantic feature generation with semantic feature review, two used semantic feature review, and the remaining seven used semantic feature generation.

From the data extraction process, there were 5,616 individual observations of session-level probe data. This included the information extracted from 11 articles as well as the raw data provided by Wambaugh and colleagues (2013). The two coders had a 75% agreement for the data extracted. The remaining 25% of observations for which there was disagreement were reconciled by the third rater.

In relation to the treatment-related variables, we first examined the baseline phase compared with the treatment phase. This model examines whether there was an overall treatment effect for SFA across all participants in the meta-analysis. There was an effect of treatment phase (estimate = 1.29, SE = 0.17, z = 7.72, $p < .001$), with a significant predicted increase in performance accuracy between the baseline and treatment phases (from 13% to 47%, an increase of 34%, collapsing across stimuli type). However, this increase in performance from the baseline to treatment phase interacted with stimulus type (treated vs. generalization items), with a positive effect of treatment for both stimuli types but with treated stimuli showing significantly more improvement than generalization stimuli (an average baseline-to-treatment accuracy increase of 46% for treated items and only 22% for generalization items; estimate = 0.94, SE = 0.11, z = 8.68, $p < .001$). See Figure 1 for an illustration of these effects.

Subsequent models focused on the treatment phase. In relation to the treatment-related variable associated with dosage, there was a positive effect of dosage during the treatment phase (estimate = 0.09, SE = 0.02 z = 3.72, $p < .001$), such that increasing the number of SFA treatment sessions led to increasing probe naming accuracy. There was also an effect of stimulus type within the treatment phase (estimate = 0.52, SE = 0.17, z = 3.04, $p < .001$), with treated items showing higher accuracy performance than generalization items. However, there was a significant interaction between these effects, with treated stimuli showing a greater effect of increasing dosage than generalization stimuli (estimate = 0.09, SE = 0.02, z = 5.21, $p < .001$). One straightforward way to interpret these effects is to use model output to make performance predictions at specific values of dosage for each stimuli type: After 15 sessions of treatment, the model predicts that probe naming accuracy would increase 52% for treated stimuli but only 32% for untreated stimuli. Figure 2 further illustrates the effects of this analysis.

Table 3. Fixed and random effects and variables coding.

Model	Fixed effects (name in the R code)	Fixed effects (target variables of interest)	Coding type	Random effects (name in R code)
1. Treatment phase	TreatmentCondition	Baseline vs. treatment phases	Dummy coding (ref = baseline)	0 + TreatmentCondition|PartCode
	TreatGen	Acquisition vs. generalization	Dummy coding (ref = generalization)	0 + TreatGen|PartCode 1|PartCode
2. Dosage	Dosage	Dosage in number of sessions for the treatment phase	Continuous variable	0 + Dosage|PartCode
	TreatGen	Acquisition vs. generalization	Dummy coding (ref = generalization)	0 + TreatGen|PartCode 1|PartCode
3. Semantic relatedness	Dosage	Dosage in number of sessions	Continuous variable	0 + Dosage|PartCode
	GenType	Related vs. unrelated for generalization stimuli	Dummy coding (ref = unrelated)	0 + GenType|PartCode 1|PartCode
4. Language impairments	TreatmentCondition	Baseline vs. treatment	Contrasts coding (–0.5 for baseline and 0.5 for treatment)	1 + TreatmentCondition|PartCode
Aphasia severity	AphasiaSeverity.z	Aphasia severity (WAB-AQ)	Aphasia severity (*z* scores)	1 + AphasiaSeverity|PartCode
	TreatGen	Acquisition vs. generalization	Contrasts coding (–0.5 for generalization and 0.5 for acquisition)	
4a. Language impairments	TreatmentCondition	Baseline vs. treatment	Contrasts coding (–0.5 for baseline and 0.5 for treatment)	1 + TreatmentCondition|PartCode
Naming	NamingImpairment.z	Naming impairment (BNT)	Naming impairment (*z* scores)	1 + NamingImpairment.z|PartCode
	TreatGen	Acquisition vs. generalization	Contrasts coding (–0.5 for generalization and 0.5 for acquisition)	
5. Demographic variables	TreatmentCondition	Baseline vs. treatment	Dummy coding (ref = baseline)	1 + TreatmentCondition|PartCode
Age	Age.cen	Age (years)	Age (centered at the mean)	0 + Age.cen|PartCode
5a. Demographic variables	TreatmentCondition	Baseline vs. treatment	Dummy coding (ref = baseline)	1 + TreatmentCondition|PartCode
MPO	MPO.z	Months postonset (*z* scores)	MPO (centered)	1 + MPO.cen|PartCode

Note. Random slopes for fixed effects were included in each model when possible. However, convergence issues required us to drop random slopes for TreatGen in Models 4 and 4a. WAB-AQ = Western Aphasia Battery–Aphasia Quotient; BNT = Boston Naming Test; ref = reference group.

The last treatment-related variable analyzed, generalization stimulus type, examined the semantic relatedness of the items used during the generalization probes. This analysis revealed a positive interaction effect between semantic relatedness (related vs. unrelated) and dosage (estimate = 0.09, $SE = 0.02$, $z = 3.81$, $p < .001$), such that the effect of dosage was significantly greater for semantically related compared with unrelated items. After 15 sessions of treatment, the model predicts that probe naming accuracy would increase 47% for semantically related generalization probes but only 12% for semantically unrelated generalization probes.

The next set of models examined person-specific variables. In relation to the language impairment variables, we looked at overall aphasia severity, as measured by WAB-AQ, and at naming impairment, as measured by performance on the BNT. There was a main effect of aphasia severity across the baseline and treatment phases (estimate = 0.42, $SE = 0.17$, $z = 2.43$, $p < .01$). This reflects that milder aphasia severity is associated with better overall naming performance. In addition, we found a three-way interaction effect between treatment phase, stimulus type, and aphasia severity (estimate = −0.35, $SE = 0.13$, $z = -2.57$, $p < .01$), such that PWA with milder aphasia showed disproportionately large generalization effects in the treatment phase, but not in the baseline phase. See Figure 3 for a depiction of this interaction.

Naming impairment was analyzed using the BNT scores across the baseline and treatment phases. For the seven participants within the bilingual studies (meta-analysis

Table 4. Participant information.

Primary study	Original participant number	Meta-analysis participant number	Age as reported by study	Months postonset as reported by study[a]	Gender as reported by study	WAB score as reported by study	BNT score as reported by study	Aphasia type as reported by study	Etiology
Boyle (2004)	1	1	70	15	M	90.6	49	Anomic	CVA
	2	2	80	14	M	61.2	27	Wernicke's	CVA
Boyle & Coelho (1995)	1	3	57	65	M	82	43	Broca's	CVA
Lowell et al. (1995)	1	4	74	16	M			Conduction	CVA
	2	5	76	9	M			Anomic	CVA
	3	6	66	30	M			Conduction	CVA
Rider et al. (2008)	1	7	73	26	M	74.6	42	Transcortical motor	CVA
	2	8	55	45	F	76.5	29	Transcortical motor	CVA
	3	9	62	126	M	65.8	20	Broca's	CVA
Coelho et al. (2000)	1	10	52	17	M	56.6	11	Fluent aphasia	TBI
Wallace & Kimelman (2013)	1	11	47	32	M	55.8[b]		Broca's	CVA
	2	12	63.5	90	M	83.8[b]		Anomic	CVA
	3	13	57.2	134	F	55.4[b]		Broca's	CVA
Wambaugh et al. (2013)	1	14	58	126	M	53.4		Broca's	CVA
	2	15	59	42	F	82		Anomic	CVA
	3	16	61	31	M	63		Broca's	CVA
	4	17	47	187	M	66		Broca's	CVA
	5	18	59	65	M	50.8		Broca's	CVA
	6	19	52	13	M	66		Broca's	CVA
	7	20	66	65	M	70.7		Broca's	CVA
	8	21	64	18	F	90.6		Anomic	CVA
	9	22	54	9	M	33.2		Wernicke's	CVA
Edmonds & Kiran (2006)	1	23	53	9	F	67.5	25[c]		CVA
	2	24	53	8	M	27	1[c]		CVA
	3	25	56	9	F	61.3	14[c]		CVA
Kiran & Roberts (2010)	1	26	55	11	F		34[d]	Expressive	CVA
	2	27	87	6	F		5[d]	Receptive	CVA
	3	28	55	33	F		22[d]	Anomic	CVA
	4	29	60	15	F		42[d]	Nonfluent	CVA
Conley & Coelho (2003)	1	30	57	96	F	46.3		Nonfluent	CVA
Hashimoto & Frome (2011)	1	31	72	120	F	35	12	Broca's	CVA
DeLong et al. (2015)	1	32	62	11	F	64.5[b]		Conduction	CVA
	2	33	54	30	M	58.3[b]		Wernicke's	CVA
	3	34	30	23	M	66.1[b]		Broca's	CVA
	4	35	53	384	F	78.4[b]		Anomic	CVA
	5	36	65	12	F	18[b]		Global	CVA

Note. WAB-AQ = Western Aphasia Battery–Aphasia Quotient (maximum score = 100); BNT = Boston Naming Test (maximum score = 60); M = male; F = female; CVA = cerebrovascular accident; TBI = traumatic brain injury.

[a]If months postonset were reported in years, the values were converted into months. [b]Western Aphasia Batter-Revised. [c]The scores were converted into raw score data; the authors reported percentages. [d]Bilingual participant, reporting BNT in English.

Figure 1. Effects of treatment phase on treated versus generalization (untreated) items, as measured on probe accuracy. Error bars reflect bootstrapped 95% confidence intervals generated via bootMer function.

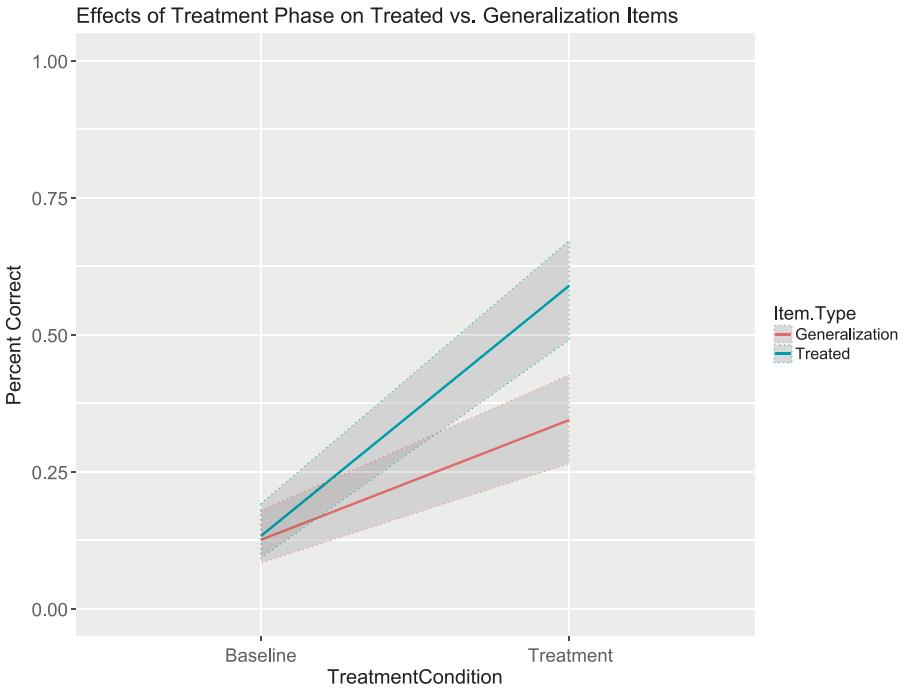


Figure 2. Effects of dosage on treated versus generalization (untreated) items. Error bars reflect bootstrapped 95% confidence intervals generated via bootMer function.

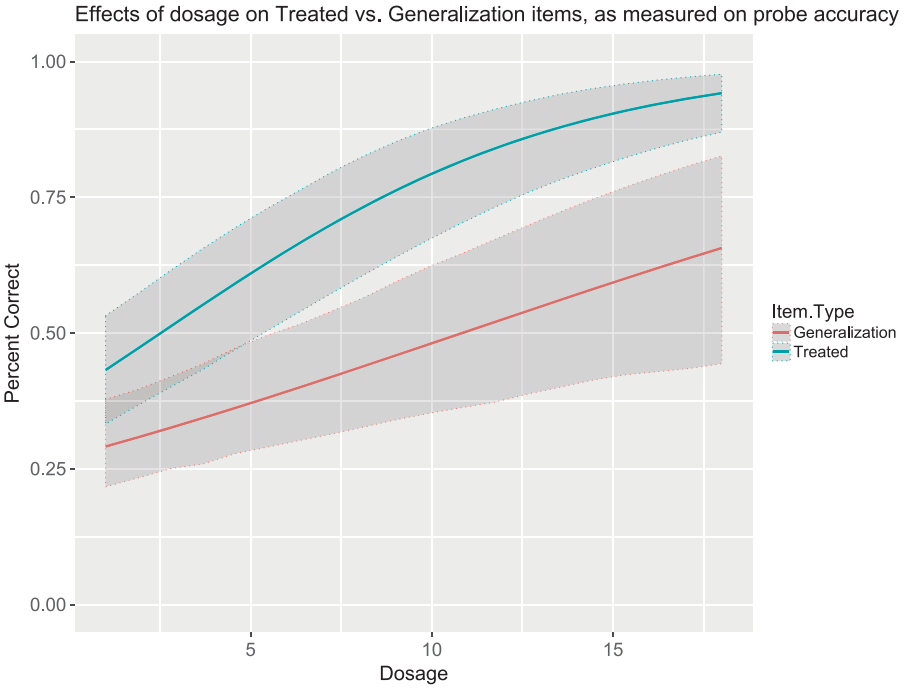
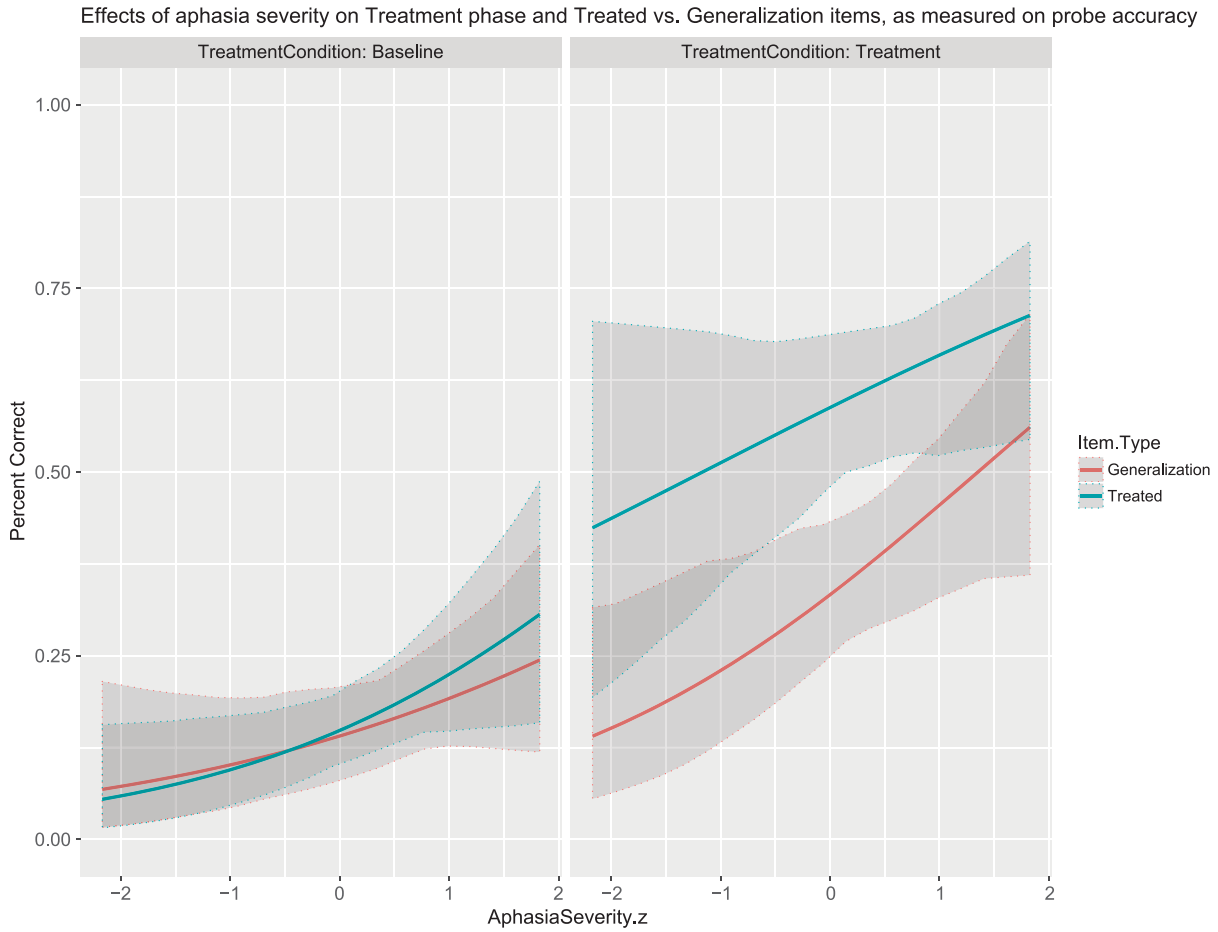


Figure 3. Effects of aphasia severity on the treatment phase and treated versus untreated items, as measured on probe accuracy. Error bars reflect bootstrapped 95% confidence intervals generated via bootMer function.



participant numbers 23–29), the authors reported BNT scores for both languages. For the most part, bilingual participants were balanced across languages with minimal differences in cross-linguistic BNT scores; therefore, we included English BNT scores in this analysis to be consistent with our larger participant group. See Table 5 for detailed information about the balance in the languages that the participants spoke. However, Participants 26 and 28 presented unbalanced BNT results for the languages that the studies reported; therefore, these two participants were dropped from the naming impairment analysis. With the remaining participants, there was an effect of naming impairment severity across the baseline and treatment phases (estimate = 0.59, $SE = 0.15$, $z = 3.78$, $p < .001$), such that

Table 5. Boston Naming Test (BNT) scores for the bilingual participants.

Primary study	Meta-analysis participant number	BNT raw score in other languages	BNT raw score in English	BNT, *M* (*SD*)
Edmonds & Kiran (2006)	23	25 Spanish	25	25 (0)
	24	0 Spanish	1	0.5 (0.71)
	25	11 Spanish	14	12.5 (2.12)
Kiran & Roberts (2010)	26	6 Spanish	34	20 (19.8)[a]
	27	3 Spanish	5	4 (1.41)
	28	39 French	22	30.5 (12.2)[a]
	29	43 French	42	42.5 (0.71)

[a]These values were excluded from the analysis because of their high standard deviation.

milder naming impairments were associated with better overall naming accuracy across the treatment phases. However, the effect of naming impairment did not moderate the relationship between treatment phases and stimulus type when tested in a three-way interaction (estimate = 0.31, $SE = 0.17$, $z = -1.79$, $p = .07$).

Finally, in relation to the demographic variables, age (estimate = −0.02, $SE = 0.02$, $z = -0.81$) and months postonset (estimate = −0.01, $SE = 0.14$, $z = -0.78$) did not show significant main effects or interactions with any of the treatment-related variables ($ps > .05$). This suggests that SFA may be an efficacious treatment for people with chronic aphasia regardless of their age or stage of recovery.

Discussion

This meta-analysis used a novel statistical approach to analyze data from single-subject controlled experimental design studies targeting SFA. Specifically, mixed-effects models were used to analyze not only pretreatment to posttreatment changes in naming accuracy (as in traditional meta-analyses) but also the full set of treatment data, including all naming probe observations in the baseline and treatment phases. This greatly increased sample size and granularity for the purposes of addressing specific study questions. In addition, the multilevel logistic regression models were able to control for the contributions of participant and study-specific differences via random-effects structures to draw stronger conclusions about the determiners of SFA treatment response in general. Together, these techniques allowed us to examine how several different variables affected the likelihood of correct naming responses during naming probes, as reported in the aggregated studies. The effects of two types of variables on naming performance were measured. The first type involved treatment-related variables, including treatment phase (baseline vs. treatment), dosage (number of treatment sessions), and stimulus type (treated vs. untreated object names, semantically related vs. unrelated generalization stimuli). The second type involved person-specific variables, including degree of language impairment and demographic variables (age, time poststroke).

The results provide group-level evidence of the efficacy of SFA, aggregating performance across a sizable ($n = 35$) and varied sample of PWA. We take *efficacy* to mean "the treatment working in ideal conditions," differing from *effectiveness* (defined as "the treatment working in everyday clinical practice") and *efficiency* (defined as "the most efficient way to apply the treatment program"; Papathanasiou & Coppens, 2016, p. 9). The single-subject results aggregated here may be seen as representing the effects of SFA as applied in ideal conditions. Results showed a robust effect of treatment phase with an increase in naming performance of 34% from the baseline to treatment phases. This supports the claim that SFA is an efficacious treatment for object naming deficits in PWA. Consistent with other observations in the aphasia treatment literature, treatment-related gains were also larger for acquisition (treated) than generalization (untreated) stimuli in this analysis.

Within the treatment phase, increasing SFA dosage was associated with increasing naming accuracy, and it appears that increasing dosage had a larger effect for directly treated items compared with untreated (generalization) stimuli. Likewise, the effects of dosage were greater for semantically related compared with unrelated items. This second finding provides novel evidence consistent with the mechanism that is supposed to underlie the positive effects of SFA on naming performance. SFA treatment activities (e.g., feature generation) target the semantic system through repeated structured practice. This stimulation should spread activation to adjacent concepts in the semantic system, thereby facilitating retrieval of lexical items that have overlapping or related features (Collins & Loftus, 1975). Therefore, this should lead to greater improvement for semantically related untreated items that share semantic features with the treated items. Interestingly, this finding also converges with recent evidence by Gravier et al. (2018) from a large ongoing SFA treatment study: They found that the number of semantic features generated per treatment trial was a reliable predictor of both acquisition (improvement on treated items) and response generalization (improvement on untreated but related items), whereas other practice-related predictors were not. The findings of the current study thus provide group-level evidence of response generalization (Thompson, 1989) after SFA.

It is worth noting that the current meta-analysis is the first study to directly examine the effects of varying SFA dosages on naming performance. The question of dosage is an active and growing area of interest in the field of aphasiology (Cherney, 2012; Cherney et al., 2011; Gravier et al., 2018; Hinckley & Craig, 1998; see also Warren, Fey, & Yoder, 2007). Quantifying the magnitude of the effect of varying dosages of different interventions (determining dose–response relationships) is a critical step in evaluating the effectiveness of aphasia treatment. The analysis approach used in the current study was able to examine the effects of dosage directly by modeling change in naming performance over the treatment phase (in response to increasing amounts of SFA treatment). Other meta-analysis approaches that ignore continuous changes during the treatment phase (e.g., those using effect size measures) are at a significant disadvantage for observing such effects. The current study was able to quantify SFA dosage effects, for the first time, providing an estimate of how much improvement in probe naming accuracy may be expected for different amounts of SFA treatment. This illustrates the potential value of the current analysis methods; they could be applied to other aphasia treatments with significant numbers of published single-subject controlled experimental design studies to quantify the effects of dosage for those treatments.

The current findings also revealed another interesting pattern related to the effects of varying SFA dosages on naming performance. Figure 2 shows a noteworthy difference in the shape of the treatment response curves for treated

and generalization (untreated) items over the treatment phase. Treated items exhibit an asymptotic curve, with the slope of naming accuracy performance decreasing toward the end of the treatment phase. In contrast, the curve for generalization items has a simpler linear shape: Naming accuracy continues to improve roughly linearly with additional treatment sessions. This contrast suggests that the two stimulus types may exhibit different relationships to increasing dosages of SFA. SFA treatment may reach a "ceiling" for treated (acquisition) items, an asymptotic point in the treatment response curve after which the rate of improvement for those items notably declines. However, generalization items may continue to show improvement in response to additional treatment even after this point, as suggested by the more linear shape of these items' treatment response curve. This finding may be consistent with the principle of overtraining (Raymer et al., 2008): To promote response generalization, it may be necessary to overtrain treated items, past the point where they exhibit continued improvement.

In relation to the person-specific variables, aphasia severity and naming impairment severity both had a significant impact in overall naming performance: Participants with milder impairments exhibited better naming performance, regardless of treatment phase (baseline or treatment). However, aphasia severity had a significant effect on treatment-related improvements for generalization stimuli in particular, as demonstrated by the significant three-way interaction of aphasia severity, treatment phase, and stimulus type. This interaction is plotted in Figure 3. Participants with milder aphasia (higher WAB-AQ scores) showed a smaller difference between treated and generalization (untreated) items, suggesting that people with milder chronic aphasia may show stronger response generalization after SFA. The aphasia treatment literature has often failed to find reliable predictors of generalization after treatment, even for studies that have successfully identified predictors of acquisition. For example, Meinzer et al. (2010) found lesion correlates of acquisition but not generalization response in a prospective naming treatment study, and Dickey and Yoo (2010) identified a behavioral predictor of stimulus acquisition but no predictors of response generalization in a meta-analysis of published sentence-production treatment studies. The current findings provide promising preliminary evidence of a person-level predictor of generalization in response to SFA treatment. This effect may ultimately reflect the dynamics of spreading activation within a compromised semantic system. It is likely that increased activation of semantic representations due to targeted stimulation (as in SFA) would have a better chance of spreading to adjacent semantic nodes in a network with less noise or less degradation (i.e., in a system with milder semantic impairment).

With respect to the other person-specific variables investigated here, it appears that demographic variables were not related to treatment-related changes in naming probe performance. This suggests that SFA may be an efficacious treatment for people with chronic aphasia regardless of their age or their point in the chronic stage of stroke recovery.

Limitations and Further Research

The current study had a number of limitations. First, although it examined both acquisition and generalization responses to SFA, the analysis was limited to only one type of generalization—response generalization. This reflects a limitation of the single-subject controlled experimental studies that were aggregated in this meta-analysis. No probes of stimulus generalization were reported in the targeted studies, with the exception of Rider, Wright, Marshall, and Page (2008). That study reported a single pre-to-post change score in the number of treated words that were produced during a story-retell procedure (one measure of stimulus generalization). However, even if this stimulus generalization measure had been collected in all the other studies, a single pre-to-post measure would suffer from many of the same limitations as effect size measures. It would provide a much reduced number of observations (one per participant), and it would also be limited in its ability to reveal effects of varying levels of dosage. Because response generalization was measured continuously throughout the baseline and treatment phases of the studies compiled for this meta-analysis, the analysis methods used here were able to address response generalization directly.

Second, the current study did not address treatment-related changes on broader measures of communicative performance, such as standardized aphasia or naming assessments. This is in contrast to the meta-analysis by Oh et al. (2016), which examined changes on the WAB and BNT. Examining such changes was not the focus of the current study. However, it is worth noting that these single pre-to-post change measures present the same issues as those noted for effect size measures above. To take full advantage of the modeling methods used in the current meta-analysis to examine changes on such measures, it would be necessary to collect standardized measures at multiple points during the baseline and treatment phases. Given the development of both short forms (Walker & Schwartz, 2012) and computer-adaptive versions (Hula, Kellough, & Fergadiotis, 2015) of standardized naming assessments, reducing both administration time and item overlap between test versions, this may be possible in future studies.

Finally, there are multiple questions remaining to be explored in the application of SFA as naming treatment for people with chronic aphasia that were not addressed in the current study. For example, this meta-analysis did not answer questions regarding the effect of typicality on acquisition or response generalization after SFA (Wambaugh et al., 2013; see also Kiran & Thompson, 2003) as well as regarding the use of few versus many exemplars in treatment (e.g., Boyle & Coelho, 1995). Because not all the aggregated studies examined these questions, the current meta-analysis lacked power to address these issues. One final question that was not addressed in the current analysis was the potential difference between feature generation

versus feature verification variants of SFA, identified by Boyle (2010) as another factor that may impact response to semantic-feature–based naming treatments. This is an interesting open question to be explored in future research.

Conclusions

This meta-analysis compiled 12 studies that used SFA as naming treatment for 35 PWA. It examined how both treatment-related and person-specific variables affected SFA treatment response, measured as changes in the likelihood of correct naming responses for acquisition (treated) and generalization (untreated) stimuli. The findings provided statistically robust evidence of a number of findings from the SFA single-subject treatment literature, but from a larger, more representative, and potentially generalizable sample of PWA. It found large and reliable treatment-related improvements in naming performance; these gains were greater for treated than untreated items and for semantically related generalization items than unrelated items. These findings provide group-level evidence of the efficacy of SFA in promoting both acquisition and stimulus generalization as well as evidence that is consistent with the spreading activation mechanism that is hypothesized to underlie the effects of SFA. The evidence to date for the positive effects of SFA has come almost exclusively from carefully conducted single-subject designs; this study is one of the first to provide group-level evidence, generalizable to the larger population of PWA. Furthermore, the analysis methods used in the current meta-analysis revealed two previously unobserved patterns in the existing SFA treatment data. First, they provided initial evidence for a person-specific predictor of SFA treatment response, in particular, generalization (overall aphasia severity). Second, they enabled us to measure and quantify dosage effects for SFA, providing preliminary estimates of how much naming performance benefit is engendered by varying dosages of SFA. These latter findings represent critical first steps toward optimizing patient selection and treatment delivery parameters for this particular aphasia treatment; they may be leveraged for future studies examining treatment effectiveness (moving beyond simple efficacy). As such, they provide an illustration of the potential value of the analysis methods used for the current study and the approach of aggregating findings from the existing single-subject aphasia treatment literature.

Acknowledgments

This research was supported by VA Rehabilitation Research and Development Award I01RX000832 to the last author and the VA Pittsburgh Healthcare System Geriatric Research Education and Clinical Center. The authors are grateful to Dr. Julie Wambaugh and colleagues for sharing original data from their 2013 article, to Dr. William Hula for discussion and input regarding the statistical analysis, and to Beth Friedmann, Gina D'Amore, Anish Kumar, and Jack Snowdon for their work in data extraction. The contents of this article do not represent the views of the Department of Veterans Affairs of the U.S. Government.

References

Abel, S., Weiller, C., Huber, W., & Willmes, K. (2014). Neural underpinnings for model-oriented therapy of aphasic word production. *Neuropsychologia, 57,* 154–165.

Antonucci, S. M. (2009). Use of semantic feature analysis in group aphasia treatment. *Aphasiology, 23*(7–8), 854–866.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

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.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16*(4), 161–169.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34*(4), 987–993.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology, 13*(3), 236–249.

Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation, 17*(6), 411–422.

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology, 4*(4), 94–98.

Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology, 21*(4), 397–414.

Cherney, L. R. (2012). Aphasia treatment: Intensity, dose parameters, and script training. *International Journal of Speech-Language Pathology, 14*(5), 424–431.

Cherney, L. R., Patterson, J. P., & Raymer, A. M. (2011). Intensity of aphasia therapy: Evidence and efficacy. *Current Neurology and Neuroscience Reports, 11*(6), 560–569.

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.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428.

Conley, A., & Coelho, C. (2003). Treatment of word retrieval impairment in chronic Broca's aphasia. *Aphasiology, 17*(3), 203–211.

Davis, L. A., & Stanton, S. T. (2005). Semantic feature analysis as a functional therapy tool. *Contemporary Issues in Communication Science and Disorders, 32,* 85–92.

DeLong, C., Nessler, C., Wright, S., & Wambaugh, J. (2015). Semantic feature analysis: Further examination of outcomes. *American Journal of Speech-Language Pathology, 24*(4), S864–S879.

Dickey, M. W., & Yoo, H. (2010). Predicting outcomes for linguistically specific sentence treatment protocols. *Aphasiology, 24*(6–8), 787–801.

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.

Falconer, C., & Antonucci, S. M. (2012). Use of semantic feature analysis in group discourse treatment for aphasia: Extension and expansion. *Aphasiology, 26*(1), 64–82.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*(2), 182–216.

Gravier, M. L., Dickey, M. W., Hula, W. D., Evans, W. S., Owens, R. L., Winans-Mitrik, R. L., & Doyle, P. J. (2018). What matters in semantic feature analysis: Practice-related predictors of treatment response in aphasia. *American Journal of Speech-Language Pathology, 27*(1S), 438–453.

Hashimoto, N., & Frome, A. (2011). The use of a modified semantic features analysis approach in aphasia. *Journal of Communication Disorders, 44*(4), 459–469.

Hinckley, J. J., & Craig, H. K. (1998). Influence of rate of treatment on the naming abilities of adults with chronic aphasia. *Aphasiology, 12*(11), 989–1006.

Hula, W. D., Kellough, S., & Fergadiotis, G. (2015). Development and simulation testing of a computerized adaptive version of the Philadelphia Naming Test. *Journal of Speech, Language, and Hearing Research, 58*(3), 878–890.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test*. Austin, TX: Pro-Ed.

Kertesz, A. (1982). *Western Aphasia Battery Test manual*. San Antonio, TX: The Psychological Corporation.

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.

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., & Roberts, P. M. (2010). Semantic feature analysis treatment in Spanish–English and French–English bilingual aphasia. *Aphasiology, 24*(2), 231–261.

Kiran, S., Sandberg, C., Gray, T., Ascenso, E., & Kester, E. (2013). Rehabilitation in bilingual aphasia: Evidence for within- and between-language generalization. *American Journal of Speech-Language Pathology, 22*(2), S298–S309.

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), 1101–1117.

Kiran, S., & Thompson, C. K. (2003). 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.

Kristensson, J., Behrns, I., & Saldert, C. (2015). Effects on communication from intensive treatment with semantic feature analysis in aphasia. *Aphasiology, 29*(4), 466–487.

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language, 83,* 1–19.

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.

Maddy, K. M., Capilouto, G. J., & McComas, K. L. (2014). The effectiveness of semantic feature analysis: An evidence-based systematic review. *Annals of Physical and Rehabilitation Medicine, 57*(4), 254–267.

Marcotte, K., & Ansaldo, A. I. (2010). The neural correlates of semantic feature analysis in chronic aphasia: Discordant patterns according to the etiology. *Seminars in Speech and Language, 31,* 52–63.

Massaro, M., & Tompkins, C. A. (1994). Feature analysis for treatment of communication disorders in traumatically brain-injured patients: An efficacy study. *Clinical Aphasiology, 22,* 245–256.

Meinzer, M., Mohammadi, S., Kugel, H., Schiffbauer, H., Flöel, A., Albers, J., … Deppe, M. (2010). Integrity of the hippocampus and surrounding white matter is correlated with language training success in aphasia. *NeuroImage, 53*(1), 283–290.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language, 99,* 1–34.

Oh, S. J., Eom, B., Park, C., & Sung, J. E. (2016). Treatment efficacy of semantic feature analyses for persons with aphasia: Evidence from meta-analyses. *Communication Sciences and Disorders, 21*(2), 310–323.

Papathanasiou, I., & Coppens, P. (2016). *Aphasia and related neurogenic communication disorders* (2nd ed.). Burlington, MA: Jones & Bartlett Learning.

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.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

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.

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.

Stadie, N., Schröder, A., Postler, J., Lorenz, A., Swoboda-Moll, M., Burchert, F., & De Bleser, R. (2008). Unambiguous generalization effects after treatment of non-canonical sentence production in German agrammatism. *Brain Language, 104*(3), 211–229.

Stanczak, L., Waters, G., & Caplan, D. (2006). Typicality-based learning and generalisation in aphasia: Two case studies of anomia treatment. *Aphasiology, 20*(2–4), 374–383.

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive aphasia test*. New York: Psychology Press.

Systat Software. (2018). SigmaPlot (Version 13.0) [Computer software]. San Jose, CA: Systat Software, Inc. Retrieved from http://www.systatsoftware.com

Tate, R., Mcdonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and *n*-of-1 trials: Introducing the single-case experimental design (SCED) scale. *Neuropsychological Rehabilitation, 18*(4), 385–401.

Thompson, C. K. (1989). Generalization research in aphasia: A review of the literature. *Clinical Aphasiology, 18,* 195–222.

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. *Journal of Communication Disorders, 39*(4), 266–291.

Walker, G. M., & Schwartz, M. F. (2012). Short-form Philadelphia naming test: Rationale and empirical evaluation. *American Journal of Speech-Language Pathology, 21*(2), S140–S153.

Wallace, S. E., & Kimelman, M. D. (2013). Generalization of word retrieval following semantic feature treatment. *Neurorehabilitation, 32*(4), 899–913.

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.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Developmental Disabilities Research Reviews, 13*(1), 70–77.

Watila, M., & Balarabe, S. (2015). Factors predicting post-stroke aphasia recovery. *Journal of the Neurological Sciences, 352*(1), 12–18.

Wickham, H., & Chang, W. (2008). ggplot2: An implementation of the Grammar of Graphics. R package (Version 2.2.1). Retrieved from http://CRAN.R-project.org/package= ggplot2