Changes in Discourse Informativeness and Efficiency Following Communication-Based Group Treatment for Chronic Aphasia

Mary Boyle[a], Christa M. Akers[b], Robert Cavanaugh[c,d], William D. Hula[c,d], Alexander M. Swiderski[c,d,e], & Roberta J. Elman[f]

Abstract

Background: Evidence regarding the effect of conversationally based communication group treatment on discourse production in aphasia is limited. Given the rich, complex communication experiences provided in these groups, it seemed plausible that participation in them could result in improvement of simpler aspects of discourse production.

Aims: To examine the effects of group communication treatment on the informativeness and efficiency of structured and conversational discourse tasks in adults with chronic aphasia.

Methods & Procedures: The data for this study were discourses elicited prospectively from 23 adults with chronic aphasia by Elman & Bernstein-Ellis (1999a) as part of their randomised controlled trial, but never previously transcribed or analyzed. We evaluated changes in discourse informativeness and efficiency at treatment exit and follow-up with Bayesian generalised linear mixed-effects models. Individual effect sizes at exit and follow-up were estimated and a region-of-practical-equivalence approach was used to evaluate whether the posterior distributions at each timepoint for each participant were clinically meaningful.

Outcomes & Results: Results at the group level revealed that structured discourses became more informative and efficient after treatment, and that this improvement was at least maintained at follow-up. Informativeness of conversational discourse did not change from treatment entry to exit, but there was modest evidence of improvement at follow-up. There was no evidence of change to the efficiency of conversational discourse at either timepoint. There was wide variability in individual response to the treatment.

Conclusions: Participation in conversation-based communication group treatment was associated with more informative and efficient structured discourse production and modestly improved informativeness in conversational discourse. Examining individual responses to treatment provided additional insight about the group-level outcomes and provided some clues about factors that might have influenced performance for some of the participants.

Discourse is the form of communication that adults use most often in their daily lives (Dietz & Boyle, 2018). It encompasses a variety of functional language activities that include conversation, sharing stories, and describing how to do things (Hallowell, 2017). Activities that typically involve discourse production have been identified as priorities by people with aphasia (Worrall et al., 2011), their families (Wallace et al., 2017a), and speech-language pathologists (Wallace et al., 2017b).

There is evidence that communication groups for aphasia, particularly those that focus on conversation, improve performance on standardised aphasia tests (Elman & Bernstein-Ellis, 1999a; Hoover et al., 2018; Hoover et al., 2020; Lanyon et al, 2013; Wertz et al., 1980), enhance psychosocial well-being (Attard et al, 2015; Elman & Bernstein-Ellis, 1999b), increase social participation (Vickers, 2010), and improve quality of life (Elman & Bernstein-Ellis, 1999b; van der Gaag et al., 2005). Evidence regarding whether participation in communication groups affects discourse production is more limited. There is some evidence that aphasia communication groups improve functional communication (DeDe et al., 2019; Elman & Bernstein-Ellis, 1999a), connected speech (DeDe et al., 2019), informativeness (Hoover et al., 2020), and word retrieval (Mason et al., 2020), at least for some participants. For example, DeDe and colleagues reported that communication group participants improved on a patient-reported measure of communication functioning in everyday life and on a picture description task that evaluated information-bearing words, grammaticality, and syntactic diversity. Hoover and colleagues reported that two of five individuals with severe aphasia showed improved informativeness as measured by correct information units. Mason and colleagues reported that one of their three participants was rated as having improved significantly in the ability to retrieve words in discourse.

Given the growing popularity of communication group treatment for aphasia (Elman, 2007; 2016), it seems important to improve our understanding of whether and in what way(s) participation in these groups affects aspects of discourse production. This information could improve expectations about treatment outcomes, allowing for improved decision making by people with aphasia and service providers.

Complexity and Communication Group Treatment

Communication groups provide an enriched environment to engage in conversation with multiple communication partners, multiple opportunities for communicative initiation and response, turn-taking, interactional and transactional uses of language, and opportunities to naturally improvise language use (Elman, 2007; 2018; 2021). These communication groups promote cross-talk among group members rather than focusing on clinician-led exchanges (Bernstein-Ellis & Elman, 2007). When language is used for human social interaction in this way, there are many factors at work in addition to the linguistic ones: thought processes, socio-pragmatic factors, shared cognition, memory, and joint action, among others. In fact, because of its role in social interaction, language has been characterised as a complex adaptive system (Beckner et al, 2009; Larsen-Freeman & Cameron, 2008).

Thompson (2007) reported that "in all language domains, cascading generalisation occurs from more to less complex structures" (p.3). This has been demonstrated in the areas of child phonological

*Corresponding Author: Mary Boyle: boylem@montclair.edu; Department of Communication Sciences and Disorders, Montclair State University, 1515 Broad Street, Bloomfield, NJ 07003

[a]Department of Communication Sciences and Disorders, Montclair State University, Bloomfield, NJ USA; [b]Department of Speech-Language Pathology, Jefferson College of Rehabilitation Sciences, Thomas Jefferson University, Philadelphia, PA USA; [c]Department of Communication Sciences and Disorders, University of Pittsburgh, Pittsburgh, PA USA; [d]Audiology and Speech Pathology Program, VA Pittsburgh Healthcare System, Pittsburgh, PA USA; [e]The Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA USA; [f]Aphasia Center of California, Oakland, CA USA

This is the accepted version which is published in the journal of Aphasiology, available here: https://doi.org/10.1080/02687038.2022.2032586.

Conflicts of Interest: There are no conflicts of interest.

Funding Statement: The original RCT was partially funded by a grant from the National Easter Seals Research Program (awarded to Roberta J. Elman). This study was partially funded by the Aphasia Center of California Research Fund.

and morphosyntactic disorders and adult syntactic and semantic impairments associated with aphasia (DeAnda et al., 2020a,2020b; Gierut, 2007; Kiran, 2007; Kiran et al., 2012; Sandberg & Kiran, 2014; Thompson, 2007; Thompson & Shapiro, 2007). Conversation is quite complex. In addition to its phonologic, semantic, syntactic, and prosodic components, it involves using them strategically, choosing appropriate speech acts, gauging the knowledge and interest of the conversational partner(s), and understanding and reacting to utterances made by the partner(s). Thus, it seems plausible that aphasia group communication treatment can result in generalised improvement to simpler, related language activities (Gierut, 2007; Kiran, 2007; Thompson, 2007; Thompson & Shapiro, 2007).

Monologic discourse, although complex, is less complex than conversational discourse because it does not involve multiple speakers, turn-taking, interactional use of language, or understanding and reacting to utterances made by conversational partners. When this kind of discourse has been elicited by clinicians in response to specific stimuli it has been characterised as "structured discourse" (Simmons-Mackie et al., 2007). Therefore, if treatment focuses at the complex level of conversation in communication groups, it is plausible to expect cascading generalisation to aspects of structured discourse (Elman, 2018).

A complicating factor in predicting generalization of treatment-induced change is that in the treatment generalization literature cited in this section, the treatments were restricted to a single component of language (e.g., phonology, semantics, or syntax) presented in carefully constructed tasks focused on the linguistic impairment. Those studies tried to limit or eliminate the influence of pragmatic and non-linguistic cognitive variables that, as discussed earlier, are integral parts of human social interaction. Thus, it is not clear whether the findings that they report for cascading generalization from more to less complex language activities in carefully controlled tasks will remain true if a treatment includes all language components and non-linguistic cognitive variables, as conversationally focused communication group treatment does.

Informativeness and Efficiency of Discourses

Aspects of discourse production that might improve as a result of conversation-focused communication groups is the amount of information that a person with aphasia can convey and how efficiently they convey it. Because all people with aphasia have problems retrieving words (Davis, 2014), their ability to convey accurate and sufficient information in discourses is frequently impaired. Simmons et al. (2007) described how word retrieval opportunities are embedded in the natural conversations that form the basis of many communication groups. Participants, therefore, have many opportunities to retrieve words as they interact in the group. When difficulties arise, skilled facilitators might model or directly suggest a particular strategy, or provide an implicit correction (i.e., repeat the utterance with the correct word and then ask a question to move the conversation forward). It seems plausible, then, that participation in communication treatment groups might improve the informativeness of the discourse production of people with aphasia by improving the accuracy and/or efficiency of word retrieval (Elman, 2007, 2018, 2021). Hoover and colleagues (2020) provide preliminary evidence for this hypothesis with their report that 2 of 5 participants with severe aphasia improved the informativeness of their narrative discourses after conversation group treatment.

A widely used measure of discourse informativeness is the correct information unit (CIU). A CIU is a word that is intelligible in context and accurate in relation to, relevant to, and informative about the topic (Nicholas & Brookshire, 1993). CIU measures are the most common discourse outcome measures used by clinicians (Bryant, Spencer, & Ferguson, 2016) and the second most common measure reported in studies of discourse in aphasia (Bryant, Ferguson, & Spencer, 2016). Pritchard et al. (2017) reported that CIU measures were among the strongest of the discourse information measures they reviewed in terms of psychometric properties. Webster and Morris (2019) reported that both the number and the percentage of CIUs produced by people with aphasia during narrative discourses correlated strongly with naïve listeners' perceptual ratings of informativeness. They concluded that the CIU measures could, therefore, be considered ecologically valid indices of informativeness. Another aspect of discourse that might improve as a result of participation in communication groups is efficiency. Nicholas and Brookshire (1993) pointed out that a person who can produce more informative words per minute than another speaker could be considered a more efficient speaker. They developed the measure of CIUs produced per minute (CIUs/min) to assess this aspect of discourse production. Participation in a communication group, with repeated opportunities to retrieve and produce words in a conversational setting, might improve efficiency by decreasing the frequency or length of pauses or by decreasing error productions or lengthy repairs.

Purpose of the Study

This study is part of a larger project aiming to examine discourse data collected prospectively by Elman and Bernstein-Ellis (1999a,b) as part of their multi-method, randomised controlled trial (RCT), but not previously transcribed or analyzed because of time and budgetary constraints. The larger project's overarching goal is to study changes in the informativeness, grammatical completeness, and coherence of discourses produced by people with aphasia who participated in communication group treatment, and to examine the ways that conversational abilities changed as a result of participating in the groups. The discourses collected as part of the Elman and Bernstein-Ellis (1999a) RCT included structured and conversational discourses. Our working hypothesis was based on the principle that using more complex language activities as a starting point in treatment promotes generalised improvement of simpler, linguistically related activities (Gierut, 2007; Kiran, 2007; Kiran et al., 2012; Thompson, 2007; Thompson & Shapiro, 2007). Specifically, we hypothesised that participation in the communication treatment groups would result in improvements in shared abilities across discourse types, like informativeness, and in conversational abilities. This report focuses on whether participants' structured and conversational discourses became more informative and/or efficient after treatment. Subsequent reports will address similar questions at the sentence and macrostructural level, as well as investigation of conversational behaviors and nonverbal communication. The specific research questions addressed in this paper are: 1) Did informativeness of structured and conversational discourse improve after treatment? 2) Did efficiency of structured and conversational discourse, as assessed by CIUs/min, improve after treatment? and 3) Was the degree of any change related to aphasia severity?

Methods

The original RCT and the current study were conducted with institutional review board approval. Aphasia-friendly forms were used to obtain informed consent from participants in the original RCT.

Participants

Twenty-eight adults with stroke-induced aphasia were enrolled in Elman and Bernstein-Ellis's (1999a) RCT. Figure 1 shows the number of recruited participants and attrition across the study, resulting in 23 participants whose outcomes are reported here.

The 23 adults (10 female, 13 male) had aphasia secondary to a single, left hemisphere cerebrovascular accident sustained more than 6 months prior to enrollment in the treatment study as documented in their medical records. All were premorbidly literate in English. Potential participants were excluded if they had major medical complications or a history of alcoholism, were more than 80 years old, or scored below the 10th or above the 90th overall percentile on the Shortened Porch Index of Communicative Ability (SPICA, Disimoni et al., 1980). Aphasia types, determined by performance profiles on the Western Aphasia Battery (Kertesz, 1982), included

Figure 1.

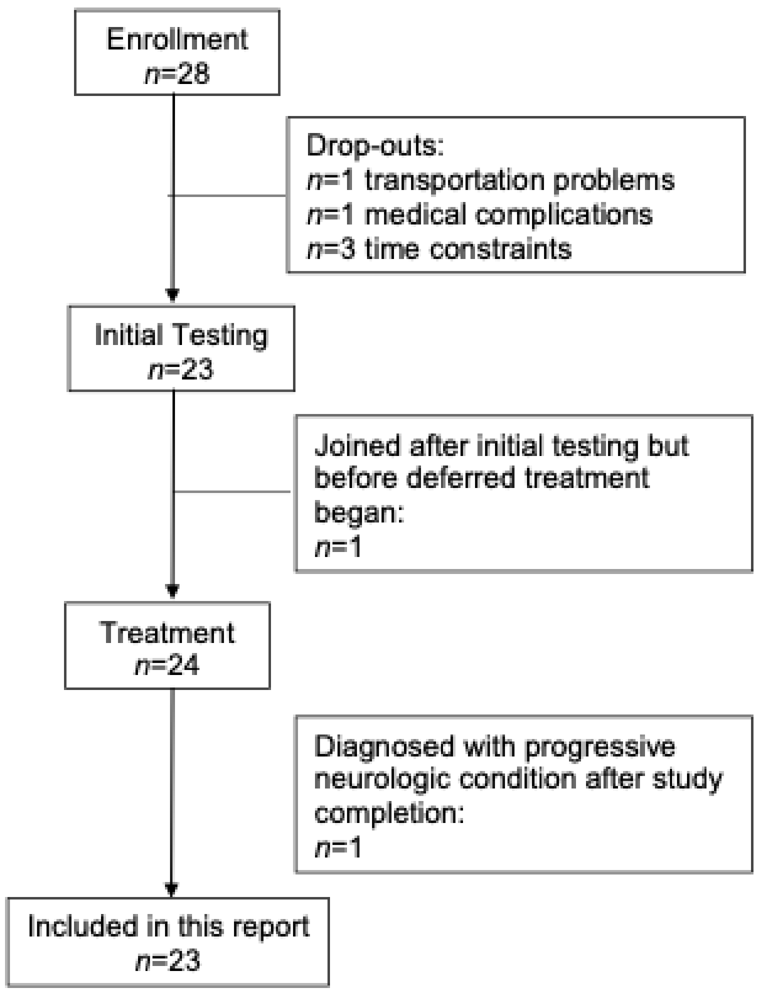

Figure 1. Participant recruitment and attrition across the study

Table 1. Participants' Demographic and Test Information

Participant*	Gender	Aphasia Type	MSD	Age	MPO	Education (years)	SPICA%	WAB AQ	CADL
Immediate Treatment Mild-Moderate									
P11	F	Anomic	None	38.0	17.00	14.00	76.00	80.80	131.00
P13	M	Unclassified	None	58.0	77.00	15.00	78.00	85.90	134.00
P14	F	Anomic	None	72.0	13.00	16.00	90.00	92.90	136.00
P17	M	Anomic	None	67.0	103.00	20.00	80.00	88.00	125.00
P18	M	Broca's	None	46.0	7.00	16.00	57.00	61.50	120.00
	Mean			56.2	43.40	16.20	76.20	81.80	129.20
	SD			14.2	43.70	2.30	12.00	12.20	6.60
Moderate-Severe									
P2	M	Trans. Motor	Mild AOS	60.0	21.00	14.00	58.00	72.80	116.00
P4	F	Trans. Motor	Moderate AOS	79.0	36.00	16.00	35.00	61.40	96.00
P6	F	Broca's	Severe AOS	58.0	33.00	15.00	35.00	13.10	57.00
P7	F	Broca's	Severe AOS	51.0	12.00	16.00	30.00	24.20	64.00
P9	M	Broca's	Severe AOS	49.0	29.00	12.00	30.00	18.90	106.00
P10	M	Broca's	Moderate AOS	63.0	16.00	12.00	61.00	45.90	124.00
P15	M	Conduction	Moderate AOS	58.0	26.00	14.00	47.00	55.90	98.00
	Mean			59.7	24.70	14.10	42.30	41.70	94.40
	SD			9.8	8.80	1.70	13.10	23.20	25.20
Deferred Treatment									
Mild-Moderate									
P3	F	Anomic	Mild dysarthria	58.0	23.00	12.00	67.00	87.70	130.00
P5	F	Anomic	None	80.0	36.00	11.00	64.00	75.10	102.00
P21	M	Anomic	None	52.0	336.00	16.00	60.00	80.20	113.00
P19	M	Anomic	Mild AOS	52.0	134.00	20.00	76.00	76.40	129.00
P24	F	Unclassified	None	70.0	43.00	16.00	78.00	94.30	129.00
	Mean			62.4	114.40	15.00	69.00	82.74	120.60
	SD			12.3	131.42	3.61	7.75	8.11	12.58
Moderate-Severe									
P1	M	Broca's	Mild AOS	47.0	10.00	18.00	42.00	57.40	121.00
P8	M	Conduction	None	48.0	19.00	14.00	54.00	67.30	107.00
P16	M	Conduction	None	55.0	7.00	16.00	54.00	54.40	123.00
P20	F	Broca's	None	71.0	137.00	18.00	50.00	63.50	104.00
P22	M	Conduction	None	59.0	42.00	16.00	46.00	65.30	114.00
P23	F	Broca's	Severe AOS	65.0	59.00	18.00	23.00	20.70	70.00
	Mean			57.5	45.70	16.70	44.80	54.80	106.50
	SD			9.5	49.00	1.60	11.70	17.40	19.40

Note. *P12 is not included in this study because it was discovered that she had a progressive neurological condition. MSD = motor speech disorder; MPO = months post onset; SPICA% = Shortened Porch Index of Communicative Ability percentile score; WAB AQ = Western Aphasia Battery Aphasia Quotient; CADL = Communicative Activities of Daily Living; SD = standard deviation; Trans. Motor = transcortical motor aphasia; AOS = apraxia of speech.

anomic (n = 7), Broca's (n = 8), transcortical motor (n = 2), conduction (n = 4), and unclassified (n = 2) aphasias. Family members reported that neither vision nor hearing interfered significantly with participants' communication, observations confirmed by the research team who viewed video recordings of the discourses. That is, the participants responded appropriately to visual stimuli during structured discourse tasks and conversational interactions, and they responded appropriately to verbal comments, questions, and requests presented at normal conversational loudness levels. The presence of coexisting motor speech disorders was determined by three of the authors (MB, CMA, RJE), all licensed, certified speech-language pathologists, using the perceptual criteria described by Duffy (2020). These include the presence of speech sound distortions or distorted substitutions; as well as phonatory, resonance, or prosodic impairments. One participant had a mild dysarthria. Ten participants had apraxia of speech: three mild, three moderate, and four severe. All participants had completed individual speech-language treatment prior to enrollment in the treatment study and none received concomitant speech-language therapy (Elman & Bernstein-Ellis, 1999a). Demographic and test information for the participants is in Table 1.

Table 2. Structured and Conversational Discourses

Type of Discourse	Structured Discourses (Monologues)
Conversational Discourse	Three picture descriptions: The Cookie Theft*, The Birthday Party*, The Argument* A personal narrative: A typical Sunday's agenda* A procedure: How to do dishes by hand*
Discourse Tasks	Participant and a familiar conversational partner of their choice conversed for 7 minutes about anything. The investigator left the room while the dyad conversed.

Note. * From Nicholas & Brookshire (1993)

Setting for Data Collection

All data were collected in an 8 ft by 10 ft (2.44 m by 3.05 m) private office in a not-for-profit outpatient rehabilitation center. The office contained a table, several chairs, a video camcorder, and an integrated video monitor/player.

Assessment Battery

In addition to the WAB, the SPICA, and the Communicative Abilities in Daily Living test (Holland, 1980) which are reported in Table 1, assessment tasks that were part of the RCT (Elman & Bernstein-Ellis, 1999a), were the Communicative Effectiveness Index (Lomas et al.,1989) and the Affect Balance Scale (Bradburn, 1969). During the RCT, the investigators video recorded structured and conversational discourse samples which were never transcribed or analyzed prior to this current report.

Group Assignment

Participants were assessed at enrollment and randomly assigned to receive immediate treatment (IT) or deferred treatment (DT) by drawing names from a bag. Based on their overall SPICA score, each participant was assigned either to a mild-moderate (SPICA scores between 50 and 90) or moderate-severe (SPICA scores between 10 and 65) communication group. Participants with moderate aphasia (SPICA scores between 50 and 65) could be assigned to either of the groups. Participants in the IT group entered treatment immediately. To control for the effects of social contact that participants in the IT group received during this period, participants in the DT group spent the 4 months after enrollment attending a minimum of 3 hours per week in a variety of community social group activities while they awaited treatment. After the 4-month period, the DT group was reassessed and then entered treatment. Thus, there were four groups: a mild-moderate and a moderate-severe group that received immediate treatment, and a mild-moderate and a moderate-severe group that received deferred treatment (Elman & Bernstein-Ellis, 1999a).

Discourse Elicitation

Structured discourses were elicited using five tasks and elicitation procedures described by Nicholas and Brookshire (1993) The conversational discourses involved the participants and a familiar conversation partner of their choice. The conversational partners did not receive any training about communicating with people who have aphasia as part of the RCT. See Table 2 for additional information about the discourse tasks.

The discourses were video recorded at three times for the IT group (treatment entry, treatment exit, and follow-up) and four times for the DT group (study enrollment, treatment entry, treatment exit, and follow-up). Follow-up samples were obtained 4-to-6 weeks after treatment ended. One of the participants, P4, produced discourses in response to all of the elicitation stimuli at treatment entry but did not respond to the prompts to tell what she usually did on Sundays or tell how to go about washing dishes by hand during the post-treatment and follow-up assessments, despite repeated encouragement to do so by the examiner. It was the examiner's impression that she was not interested in responding to these prompts, rather than that she lacked the ability to respond to them. Therefore, these two tasks were not included in the analysis for P4 at post-treatment and follow-up. P11 did not attend the follow-up session for medical reasons, and one task (how to wash dishes by hand) was not recorded for P22 in the follow-up session because of experimenter error. P23's partner did not attend the exit or follow-up sessions, so P23's data was not included in the conversational discourse analyses.

Communication Treatment

Participants received 5 hours of group communication treatment per week in two 2.5-hour sessions over a 4-month period for a total of 160 hours of treatment. The group facilitator was a licensed, certified speech-language pathologist who had more than 10 years of post-clinical-fellowship experience providing aphasia treatment and about 1 year of experience in facilitating communication-based aphasia groups. The treatment groups were consistent with the Life Participation Approach to Aphasia (LPAA Project Group, 2000). Treatment sessions were conversationally-based and focused on facilitated discussion of current events and activities, usually without preselected topics. Treatment emphasised exchange of information using whatever means possible (e.g., speech, gesture, drawing, writing, facial expressions, etc.) and encouraged both initiation of conversational turns and continuation of others' conversational turns. Occasionally, social games like Uno or blackjack, reading and writing practice, or arts-based activities made up the last 60 minutes of the session. The clinician managed the group implicitly by using modeling and scaffolding with natural feedback, but also by using explicit direction to guide participants in using strategies or initiating conversational exchanges when necessary (Elman & Bernstein-Ellis, 1999a; Simmons-Mackie et al., 2014). A minimum attendance rate of 80% of all treatment sessions was required for inclusion in the study, and all participants reported here met this criterion. More details about the group activities and the facilitative techniques that were used can be found in Elman & Bernstein-Ellis (1999a) and Bernstein-Ellis & Elman (2007).

Blinding

The original Elman and Bernstein-Ellis study was unblinded. For the current study, discourse transcribers and analysts were blind to treatment timepoint and treatment condition (immediate versus delayed treatment). The first author of this paper used a free, online tool (https://www.randomizer.org/) to randomly assign a code to each of the three or four assessment time points for each participant. Videos and transcripts were labeled with those randomly-assigned codes until transcription and coding were complete.

Data Preparation and Analysis

The discourse samples were transcribed from the video recordings using the CHAT format that is part of the Computerized Language Analysis (CLAN) system (MacWhinney, 2000). This system time-links the video to the transcript and can generate a duration for each speaker in the sample. Transcription training was provided to graduate student research assistants verbally, supplemented with written instructions that included directions for segmenting utterances into T-units and fragments (incomplete clauses; German, 1991) and on coding communicative gestures, writing or drawing attempts, and pauses, although pauses and nonverbal communication attempts do not enter into the discourse analyses used in this study. After instruction, the assistants transcribed practice samples of non-study discourses elicited with the Nicholas and Brookshire (1993) stimuli.

Practice continued until agreement with an experienced transcriber (the first author) on at least 80% of utterances was achieved. The transcriptions of the discourse samples that were elicited as part of this study were reviewed by the first author while viewing the video recording. Discrepancies were resolved by consensus prior to analysis.

The structured discourses were combined and analyzed for number of words and number of correct information units (CIUs) according to the rules published by Nicholas & Brookshire (1993). These rules also formed the basis for analyzing words and CIUs in the conversation samples, but we followed the adaptations made by Leaman and Edmonds (2019) for analyzing CIUs in conversation. These adaptations included: not counting verbatim repetitions of the partner as CIUs except when used to respond to an explicit question or request clarification; counting vague language such as "whatever" as a CIU if it is acceptable in informal conversation; and counting words spelled out loud as CIUs, if accurate. Please refer to Leaman and Edmonds (2019) for operational definitions and additional examples. Additionally, we followed Leaman's recommendation (M. Leaman, personal communication, February 3, 2020) not to count "yes", "no" or their synonyms (e.g., "mhm", "unhunh") as words or CIUs in conversation because of the difficulty in determining whether they are being used in an informative way. The research assistants who analyzed the utterances received verbal and written instruction on preparing the transcript, counting words, and counting CIUs. They practiced on discourse samples that were not part of this study until they achieved at least 80% agreement with the first author.

To assess informativeness across samples of different lengths, the proportion of CIUs was calculated (total CIUs/total words). To assess how efficiently information was conveyed, the number of CIUs produced per minute was calculated by dividing the number of CIUs by the participant's speaking duration provided by the CLAN program. To assess reliability of word and CIU identification, a second rater independently analyzed 25% of the discourse samples chosen randomly, and the first rater analyzed a randomly-chosen 25% of the samples a second time, at least 4 weeks after the first analysis, without consulting the initial analysis. The intraclass correlation coefficient (ICC) was used to assess the extent of agreement between the two raters and the two times. ICC estimates and their 95% confidence intervals were calculated using the R package irr (Gamer et al., 2019) with a two-way mixed model (Koo & LI, 2016). In structured discourse, the ICC values obtained for interrater reliability of identifying words and CIUs were 0.99 ($p < 0.05$) for each variable. The ICC values for intrarater reliability of identifying words (0.99, $p < 0.05$) and CIUs (0.99, $p < 0.05$) was similarly high. In conversational discourse, the ICC value obtained for interrater reliability of identifying words was 0.99 ($p < 0.05$) and of identifying CIUs was 0.99 ($p < 0.05$). The ICC value for intrarater reliability of identifying words was 0.99 ($p < 0.05$), and of identifying CIUs was 0.99 ($p < 0.05$). Together, these ICC values indicate reliable scoring of CIUs and words in structured and conversational discourse samples.

Statistical Analysis

Changes in discourse informativeness (proportion of CIUs) and efficiency (CIUs per minute) were evaluated with Bayesian generalised linear mixed-effects models in Stan (Carpenter et al., 2017; Stan Development Team, 2020) accessed through the R package BRMS. This modeling approach is well-suited to this study, as it can accommodate repeated measurements within participants and leverages partial pooling to make more generalizable estimates (McElreath, 2020). The Bayesian framework is advantageous because it permits intuitive interpretation of results (e.g. a 90% credible interval is interpreted as a 90% probability that the effect falls within the interval, based on the data and prior assumptions, which were purposely vague) and permits sampling of the posterior distributions (the distribution of possible effects, based on the data and prior assumptions) of group-level effects to estimate individual effect sizes and the degree to which they are clinically meaningful (see Nalborczyk et al., 2019 for a review). A detailed description of the statistical analysis is described in Appendix A.

First, we evaluated whether the participants randomised to the deferred treatment group demonstrated any changes in discourse informativeness or efficiency after four months of social activities prior to initiating treatment. For both discourse measures, models included a categorical effect of timepoint (enrollment to treatment entry), with intercepts and slopes varying by participant.

Second, we evaluated whether participants improved in discourse informativeness and efficiency from entry to exit and from entry to follow-up, and whether such change was moderated by aphasia severity. For both discourse measures, models included effects of timepoint, aphasia severity (as measured by the SPICA), and their interaction. Timepoint was structured as a categorical (factor) variable using treatment (dummy) contrast coding with treatment entry as the reference value, thus describing change from treatment entry to exit and treatment entry to follow-up. Change between exit and follow-up timepoints was examined by comparing the posterior distributions at these time points post-hoc. Aphasia severity was centered and standardised across participants (scaled by its standard deviation). The model structure included varying intercepts by participant and varying slopes by participant for timepoint.

Individual effect sizes at exit and follow-up were estimated by taking the difference of the model's posterior distributions at each timepoint for each participant.[1] A region of practical equivalence (ROPE; Makowski et al., 2019) approach was used to evaluate whether the resulting distributions were clinically meaningful. Following this approach, we defined a ROPE (that is, a region where any change, even if statistically reliable, would be too small to be practically meaningful) as being within +/- 5 percentage points of treatment entry for the proportion of CIUs and +/- 3 CIU/min of treatment entry for efficiency. As is recommended, the region of practical equivalence was set by establishing consensus among the research team based on our research and clinical experience with both discourse and aphasia treatment. We considered an individual's effect size to be clinically meaningful if > 90% of their posterior distribution (the distribution of possible effect sizes for that individual, given the data) was greater than this range. Similarly, if > 90% of an individual's posterior distribution fell within the region of practical equivalence, we interpreted this as evidence that they did not show clinically meaningful change in discourse informativeness. Posterior distributions that did not meet either of these conditions were categorised as "uncertain", meaning that there was not clear evidence as to whether the participant changed in a clinically meaningful way as a result of the intervention.

Because individual effect sizes were evaluated relative to our best judgement of a clinically meaningful benchmark rather than relative to zero, this approach to examining individual participants' results is more conservative than is typically found in frequentist analyses of RCT aphasia treatment studies. Typically, if individual participant results are reported in frequentist analyses, any improvement, no matter how small, is considered to be a positive response to treatment. For example, Wertz et al. (1981) described all participants who made even a 1-point improvement on their outcome measures as having responded positively to treatment. In contrast, the ROPE approach to examining individual outcomes allows researchers to specify what amount of change is likely to be clinically meaningful. As a more traditional metric, we also reported 90% credible intervals for each individual's effect size, to further provide information about whether their change was reliably different from zero.

Results

Overall Changes Across Timepoints

Structured discourse

[1] 1 For one instance of missing data at follow-up in the structured discourse samples (P11; medical reasons), model predictions of the individual effect size estimate included the residual (observation-level) variance. This approach is equivalent to generating model predictions for "new" data and can be used to impute values when data are only missing for the outcome variable.

In structured discourse, on average, both the mean and median CIUs increased from entry (mean: 131, median: 77) to exit (mean: 136, median: 132) to follow-up (mean: 154, median: 117). The mean number of words produced fell from entry (312) to exit (301) but increased overall by follow-up (325). The median number of words produced increased across all timepoints (entry: 221, exit: 262, follow-up: 257). The mean and median number of minutes (i.e., total talking time) remained relatively stable across all timepoints (mean: entry 6.39, exit 6.37, follow-up 6.16; median: entry 5.12, exit 5.72, follow-up 5.04). Participant performance in terms of the proportion of CIUs and CIUs per minute in structured discourse across timepoints is displayed in Figure 2, with descriptive statistics for these measures in Table 3.

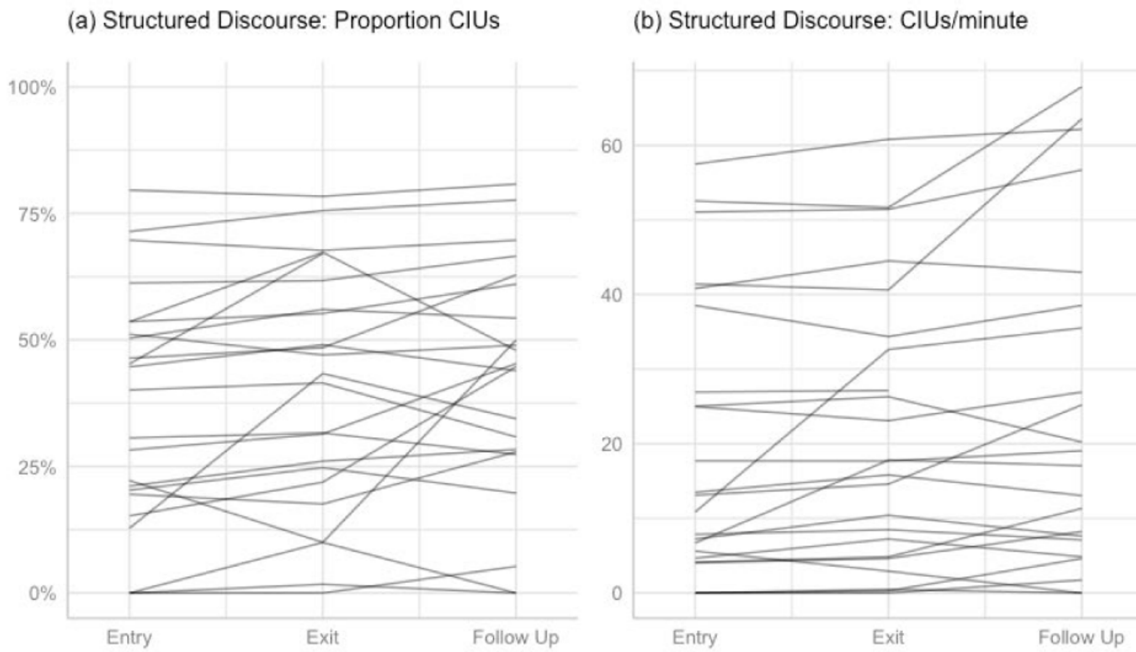

Figure 2. Change in (a) Informativeness (Proportion CIUs) and (b) Efficiency (CIUs Per Minute) in Structured Discourse

Conversational discourse

The overall CIU changes in conversational discourse followed a generally similar pattern from entry (mean: 148, median: 130) to exit (mean: 158, median: 93) to follow-up (mean: 177, median: 101). The mean and median number of words produced fell from entry (mean: 256, median: 183) to exit (mean: 248, median: 179) but increased at follow-up (mean: 267, median: 199). The mean and median number of minutes remained relatively stable across timepoints (mean: entry 4.06, exit 3.80, follow-up 3.83; median: entry 3.87, exit 3.72, follow-up 4.05). Participant performance in terms of the proportion of CIUs and CIUs per minute in conversational discourse across timepoints is displayed in Figure 3.

Test-retest stability

Structured discourse

For structured discourse, there was no evidence of substantial change from enrollment to treatment entry for the DT group for informativeness (β = 0.05, 90% Credible Interval: [−0.30,

Table 3. Descriptive Statistics for Outcome Measures

	n	Mean	Standard Deviation	Min	Max
Structured Discourse					
Proportion Correct Information Units					
Treatment entry	23	0.37	0.23	0	0.80
Treatment exit	23	0.40	0.23	0	0.78
Follow-up	22	0.42	0.23	0	0.81
CIUs/min					
Treatment entry	23	23.80	20.37	0	67.88
Treatment exit	23	20.69	17.98	0	60.79
Follow-up	22	19.94	19.68	0	63.51
Conversational Discourse					
Proportion Correct Information Units					
Treatment entry	23	0.52	0.21	0	0.85
Treatment exit	22	0.54	0.23	0	0.85
Follow-up	21	0.56	0.24	0	0.83
CIUs/min					
Treatment entry	23	34.70	28.30	0	107.29
Treatment exit	22	36.85	30.82	0	103.63
Follow-up	21	39.84	30.44	0	104.79

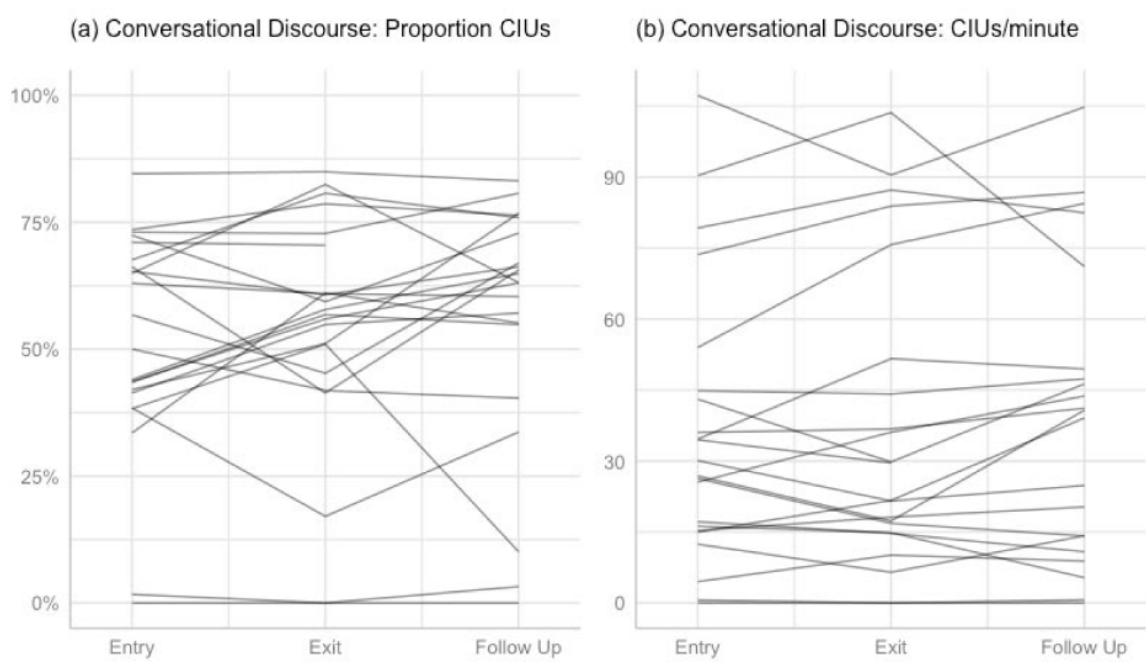

Figure 3. Change in (a) Informativeness (Proportion CIUs) and (b) Efficiency (CIUs Per Minute) in Conversational Discourse

Model Results for Informativeness (Proportion CIUs) in Structured Discourses

Parameter	Estimate	Std. Error	90% CI
Population level effects			
Intercept	-0.90	0.24	[-1.28, -0.50]
Timepoint: entry to exit	0.28	0.13	[0.07, 0.49]
Timepoint: entry to follow up	0.35	0.15	[0.10, 0.59]
Severity	1.26	0.25	[0.85, 1.69]
Timepoint: entry to exit: severity	-0.10	0.15	[-0.34, 0.14]
Timepoint: entry to follow up: severity	-0.11	0.17	[-0.39, 0.18]
Group level effects (participant)			
sd: intercept	1.10	0.20	[0.77, 1.40]
sd: timepoint: entry to exit	0.47	0.10	[0.32, 0.63]
sd: timepoint: entry to follow up	0.57	0.14	[0.36, 0.78]

Note: 90% CI refers to the 90% quantile equal tailed credible interval.
Treatment entry was the reference level.

0.36]) or efficiency (β = 0.00, 90% Credible Interval: [-0.31, 0.29]). Only one participant, P1, appeared to demonstrate meaningful change in structured discourse informativeness (31% CIU to 56% CIU) after the 4-month period of community-based social involvement.

Conversational discourse

For the conversational discourses, there was no evidence of substantial change from enrollment to treatment entry for the DT group for informativeness ($\beta = 0.04$, 90% Credible Interval (CI) [-0.31, 0.42]) or efficiency (β = 0.00, 90% CI [-0.31, 0.29]).

Informativeness

Structured Discourse

The main effect of timepoint from both treatment entry to exit ($\beta = 0.28$, 90% CI [0.07, 0.49]) and treatment entry to follow-up ($\beta = 0.35$, 90% CI [0.10, 0.59]) was reliably positive, suggesting that informativeness improved immediately after treatment, and between treatment entry and follow-up. The estimate for the change from exit to follow-up was was not reliably different from zero, suggesting that the treatment effect was unchanged at follow-up. The main effect of severity was large ($\beta = 1.26$, 90% CI [0.85, 1.69]), indicating that aphasia severity was strongly associated with informativeness of structured discourses at treatment entry. However, neither the interaction between severity and change from entry to exit ($\beta = -0.10$, 90% CI [-0.34, 0.14]) or between severity and change from entry to follow-up ($\beta = -0.11$, 90% CI [-0.39, 0.18]) were reliably different from zero. The posterior probabilities that the timepoint by severity interactions were less than zero were 76.4% and 74.7%, for exit and follow-up, respectively. These results provide uncertain evidence for the presence or absence of the interaction. Full model results for informativeness in structured discourse are reported in Table 4.

Conversational Discourse

For informativeness in conversation, the main effect of timepoint for entry to exit was not meaningfully different than zero (β = 0.02, 90% CI [-0.24, 0.30]). However, change from entry to follow-up was positive (β = 0.17, 90% CI [-0.09, 0.42]), with a posterior probability of 86.2% in favor of this small, albeit positive effect. The posterior probability of a positive change from exit to follow-up was 78.03% (Estimate = 0.15, 90% CI [-0.17, 0.48]). Aphasia severity was strongly associated with performance at baseline (β = 0.72, 90% CI [0.33, 1.12]). There was an interaction

Model Results for Informativeness (Proportion CIUs) in Conversational Discourse

Parameter	Estimate	Std. Error	90% CI
Population level effects			
Intercept	-0.90	0.24	[-1.28, -0.50]
Timepoint: entry to exit	0.28	0.13	[0.07, 0.49]
Timepoint: entry to follow up	0.35	0.15	[0.10, 0.59]
Severity	1.26	0.25	[0.85, 1.69]
Timepoint: entry to exit: severity	-0.10	0.15	[-0.34, 0.14]
Timepoint: entry to follow up: severity	-0.11	0.17	[-0.39, 0.18]
Group level effects (participant)			
sd: intercept	1.10	0.20	[0.77, 1.40]
sd: timepoint: entry to exit	0.47	0.10	[0.32, 0.63]
sd: timepoint: entry to follow up	0.57	0.14	[0.36, 0.78]

Note: 90% CI refers to the 90% quantile equal tailed credible interval. Treatment entry was the reference level.

between aphasia severity and both timepoint coefficients (entry to exit: $\beta = 0.24$, 90% CI [-0.07, 0.55]; entry to follow-up: $\beta = 0.22$, 90% CI [-0.07, 0.53]), such that conversational discourse informativeness improved more for people with milder aphasia. The posterior probability that the timepoint by severity interactions were greater than zero were 91.0% and 89.3% respectively, providing modest evidence that these estimates may be reliably greater than zero. Full model results for informativeness in structured discourse are reported in Table 5.

Efficiency

Structured Discourse

For efficiency, the main effect of timepoint for both entry to exit ($\beta = 0.24$, 90% CI [0.09, 0.38]) and entry to follow-up ($\beta = 0.34$, 90% CI [0.14, 0.54]) was robustly positive, suggesting that discourse efficiency improved as a result of treatment. The posterior probability of a positive change from exit to follow-up was 82.88%, providing modest evidence that these estimates may be reliably greater than zero (Estimate = 0.10, 90% CI [-0.08, 0.28]). Again, aphasia severity was strongly associated with performance at baseline ($\beta = 1.40$, 90% CI [0.98, 1.81]). There was a negative interaction between aphasia severity and both timepoint coefficients (entry to exit: $\beta = -0.20$, 90% CI [-0.35, -0.03]; entry to follow-up: $\beta = -0.21$, 90% CI [-0.43, 0.02]), such that efficiency improved more for people with more severe aphasia. Full model results for efficiency in structured discourse are reported in Table 6.

Conversational Discourse

For conversational efficiency, neither the main effect of timepoint for entry to exit ($\beta = -0.08$, 90% CI [-0.23, 0.07]). or entry to follow-up ($\beta = 0.06$, 90% CI [-0.16, 0.24]) were meaningfully different than zero. Aphasia severity was strongly associated with performance at baseline ($\beta = 0.87$, 90% CI [0.41, 1.31]). There was a reliable interaction between aphasia severity and timepoint from entry to exit ($\beta = 0.16$, 90% CI [0.00, 0.31] and a more modest one from entry to follow-up: $\beta =$ 0.09, 90% CI [-0.13, 0.30]), such that conversational efficiency improved more for people with milder aphasia. The posterior probabilities that the timepoint by severity interactions were greater than zero were 95.8% from entry to exit and 77.0% from entry to follow-up, providing modest evidence that this interaction was greater than zero between entry and exit but uncertain evidence that aphasia severity moderated improvement from entry to follow-up. Full model results for efficiency in conversational discourse are reported in Table 7.

Model Results for Efficiency (CIUs per minute) in Structured Discourse

Parameter	Estimate	Std. Error	90% CI
Population level effects			
Intercept	2.14	0.24	[1.75, 2.54]
Timepoint: entry to exit	0.24	0.09	[0.09, 0.38]
Timepoint: entry to follow up	0.34	0.12	[0.14, 0.54]
Severity	1.40	0.26	[0.98, 1.81]
Timepoint: entry to exit: severity	-0.20	0.10	[-0.35, -0.03]
Timepoint: entry to follow up: severity	-0.21	0.14	[-0.43, -0.02]
Group level effects (participant)			
sd: intercept	1.10	0.21	[0.78, 1.42]
sd: timepoint: entry to exit	0.27	0.07	[0.15, 0.38]
sd: timepoint: entry to follow up	0.42	0.11	[0.25, 0.59]

Note: 90% CI refers to the 90% quantile equal tailed credible interval. Treatment entry was the reference level.

Model Results for Efficiency (CIUs per minute) for Conversational Discourse

Parameter	Estimate	Std. Error	90% CI
Population level effects			
Intercept	2.95	0.28	[2.47, 3.40]
Timepoint: entry to exit	-0.08	0.09	[-0.23, 0.07]
Timepoint: entry to follow up	0.06	0.12	[-0.16, 0.24]
Severity	0.87	0.28	[0.41, 1.31]
Timepoint: entry to exit: severity	0.16	0.10	[0.00, 0.31]
Timepoint: entry to follow up: severity	0.09	0.13	[-0.13, 0.30]
Group level effects (participant)			
sd: intercept	1.29	0.25	[0.89, 1.67]
sd: timepoint: entry to exit	0.30	0.08	[0.19, 0.42]
sd: timepoint: entry to follow up	0.44	0.10	[0.28, 0.59]

Note: 90% CI refers to the 90% quantile equal tailed credible interval. Treatment entry was the reference level.

Individual Effect Estimates

Structured Discourse

Individual estimates for informativeness and efficiency in structured discourse are presented in Figures 4 and 5. Tables of values for the effect sizes, credible intervals, and the ROPE are in Appendix B. A high variability in response is apparent across all study participants and is either null or positive in almost all cases. For informativeness, according to our stringent criteria, two participants showed meaningful change at exit and five at follow-up. One participant showed meaningful decline at follow-up. That is, his informativeness in structured discourse decreased by more than 5% at the follow-up session. Six participants showed clear evidence of no meaningful change by the ROPE criteria at exit and three participants did so at follow-up. The remaining participants made changes in informativeness between entry and exit (n = 15) or follow-up (n = 14), but the change straddled the ROPE limits, making their response to treatment uncertain. Appendix B provides individual estimates for each timepoint and each measure, 90% credible intervals, posterior probabilities, and treatment response of each participant. Figure 4 provides a visual representation of the individual effects for informativeness in structured discourse.

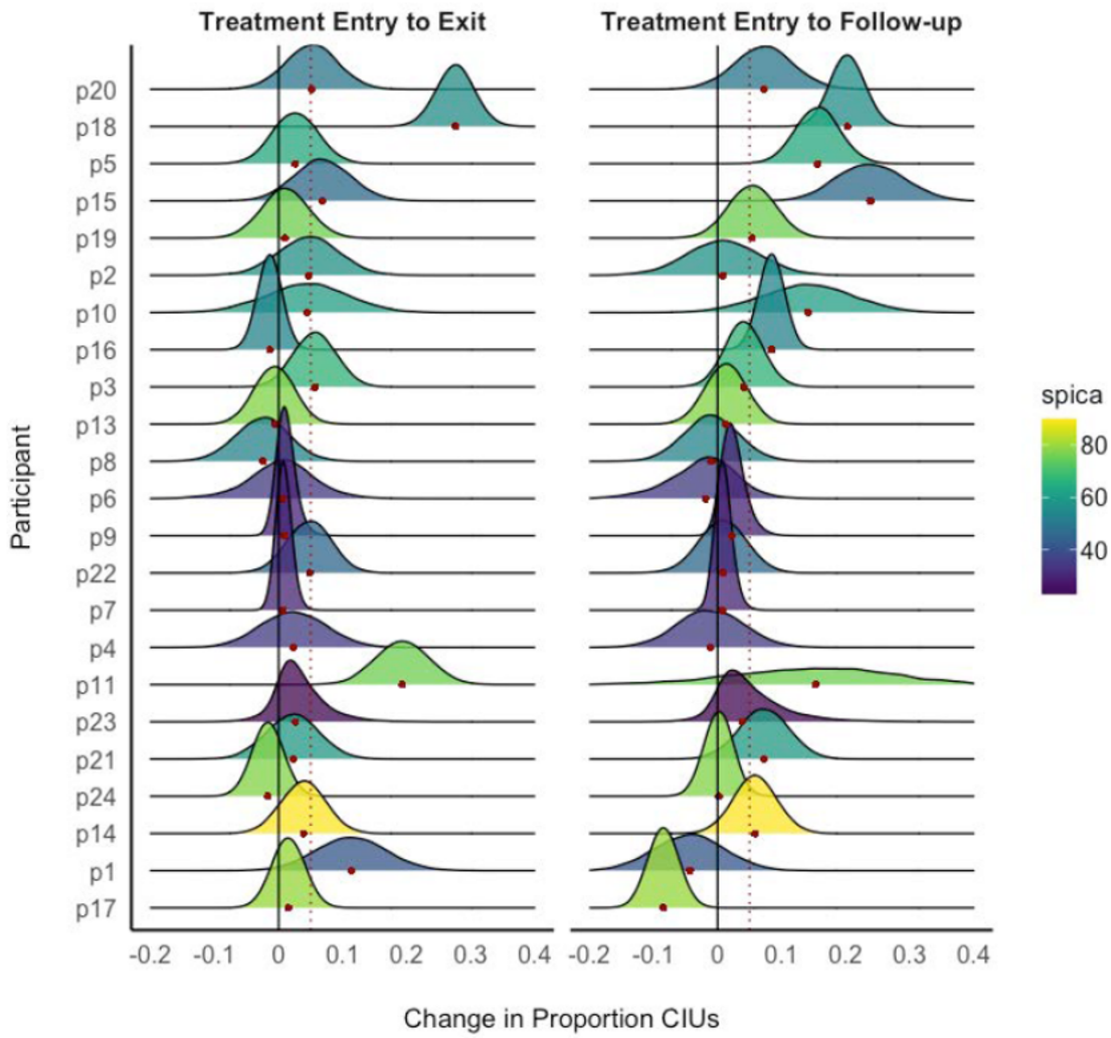

Figure 4. Individual Effect Estimates for Informativeness (Proportion CIUs) in Structured Discourse. Note. Color indicates aphasia severity (SPICA), with lighter colors indicating milder and darker colors indicating more severe aphasia. Red dots indicate the median point estimate for each effect size.

For efficiency, two participants, met the criteria for improved efficiency in structured discourse at exit. Six participants met the criteria at follow-up. Nine participants at exit and seven at follow-up were within the ROPE suggesting no clinically meaningful change in response to treatment. Twelve participants at exit and 10 at follow-up had uncertain outcomes. Individual estimates for changes in efficiency in structured discourse are in Appendix B, and Figure 5 provides a visual representation.

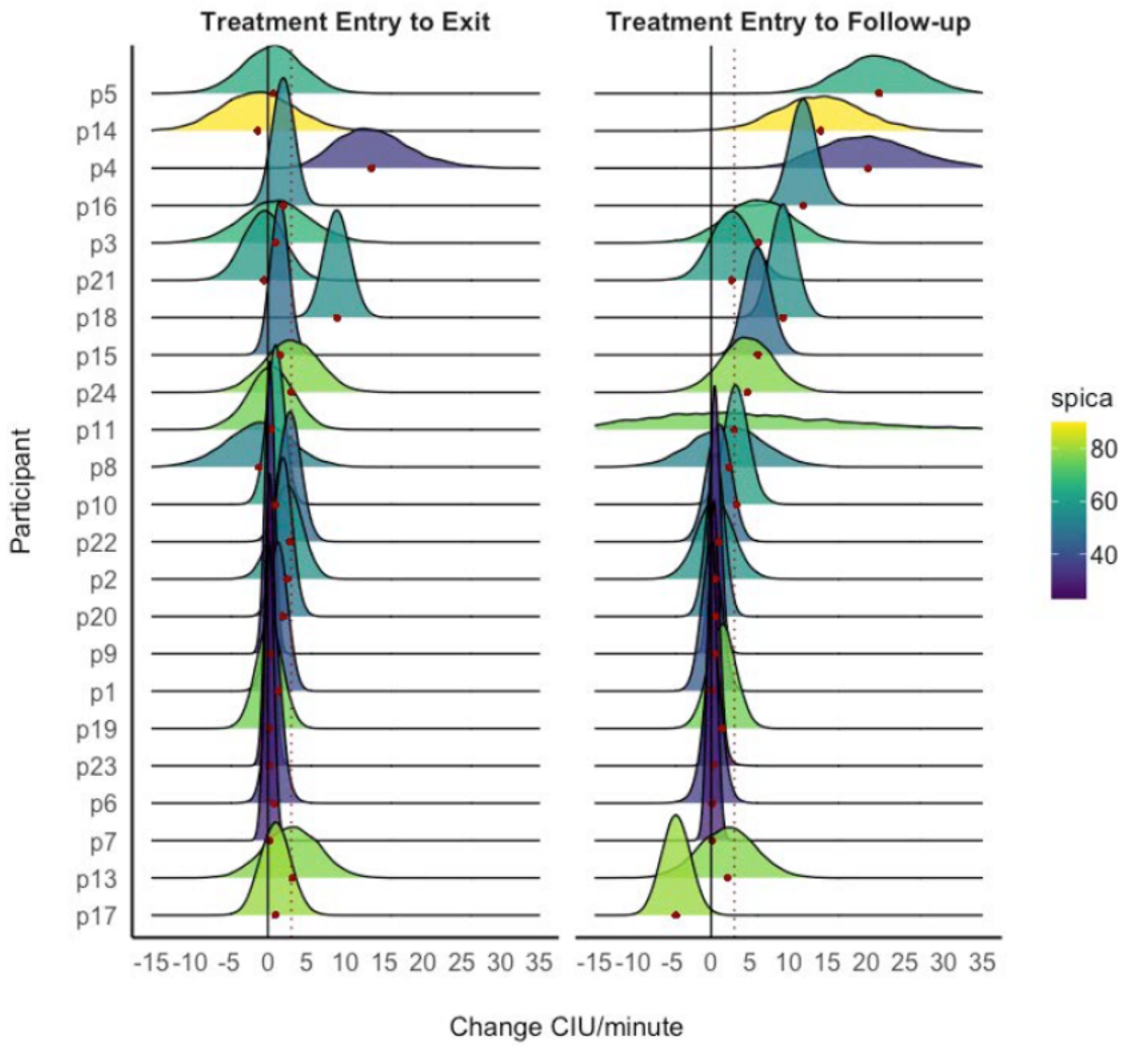

Figure 5. Individual Effect Estimates for Efficiency (CIUs Per Minute) in Structured Discourse. Note. Color indicates aphasia severity (SPICA), with lighter colors indicating milder and darker colors indicating more severe aphasia. Red dots indicate the median point estimate for each effect size.

Conversational Discourse

A high variability in response is apparent across all study participants in the conversational discourse outcomes. For informativeness, five participants showed meaningful positive change at exit and five did so at follow-up according to our stringent criteria. Four participants had clinically meaningful negative change at exit and one participant became less informative at follow-up. Three participants at exit and two participants at follow-up showed no response to treatment based on the ROPE criteria. The outcomes of 10 participants at exit and 13 participants at follow-up were judged to be uncertain. Individual estimates for changes in informativeness in conversational discourse are in Appendix C, with a visual representation in Figure 6.

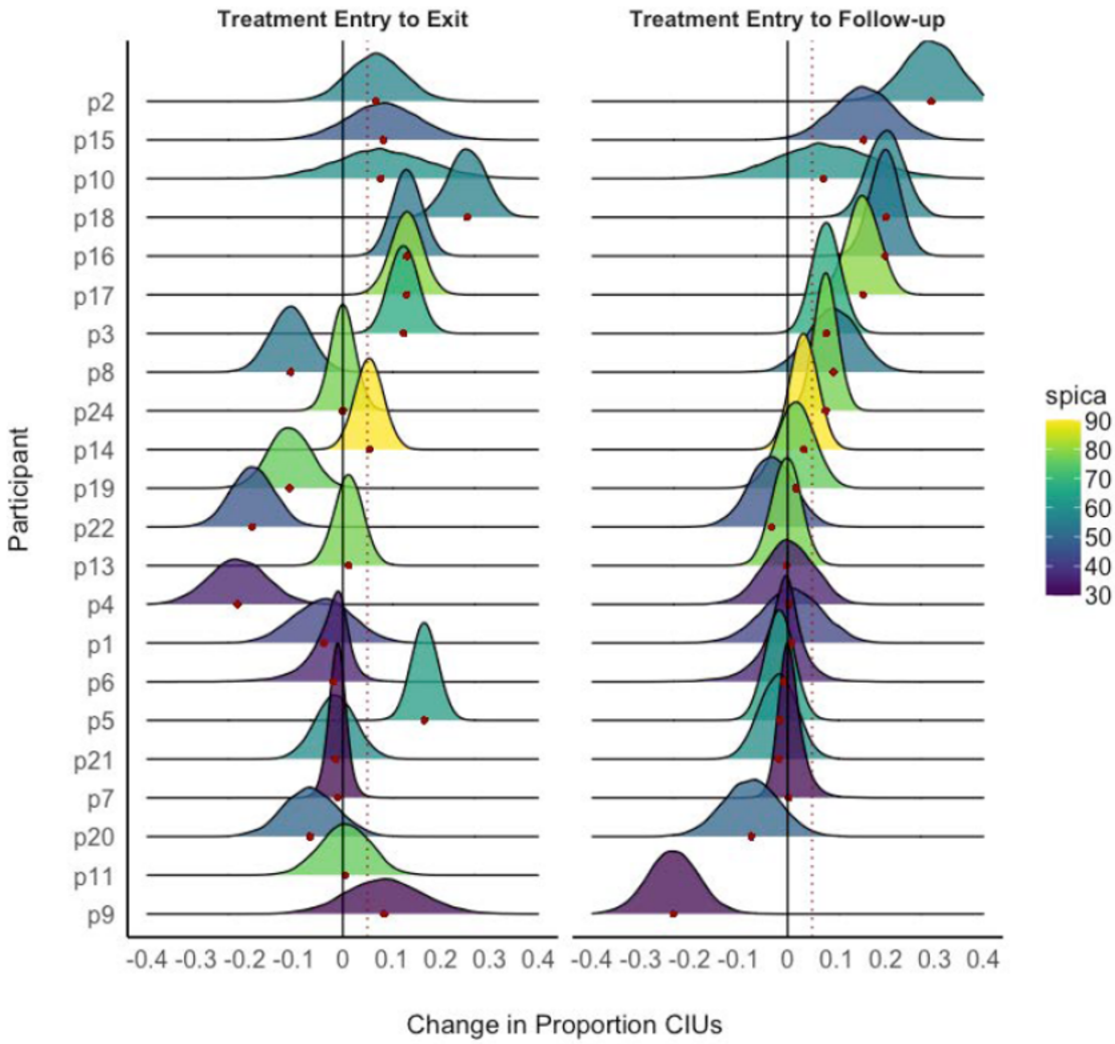

Figure 6. Individual Effect Estimates for Informativeness (Proportion CIUs) in Conversational Discourse. Note. Color indicates aphasia severity (SPICA), with lighter colors indicating milder and darker colors indicating more severe aphasia. Red dots indicate the median point estimate for each effect size.

Five participants responded positively to treatment in terms of efficiency in conversational discourse at exit and 6 did so at follow-up. Four participants at exit and three participants at follow-up showed clinically meaningful negative change. Two participants at exit and three at follow-up obtained effect sizes indicating no meaningful change. Eleven participants at exit and 10 participants at follow-up made changes that were uncertain. Individual estimates for changes in conversational efficiency are in Appendix C, with a visual representation in Figure 7.

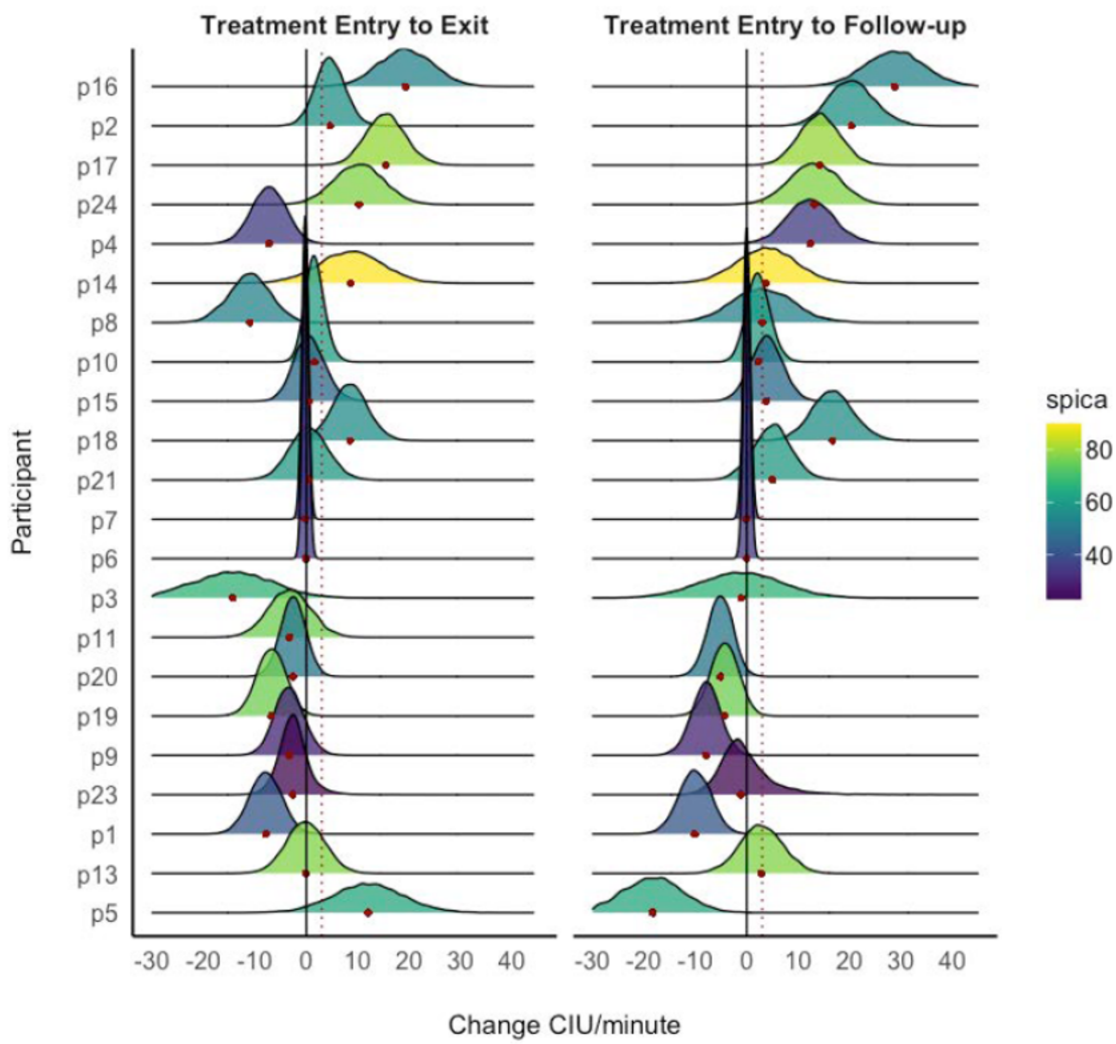

Figure 7. Individual Effect Estimates for Efficiency (CIUs Per Minute) in Conversational Discourse. Note. Color indicates aphasia severity (SPICA), with lighter colors indicating milder and darker colors indicating more severe aphasia. Red dots indicate the median point estimate for each effect size.

Discussion

In this study we examined the effects of group communication treatment on structured and conversational discourse tasks in adults with chronic aphasia. Of specific interest were changes in the informativeness and efficiency of structured and conversational discourse as a function of participation in the group communication therapy. We hypothesised that the participants' discourses would be more informative and efficient relative to their performance prior to entry in the treatment.

The fact that there was no evidence of improvement on either outcome measure for the DT group while they awaited treatment, despite their participation in a minimum of 3 hours of social activities per week in the interval, suggests that any changes at the outcome timepoints are related to participation in the communication group rather than to variability of the outcome measures or of the behaviors. The results at the group level revealed that structured discourses became more informative (as measured by the proportion of CIUs) and efficient (as measured by CIUs/min) after conversation-focused communication group treatment. Additionally, there is weak evidence that informativeness and efficiency in structured discourse was at least maintained or continued to improve from the end of treatment to the follow-up session. The group-level results for conversational discourse revealed that informativeness did not change from treatment entry to exit, but there was weak evidence of improvement at follow-up. There was no evidence of change to the efficiency of conversational discourse at either timepoint at the group level.

While it is not typical for improvement to continue after treatment has ended, this pattern has been reported previously. For example, Edmonds et al. (2014) reported that Verb Network Strengthening Treatment did not significantly improve control sentences at the end of treatment, but did significantly improve them at the maintenance timepoint 3 months after treatment ended. Silkes et al. (2018, 2020) reported that people with aphasia who received phonomotor treatment continued to improve three months after treatment ended. The latter research group speculated that improvement after treatment ends might reflect continued consolidation of the changes produced by treatment as participants use language in their daily lives. Our data are consistent with this idea.

It is important to recall that the communication treatment group did not explicitly focus on word retrieval, which is thought to be related to the informativeness of discourse production (Boyle, 2014). There were no drills or exercises aimed at improving word retrieval and participants were not asked to practice word retrieval tasks at home, although the group facilitator provided support for problematic word retrieval attempts as they occurred during group conversations. None of the participants were receiving concomitant speech-language therapy while they participated in the study. The change in the informativeness and efficiency of discourses, therefore, may be related to the rich, complex, supported environment of the communication groups, which provided multiple opportunities to retrieve words as the participants engaged in conversational activities focused on maintaining social connections and exchanging information with multiple partners. Thus, the results of this study can be considered as additional evidence for the theory that targeting more complex aspects of language can result in generalised improvement of simpler aspects (De Anda et al., 2020a, 2020b; Gierut, 2007; Kiran, 2007; Kiran et al., 20120; Thompson, 2007; Thompson & Shapiro, 2007), since providing repeated opportunities to engage in the complex activity of conversation resulted in improvement of more basic components (informativeness and efficiency). This result was greater in structured than in conversational discourse.

In this study, evidence of improved informativeness was more reliable in structured than in conversational discourse. This might seem surprising since we typically expect to see changes in the genre that more closely resembles the treatment context. However, during treatment participants received support from the facilitator via modeling, scaffolding, and explicit direction to try verbal and nonverbal strategies. The familiar conversational partners had not been trained in these supportive techniques and very few provided such support. Simmons-Mackie (2018b) suggested that, based on communication accommodation theory, when speaking partners fail to use or positively orient to supportive strategies, people with aphasia are likely to accommodate to the standard communication style of the partner. This might explain why we did not obtain strong evidence of improvement in conversational discourse with the familiar partner. It also highlights the importance of training conversational partners in supportive communication techniques. It is also possible that the lack of observed improvement in communication is related to the CIU measure, which does not capture alternative methods of communication.

These results add to the limited evidence of improved informativeness following conversation-based group treatment provided by Hoover et al. (2020). They reported that two of five participants with severe aphasia demonstrated such improvement after participating in either dyadic or large-group conversation treatment. The results also align with findings of Mason and colleagues (2020), who reported that one of their three participants demonstrated improved word retrieval in discourse after participating in group treatment.

In addition to the finding that treating at the level of conversation resulted in cascading generalisation to the less complex elements of informativeness and efficiency, the results of this study have clinical implications. Given the decline in the number of funded sessions for aphasia treatment (Simmons-Mackie, 2018a), less expensive group treatment options have become an increasingly popular service delivery approach (Elman, 2007, 2016). For clinicians who are reluctant to move from an impairment-based form of treatment to a participation-focused form of treatment (Laliberté et al., 2016; Simmons-Mackie, 2018; Torrence et al., 2016), this study adds to previous evidence that conversation-based communication treatment groups result in improvements in the language impairment without directly focusing treatment on the impairment level (DeDe et al., 2019; Elman & Bernstein-Ellis, 1999a; Hoover et al., 2020; Wertz et al., 1980).

As expected, aphasia severity was strongly related to informativeness and efficiency of structured and conversational discourses at treatment entry. For informativeness of structured discourse, there was no evidence of an interaction between severity and change across timepoints, suggesting that participants improved regardless of aphasia severity. For informativeness of conversational discourse, the interaction of severity and change at both timepoints suggests that the small improvements were greater for people with mild aphasia. The relatively small effect sizes and estimated uncertainty warrant further investigation. Regarding aphasia severity and efficiency, there is modest evidence that people with more severe aphasia improved more than people with milder aphasia in structured discourse. In conversational discourse, efficiency improved more for people with milder aphasia at both timepoints, although the improvement was very modest. The reason for these contrasting findings is not clear. We can speculate that people with more severe aphasia might have done better in structured discourse because, as discussed in the introduction to this paper, it is less cognitively and communicatively complex than conversational discourse. However, why people with severe aphasia improved more than people with milder aphasia in the efficiency with which they produced structured discourse is not clear and could be a fruitful focus of future research.

Thus far this discussion has focused on the changes that occurred at the group level. As with most aphasia treatment studies, we found that there was wide variability in response to the treatment, with some participants demonstrating robust and meaningful positive change and a smaller number showing clear declines in performance. There were also many participants who improved, but whose effect sizes were not reliably outside the ROPE limits that we set, rendering their response to treatment uncertain. Table 8 summarises the treatment response of each participant on each outcome variable in each discourse genre. Examining the individual responses to treatment can help to clarify the group-level results and might serve to set realistic expectations about treatment outcomes.

Only one participant, P18, demonstrated robust, clinically meaningful change on both

Summary of Individual Treatment Response Across Outcome Variables and Discourse Genres at each Timepoint

Participant	Timepoint	Informativeness Structured Discourse	Efficiency Structured Discourse	Informativeness Conversational Discourse	Efficiency Conversational Discourse
P1	exit	uncertain	no response	uncertain	negative response
P1	follow-up	uncertain	no response	uncertain	negative response
P2	exit	uncertain	uncertain	uncertain	uncertain
P2	follow-up	uncertain	uncertain	positive response	positive response
P3	exit	uncertain	uncertain	positive response	negative response
P3	follow-up	uncertain	uncertain	uncertain	uncertain
P4	exit	uncertain	positive response	negative response	uncertain
P4	follow-up	uncertain	positive response	uncertain	positive response
P5	exit	uncertain	uncertain	positive response	positive response
P5	follow-up	positive response	positive response	uncertain	negative response
P6	exit	uncertain	no response	uncertain	no response
P6	follow-up	uncertain	no response	uncertain	no response
P7	exit	no response	no response	no response	no response
P7	follow-up	no response	no response	no response	no response
P8	exit	uncertain	uncertain	negative response	negative response
P8	follow-up	uncertain	uncertain	uncertain	uncertain
P9	exit	no response	no response	uncertain	uncertain
P9	follow-up	no response	no response	negative response	negative response
P10	exit	uncertain	no response	uncertain	uncertain
P10	follow-up	positive response	uncertain	uncertain	uncertain
P11	exit	positive response	uncertain	uncertain	uncertain
P11	follow-up	uncertain	uncertain		
P13	exit	no response	uncertain	no response	uncertain
P13	follow-up	uncertain	uncertain	no response	uncertain
P14	exit	uncertain	uncertain	uncertain	uncertain
P14	follow-up	uncertain	positive response	uncertain	uncertain
P15	exit	uncertain	no response	uncertain	uncertain
P15	follow-up	positive response	positive response	positive response	uncertain
P16	exit	no response	uncertain	positive response	positive response
P16	follow-up	positive response	positive response	positive response	positive response
P17	exit	no response	no response	positive response	positive response
P17	follow-up	negative response	uncertain	positive response	positive response
P18	exit	positive response	positive response	positive response	positive response
P18	follow-up	positive response	positive response	positive response	positive response
P19	exit	uncertain	no response	negative response	negative response
P19	follow-up	uncertain	uncertain	uncertain	uncertain
P20	exit	uncertain	uncertain	uncertain	uncertain
P20	follow-up	uncertain	no response	uncertain	uncertain
P21	exit	uncertain	uncertain	uncertain	uncertain
P21	follow-up	uncertain	uncertain	uncertain	uncertain
P22	exit	uncertain	uncertain	negative response	uncertain
P22	follow-up	uncertain	no response	uncertain	uncertain
P23	exit	uncertain	no response		
P23	follow-up	uncertain	no response		
P24	exit	no response	uncertain	no response	positive response
P24	follow-up	no response	uncertain	uncertain	positive response

Note: Positive response = changes reliably exceeded the ROPE in a positive direction. No response = any change reliably fell within the ROPE. Uncertain = changes were not reliably outside the ROPE. Negative response = changes reliably exceeded the ROPE in a negative direction. Empty cells = missing data (P11 did not attend the follow-up session for medical reasons; P23 is not included in the analysis of conversational discourse because her partner did not attend the exit or follow-up assessment sessions).

outcome variables in both discourse genres at exit and follow-up. This participant was relatively early post-stroke (7 months post-onset). The next most successful participant in terms of treatment response was P16, who showed meaningful positive change on both informativeness and efficiency of structured discourse at follow-up (but not at exit). He demonstrated meaningful positive change on both outcome variables at both timepoints in conversational discourse. P16 was also 7 months post-onset. Both of these participants were the earliest post-onset in the group (range: 7 to 336 months), suggesting that time post onset might influence overall improvement across the outcome variables and discourse genres.

Only one participant, P7, showed clear evidence of no positive change on both outcome variables and discourse genres at exit and follow-up. P7's aphasia was among the most severe (SPICA: 30th percentile; WAB AQ: 24.2) of the group (SPICA percentile range: 23rd to 90th; WAB AQ range: 13.1 to 94.3), and she had severe apraxia of speech. Two participants who demonstrated no change at both timepoints on one of the discourse genres (P9 in structured discourse at both timepoints) or on one of the outcome variables (P6: no change in efficiency in both discourse genres at both timepoints) were also among those with the most severe aphasia and with severe apraxia of speech. P9 scored at the 30th percentile on the SPICA and had a WAB AQ of 18.9, and P6 scored at the 35th percentile of the SPICA, with a WAB AQ of 13.1. This might suggest that despite the group results, which suggested that for the majority of participants severity did not influence treatment response in structured discourse, people with the most severe aphasia and apraxia of speech might not improve spoken discourse informativeness or efficiency by participating in conversationally based communication group treatment. Of the 10 participants who had apraxia of speech, only three (P2 with mild apraxia of speech, and P4 and P15 with moderate apraxia of speech) demonstrated clinically meaningful change after treatment. It may be that there is a minimal level of spoken language ability necessary for improvement of spoken discourse production via participation in conversation-based communication groups. Additional research exploring this possibility would be informative, and as a result of these findings, evaluating how the presence and severity of apraxia of speech moderate discourse outcomes will be a focus of our larger project examining discourse outcomes from the Elman and Bernstein-Ellis (1998a,b) study. It is possible that the individuals with very severe aphasia and apraxia of speech improved their communication abilities in ways that this study did not measure. For example, they may have increased their use of non-verbal communication strategies, like gesturing, drawing, or using writing to supplement speech. These augmentative and alternative modes of communication were encouraged and supported as part of the treatment. Examining the possibility that these non-verbal aspects of communication changed is a focus of another part of the larger project.

It is less clear whether demographic or test variables might have influenced participants who showed performance declines on some variables. These participants included people who spanned the range of chronicity of aphasia and of aphasia severity. In fact, of the eight participants with a negative response on at least one outcome measure, four (P3, P4, P5, and P17) showed clear positive response on one or more of the other outcome measures. Therefore, it is difficult to make any general statements about what might underlie a negative response to treatment in this study.

The majority of the participants responded to the treatment in a way that did not allow them to be judged unambiguously as either demonstrating or failing to demonstrate change. This is because the changes that they made did not unequivocally exceed the ROPE limits that we set. Though we based the ROPE limits on our clinical and research experience, these limits are somewhat arbitrary. Another team of investigators might establish a different ROPE. Nevertheless, this result suggests that the group-level positive outcomes for informativeness and efficiency in structured discourse were driven more by the number of participants showing uncertain response than by the number showing an unambiguous positive response. In conversational discourse, the mixed results for informativeness and the null results for efficiency seem related to the number of participants who had negative treatment responses in this discourse genre. As we try to use results of RCTs to make treatment decisions for people with aphasia, it is useful to remember that positive group-level outcomes might not have resulted in clinically meaningful differences for some (or even most) participants, and that null group-level outcomes might mask positive responses by some participants. If the reporting of individual effects in RCTs became standard practice, the field might draw closer to discovering which treatments are best for which patients in which conditions.

This study, which is part of a larger project, examined only two aspects of the many communication abilities that might have changed as a result of the treatment. Other project studies will examine outcomes at the sentence level, in discourse macrostructure, in nonverbal communication, and in conversational behaviors. We hope this approach will allow us to build on the knowledge that the current study provides, leading to a fuller understanding of the treatment's outcomes. The treatment was provided for a relatively long period (4 months) at a relatively high intensity (two 2.5 hour sessions per week), resulting in 160 hours of treatment. It is not clear to what degree the treatment dosage contributed to any positive outcomes in this study, and it is possible that shorter or less intense treatment exposure might yield different results. The treatment was administered by a speech-language pathologist who was an experienced aphasia clinician, and this might also have influenced treatment outcome. Finally, the group of participants with aphasia in this study spanned the aphasia severity range from very severe to very mild aphasia. It is likely that confining the treatment to a more homogeneous group of participants in terms of severity would lead to more robust, or at least more easily interpretable, outcomes. However, such a restricted sample would be less representative of the population of individuals with aphasia than the sample in this study.

Conclusion

Participation in conversation-based communication group treatment was associated with more informative and efficient structured discourse production and modestly improved informativeness in conversational discourse for the group of participants in this study. This supports the notion that treating in the context of a complex language activity (a group conversational exchange) can result in cascading generalisation to simpler aspects of language (discourse informativeness and efficiency). It also adds to evidence that focusing treatment at the level of participation can improve performance at the impairment level without specifically targeting such improvement. The weaker findings for conversational than for structured discourse observed in this study could be due to the inability of the CIU measure to capture alternative communication methods or to the lack of support provided by the untrained conversation partners. Examining the effects of treatment on individual participants provided additional insight about the group-level outcomes and provided some clues about factors that might have influenced performance for some of the participants. This study explored the effects of the treatment on two of many aspects of communication that might have been affected by the treatment. Future reports will explore the effect of conversationally-based treatment on other aspects of communication.

Acknowledgments

We thank Ellen Bernstein-Ellis, M.A., CCC-SLP and the people with aphasia who participated in the study. Thanks also to the following students who contributed to various phases of this project: Alexa Bergenfeld, Alexandra Buchanan, Abby Cohen, Anne DeCaro, Vanessa Gomez, Sophia Hendrix, Sara Manzon, and Carly Truglio.

Declaration of Interest: The original RCT was partially funded by a grant from the National Easter Seals Research Program (awarded to Roberta J. Elman). This study was partially funded by the Aphasia Center of California Research Fund.

References

Attard, M. C., Lanyon, L., Togher, L., & Rose, M. L. (2015). Consumer perspectives on community aphasia groups: A narrative literature review in the context of psychological well-being. *Aphasiology, 29*(8), 983-1019.

Beckner, C., Blythe, R., Bybee, J. Christiansen, M.H., Croft, W., Ellis, N.C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position Paper. In Ellis, N.C. & Larsen-Freeman, D. (Eds.). *Language as a complex adaptive system.* John Wiley & Sons.

Bernstein-Ellis, E. & Elman, R.J. (2007). Aphasia group communication treatment: The Aphasia Center of California approach. In Elman, R.J. (Ed.). *Group treatment of neurogenic communication disorders: The expert clinician's approach* (2nd ed., pp. 71-94). Plural Publishing.

Bradburn, N. (1969). *The structure of psychological well-being.* Aldine.

Bryant, L., Ferguson, A., & Spencer, E. (2016a). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics, 30*(7), 489-518. http://dx.doi.org/10.3109/02699206.2016.1145740

Bryant, L., Spencer, E., & Ferguson, A. (2016b). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology, 31*(10), 1105-1126. https://dx.doi.org/10.1080/02687038.2016.1239013

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1). https:// doi:10.18637/jss.v076.i01

Davis, G.A. (2014). *Aphasia and related cognitive-communicative disorders.* Boston, MA: Pearson.

De Anda, S., Blossom, M., & Abel, A. D. (2020a). A complexity approach to treatment of tense and agreement deficits: A case study. *Communication Disorders Quarterly, 41*(4), 250–260. https://doi.org/10.1177/1525740118822477

De Anda, S., Blossom, M., & Abel, A. D. (2020b). Cross-morpheme generalization using a complexity approach in school-age children. *Journal of Speech, Language & Hearing Research, 63*(10), 3501–3524. https://doi.org/10.1044/2020_JSLHR-19-00173

DeDe, G., Hoover, E., & Maas, E. (2019). Two to tango or the more the merrier: A randomized controlled trial of the effects of group size in aphasia conversation treatment on standardized tests. *Journal of Speech and Hearing Research, 62*, 1437-1451.

Dietz, A. & Boyle, M. (2018). Discourse measurement in aphasia research: Have we reached the tipping point? *Aphasiology, 32*(4), 459-464. https://doi.org/10.1080/02687038.2017.1398803

Disimoni, F.G., Keith, R.L., & Darley, F.L. (1980). Prediction of PICA overall score by short version of the test. *Journal of Speech and Hearing Research,23*, 511-516.

Duffy, J.R. (2020). *Motor speech disorders: Substrates, differential diagnosis, and management, fourth edition.* Elsevier.

Edmonds L A, Mammino K, and Ojeda J. (2014). Effect of Verb Network Strengthening Treatment (VNeST) in persons with aphasia: Extension and replication of previous findings. *American Journal of Speech-Language Pathology, 23*(2 S312-S329), 10.1044/2014_AJSLP-13-0098

Elman, R. J. (2007). The importance of aphasia group treatment for rebuilding community and health. *Topics in Language Disorders, 27*(4), 300-308.

Elman, R. J. (2016). Aphasia centers and the life participation approach to aphasia: A paradigm shift. *Topics in Language Disorders, 36*(2), 154-167.

Elman, R.J. (September, 2018). Integrating the Life Participation Approach to Communication Disorders with Impairment-based Disorders: Making Connections. Invited keynote address to the 13th Annual Eleanor M. Saffran Conference on Cognitive Neuroscience & Rehabilitation of Communication Disorders, Temple University, Philadelphia, Pennsylvania.

Elman, R. J. (2021). C.A.P.E.: A Checklist of Four Essential and Evidence-based Categories for Aphasia Intervention. In A.L. Holland & R.J. Elman (Eds). *Neurogenic Communication Disorders and the Life Participation Approach: The Social Imperative in Supporting Individuals and Families* (pp. 21-52). Plural Publishing.

Elman, R. J. & Bernstein-Ellis, E. (1999a). The efficacy of group communication treatment in adults with chronic aphasia. *Journal of Speech Language and Hearing Research, 42*, 411-419. https://doi.org/10.1044/jslhr.4202.411

Elman, R. J. & Bernstein-Ellis, E. (1999b). Psychosocial aspects of group communication treatment: Preliminary findings. *Seminars in Speech and Language, 20*(1), 65-72.

Gamer, M. Lemon, J., Fellows, I. & Singh, P. (2019). irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.

Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

German, D.J. (1991). *Test of Word Finding in Discourse (TWFD): Administration, scoring, interpretation, and technical manual.* Pro-Ed.

Gierut, J. A. (2007). Phonological complexity and language learnability. *American Journal of Speech-Language Pathology. 16*(1), 6-17. https://doi.org/10.1044/1058-0360(2007/003)

Hallowell, B. (2017). *Aphasia and other acquired neurogenic language disorders: A guide for clinical excellence.* Plural Publishing.

Hartig, F. (2020). DHARMa: *Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models.* https://CRAN.R-project.org/package=DHARMa

Holland, A. (1980). *Communicative abilities in daily living.* University Park Press.

Hoover, E., DeDe, G., & Maas, E. (2018). Effects of group size on conversation treatment outcomes: Results of standardized testing. *Aphasiology*, 32(S1), 93-95. https://doi.org/10.1080/02687038.2018.1487004

Hoover, E., McFee, A., & DeDe, G. (2020). Efficacy of group conversation intervention in individuals with severe profiles of aphasia. *Seminars in Speech and Language, 41*, 71-82. https://doi.org/10.1055/s-0039-3400991.

Kertesz, A. (1982). *The Western Aphasia Battery.* Grune & Stratton.

Kiran, S. (2007). Complexity in the treatment of naming deficits. *American Journal of Speech-Language Pathology, 16*(1), 18-29. https://doi.org/10.1044/1058-0360(2007/004)

Kiran, S., Caplan, D., Sandberg, C., Levy, J., Berardino, A., Ascenso, E., Villard, S., & Tripodis, Y. (2012). Development of a theoretically based treatment for sentence comprehension deficits in individuals with aphasia. *American Journal of Speech-Language Pathology, 21*(2), S88–S102. https://doi.org/10.1044/1058-0360(2012/11-0106)

Koo, T.K. & Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15,* 155-163. http://dx.doi.org/10.1016/j.jcm.2016.02.012

Laliberté, M-P., Gauvrau, C.A., & Le Dorze, G. (2016). A pilot study on how speech-language pathologists include social participation in aphasia rehabilitation. *Aphasiology, 30*(10), 1117-1133. https://doi.org/10.1080/02687038.2015.1100708

Lanyon, L. E., Rose, M. L., & Worrall, L. (2013). The efficacy of outpatient and community-based aphasia group interventions: A systematic review. *International Journal of Speech-Language Pathology, 15*(4), 359-374. https://doi.org/ 10.3109/17549507.2012.752865

Larsen-Freeman, D. & Cameron, L. (2008). *Complex systems and applied linguistics.* Oxford University Press.

Leaman, M.C. & Edmonds, L.A. (2019). Revisiting the correct information unit: Measuring informativeness in unstructured conversations in people with aphasia. *American Journal of Speech-Language Pathology, 28,* 1099-1114. https://doi.org/ 10.1044/2019_AJSLP-18-0268

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis, 100*(9), 1989-2001.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Development and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders, 54*, 113-124.

LPAA Project Group (2000). (Authors in alphabetical order: Chapey, R., Duchan, J., Elman, R., Garcia, L., Kagan, A., Lyon, J., & Simmons-Mackie, N.). Life Participation Approach to Aphasia: A statement of values for the future. *ASHA Leader 5*(3), 4-6. https://doi.org/10.1044/leader.FTR.05032000.4

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

Makowski D., Ben-Shachar M.S., & Lüdecke D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *The Journal of Open Source Software, 4*(40), 1541. https://doi.org/10.21105/joss.01541

Mason, C., Nickels, L., & McDonald, B. (2020). An exploration of the impact of group treatment for aphasia on connected speech. *Journal of the International Neuropsychological Society, 26,* 72-85. https://doi.org/10.1017/S1355617719001012

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan.* CRC press.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods,* 16(4), 406. https://doi.org/10.1037/a0024377

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P. C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research, 62*(5), 1225-1242.

National Aphasia Association. https://aphasia.org

Nicholas, L., & Brookshire, R. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36,* 338–350. https://doi.org/10.1044/1092-4388(2012/12-0065)

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measure in aphasia. *International Journal of Language and Communication Disorders, 52*(6), 689-732.

Sandberg, C. & Kiran, S. (2014) How justice can affect jury: Training abstract words promotes generalisation to concrete words in patients with aphasia. *Neuropsychological Rehabilitation, 24*(5), 738-769. http://dx.doi.org/10.1080/09602011.2014.899504

Silkes, J.P., Fergadiotis, G., Graue, K., & Kendall, D.L. (2021). Effects of phonomotor treatment and semantic feature analysis on discourse production. *American Journal of Speech-Language Pathology. 30*(1S), 441-454. https://doi.org/10.1044/2020_AJSLP-19-00111

Silkes, J.P., Fergadiotis, G., Hunting Pompon, R., Torrence, J., & Kendall, D.L. (2018). Effects of phonomotor treatment on discourse production. *Aphasiology, 33*(2), 125-129. https://doi.org/

10. 1080/02687038.2018.1512080

Simmons-Mackie, N. (2018a). *Aphasia in North America: Frequency, demographics, impact of aphasia, communication access, services and service gaps.* https://www.aphasiaaccess.org/white-papers

Simmons-Mackie, N. (2018b). Communication partner training in aphasia: Reflections on communication accommodation theory. *Aphasiology, 32*(10), 1215-1224. https://doi.org/10.1080/02687038.2018.1428282

Simmons-Mackie, N., Elman, R.J., Holland, A.L., & Damico, J.S. (2007). Management of discourse in group therapy for aphasia. *Topics in Language Disorders, 27*(1), 5-23.

Simmons-Mackie, N., Savage, M.C., Worrall, L. (2014). Conversation therapy for aphasia: A qualitative review of the literature. *International Journal of Language and Communication Disorders, 49*(5), 511-526. https://doi.org/10.1111/1460-6984.12097

Stan Development Team (2020). "RStan: The R interface to Stan." R package version 2.21.2, http://mc-stan.org

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American journal of speech-language pathology, 28*(3), 1067-1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Thompson, C.K. (2007). Complexity in language learning and treatment. *American Journal of Speech-Language Pathology, 16*, 3-5.

Thompson, C. K., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology, 16*(1), 30-42. https://doi.org/10.1044/1058-0360(2007/005)

Torrence, J.M., Baylor, C.R., Yorkston, K.M., and Spencer, K.A. (2016) Addressing communicative participation in treatment planning for adults: A survey of U.S. speech-language pathologists. *American Journal of Speech-Language Pathology, 25* (3) 355-370. doi:10.1044/2015_AJSLP-15-0049

van der Gaag, Davis, S., Moss, B., Cornelius, V., Laing, S., & Mowles, C. (2015). Therapy and support services for people with long-term stroke and aphasia and their relatives: A six-month follow-up study. *Clinical Rehabilitation, 19*, 372-380.

Vickers, C.P. (2010). Social networks after the onset of aphasia: The impact of aphasia group attendance. *Aphasiology, 24*(6-8), 902-913. https://doi.org/10.1080/02687030903438532

Wallace, S., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., Kong, A., Simmons-Mackie, N., Scarinci, N. & Gauvreau, C. (2017a). Which outcomes are most important to people with aphasia and their families? An international nominal group technique study framed within the ICF. *Disability and Rehabilitation, 39*(14), 1364-1379. https://doi.org/10.1080/09638288.2016.1194899

Wallace, S.J., Worrall, L., Rose, T., & Le Dorze, G. (2017b). Which treatment outcomes are most important to aphasia clinicians and managers? An international e-Delphi consensus study, *Aphasiology, 31*(6), 643-673. https://doi.org/10.1080/02687038.2016.1186265

Webster, J. & Morris, J. (2019). Communicative informativeness in aphasia: Investigating the relationship between linguistic and perceptual measures. *American Journal of Speech-Language Pathology, 28*, 1115–1126. https://doi.org/10.1044/2019_AJSLP-18-0256

Wertz, R.T., Collins, M.J., Weiss, D., Kurtzke, J.F., Friden, T., Brookshire, R.H., Pierce, J., Holtzapple, P., Hubbard, D.J., Porch, B.E., West, J.A., Davis, L., Matovitch, V., Morley, G.K., Resurreccion, E. (1981). Veterans Administration cooperative study on aphasia: A comparison of individual and group treatment. *Journal of Speech & Hearing Research, 24*(4), 580–594. https://doi.org/10.1044/jshr.2404.580

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology, 25,* 309-322.

Appendix A

Model Fitting Details

The following describes the model fitting and evaluation procedures for the Bayesian generalized mixed-effects models conducted for this study. We refer the reader to Nalborczyk et al. (2019) for an excellent and accessible tutorial on the use, evaluation, and interpretation of Bayesian mixed-effects models in the context of communication science and disorders.

A binomial probability distribution with a logit link function was used to model the proportion of words that were CIUs. CIUs per minute were initially modelled as a continuous rate variable using a lognormal distribution with an identity link function. However, this approach did not provide adequate model fit as evidenced by inadequate posterior predictive plots. Subsequently, CIUs were modelled using a Poisson probability distribution with an additional offset variable to adjust for varying time in each observation (Gelman & Hill, 2006). For both the proportion of CIUs and CIUs per minute, models were evaluated for overdispersion via simulation using the R package DHARMa (Hartig, 2020); overdispersion was not present in either case ($p > .05$).

For each model, four independent Hamiltonian Markov Chain Monte Carlo (MCMC) chains were run with 4000 iterations. The initial 1000 iterations were discarded as warmup and not included in parameter estimation. Models were run with weakly informative, regularizing priors to improve sampling efficiency. Prior predictive checks were used to ensure priors reflected a reasonable sampling space. The priors on the intercept and beta-coefficients were student's t-distributions with 3 degrees of freedom, location of 0 and scale of 2. The location is defined by BRMS as the median of the estimate of interest; the priors used were centered around a 50/50 odds that a word would be a CIU and zero change, and zero effect of severity. With a scale parameter of 2; most of the prior distributions fell within +/- 5 logits of the location parameter reflecting a range of reasonable values for models using a logit-link function. The standard deviation was modelled with a half-normal distribution with a standard deviation of 2 and mean of zero. The default LKJ (Lewandowski, Kurowicka, and Joe) prior (Lewandowski et al., 2009) with = 1 was utilized for the correlation parameter. With 1, extreme correlation values become less likely.

Prior to interpreting the model results, we assessed the model for convergence by ensuring the split-half potential scale reduction factor ($\hat{R}$), was less than 1.01 and the effective samples were greater than 400 for each parameter estimate (Gelman et al., 2013). The $\hat{R}$ statistic can be interpreted as the ratio of the variance within each chain to the variance pooled across chains, with values exceeding 1.01 indicating lack of convergence and unhealthy MCMC mixing. The number of effective samples is the number of independent draws that were not autocorrelated with previous draws. Therefore, the greater number of effective samples, the more stable the distributions will be from which parameter estimates are derived. In all the models the estimated $\hat{R}$ values were less than 1.01, and the number of effective sample sizes exceeded 400 (min > 1000) for all parameters. We therefore concluded that each parameter estimate had adequately converged on the expectation values.

Next, we conducted posterior predictive checks to confirm the models adequately fit the data. Posterior predictive checking is the process of estimating many distributions of estimated parameters and comparing these to the distribution of the observed data. We estimated 1,000 predictive distributions of each dependent variable and then compared these to our observed dependent variable. The estimated distributions closely followed the pattern of the density distribution for the observed data. Thus, we concluded that the estimated models adequately fit the data.

Appendix B

Individual Effect Size Estimates for Structured Discourse

Table B1. Individual Effect Size Estimates for Change in Informativeness (Proportion CIUs) at Exit and Follow-up in Structured Discourse

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P1	exit	0.11	0.03	0.20	0.12	uncertain
P1	follow-up	-0.04	-0.13	0.04	0.51	uncertain
P2	exit	0.05	-0.03	0.12	0.50	uncertain
P2	follow-up	0.01	-0.08	0.10	0.62	uncertain
P3	exit	0.06	0.01	0.11	0.42	uncertain
P3	follow-up	0.04	0.00	0.09	0.62	uncertain
P4	exit	0.02	-0.06	0.10	0.63	uncertain
P4	follow-up	-0.01	-0.09	0.07	0.66	uncertain
P5	exit	0.03	-0.03	0.08	0.74	uncertain
P5	follow-up	0.16	0.10	0.21	0.00	positive response
P6	exit	0.01	-0.08	0.09	0.69	uncertain
P6	follow-up	-0.02	-0.11	0.05	0.66	uncertain
P7	exit	0.01	0.00	0.02	1.00	no response
P7	follow-up	0.01	0.00	0.03	1.00	no response
P8	exit	-0.02	-0.09	0.04	0.70	uncertain
P8	follow-up	-0.01	-0.08	0.06	0.75	uncertain
P9	exit	0.01	0.00	0.03	0.99	no response
P9	follow-up	0.02	0.00	0.05	0.93	no response
P10	exit	0.04	-0.06	0.15	0.46	uncertain
P10	follow-up	0.14	0.03	0.25	0.10	positive response
P11	exit	0.19	0.13	0.26	0.00	positive response
P11	follow-up	0.15	-0.05	0.34	0.15	uncertain
P13	exit	-0.01	-0.05	0.04	0.90	no response
P13	follow-up	0.01	-0.04	0.06	0.86	uncertain
P14	exit	0.04	-0.02	0.09	0.63	uncertain
P14	follow-up	0.06	0.01	0.11	0.40	uncertain
P15	exit	0.07	0.00	0.14	0.33	uncertain
P15	follow-up	0.24	0.16	0.32	0.00	positive response
P16	exit	-0.01	-0.04	0.01	0.98	no response
P16	follow-up	0.08	0.05	0.11	0.03	positive response
P17	exit	0.01	-0.02	0.05	0.92	no response
P17	follow-up	-0.08	-0.12	-0.05	0.06	negative response
P18	exit	0.28	0.23	0.32	0.00	positive response
P18	follow-up	0.20	0.16	0.24	0.00	positive response
P19	exit	0.01	-0.05	0.07	0.82	uncertain
P19	follow-up	0.05	0.00	0.11	0.45	uncertain

Table B1. Individual Effect Size Estimates for Change in Informativeness (Proportion CIUs) at Exit and Follow-up in Structured Discourse (continued)

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P20	exit	0.05	-0.01	0.12	0.47	uncertain
P20	follow-up	0.07	0.00	0.15	0.31	uncertain
P21	exit	0.02	-0.04	0.09	0.72	uncertain
P21	follow-up	0.07	0.01	0.13	0.28	uncertain
P22	exit	0.05	-0.01	0.10	0.52	uncertain
P22	follow-up	0.01	-0.05	0.07	0.82	uncertain
P23	exit	0.03	-0.02	0.10	0.74	uncertain
P23	follow-up	0.04	-0.01	0.15	0.59	uncertain
P24	exit	-0.02	-0.05	0.02	0.93	no response
P24	follow-up	0.00	-0.03	0.04	0.98	no response

Note: Note: Lower90 and upper90 refer to bounds on 90% highest density interval. ROPE = region of practical equivalence. ROPE allows researchers to define a range of responses that are considered to be equivalent to the frequentist null hypothesis The values in the ROPE column are the percentage of the Credible Interval that is within the ROPE. The lower the value, the more confident one can be that the participant's response is outside the ROPE.

Table B2. Individual Effect Size Estimates for Change in Efficiency (CIUs per minute) at Exit and Follow-up in Structured Discourse

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P1	exit	1.31	-0.51	3.10	0.98	no response
P1	follow-up	0.08	-1.77	1.87	1.00	no response
P2	exit	2.47	-0.78	5.57	0.63	uncertain
P2	follow-up	0.54	-3.26	4.38	0.88	uncertain
P3	exit	0.99	-6.23	8.36	0.56	uncertain
P3	follow-up	6.11	-0.95	12.97	0.20	uncertain
P4	exit	13.35	6.68	22.37	0.00	positive response
P4	follow-up	20.23	11.34	31.36	0.00	positive response
P5	exit	0.67	-5.82	6.98	0.62	uncertain
P5	follow-up	21.63	13.74	29.93	0.00	positive response
P6	exit	0.75	-0.61	2.36	1.00	no response
P6	follow-up	0.19	-1.74	1.95	1.00	no response
P7	exit	0.14	0.00	0.40	1.00	no response
P7	follow-up	0.13	-0.05	0.50	1.00	no response
P8	exit	-1.15	-8.38	5.54	0.58	uncertain
P8	follow-up	2.25	-5.31	9.79	0.49	uncertain
P9	exit	0.22	-0.02	0.59	1.00	no response
P9	follow-up	0.45	0.06	1.27	1.00	no response
P10	exit	0.98	-0.76	2.63	1.00	no response
P10	follow-up	3.26	0.93	5.80	0.42	uncertain

Table B2. Individual Effect Size Estimates for Change in Efficiency (CIUs per minute) at Exit and Follow-up in Structured Discourse (continued)

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P11	exit	0.41	-4.52	5.16	0.77	uncertain
P11	follow-up	3.00	-15.00	30.00	0.25	uncertain
P13	exit	3.11	-2.77	8.89	0.49	uncertain
P13	follow-up	2.14	-3.98	8.26	0.57	uncertain
P14	exit	-1.32	-8.99	6.53	0.50	uncertain
P14	follow-up	14.11	5.43	23.19	0.00	positive response
P15	exit	1.52	-0.20	3.22	0.97	no response
P15	follow-up	6.02	3.48	8.81	0.00	positive response
P16	exit	1.96	-0.21	4.12	0.81	uncertain
P16	follow-up	11.87	9.14	14.65	0.00	positive response
P17	exit	0.97	-2.11	4.12	0.90	no response
P17	follow-up	-4.52	-7.41	-1.57	0.16	uncertain
P18	exit	8.90	6.35	11.57	0.00	positive response
P18	follow-up	9.29	6.77	11.93	0.00	positive response
P19	exit	0.16	-2.54	2.83	1.00	no response
P19	follow-up	1.44	-1.43	4.30	0.86	uncertain
P20	exit	1.97	0.37	3.70	0.89	uncertain
P20	follow-up	0.56	-1.10	2.14	1.00	no response
P21	exit	-0.49	-4.77	3.77	0.83	uncertain
P21	follow-up	2.66	-1.64	6.97	0.56	uncertain
P22	exit	2.87	0.74	5.05	0.55	uncertain
P22	follow-up	1.03	-1.36	3.38	0.96	no response
P23	exit	0.21	0.01	0.61	1.00	no response
P23	follow-up	0.33	0.00	1.32	1.00	no response
P24	exit	3.01	-2.79	8.65	0.49	uncertain
P24	follow-up	4.66	-0.88	10.23	0.28	uncertain

Note: Lower90 and upper90 refer to bounds on 90% highest density interval. ROPE = region of practical equivalence. ROPE allows researchers to define a range of responses that are considered to be equivalent to the frequentist null hypothesis The values in the ROPE column are the percentage of the Credible Interval that is within the ROPE. The lower the value, the more confident one can be that the participant's response is outside the ROPE.

Appendix C

Individual Effect Size Estimates for Conversational Discourse

Table C1. Individual Effect Size Estimates for Change in Informativeness (Proportion CIUs) at Exit and Follow-up in Conversational Discourse

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P1	exit	-0.04	-0.14	0.06	0.50	uncertain
P1	follow-up	0.01	-0.09	0.11	0.59	uncertain
P2	exit	0.07	-0.02	0.16	0.36	uncertain
P2	follow-up	0.29	0.20	0.38	0.00	positive response
P3	exit	0.12	0.08	0.17	0.00	positive response
P3	follow-up	0.08	0.03	0.13	0.15	uncertain
P4	exit	-0.22	-0.31	-0.12	0.00	negative response
P4	follow-up	0.00	-0.08	0.08	0.66	uncertain
P5	exit	0.17	0.13	0.21	0.00	positive response
P5	follow-up	-0.02	-0.06	0.03	0.87	uncertain
P6	exit	-0.02	-0.11	0.02	0.76	uncertain
P6	follow-up	-0.01	-0.09	0.05	0.79	uncertain
P7	exit	-0.01	-0.04	0.01	0.99	no response
P7	follow-up	0.00	-0.03	0.05	0.95	no response
P8	exit	-0.11	-0.17	-0.04	0.08	negative response
P8	follow-up	0.09	0.01	0.17	0.20	uncertain
P9	exit	0.08	-0.04	0.21	0.29	uncertain
P9	follow-up	-0.23	-0.32	-0.15	0.00	negative response
P10	exit	0.08	-0.07	0.23	0.30	uncertain
P10	follow-up	0.07	-0.08	0.23	0.30	uncertain
P11	exit	0.01	-0.08	0.08	0.66	uncertain
P11	follow-up					
P13	exit	0.01	-0.03	0.06	0.90	no response
P13	follow-up	0.00	-0.05	0.04	0.91	no response
P14	exit	0.05	0.01	0.10	0.45	uncertain
P14	follow-up	0.03	-0.01	0.08	0.73	uncertain
P15	exit	0.08	-0.03	0.02	0.30	uncertain
P15	follow-up	0.15	0.05	0.26	0.05	positive response
P16	exit	0.13	0.08	0.18	0.00	positive response
P16	follow-up	0.20	0.15	0.25	0.00	positive response
P17	exit	0.13	0.08	0.18	0.01	positive response
P17	follow-up	0.15	0.10	0.21	0.00	positive response
P18	exit	0.25	0.19	0.31	0.00	positive response
P18	follow-up	0.20	0.14	0.26	0.00	positive response
P19	exit	-0.11	-0.18	-0.04	0.08	negative response
P19	follow-up	0.02	-0.04	0.08	0.78	uncertain

Table C1. Individual Effect Size Estimates for Change in Informativeness (Proportion CIUs) at Exit and Follow-up in Conversational Discourse (continued)

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P20	exit	-0.07	-0.15	0.02	0.36	uncertain
P20	follow-up	-0.07	-0.17	0.02	0.32	uncertain
P21	exit	-0.02	-0.08	0.05	0.74	uncertain
P21	follow-up	-0.02	-0.08	0.04	0.76	uncertain
P22	exit	-0.18	-0.25	-0.12	0.00	negative response
P22	follow-up	-0.03	-0.11	0.04	0.60	uncertain
P24	exit	0.00	-0.04	0.04	0.96	no response
P24	follow-up	0.08	0.04	0.11	0.11	uncertain

Note: Lower90 and Upper90 refer to bounds on 90% highest density interval. ROPE = region of practical equivalence. ROPE allows researchers to define a range of responses that are considered to be equivalent to the frequentist null hypothesis. The values in the ROPE column are the percentage of the Credible Interval that is within the ROPE. The lower the value, the more confident one can be that the participant's response is outside the ROPE. P11 did not attend the follow-up session for medical reasons. P23 is not included in the analysis of conversational discourse because her partner did not attend the exit or follow-up assessment sessions.

Table C2. Individual Effect Size Estimates for Change in Efficiency (CIUs per minute) at Exit and Follow-up in Conversational Discourse

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P1	exit	-7.93	-13.08	-2.82	0.06	negative response
P1	follow-up	-10.10	-15.28	-5.04	0.01	negative response
P2	exit	4.60	-0.01	9.46	0.28	uncertain
P2	follow-up	20.34	13.25	28.03	0.00	positive response
P3	exit	-14.32	-27.32	-1.66	0.06	negative response
P3	follow-up	-1.13	-14.43	11.97	0.29	uncertain
P4	exit	-7.25	-12.88	-1.78	0.10	uncertain
P4	follow-up	12.34	5.05	20.09	0.02	positive response
P5	exit	12.01	0.71	23.45	0.08	positive response
P5	follow-up	-18.30	-28.43	-8.72	0.00	negative response
P6	exit	-0.07	-0.38	0.09	1.00	no response
P6	follow-up	-0.01	-0.33	0.40	1.00	no response
P7	exit	-0.19	-0.66	0.09	1.00	no response
P7	follow-up	-0.03	-0.55	0.62	1.00	no response
P8	exit	-11.00	-17.90	-4.42	0.02	negative response
P8	follow-up	3.02	-6.66	12.92	0.34	uncertain
P9	exit	-3.36	-7.95	1.49	0.43	uncertain
P9	follow-up	-7.91	-12.49	-3.44	0.04	negative response
P10	exit	1.58	-1.25	4.84	0.77	uncertain
P10	follow-up	2.21	-1.31	6.10	0.64	uncertain

Table C2. Individual Effect Size Estimates for Change in Efficiency (CIUs per minute) at Exit and Follow-up in Conversational Discourse (continued)

Participant	Timepoint	Effect Size	Lower90	Upper90	ROPE	Interpretation
P11	exit	-3.27	-9.96	3.36	0.42	uncertain
P11	follow-up					
P13	exit	-0.16	-6.35	6.05	0.57	uncertain
P13	follow-up	2.76	-4.07	9.82	0.44	uncertain
P14	exit	8.64	-1.76	18.75	0.15	uncertain
P14	follow-up	3.82	-5.85	13.16	0.32	uncertain
P15	exit	0.56	-4.15	5.73	0.68	uncertain
P15	follow-up	3.84	-1.26	8.98	0.38	uncertain
P16	exit	19.31	10.70	27.98	0.00	positive response
P16	follow-up	28.79	19.27	38.41	0.00	positive response
P17	exit	15.41	9.26	21.91	0.00	positive response
P17	follow-up	14.16	7.69	20.70	0.00	positive response
P18	exit	8.48	2.81	14.26	0.06	positive response
P18	follow-up	16.66	10.09	23.59	0.00	positive response
P19	exit	-6.79	-11.49	-2.17	0.09	negative response
P19	follow-up	-4.29	-8.78	0.06	0.31	uncertain
P20	exit	-2.56	-6.54	1.39	0.56	uncertain
P20	follow-up	-5.16	-9.22	-1.27	0.18	uncertain
P21	exit	0.44	-5.61	6.62	0.58	uncertain
P21	follow-up	5.01	-1.30	11.10	0.28	uncertain
P22	exit	-2.62	-6.86	2.35	0.53	uncertain
P22	follow-up	-1.17	-6.90	9.48	0.50	uncertain
P24	exit	10.23	2.30	18.14	0.06	positive response
P24	follow-up	13.14	5.08	21.10	0.02	positive response

Note: Lower90 and Upper90 refer to bounds on 90% highest density interval. ROPE = region of practical equivalence. ROPE allows researchers to define a range of responses that are considered to be equivalent to the frequentist null hypothesis. The values in the ROPE column are the percentage of the Credible Interval that is within the ROPE. The lower the value, the more confident one can be that the participant's response is outside the ROPE. P11 did not attend the follow-up session for medical reasons. P23 is not included in the analysis of conversational discourse because her partner did not attend the exit or follow-up assessment sessions.