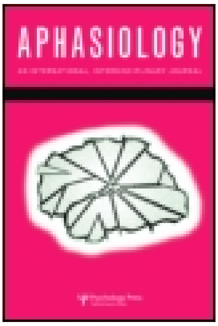

Aphasiology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/paph20

A comparison of features and categorical cues to improve naming abilities in aphasia

Naomi Hashimoto [a], Brooke Widman [b], Swathi Kiran [b] & Meredith A. Richards [a]

[a] Communicative Disorders, University of Wisconsin-River Falls, WEB 222, 410 South Third Street, River Falls, WI, 54022, United States

[b] Speech, Language, and Hearing Sciences, Sargent College, Boston University, 635 Commonwealth Avenue 326, Boston, MA, United States

Published online: 22 Jul 2013.

To cite this article: Aphasiology (2013): A comparison of features and categorical cues to improve naming abilities in aphasia, Aphasiology, DOI: 10.1080/02687038.2013.814760

To link to this article: http://dx.doi.org/10.1080/02687038.2013.814760

Aphasiology, 2013
http://dx.doi.org/10.1080/02687038.2013.814760

Routledge
Taylor & Francis Group

A comparison of features and categorical cues to improve naming abilities in aphasia

Naomi Hashimoto[1], Brooke Widman[2], Swathi Kiran[2], and Meredith A. Richards[1]

[1]Communicative Disorders, University of Wisconsin-River Falls, WEB 222, 410 South Third Street, River Falls, 54022 WI, United States
[2]Speech, Language, and Hearing Sciences, Sargent College, Boston University, 635 Commonwealth Avenue 326, Boston, MA, United States

Background: Evidence from the picture–word interference literature reveals that picture–word pairs bearing a non-categorical relationship (e.g., *RING—expensive*) will facilitate naming more than picture–word pairs bearing a categorical relationship (e.g., *BRACELET—earrings*). It is not known whether these differential effects would be found within a naming treatment paradigm for aphasia; however, if it is the case that one type of semantic relations will yield more robust treatment effects than another, this would provide a more efficient and effective delivery of treatment. Moreover, since semantic errors are commonly produced by individuals with aphasia, an approach which helps strengthen the semantic network will, in turn, strengthen lexical retrieval and access processes.
Aim: The aim of the study was to compare a features condition, or a condition in which attributes were used, to a categorical condition, or a condition in which categorical members were used to see which would yield greater naming improvements in aphasia.
Methods & Procedures: Eight individuals with aphasia were recruited for the study. A multiple-baseline design across behaviours with a crossover component was used. A categorical–features sequence was used in four participants, and a features–categorical sequence was used in four other participants. Accuracy of correct naming was calculated to determine improvements in treatment. The types of naming errors produced by individuals with positive treatment effects in the trained condition were compared to the types of naming errors produced in the control condition.
Outcome & Results: A range of small to large treatment effect sizes was obtained in six of the eight participants when considering both treated conditions. However, no overwhelming advantage was found for either condition. Analyses of naming error patterns indicated increased lexical access and retrieval of the targeted picture name.
Conclusions: Approaches that use cues focusing either on categorical membership or attributes will facilitate naming abilities in individuals with aphasia whether naming per cent accuracy is calculated or naming error types are tallied. However, the current results did not indicate an overwhelming advantage in using one or the other condition. Future studies should specify the type of features used (e.g., associative cues, thematic cues or perceptual cues) to determine whether such a differentiation will yield clearer differential treatment effects.

Keywords: Aphasia; Naming treatment; Features; Categorical.

Address correspondence to: Naomi Hashimoto, Communicative Disorders, University of Wisconsin-River Falls, WEB 222, 410 South Third Street, River Falls, 54022 WI, USA. E-mail: naomi.hashimoto@uwrf.edu

Current word production models stipulate activation of semantic and phonological processes so that lexical access and retrieval processes can take place (Caramazza, 1997; Dell & O'Seaghdha, 1991, 1992; Levelt, Roelofs, & Meyer, 1999; Starreveld & La Heij, 1996). Of primary interest in this study is the semantic processing level, which is involved in those operations that lead ultimately to the selection of a lexical concept. In models where semantic representations are considered decompositional (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), the semantic network consists of conceptual features. A set of features would therefore be activated to retrieve the lexical concept. The types of semantic features represented at this level typify many of the relationships encountered in day-to-day life, including ones that are formed on the basis of categorical membership or ones formed on the basis of encyclopaedic functions, functional properties and perceptual properties, to name a few (Cree & McRae, 2003; Garrard, Lambon Ralph, Hodges, & Patterson, 2001; McRae, De Sa, & Seidenberg, 1997; Rogers et al., 2004; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Vigliocco, Vinson, Lewis, & Garrett, 2004). As an example, the conceptual representation (e.g., DOG), could be represented by feature representations (e.g*., is an animal, is a pet, can bark, has a tail*), which would then be used to retrieve the lexical concept, *dog*, (Dell et al., 1997). Thus, the semantic network consists of conceptual feature nodes whereby closely related feature nodes are linked to one another. Activation would spread throughout this network via related semantic features. Selection of the targeted lexical concept occurs when the appropriate set of semantic features are activated.

Although feature-based representations have been used to explain category-related deficits via use of feature types or featural correlations (see Cree & McRae, 2003, for recent review), feature-based naming treatment protocols have not been widely reported in the aphasia literature. One exception is the semantic features analysis approach, an approach which typically uses six different features during naming treatment. The premise of the approach is that using relevant, appropriate semantic features will result in a convergence of activation onto the targeted picture representation. A certain threshold of activation is then reached and naming of the picture will occur. Studies that have used feature-based protocols have reported improvements of trained items and untrained items (see Boyle, 2010, for review). However, since certain features are not universally relevant across all concepts, it may be that improvements occurred because of one or two highly relevant features (Boyle, 2004). If this is the case, it would be worthwhile examining which specific feature cues best facilitate naming improvements in aphasia. In doing so, there would be a quicker progression through therapy using semantic cues that maximise the benefits of treatment. In terms of which cues to select, a comparison of categorical and non-categorical cues may be appropriate given the naming error types found in aphasia, as well as the empirical evidence from the picture–word interference literature.

Naming errors in aphasia

Among the types of naming errors produced when naming pictures, most speakers will overwhelmingly produce semantic errors that bear either a categorical relationship (e.g., *rabbit* for the picture of a *dog*) or an associative relationship (e.g., *nuts* for the picture of a *squirrel*) with the picture name. This is observed in neurologically intact individuals who are induced to produce naming errors, either through the use of speeded response paradigms (Hodgson & Lambon Ralph, 2008; Vitkovitch & Humphreys, 1991) or retrieval competition paradigms (Starreveld & La Heij, 1999) as well as in individuals with aphasia (Dell et al., 1997; Jefferies & Lambon Ralph, 2006; Schwartz, Dell, Martin, Gahl, & Sobel, 2006; Schwartz et al., 2011). If semantic features are connected to other semantically related features, disruptions or breakdowns within the semantic network can lead to mis-selection of these shared feature representations, which could result in the production of semantic errors (Dell et al., 1997; Schwartz et al., 2006; see Howard & Gatehouse, 2006, for review of case studies). Alternatively, these errors may signify a post-semantic origin, particularly in individuals who demonstrate relatively intact semantic/comprehension abilities in conjunction with production of semantic errors on oral or written production tasks. It has been argued that these semantic errors occurred because target phonological representations had become inaccessible, allowing the most highly activated, semantically related phonological representation to be produced instead (Caramazza & Hillis, 1990; Franklin, Howard, & Peterson, 1995; Hillis & Caramazza, 1995; Kay & Ellis, 1987). There are also reports of individuals who demonstrate intact semantic comprehension and speech output production in reading and repetition, but also present with significant anomia. These individuals typically indicate, through detailed accurate circumlocutions, adequate knowledge of the concept, but an inability to provide the correct name unless aided by phonemic cues. Moreover, these individuals are rarely miscued (e.g.,/v/[for violin] provided for the picture *guitar* is rejected). In these cases, the hypothesised breakdown lies in the mapping between semantics and phonology (Lambon Ralph, Sage, & Roberts, 2000).

If the presence of semantic errors signifies a disruption of lexical access and retrieval processes, cues that are categorical in nature as well as cues that highlight semantic attributes of a concept should strengthen semantic connections, resulting in improved naming abilities. This should be true regardless of the locus of the semantic error if, as posited in word productions models, initial activation at the semantic processing level feeds forward to subsequent syntactic and phonological processing levels (Dell & O'Seaghdha, 1991, 1992; Levelt et al., 1999); in other words, a strengthening of connections in the initial stages of lexical access should have positive effects on subsequent processing levels through feedforward or feedforward–feedback patterns of activation.

Picture–word interference studies

Another reason to compare cues along categorical and non-categorical lines is the empirical evidence found in the picture–word interference (PWI) literature. The PWI paradigm, which has primarily been used with neurologically intact individuals, involves the presentation of a visual competitor word with the picture to be named. Participants name the picture while ignoring the competitor word typically superimposed on the picture. Naming is *slowed* if the competitor word bears a semantic-coordinate relationship to the picture (e.g., *horse*—DOG) compared to when the competitor word has no relationship to the picture (e.g., *pencil*—DOG) (see Spalek, Damian, & Bölte, 2013). These effects are thought to reflect the lexical competition processes between the picture name and the competitor word (Levelt et al., 1999; Roelofs, 1992), or the competition at the level of a post-lexical response buffer (Costa, Mahon, Savova, & Caramazza, 2003; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007).

Interestingly, naming is faster if the competitor word bears a non-categorical relationship to the picture compared to an instance when the competitor word has no relationship to the picture (e.g., *pencil*—DOG). Non-categorical relationships cover a wide gamut of relationships, including those that are functional (e.g., *sweep*—BROOM), associative (e.g., *carrot*—RABBIT), part-whole (e.g., *tail*—DOG) and contextual (e.g., *garden*—BEE). Facilitation effects are usually obtained when the non-categorical word is presented well in advance of picture presentation (see Spalek et al., 2013). Interpretations of semantic facilitation effects have varied. One account proposes excitatory connections at the phonological processing level where associative relations are thought to be localised (Cutting & Ferreira, 1999), while other accounts propose either relatively less competition for associative relations, compared to categorical relations, due to less convergence onto other related concepts (Rahman & Melinger, 2009) or an absence of competition during lexical selection processes (Costa et al., 2003; Mahon et al., 2007).

Given the findings from the PWI literature, it seems feasible to incorporate different types of semantic relations in a naming treatment paradigm to examine their relative benefits as cues in treating naming deficits in aphasia. Should differential effects emerge, the findings would provide a more directed approach in remediating naming deficits in aphasia.

Purpose of the study

One aim of the study was to examine the effects of two different cue types in improving naming abilities in eight individuals with aphasia. A features (FEAT) condition, which involved the use of various attributes as cues to facilitate picture naming, was compared to a categorical (CATG) condition, which involved the use of categorical members as cues to facilitate picture naming. A control (CONT) condition was also included, which consisted of untrained items. A multiple-baseline design with a crossover component was used. A categorical–features sequence was used in four of the participants, while a features–categorical sequence was used in the other four participants. Measures other than accuracy in naming can indicate positive treatment outcomes. Another aim of the study, therefore, was to analyse the types of naming errors produced by conditions to determine if significant differences in naming error patterns could be discernible across conditions, and if so, whether differences in naming error patterns were indicative of increased lexical access and retrieval. Of relevance to the current study were omissions or no responses and semantic paraphasias, which were the two most commonly produced error types in the majority of participants of the current study.

Based on the literature review, it was predicted that larger treatment effects would be obtained for the features condition relative to the categorical condition. Although the PWI and picture naming literature would suggest, *de facto,* that the categorical cues will interfere with naming efforts (and thereby result in minimal to no treatment effects), there are sufficient differences between the current treatment paradigm and the PWI paradigms to warrant a more tempered prediction. The treatment paradigm used in the study does not involve quickly naming a picture while an interfering (categorically related) competitor word is being presented during the naming process. Therefore, the stressors associated with naming under timed conditions are not found in the treatment paradigm. Nevertheless, categorical cues may not improve naming abilities to the same degree as feature cues since lexical competition processes may attenuate improvements initially. With regard to naming error patterns, there is an assumption that treatment will strengthen connections within the semantic network, resulting in increased accurate naming of trained items. Therefore, it is predicted that naming errors of all types should decline over the course of treatment. Increased lexical access would be indicated if there were declines in no responses and semantic errors along with increased naming accuracy.

METHODS

Participants

Five males and three females participated in the study. Two participants who initially started the treatment study dropped out due to lengthy absences. These participants were subsequently replaced. The participants were recruited from the University Wisconsin-River Falls (UWRF) Aphasia Research Laboratory subject pool, Twin Cities stroke clubs/groups and Boston University (BU) Aphasia Research Laboratory. All individuals met the following inclusionary or exclusionary criteria: (1) completion of high school; (2) normal or corrected-to-normal vision; (3) adequate hearing acuity for 1:1 conversational exchanges; (4) monolingual English speakers; (5) no previous history of neurological-or psychiatric-based illnesses or diseases; (6) no history of language or learning difficulties; (7) no history of alcohol/substance abuse; and (8) documentation of a vascular lesion in the dominant left hemisphere at least 6 months old as documented by a medical and/or physician report. Further demographic details are provided in Table 1.

The Western Aphasia Battery Aphasia Quotient (WAB AQ) (Kertesz, 1982) was used to determine overall aphasia type and severity. Further testing was conducted using the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 2001), the Pyramids and Palm Trees Test (PPTT) (Howard & Patterson, 1992), and subtests of the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA) (Kay, Lesser, & Coltheart, 1992). Table 2 provides a summary of the test results.

TABLE 1
Participant characteristics

Participant	*Age (years)*	*Education (years)*	*Post-stroke (years)*	*Sex*	*Aetiology*[a]	*Race*	*Aphasia type*[b]
1	59	14	4	M	LMCA CVA	Caucasian	Broca
2	58	16	1	M	LMCA CVA	Caucasian	Anomia
3	50	20+	6	M	LMCA CVA	Caucasian	Anomia
4	53	18	5	F	LMCA CVA	Caucasian	Anomia
5	64–65	18	4	F	LBGH	Caucasian	Broca
6	75	18	13	M	LMCA CVA; TBI[c]	Caucasian	Wernicke
7	62	16	6	F	LBGH	Caucasian	Anomia
8	52	11	20	M	LMCA CVA	Black	Anomia

LMCA CVA = left middle cerebral artery cerebrovascular accident; LBGH = left basal ganglia haemorrhage; TBI = traumatic brain injury.
[a]Based on HCT/MRI scans and/or neurological reports. [b]Based on WAB results (Kertesz, 1982). [c]Participant also sustained a TBI 9 years-post CVA.

TABLE 2
Summary of pre-treatment performance on various language and psycholinguistic tests

Subtests	*N*	*Norms M; SD*	*P1*	*P2*	*P3*	*P4*	*P5*	*P6*	*P7*	*P8*
WAB AQ			33.2	62.8	74.2	78.7	28.4	65.1	65.1	61.3
BNT[a]	60		1	7	44	33	3	19	18	23
PPTT										
Picture association match	52	47	41	45	49	49	38	40	46	43
PALPA #53										
Picture naming—spoken	40	39.8; 0.35	NT	13	31	36	3	27	27	22
Picture naming—written	40	39; 1.33	NT	0	16	13	1	6	33	NT
Picture naming—reading	40	39.96; 0.19	NT	37	37	39	NT	23	23	NT
Picture naming—repetition	40	39.79; 0.83	32	40	40	40	NT	28	40	40
Picture naming—spelling	40	NA	NT	0	18	13	NT	11	NT	NT
PALPA #47/48										
Spoken word picture match	40	39.2; 1.07	25	37	37	37	33	34	40	39
Written word picture match	40	39.4; 1.01	25	35	39	36	30	34	37	NT
PALPA #51										
Semantic associates—HI	15	13.4; 1.26	NT	10	9	10	5	7	11	NT
Semantic associates—LI	15	12.2; 1.82	NT	4	9	12	8	6	8	NT
PALPA #49/50										
Written synonym—HI	30	28.9; 0.85[b]	30	29	29	29	29	30	26	30
Written synonym—LI	30	27.8; 1.69[b]	22	27	28	28	30	29	26	30
PALPA #33										
3-letter reading	6	NA	0	NT	NT	NT	NT	NT	3	NT
4-letter reading	6	NA	0	NT	NT	NT	NT	NT	4	NT
5-letter reading	6	NA	0	NT	NT	NT	NT	NT	2	NT
6-letter reading	6	NA	0	NT	NT	NT	NT	NT	1	NT
PALPA # 31										
Image × Freq reading										
HI, HF	20	19.94; 0.25	NT	17	19	19	0	12	NT	NT
HI, LF	20	19.94; 0.07	NT	17	20	20	1	12	NT	NT
LI, HF	20	20; 0	NT	16	18	16	0	5	NT	NT
LI, LI	20	19.52; 0.68	NT	15	15	16	0	5	NT	NT

WAB AQ = Western Aphasia Battery Aphasia Quotient (Kertesz, 1982); Boston Naming Test = BNT (Kaplan et al., 2001); PPTT = Pyramids and Palm Trees Test (Howard & Patterson, 1992); PALPA = Psycholinguistic Assessment of in Aphasia (Kay et al., 1992); NT = not tested; NA = no norms available; HI = high imageability; LI = low imageability; Image × Freq = Imageability × Frequency.
[a]All participants' BNT scores were below BNT age norms. [b]Norms taken from Nickels and Cole-Virtue (2004).

According to the classification scheme used by the WAB, five of the participants presented with anomic aphasia, one presented with Wernicke's aphasia and two presented with Broca's aphasia. All of the participants had naming impairments as indicated by scores that were below the norm for their given age (Kaplan et al., 2001). All but two of the participants (P3, P4) were below the cut-off score considered to be within normal performance on the PPTT. Performance on the PALPA Spoken Word–Picture Matching subtest revealed that all but two of the participants (P7, P8) performed below the cut-off score while performance on the PALPA Written Word–Picture Matching subtest revealed that all but one of the participants (P3) performed below cut-off scores. On the PALPA Semantic Association subtest, all but one of the participants (P4) scored below the cut-off scores. No overwhelming trends were noted when comparisons were made between high- and low-imageability items. On the PALPA Written Synonym Judgements subtest, only one participant (P1) performed below the cut-off scores for the high-imageability, low-imageability items or both. There was a trend towards better performance on high-imageability items compared to low-imageability items. Performance on the PALPA Picture Naming—Repetition subtest revealed adequate repetition skills for all but three of the participants (P1, P5, P6). Performance on either the PALPA Letter-Length or the PALPA Imageability × Frequency Reading revealed impairments for most of the participants; only two of the participants (P3, P4) demonstrated adequate reading abilities but only with high-imageability items while all participants demonstrated impaired oral reading abilities across imageability items. P1, P5 and P7 also demonstrated apraxia of speech using Duffy's (2013) tasks for assessing apraxia of speech. The overall pattern of relative strengths and weaknesses across participants suggested that naming deficits, as in the majority of participants reported in the literature, were due to impaired access or retrieval at all levels involved in the naming process (Martin, Fink, Renvall, & Laine, 2006; Schwartz et al., 2006).

Materials

Seventy-four black-and-white line drawings were obtained from online clip art websites. Picture names were normed using five to ten English-speaking volunteers who were asked to provide the names of the pictures. Familiarity ratings were obtained from a semantic features normative database (McRae, Cree, Seidenberg, & McNorgan, 2005). If the concept was not included in the database, the same instructions and rating scales described in the article were used to obtain ratings from a group of 23 individuals who volunteered or who were undergraduate students receiving course credit for their participation. Visual complexity ratings were also obtained from the same group of 23 individuals using the instructions and rating scales as reported in Snodgrass and Vanderwart's (1980) article.

Word cues were taken from the same database (McRae et al., 2005) used to obtain familiarity ratings. This database provided not only semantic features information, but categorical information as well. If a certain concept was not provided in the database, the norms for a selected concept were obtained from a group of 20 individuals who volunteered to complete the ratings or who were undergraduate students receiving course credit for their participation. The same instructions and cut-offs provided in the database were used. If categorical information was not provided in the database, 10 individuals, none of whom were involved in providing the previously obtained conceptual normative data, were asked to provide the category to which each word belonged. The category names formed coherent categories (e.g., *transportation* for *ship* and *gardening tool* for *wheelbarrow*), so were included in the study. One-way analyses of variances (ANOVAs) calculated across the stimuli sets of the three conditions (CATG, FEAT, CONT) revealed no differences in visual complexity, $F(2, 0.09) = 0.11$, $p = .896$ or in familiarity ratings $F(2, 0.08) = .02$, $p = 0.978$. Once the stimuli were finalised, the pictures were placed on 3×4 laminated cards. Cues (words) that were paired with the pictures were printed on 3.5×1.75 laminated cards. There were also blank cards on which alternate cues were written. Appendix A provides the stimuli sets.

Conditions

The CATG condition ($n = 25$) consisted of pictures paired with three words that belonged to the same category as the picture. The FEAT condition ($n = 25$) consisted of pictures paired with three words that represented a feature common across all stimuli (e.g., the feature, *explodes,* was selected since it was common not only to the targeted picture names, GRENADE and CANNON, but also to the three weapons, *missile, gun, bomb*, used in the CATG condition). Although slightly more functional and encyclopaedic attributes were used, the features represented the variety of attributes found in the database. The CONT condition, ($n = 24$ pictures) consisted of pictures that were from the same categories or shared the same features as the treated items. These pictures were administered during probe sessions only.

Design

A multiple-baseline design across behaviours was used in which there was a crossover component. Conditions were counter-balanced across participants so that four participants received a CATG–FEAT treatment sequence, while another four participants received the FEAT–CATG sequence. This design was used in order to allow a direct comparison of two treatments in a time-efficient manner. The three features and three categorical members that were paired to each of the pictures were common to both the CATG and FEAT pictures; therefore, the CATG and the FEAT conditions were rotated across participants so that pictures used in the categorical condition in four of the participants were then used in the features condition in the other four participants.

Probes

Baseline probes were obtained over three consecutive sessions across all conditions in order to ensure stability of performance prior to initiation of treatment. Once the treatment was initiated, probe schedules varied from participant to participant since probes were obtained whenever the participant had gone through the 25-item set; this schedule was adopted in order to minimise exposure effects that might come from repeated measures. Typically, probes for all conditions were obtained every third to fifth session right before a treatment session. Maintenance probes were obtained once the initial condition had been completed. Follow-up probes, which were carried out on all items to assess the long-term effects of treatment, were obtained 1, 2 and 4 weeks after the last treatment session.

During probe sessions, participants were instructed to name a picture that was chosen randomly from the pile. No feedback or cueing was provided, although intermittent encouragement was given. No response time was imposed during the probe sessions. Correct responses included the correct name for the picture, acceptable substitutions (e.g., *jeans* for *pants* or *bookshelf* for *bookcase*) and minor distortions/omissions that were phonetic in nature. Naming errors were classified into one of the following types: (a) semantic paraphasias (SEM): real word errors that bore a semantic relationship to the target name (e.g., *flower* for *tulip* or *cardinal* for *blue jay*); (b) phonological paraphasias (PHO): real word errors that bore a phonological relationship to the target name (e.g., *toast* for *toaster*) or non-word errors which had $\geq$50% overlapping phonemes in the same structural position (e.g., *pijoaemo* for *piano*); (c) no response (NR); (d) unrelated responses (UNR): real word responses that bore no relationship to the target name; (e) neologisms (NEO): non-word errors that bore no semantic or phonological relationship to the target name; and (f) other (OTH): a multi-word response that described the target name (e.g., "cut the wood" for *axe*) or appeared to be a filler (e.g., "oh come on, hang on"; "1-2-3-4"). Since no time limit was imposed, some participants had a tendency to provide multiple responses; therefore, the most frequent response constituted the final response. If an equal number of error types were represented, the last error type recorded was taken as the final response.

Either the primary author or two graduate students, trained in the protocol, conducted these probe sessions. Students were trained using a videotaped recording of a session using a participant already enrolled in the study. While watching the recording, the students followed along reading from the protocol detailing the steps involved (Appendix B). Questions that were raised regarding scoring and treatment steps were answered.

Treatment sessions

All participants received twice-weekly treatment sessions, each session lasting approximately 45 minutes to 1 hour (P5 received three sessions weekly for the first FEAT condition, but this was dropped to a twice-weekly schedule for the second CATG condition to match the other participants' schedules). Sessions were conducted either in the participant's home or in the research labs. The first author or graduate students, trained in the protocol, conducted the sessions. The number of items completed per session varied by participant; a range of 6 to 20 items was usually completed. Criteria for completion were either $\geq$80% accuracy over two consecutive sessions or once a maximum of 20 sessions had been completed, whichever came first. Consequently, the number of sessions varied for each participant (range = 3–20), depending on whether performance accuracy or number of sessions criteria were reached first. Once the criterion was reached, a 1-week break was provided to reduce any potential carry-over effects. Following that week, the other condition was presented for treatment. Once criterion was reached for the second condition, treatment ended.

Analyses

Visual analyses of treatment effects were accomplished using the conservative dual criterion (CDC) method (Fisher, Kelley, & Lomas, 2003). Mean (level) lines, represented by long dashed lines, and trend lines, represented by short dashed lines, were calculated based on baseline data. These lines were raised by .25 standard deviations (based on baseline data) and subsequently superimposed on each treatment phase. Improvements occurred if a majority or all data points were found above each of the lines according to a binomial equation. Note that a minimum of five data points are needed, so for P3, P4, P7 (FEAT condition), the CDC method was not applicable.

Effect sizes (ES) were calculated, using *d* statistics (Beeson & Robey, 2006; Busk & Serlin, 1992), to determine the magnitude of treatment. For the initial treatment condition, the mean of the first three post-treatment probes was subtracted from the mean of the first three baseline probes. This value was then divided by the standard deviation of the baseline probes (post-treatment mean—mean of baseline probes/standard deviation of baseline probes). For the subsequent, second treatment condition, the mean of follow-up probes was subtracted from the mean of the last three pre-treatment probes (i.e., probes obtained just prior to initiation of the second treatment condition). This value was then divided by the standard deviation of the pre-treatment probes (pre-treatment mean—mean of follow-up probes/standard deviation of pre-treatment probes). Generalisation to the untrained set was thought to have occurred if CONT ES were obtained, using the same time frames as the treatment conditions. The ES were interpreted using benchmarks proposed by Beeson and Robey (2006) for lexical retrieval treatments: small ES = 4.0, medium ES = 7.0, large ES = 10.1.

Reliability

Approximately 20% of each of the participant's sessions were observed to measure treatment integrity. One graduate student, who was not involved in the study, was provided with the step-by-step treatment protocol (Appendix B). She scored whether each of the steps had been completed, using a binary +/– system, while watching either live or videotaped sessions. No further training was needed since each step was either scored as present (+) or absent (–). Using this procedure, point-by-point agreement was found to be 100%.

Reliability in scoring naming error types was obtained by the graduate student who had collected treatment integrity measures. General guidelines and specific examples were provided. Both the primary author and student used responses recorded on probe sheets to classify responses as correct or as a naming error. Total per cent accuracy and naming error types were then compared. If there was a discrepancy in coding, responses were re-scored independently and then reviewed together. Any discrepancies remaining after the second comparison were resolved through discussion. Approximately 20% of all participants' responses, with the exception of P1 and P2, were coded and reviewed. In the case of P1 and P2, all responses were coded and reviewed to insure accuracy in coding their multiple responses. Point-by-point agreement for the first coding attempt was 97%, while point-by-point agreement for the second coding attempt was 99%.

RESULTS

CATG–FEAT sequence

Participants P2, P4, P6 and P7 were treated using the CATG–FEAT sequence. Figures 1–4 provide a graphic display of treatment performances. Table 3 provides the treatment ES.

Figure 1 displays P2's response to treatment. Visually, there were no changes due to the CATG treatment; however, there were improvements demonstrated in the second, treated FEAT condition. Likewise, he demonstrated no meaningful CATG effects ($d = 1.0$), but small FEAT effects ($d = 4.66$). Figure 2 displays P4's response to treatment. The CDC method was not applied due to too few data points. However, generalised learning was visually noted, as evidenced by increasing performance over time across the untrained FEAT and CONT conditions. A large CATG ES was obtained ($d = 14.42$), as well as a small FEAT ES, ($d = 5.89$). P6 demonstrated a great deal of variability in performance across conditions, as can be seen in Figure 3. No improvements were found in either condition using the CDC method (Note that both mean and trend lines are superimposed). Additionally, he did not demonstrate any meaningful ES in either the FEAT ($d = -0.577$) or the CATG condition

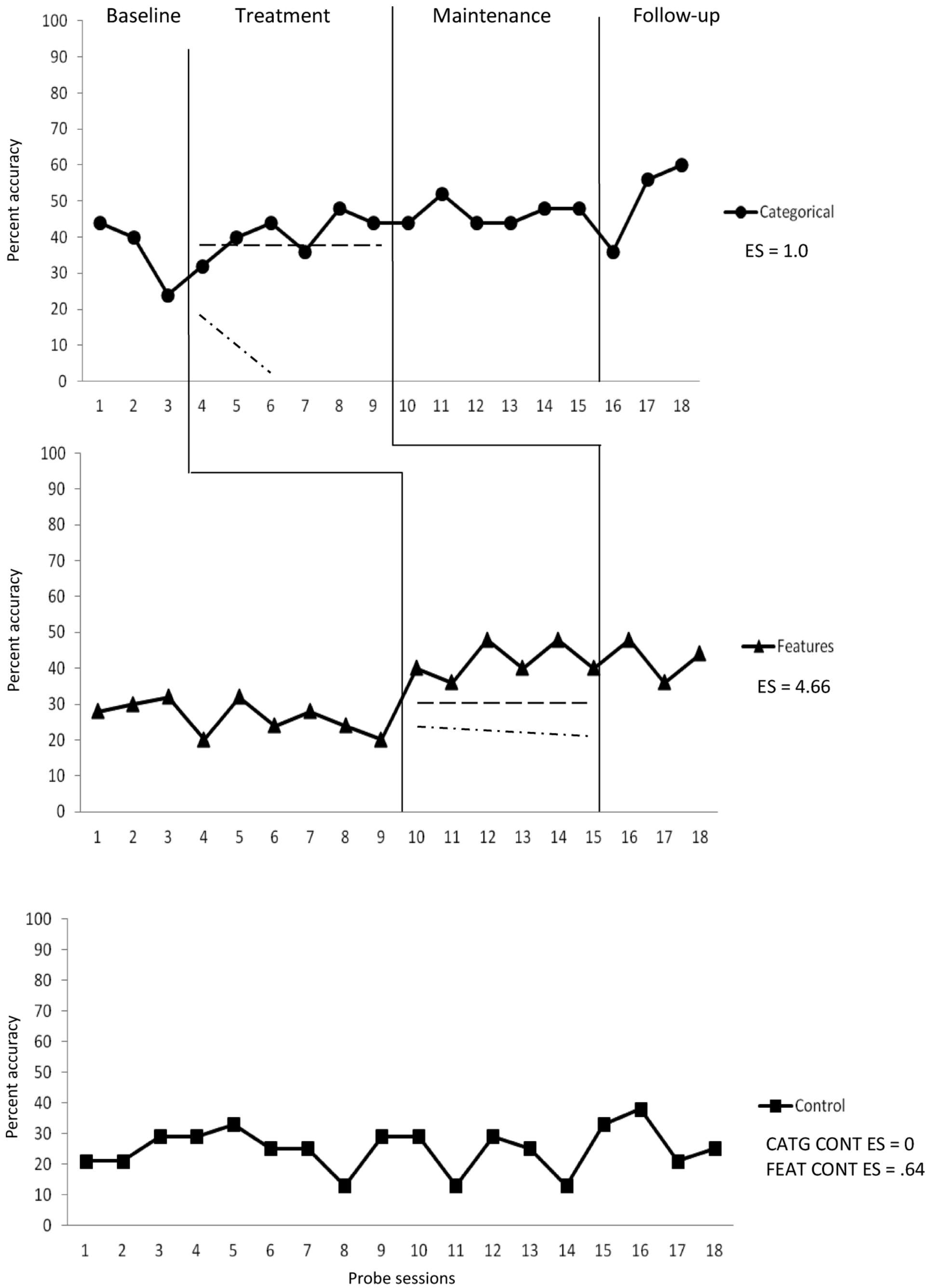

Figure 1. P2's naming accuracy performance. Long dashed lines represent the level (mean) line and short dashed lines represent the trend line.

($d = -0.57$). It should be noted that this participant performed poorly during probe session 19 (FEAT treatment); his physician later confirmed that he had likely sustained a seizure. Consequently, probes for that session were not calculated in the analyses; instead, probes that were taken at the next session replaced the probes obtained

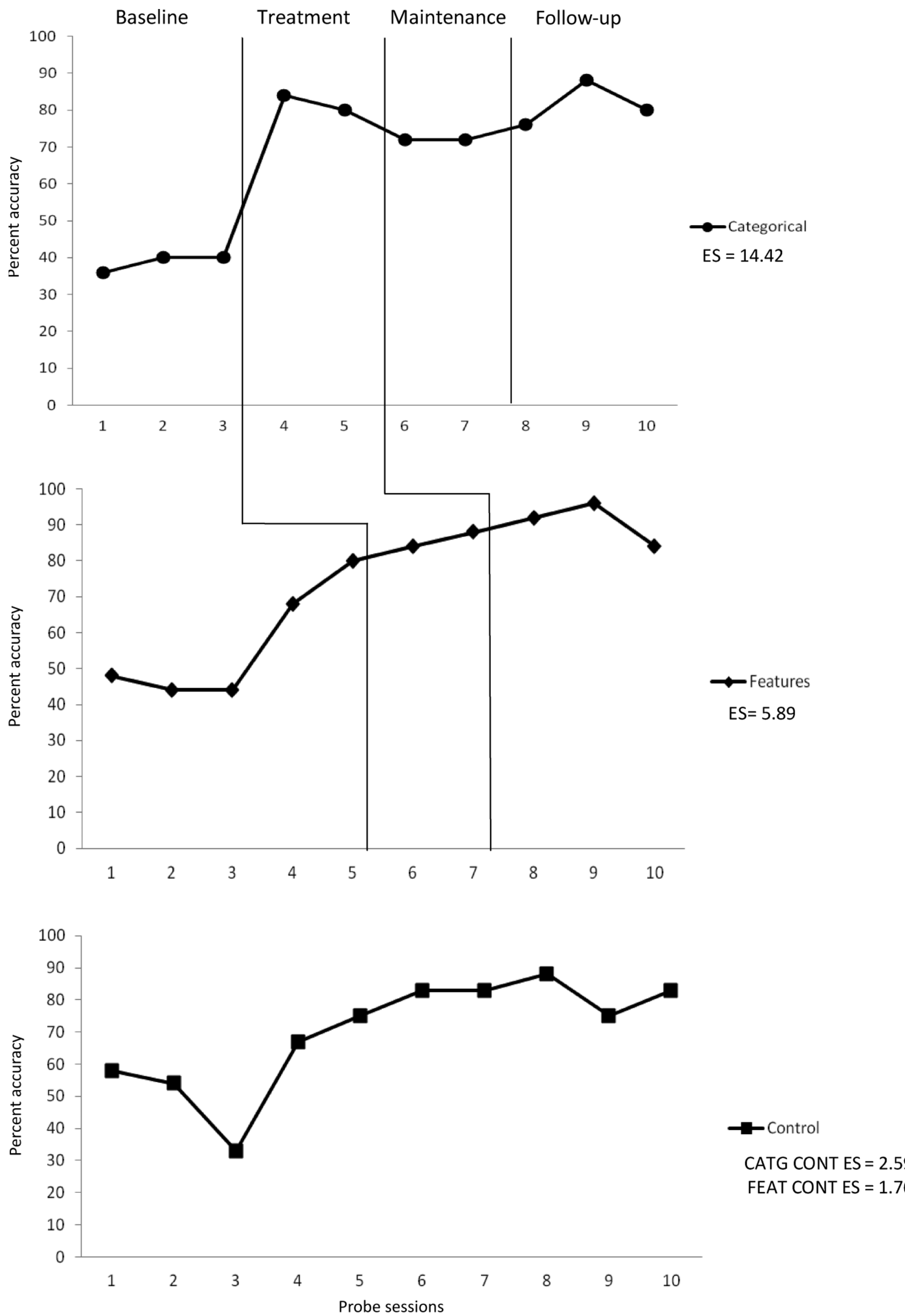

Figure 2. P4's naming accuracy performance.

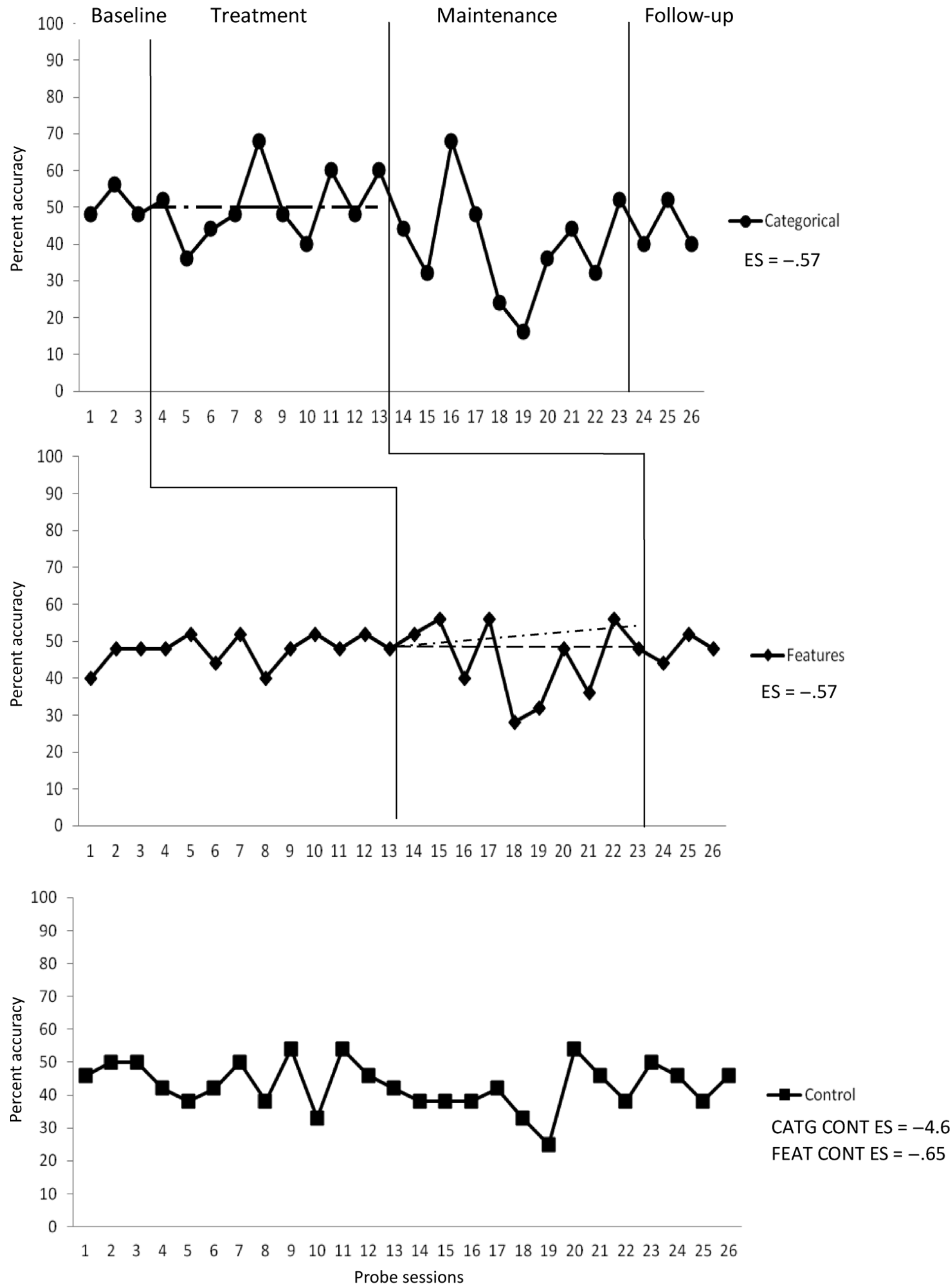

Figure 3. P6's naming accuracy performance. Level and trend lines are superimposed on one another in CATG approach. Long dashed lines represent the level (mean) line and short dashed lines represent the trend line in the FEAT approach.

in session 19. Although the seizure may have contributed to the variability seen in later probe sessions, his performance had been quite variable prior to this session. Figure 4 displays P7's response to treatment. P7 demonstrated, overall, good response to CATG treatment as indicated by visually improved treatment performance (Note

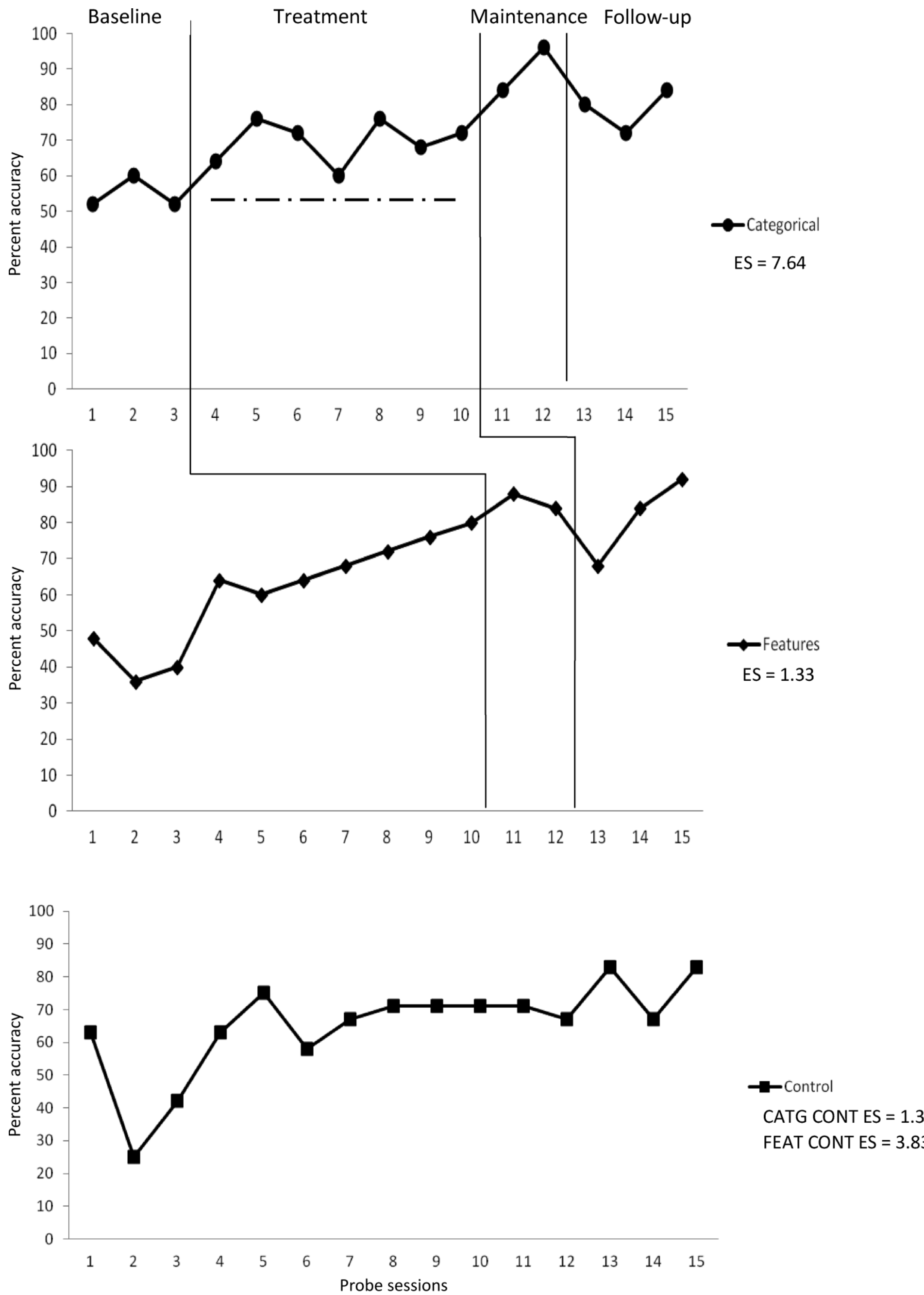

Figure 4. P7's naming accuracy performance. Level and trend lines are superimposed on one another in CATG approach.

that both mean and trend lines are superimposed). This was also reflected in a medium CATG ES ($d = 7.64$). Although the CDC method was not used for the second, FEAT condition (due too few data points), visual inspection of P7's baseline performance revealed generalised learning, as evidenced by a steady increase in performance prior

TABLE 3
Effect sizes (ES) of participants by treatment sequence

	CATG–FEAT sequence			
	P2	*P4*	*P6*	*P7*
CATG ES	1.00	14.42	−0.577	7.64
CATG CONT ES	0	2.59	−4.61	1.36
FEAT ES	4.66	5.89	−0.57	1.33
FEAT CONT ES	0.64	1.76	−0.65	3.83
	FEAT–CATG sequence			
	P1	*P3*	*P5*	*P8*
FEAT ES	1.66	6.11	4.61	2.26
FEAT CONT ES	2.30	1.19	0.454	0
CATG ES	3.46	2.3	0.86	12.83
CATG CONT ES	0.36	0.65	−1.74	1.15

CATG–FEAT sequence = treatment was provided in categorical–feature sequence; FEAT–CATG sequence = treatment was provided in feature–categorical sequence; CATG = categorical; CONT = control; FEAT = features.

to initiation of the FEAT treatment. However, no meaningful ES (d = 1.33) was obtained for the FEAT condition. Across all participants treated with the CATG–FEAT sequence, no meaningful effects were found in the CONT conditions. Thus, the CATG–FEAT sequence did not result in generalisation effects.

FEAT–CATG sequence

Participants P1, P3, P5 and P8 were treated using the FEAT–CATG sequence. Figures 5–8 provide a graphic display of the treatment performances. Table 3 provides the treatment ES.

As seen in Figure 5, P1 did not demonstrate any improvements in either condition when the CDC method was applied. Although the treatment effects were negligible, P1 demonstrated differential effects; the second approach, the CATG condition, resulted in larger effects (d = 3.46) relative to the FEAT condition (d = 1.66). P3's performances are displayed in Figure 6. The CDC method was not applied due to too few data points. He demonstrated a small ES in the FEAT condition (d = 6.11) but no meaningful ES in the CATG condition (d = 2.3). Figure 7 displays P5's treatment performance. The CDC method was not applied due to too few data points. Small treatment effects were obtained during the FEAT condition (d = 4.61) while no meaningful ES were obtained for the CATG condition (d = 0.86). P8, whose performance is displayed in Figure 8, demonstrated visual improvements across conditions using the CDC method. Large effects were found during the CATG condition (d = 12.83), while no meaningful FEAT effects (d = 2.26) were found. As was the case with the participants treated with the CATG–FEAT sequence, no meaningful effects were found in the CONT conditions. Thus, the sequence elicited treatment-specific effects only.

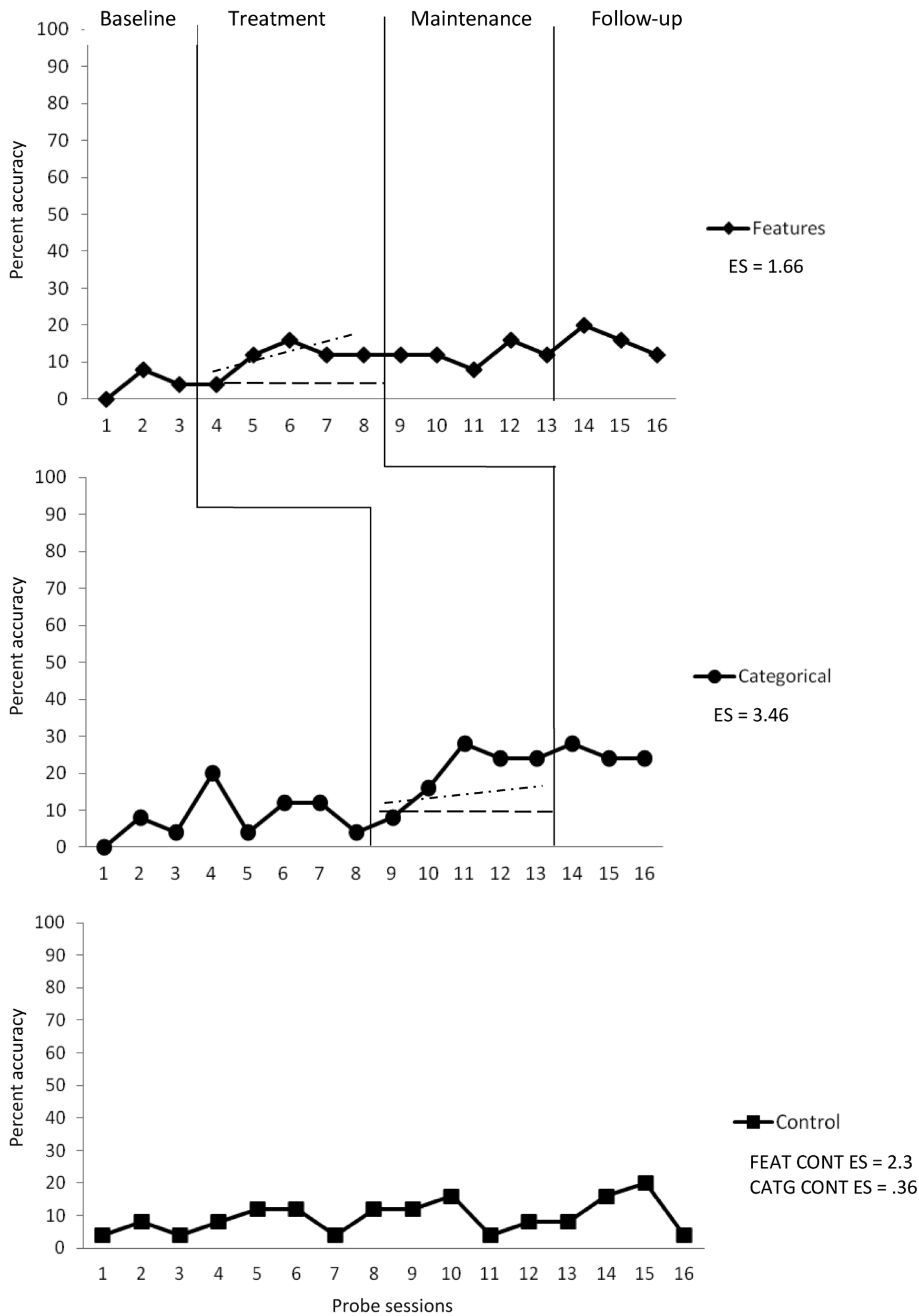

Figure 5. P1's naming accuracy performance. Long dashed lines represent the level (mean) line and short dashed lines represent the trend line.

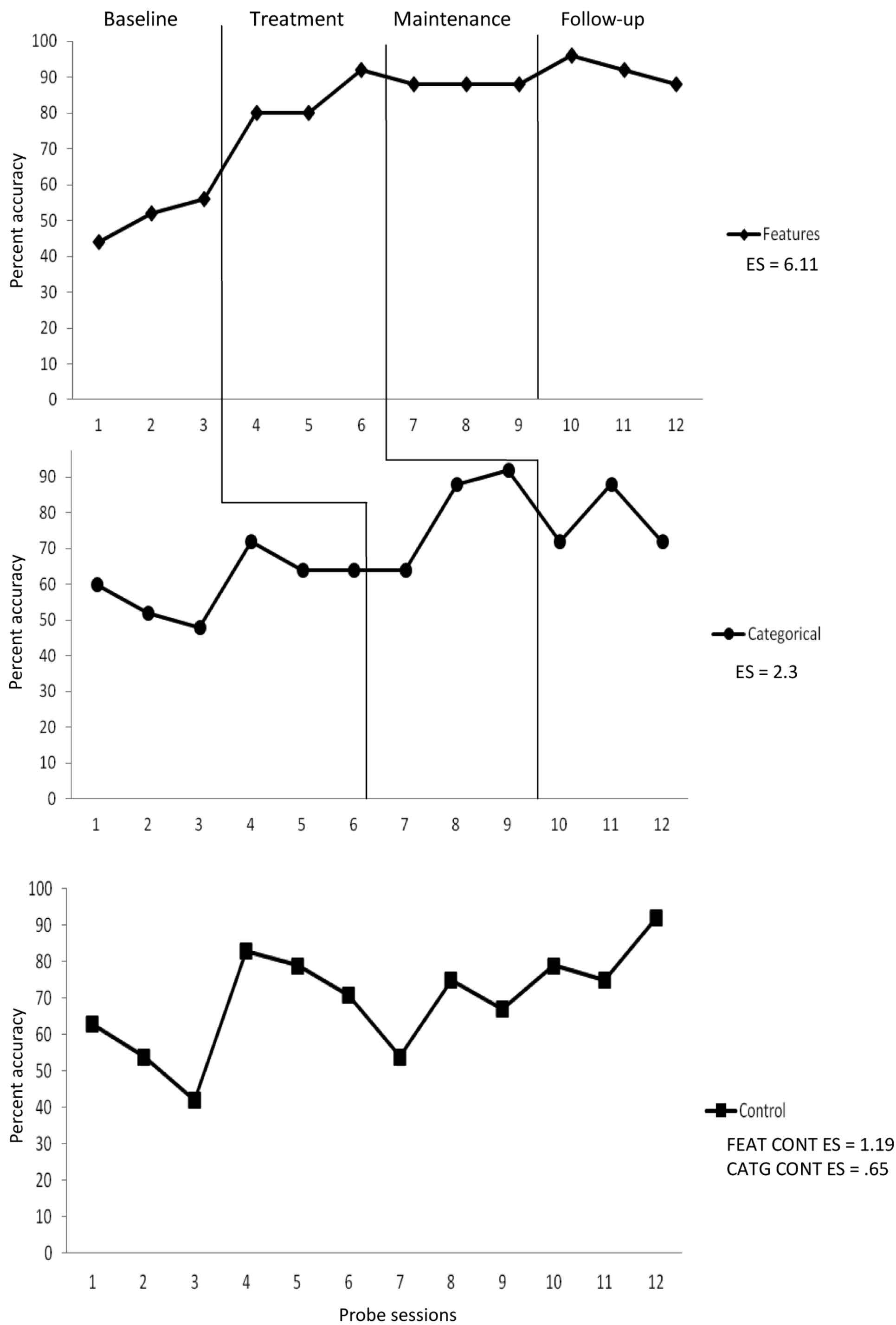

Figure 6. P3's naming accuracy performance.

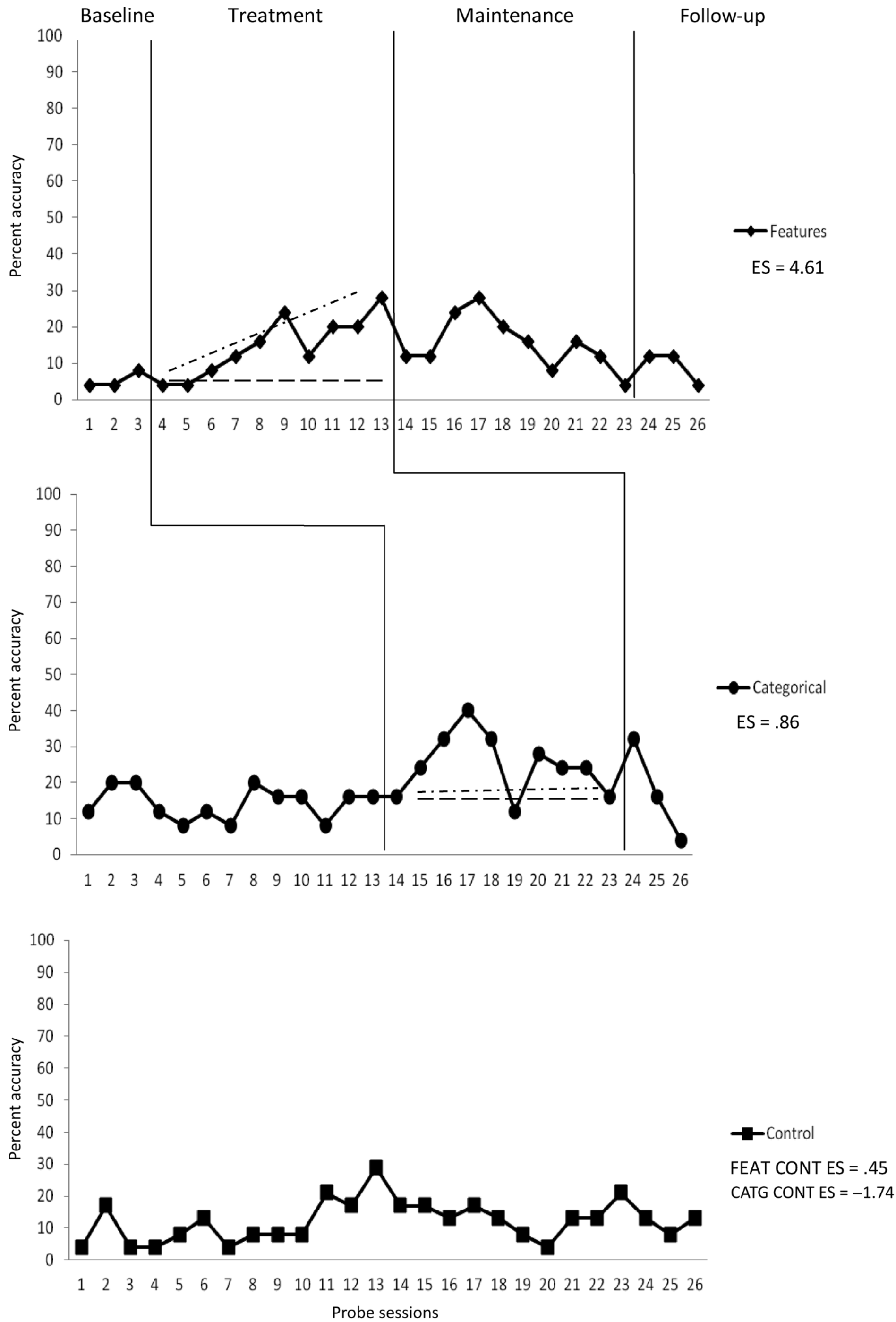

Figure 7. P5's naming accuracy performance. Level and trend lines are superimposed on one another in CATG approach.

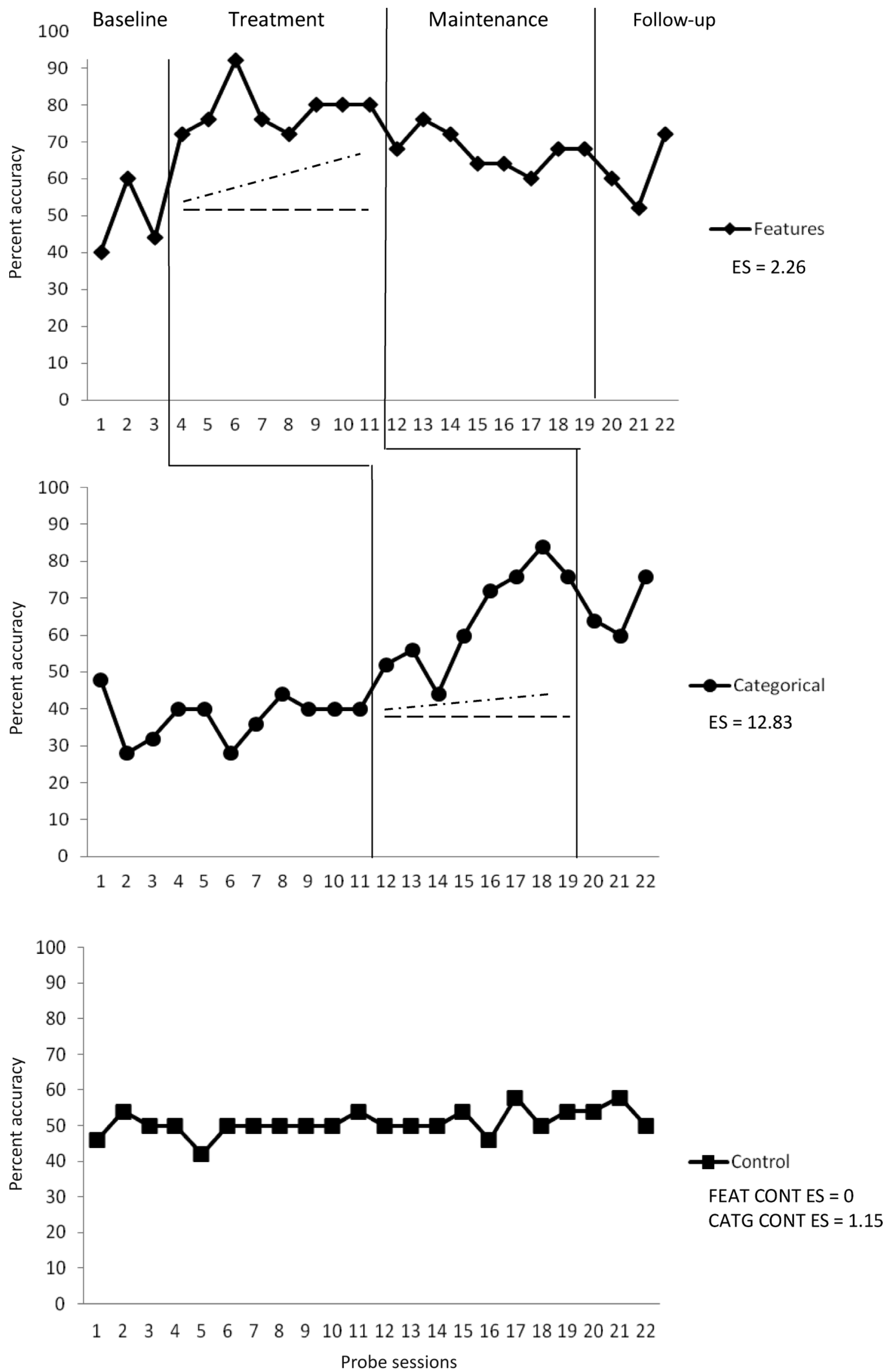

Figure 8. P8's naming accuracy performance. Long dashed lines represent the level (mean) line and short dashed lines represent the trend line.

Naming errors

To examine differences in naming error patterns across conditions, the number and types of naming errors were tallied across the different treatment conditions and compared to the control condition. Only the SEM and NR errors produced in conditions that yielded small–large ES were then compared to errors produced in the CONT condition. Participants who demonstrated a meaningful CATG ES included P4, P7 and P8. The number of errors produced during the CATG condition revealed more SEM ($n = 96$) than NR errors ($n = 47$), while the CONT condition in that same time period yielded more NR ($n = 104$) than SEM ($n = 71$) errors. Analyses of the responses revealed an overall difference in the distribution of error types, $\chi^2(11) = 128.68$, $p < .001$. Follow-up Wilcoxon paired ranks tests were carried out comparing the SEM error rates in the CATG and CONT conditions as well as the NR rates in the CATG and CONT conditions. There were significantly less NR ($z = –3.16$, $p < .01$) but significantly more SEM errors ($z = –2.03$, $p > .05$) in the CATG condition compared to the CONT condition. Participants who demonstrated meaningful FEAT ES included P2, P3 and P5 (Note that P4 demonstrated significant ES across both conditions but her data were analysed only in the CATG condition). The FEAT condition yielded more SEM ($n = 64$) than NR errors ($n = 10$); and the CONT condition also yielded more SEM ($n = 63$) than NR errors ($n = 39$). There was a reduction in NR errors only when the FEAT condition was compared to the CONT condition. Analyses of the error rates revealed an overall difference in the distribution of error types, $\chi^2(11) = 38.91$, $p < .001$. Follow-up Wilcoxon paired ranks tests revealed significantly less NR errors in the significant FEAT condition compared to the CONT condition ($z = –2.39$, $p < .05$). This series of analyses revealed differing naming error patterns depending on the condition that resulted in small–large ES. Participants who responded to the CATG condition (i.e., participants who had meaningful CATG ES) demonstrated significant reductions in NR errors relative to the CONT condition. Interestingly, these participants also produced significantly *more* SEM errors when treated with the CATG condition relative to the CONT condition. Participants who responded to the FEAT condition (i.e., participants who had meaningful FEAT ES) demonstrated significant reductions in NR errors relative to the CONT condition.

DISCUSSION

The overall aim of the study was to compare two different conditions that paired either categorical or feature cues with a picture to improve naming abilities in eight individuals with aphasia. Level and trend lines were used to objectively determine whether improvements could be visually seen in the treatment performance. Naming accuracy measures were calculated to determine the treatment effects. Finally, naming error rates and naming error types produced across treated and untreated conditions were analysed to determine if significant changes in naming error patterns indicated increased lexical access.

Naming accuracy performance

Both approaches improved naming of trained items in individuals with varying degrees of anomia. Of the eight participants involved in the study, only one individual, P6, did not demonstrate improvements after being treated with either condition. Additionally, all participants, with the exception of P4 and P7, demonstrated relatively stable baseline performance for the untreated condition; thereby demonstrating good experimental control. For P4 and P7, generalised learning appeared to be occurring as evidenced by increased performance during the baseline phases. However, this did not appear to impact treatment ES. P7 demonstrated an initial medium CATG ES, but no meaningful second FEAT ES; P4 demonstrated an initial large CATG ES followed by a small FEAT ES.

Visual analyses using the CDC method (Fisher et al., 2003) revealed improvements for P2 (FEAT), P5 (CATG), P7 (CATG) and P8 (FEAT/CATG). Thus, outcomes were similar to ES calculations with two exceptions. According to the CDC method, P8 demonstrated improvements across both conditions, whereas using ES calculations, he demonstrated only a CATG ES. P5 demonstrated improvements in the CATG condition according to CDC methods but no improvements in the FEAT condition. This was opposite to the ES calculations, which yielded a FEAT ES but no CATG ES. These differences are likely due to the use of different data sets used to determine changes as a result of treatment.

Although it was encouraging to see that a majority of the participants responded positively to one of the two approaches, the predicted differential, greater FEAT than CATG treatment effects, was not found. This prediction was based, in part, on empirical evidence gleaned from the PWI literature, which reports interference effects, or slowed naming response times when the picture–word pair bears a categorical relationship (e.g., *horse*—DOG), relative to when the picture–word pair has no relationship (e.g., *pencil*—DOG) and facilitative effects, or faster naming response times when the picture–word pair has a semantically related, non-categorical relationship (e.g., *tail*—DOG), relative to when the picture–word pair has no relationship (e.g., *pencil*—DOG) (see Spalek et al., 2013, for review). Based on these findings, it was predicted that feature cues would prove to be more facilitative, and therefore, more beneficial, than categorical cues in improving naming performance. However, this was not supported by the data. The CATG approach produced medium–large ES in three of the participants (P4, P7, P8) while the FEAT approach produced a range of small–large ES in four of the participants (P2, P3, P4, P5). The lack of an overwhelming FEAT benefit may have to do with some key differences between the PWI paradigm and the study's treatment paradigm: First, the timed pressures associated with the PWI paradigm were eliminated in the treatment paradigm. Second, although words that bore a categorical relationship to the picture name were used, the words acted as cues rather than as competitors during treatment. Thus, repeated exposures to members of the same category as the targeted picture appeared to strengthen connections between the semantically related concepts in the semantic network.

Another factor that may have played a role in the lack of differential treatment effects may have been the features used in treatment. Although faster naming response times are found in non-brain-damaged participants when the pictured concepts have strongly correlated distinctive features (Taylor, Devereux, Acres, Randall, & Tyler, 2012), it would have been difficult to find three, strongly correlated distinctive features for each picture unless the stimuli sets were severely limited in number. Additionally, it would have been difficult to have fulfilled the correlated, distinctive criteria for the CATG treatment condition. Therefore, the features that were chosen for each picture were the ones that were shared across similar conceptual categories (e.g., the weapons, GRENADE, CANNON, *missile*, *gun* and *bomb* had the features of *war*, *loud* and *explode* in common). Since these features represented ones that were common to a number of concepts, activation converging onto the targeted concept may have been weak. By contrast, the CATG approach used categorical members that were more specific to the targeted picture, which would have resulted in a stronger convergence onto the targeted concept.

Although significant treatment effects were found for trained items, there was no generalisation to untrained items, even though these items were taken from the same category and shared the same features as the treated items. One reason may be have to do with the fact that the training protocol did not explicitly train self-generation of cues, a step that may be needed in order to see generalisation to untrained items (Boyle, 2004, 2010; Edmonds, Nadeau, & Kiran, 2009). Although it was not discouraged, participants were not directly instructed to try to self-generate cues associated with the picture. Yet another reason may have to do with the lack of repeated exposure, and therefore, lack of repeated attempts to name the untreated stimuli. Participants in this study were probed once he/she had gone through the 25-item set. Consequently, individuals were probed at variable rates, typically every third to fifth session. None were probed every session. Some researchers (Boyle, 2010; Howard, 2000; Nickels, 2002) have raised the possibility that the generalisation effects reported in aphasic naming treatment studies may in fact be treatment effects, that is, effects due to repeated attempts to name the untreated items during treatment. Studies that have either re-analysed previous data (Howard, 2000), implemented a limited exposure probe schedule (Rider, Wright, Marshall, & Page, 2008) or have contrasted amounts of exposure to an untreated stimuli set (Nickels, 2002) have found that generalisation does not occur unless there are repeated attempts to name pictures. This may partially explain the lack of generalisation effects for the participants in the current study; however, it should be noted that it cannot be the entire reason since there were participants who were probed every third session, a probe schedule similar to what other individuals have been provided who did demonstrate generalisation effects (Boyle, 2004; Boyle & Coelho, 1995; Coelho, McHugh, & Boyle, 2000).

Naming error patterns

Naming error analyses were performed on the SEM and NR errors produced in conditions that yielded small–large ES and then compared to errors produced in the CONT condition. Participants who had demonstrated meaningful CATG ES demonstrated significant reductions in NR as well as significantly more SEM errors relative to the CONT condition. Participants who had meaningful FEAT ES demonstrated significant reductions in NR errors relative to the CONT condition. Increased lexical access and retrieval processes were indicated by a decline in NR responses across both groups. Interestingly, there was a concomitant increase in SEM errors in the CATG group. Since naming error rates were not analysed over time, it is hard to interpret the latter findings. However, it may be that use of cues in the CATG approach initially resulted in mis-selections of semantically related neighbours, but as treatment progressed, naming became more accurate. Hence, the concomitant findings of increased SEM errors with reduced NR errors.

Limitations and future directions

One limitation of the study was the absence of any generalisation effects even though control items were similar to trained items. Since generalisation has been reported when participants self-generate cues learned in treatment (Boyle, 2004, 2010; Edmonds et al., 2009), a protocol that incorporates self-cueing during treatment may be a fruitful area to explore in future studies. Alternatively, manipulation of the probe schedule may help further determine if it is, in fact, the repeated exposure to items that gives rise to generalisation effects. Another limitation was the inability to discern an overwhelming FEAT advantage. Since the cues used in the FEAT approach were common, shared features, these cue types may have been insufficient in producing strong treatment effects. Therefore, future studies should further refine the type of stimuli used in the FEAT approach. The use of non-categorical cues such as whole–part relations (e.g., *truck–bumper*), functional relations (e.g., *sweep–broom*), contextual (e.g., *garden–bee*) or associative relations (e.g., *carrot–rabbit*) may yield stronger effects. Since these relation types facilitate naming in neurologically intact individuals (Moss, Ostrin, Tyler, & Marslen-Wilson, 1995; Rahman & Melinger, 2007; Sailor, Brooks, Bruening, Seiger-Gardner, & Guterman, 2009; see Spalek et al., 2013), there is merit in considering their use within an aphasic naming treatment paradigm. Another area to explore is the type of aphasia an individual must have in order to benefit the most from such approaches. Since participants in the current study demonstrated a range of severity, matching individuals in terms of anomia severity may help resolve this issue. Finally, one of the drawbacks of the crossover design is the potential for carryover effects from the first to the second treated condition. To mitigate these effects, a 1-week break was provided between conditions. Nevertheless, the ES obtained during the second condition for P2 and P8 may have been partly due to carryover effects from the first treated condition.

Manuscript received 21 November 2012
Manuscript accepted 10 June 2013
First published online 18 July 2013

REFERENCES

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

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

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

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

Busk, P. L., & Serlin, R. (1992). Meta-analysis for single case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex*, *26*, 95–122.

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology*, *14*, 133–142.

Costa, A., Mahon, B., Savova, V., & Caramazza, A. (2003). Level of categorisation effect: A novel effect in the picture word interference paradigm. *Language & Cognitive Processes*, *18*, 205–234.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of *chipmunk, cherry, chisel, cheese,* and *cello* (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*, 163–201.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 318–344. doi: 10.1037/0278-7393.25.2.318

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, *98*, 604–614.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.

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*, 801–838.

Duffy, J. R. (2013). Examination of motor speech disorders. In R. Duffy (Ed.), *Motor speech disorders: Substrates, differential diagnosis, and management* (3rd ed., pp. 61–92). St. Louis, MO: Elsevier Mosby.

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of verb network strengthening treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, *23*, 402–424.

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*, *36*, 387–406.

Franklin, S., Howard, D., & Peterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology*, *12*, 549–566.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Journal of Cognitive Neuroscience*, *18*, 125–174.

Hillis, A. E., & Caramazza, A. (1995). The compositionality of lexical semantic representations: Clues from semantic errors in object naming. *Memory*, *3*, 333–358.

Hodgson, C., & Lambon Ralph, M. A. (2008). Mimicking aphasic semantic errors in normal speech production: Evidence from a novel experimental paradigm. *Brain and Language*, *104*, 89–101.

Howard, D. (2000). Cognitive neuropsychology and aphasia therapy: The case of word retrieval. In I. Papathanasiou (Ed.), *Acquired neurogenic communication disorders: A clinical perspective* (pp. 76–100). London: Whurr.

Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology*, *20*, 921–950.

Howard, D., & Patterson, K. (1992). *Pyramids and palm trees test: A test of semantic access from picture and words*. Bury St. Edmunds: Thames Valley Test Company.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*, 2132–2147.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Kay, J., & Ellis, A. (1987). A cognitive neuropsychological case study of anomia. *Brain*, *110*, 613–629.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessment of language processes in aphasia (PALPA)*. London: Lawrence Erlbaum Associates.

Kertesz, A. (1982). *Western Aphasia Battery*. Sydney: Pearson Psychcorp.

Lambon Ralph, M. A., Sage, K., & Roberts, J. (2000). Classical anomia: A neuropsychological perspective on speech production. *Neuropsychologia*, *38*, 186–202.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, *22*, 1–75.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *33*, 503–535.

Martin, N., Fink, R. B., Renvall, K., & Laine, M. (2006). Effectiveness of contextual repetition priming treatments for anomia depends on intact access to semantics. *Journal of the International Neuropsychological Society*, *12*, 853–866.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*, 547–559.

McRae, K., De Sa, V., & Seidenberg, M. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 863–883.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*, 935–979.

Nickels, L., & Cole-Virtue, J. (2004). Reading tasks from PALPA: How do controls perform on visual lexical decision, homophony, rhyme, and synonym judgements? *Aphasiology*, *18*, 103–126.

Rahman, R. A., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 604–614.

Rahman, R. A., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*, 713–734.

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*, 161–172.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*, 205–235.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Sailor, K., Brooks, P. J., Bruening, P. R., Seiger-Gardner, L., & Guterman, M. (2009). Exploring the time course of semantic interference and associative priming in the picture-word interference task. *The Quarterly Journal of Experimental Psychology*, *62*, 789–801.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, *54*, 228–264.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., . . . Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 8520–8524.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language & Cognitive Processes*, *28*, 597–614.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 869–918.

Starreveld, P. A., & La Heij, W. (1999). Word substitution errors in a speeded picture—word task. *American Journal of Psychology*, *112*, 521–553.

Taylor, K. I., Devereux, B. J., Acres, K., Randall, B., & Tyler, L. K. (2012). Contrasting effects of feature-based statistics on the categorization and basic-level identification of visual objects. *Cognition*, *122*, 363–374.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.

Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 664–680.

APPENDIX A

TABLE A1

Categorical (CATG), features (FEAT) and control (CONT) conditions

Pictures	*CATG condition Word cues*			*FEAT condition Word cues*		
ankle thumb	knee	foot	wrist	bones	breakable	bends
bluejay cardinal	sparrow	finch	robin	sings	nests	small
bookcase desk	cabinet	dresser	cupboard	storage	shelves/ drawers	wood

(*Continued*)

TABLE A1
(Continued)

	CATG condition			*FEAT condition*		
bracelet ring	necklace	earrings	cufflinks	expensive	gold	silver
bus van	train	airplane	ship	engine	large	passengers
cannon grenade	missile	gun	bomb	war	loud	explodes/ fires
carrot celery	cucumber	lettuce	radish	crunchy	salads	gardens
clarinet trombone	tuba	trumpet	flute	blow air	orchestras	bands
deer fox	bear	wolf	moose	forest/wilderness	fur	hunted
donut brownie	cookie	cake	muffins	oven baked	sweet	chocolate
dress skirt	camisole	blouse	shawl	women	colours	fabrics
elephant tiger	rhino	giraffe	lion	large	zoo	Africa
fly mosquito	cockroach	wasp	flea	black	annoying	small
grapefruit pineapple	banana	pear	lemon	Trees	skin	yellow
grater corkscrew	ladle	spatula	tongs	Kitchens	handle	utensil
hoe spade	rake	wheelbarrow	shovel	Gardening	handle	metal
lobster shrimp	octopus	eel	tuna	Swims	oceans	edible
owl eagle	vulture	hawk	crow	Beak	flies	meat eater
peach grape	strawberry	apple	cantaloupe	juicy	sweet	seeds
peas asparagus	broccoli	spinach	cabbage	soups	green	nutritious
piano guitar	harp	cello	violin	music	wood	strings
rabbit chipmunk	squirrel	hamster	mouse	brown	small	tail

(*Continued*)

TABLE A1
(Continued)

	CATG condition			*FEAT condition*		
saw axe	knife	sword	scissors	sharp	blade	cutting
tulip rose	petunia	daisy	lily	colourful	gardens	petals
turtle frog	alligator	toad	salamander	eats insects/animals	green	swims

CONT condition	
Picture	
ant	oven
cauliflower	pants
cow	penguin
fork	pheasant
goat	plate
goose	porcupine
hammer	screwdriver
ladybug	seal
leg	socks
mole	stork
mushroom	toaster
otter	toe

APPENDIX B
Treatment protocol

(1) The participant was asked to name the picture. If the response was correct, the participants' correct response was acknowledged. If the response was incorrect, the correct name was provided. Regardless of whether the answer was correct or incorrect, the treatment protocol was initiated.

(2) The participant was asked to provide a categorical or feature cue, depending on the approach that was being treated. ("What does this make you think of?" [FEAT condition] or "Can you think of another [category label]?" [CATG condition].) If the participant provided a cue that had been pre-selected, the corresponding card (with the written cue) was provided. If the participant provided an alternate but appropriate response, that response was written on blank cards. Participants could provide up to two alternate but appropriate responses. If he/she provided an alternate but inappropriate response, feedback was provided, and then a pre-selected response was provided. If the participant was unable to provide any cues, a pre-selected response was provided.

(3) As the cue was presented, the participant was encouraged to verbalise the response. If he/she was unable to do this, the response was verbalised by the treating clinician, while the participant repeated after the clinician. This step was repeated for all of the cues, including ones that had been generated by the participant. Thus, a range of three cues to five cues were reviewed.

(4) Once all the cues had been reviewed, all were removed from the table.

(5) The cues that had been provided by the clinician or participant were then presented on the table along with three foils that were randomly selected from a pile

of cards. These foils consisted of other categories or features, depending on the approach being used at the time.

(6) The participant was then asked to choose his/her own responses as well as pre-selected responses. If the participant selected incorrectly, this was brought to his/her attention and errors were either self-corrected or the clinician provided the correct response.

(7) Once all the appropriate cues were selected, each of the cues were verbalised by the participant if he/she could do so or repeated after the clinician if he/she could not. The clinician then prompted the participant to name the target picture again. If the participant was unable to provide the correct name, the name was provided.