ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

The Benefits of Conversation Group Treatment for Individuals With Chronic Aphasia: Updated Evidence From a Multisite Randomized Controlled Trial on Measures of Language and Communication

Elizabeth Hoover,[a] Gretchen Szabo,[b] Francine Kohen,[c] Sarah Vitale,[a] Nicholas McCloskey,[c] Edwin Maas,[c] Varsha Kulkarni,[d] and Gayle DeDe[c]

[a] Department of Speech, Language & Hearing Sciences, Boston University, MA [b] Adler Aphasia Center, Maywood, NJ [c] Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA [d] Department of Mathematics and Statistics, Boston University, MA

ARTICLE INFO

Article History:
Received July 22, 2024
Revision received November 7, 2024
Accepted January 10, 2025

Editor-in-Chief: Rita R. Patel
Editor: Natalie F. Douglas

https://doi.org/10.1044/2025_AJSLP-24-00279

ABSTRACT

Purpose: Aphasia is a communication disorder that affects up to 30% of stroke survivors. Insufficient access to communication services creates personal, social, and financial costs to people with aphasia (PwA), care partners, and the community. Group conversation treatment has the potential to improve communication and reduce social isolation in a cost-effective manner, but little is known about its critical ingredients. This multicenter randomized controlled trial examined the effects of conversation treatment and whether the pattern of changes on outcome measures differed when treatment was delivered in large groups compared to dyads.

Method: One hundred four PwA were randomly assigned to a dyad, large group, or delayed control condition. Conversation group treatment was 1 hr, twice weekly, over 10 weeks. Individual communication goals were addressed within thematically oriented conversation treatment. To evaluate treatment effects, primary (Aphasia Communication Outcome Measure [ACOM]) and secondary outcome measures were examined at pretreatment, posttreatment, and 6 weeks posttreatment.

Results: The ACOM did not show significant changes in the planned omnibus analyses. Post hoc analyses suggested that the large group, but not dyad, treatment condition showed a treatment effect on the ACOM from pre- to posttreatment. Both treatment conditions showed changes on a measure of naming, and the dyads also showed improvement on a measure of repetition.

Conclusions: The study failed to show the effects of conversation treatment in the omnibus analysis, but there was evidence that conversation group treatment, delivered in a large group, is effective for people with chronic aphasia. This study also illustrated how manipulating the size of the group may alter the outcomes for individuals. The results of this study offer support for a *cost-effective* treatment option for PwA across the continuum of care.

Supplemental Material: https://doi.org/10.23641/asha.28719578

Aphasia is a communication disorder that affects up to 30% of stroke survivors. Conservative estimates suggest that 2.5 million Americans are living with this condition (Simmons-Mackie & Cherney, 2018). Aphasia is associated with greater disability compared to stroke patients without aphasia (Flowers et al., 2016; Gialanella et al., 2011) and has a negative impact on life satisfaction and quality of life (Cruice et al., 2003; Ellis & Peach, 2017; Hilari et al., 2012; Koleck et al., 2017; Lam & Wodchis, 2010; Shadden, 2005). In comparison to over 60 diseases and 15 health conditions in a large cohort of individuals living in long-term care, aphasia was reported as having the largest negative impact on health-related quality of life (Lam &

Correspondence to Gayle DeDe: gayle.dede@temple.edu. Elizabeth Hoover and Gayle DeDe contributed equally to this work. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Wodchis, 2010). This disparity in health-related outcomes is also observed specifically between stroke survivors with and without aphasia and is likely due to the inherent impact of the communication difficulties on the ability to return to work, to participate in hobbies, and to maintain social relationships (Simmons-Mackie & Cherney, 2018). Aphasia also has significant financial consequences to an individual, including lost wages, and to the health care system (Ellis et al., 2012). According to Jacobs and Ellis (2023), the economic burden of aphasia is more than $15 billion annually in the United States alone. Given these personal, social, and financial costs of aphasia, it is critical that cost-effective treatment approaches be identified.

Services for aphasia in North America vary greatly. In a comprehensive review of the state of aphasia, Simmons-Mackie and Cherney (2018) identified several gaps in care for people with aphasia (PwA), including either a discharge from emergency services without a referral for follow-up care or a significant lack of services for communication access, education, and treatment across the health care continuum. Traditional rehabilitation for aphasia is often limited to the first months following onset of the condition (2–5 weeks inpatient, 8–12 weeks outpatient). For most PwA, this amount of intervention is insufficient. Indeed, this is the point at which most individuals with aphasia and their loved ones are beginning to understand the challenges that lie ahead (Simmons-Mackie & Cherney, 2018). Depression and withdrawal from activities frequently deteriorate as rehabilitation services end and individuals must learn to renegotiate a new life and rebuild identity without skilled support (Shadden, 2005). Simmons-Mackie and Cherney argue that specialized services are needed to help people live successfully with aphasia and that our health care system must shift from treating aphasia as a temporary, curable impairment to recognizing aphasia as a chronic condition that requires long-term support.

An emerging body of literature addresses the need for long-term care for people with chronic conditions and suggests that the focus should be on promoting social connections, well-being, and community participation based on life interests and preferences (National Quality Forum, 2016). A recent study investigating outcomes of a community program designed to increase life participation after stroke found that participants increased the number of hours of activity occurring outside of the home and improved on a range of other outcomes (Mayo et al., 2015). Furthermore, the number of hours of meaningful activity that occurred outside the house was judged to be an important outcome by stroke survivors in the study and was associated with changes in overall life satisfaction levels; health-related quality of life ratings; and improvement in mood, apathy, and mobility. Interestingly, these changes in meaningful activity took 1 year to achieve, further justifying the need for long-term access to services for individuals with aphasia (Mayo et al., 2015).

As stated previously, current reports of clinical services for aphasia not only fall short of this year time frame but are also largely below the optimum therapeutic ranges in both duration and dosage of treatment compared to the research base. The REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE Collaborators, 2022) collaborators reviewed the individual participant data of 959 individuals across 25 clinical trials on aphasia language outcomes and time post-onset. This systematic review explored patterns of interaction between speech-language treatment frequency, intensity, dosage, and language outcomes of groups of participants. Results showed that greatest gains in overall language and comprehension were associated with a dosage of 20 to 50 hr of speech-language therapy. The greatest clinical gains in overall language and functional communication were associated with a dosage of 2–4 hr/week, and changes in auditory comprehension were noted with > 9 hr/week. Although small differences were noted in subgroups related to time post-onset, gender, and age, therapy that addressed both comprehension and expression skills, and that was functionally tailored and prescribed with home practice, was associated with the greatest overall gains.

The findings in the RELEASE report support clinical service delivery at a greater intensity and duration than reports of usual clinical services internationally and in the United States (Cavanaugh et al., 2021, 2023; Guo et al., 2014; Kong & Tse, 2018; Rose et al., 2014). Cavanaugh et al. (2021) reported estimates of clinical treatment dosage for PwA in an outpatient setting from the greater Western Pennsylvania region. The median dose of treatment sessions was 10 (M = 14.8; interquartile range: 5–20). The median total number of hours of treatment was 7.5 hr. Cavanaugh et al. (2023) replicated this study on a national scale using closed commercial claims data. Their findings revealed that of all people in the database who were diagnosed with aphasia (N = 1,968) and received both evaluation and treatment visits, the average number of treatment sessions was 14.5 (Mdn = 10), totaling 10.6 hr (Mdn = 7.5) over 10.6 weeks (Mdn = 7.7). Only 6% of PwA received more than 50 visits per episode of care and only 0.07% (n = 14) received more than 100 visits with more than three visits per week. These statistics fall significantly short of the minimum recommended dosage identified in the RELEASE (RELEASE Collaborators, 2022) systematic review.

In sum, insufficient access to communication services across the continuum of care, particularly in the chronic stages, creates considerable personal, social, and financial costs to PwA, care partners, and the community at large (Simmons-Mackie & Cherney, 2018). It is critical that evidence-based, cost-effective, and participation-centered interventions that focus on functional communication be available to people living with aphasia at all stages of care.

Group conversation treatment, which is typically more affordable than individual treatment, is offered at aphasia centers throughout the United States (Elman, 2007a). Given its relatively low cost, it may be a long-term solution for PwA in the chronic stages of recovery. There are several theoretical constructs that motivate conversation group treatment; for a review of these foundational theories, please see Elman (2007b) and DeDe et al. (2019). Group conversation treatment has the potential to improve communication and reduce social isolation in a cost-effective manner (Boyle et al., 2023; DeDe et al., 2019; Elman, 2007b; Elman & Bernstein-Ellis, 1999a; Hoover et al., 2021). However, evidence to date is based on data from two relatively small-sized randomized controlled trials (RCTs), and little is known about the mechanisms underlying the intervention (DeDe et al., 2019; Elman & Bernstein-Ellis, 1999a).

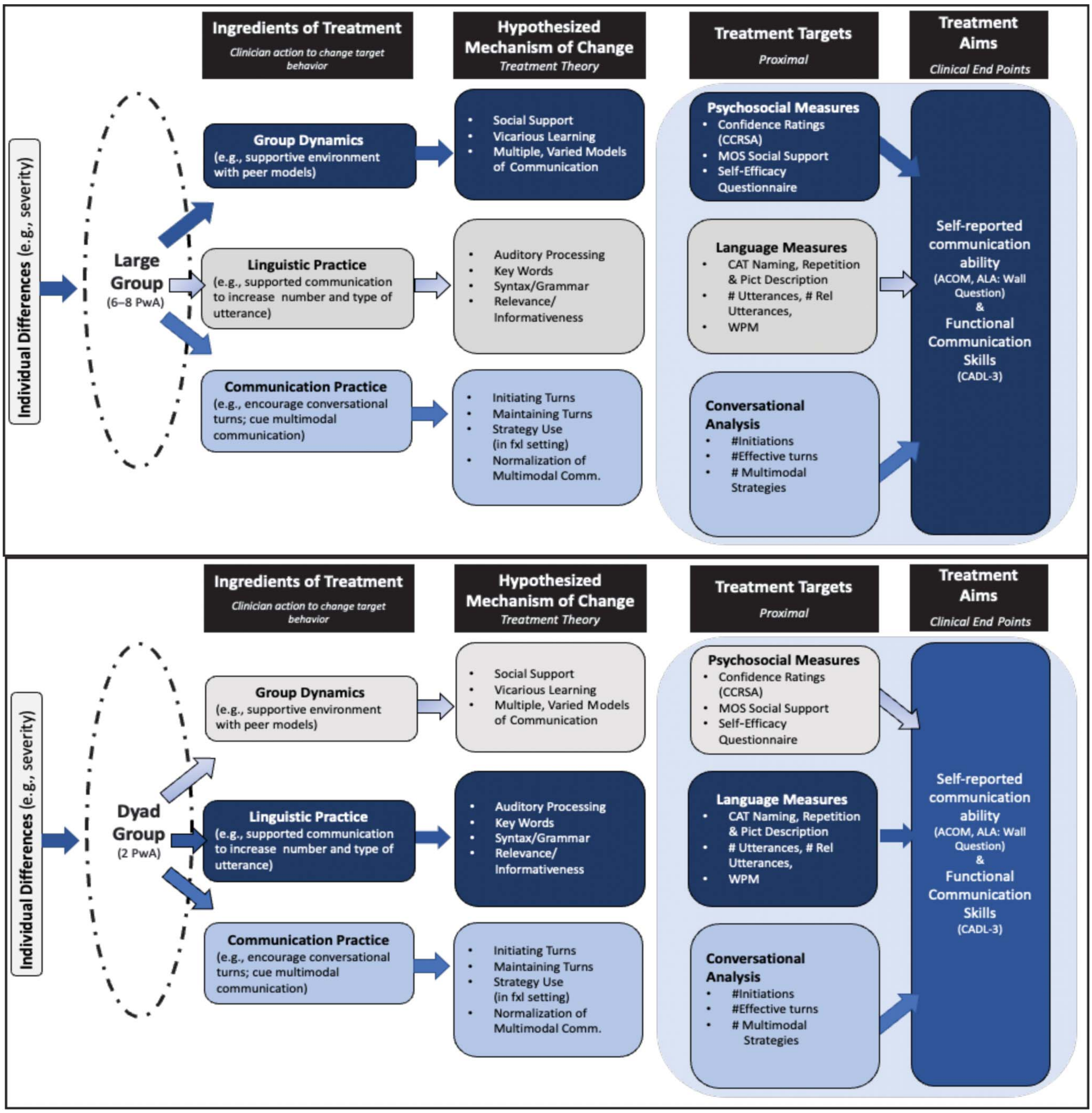
One challenge is that group conversation treatment has many different implementations, with fundamental variations surrounding the number of people included in the group, the group facilitation methods, the topics of conversation, and the degree of structure (e.g., Archer et al., 2019; Simmons-Mackie et al., 2014). Elman and Bernstein-Ellis (1999a, 1999b) compared a communication group treatment to a deferred treatment control group. Treatment comprised 80 hr over 16 weeks and focused on improving the ability to convey a message using any modality in addition to promoting confidence for meaningful conversations. The deferred treatment group participated in nonskilled socialization groups to control for the effects of social contact. Results showed that only the treatment group made significant improvements on language and functional measures of communication (Porch Index of Communication Ability, Western Aphasia Battery [WAB]–Aphasia Quotient, and Communication Activities of Daily Living [CADL]), demonstrating that changes were associated with participation in treatment rather than socialization. Qualitative interviews after treatment revealed positive psychosocial changes including greater levels of confidence and motivation to improve and to socialize. Using the same participant data, Boyle et al. (2023) reported changes in discourse informativeness and efficiency on five structured discourse tasks following treatment for all the participants. Specifically, results revealed improvement in the proportion of correct information units (CIUs; Nicholas & Brookshire, 1993) and the number of CIUs per minute. No improvement was found in conversational discourse.

In our implementation of group conversation treatment, communication and linguistic goals are addressed in the context of naturalistic conversation. We have described this group conversation treatment within the Rehabilitation Treatment Specification System (RTSS; Hart et al., 2019; Hoover et al., 2021; Van Stan et al., 2019; Zanca et al., 2019). The RTSS provides a common terminology for explicating the pathway by which active ingredients will effect changes in specific behaviors (treatment targets) and how those specific behaviors may lead to change on long-term goals (treatment aims). The RTSS may be used to describe the mechanisms of action underlying rehabilitation treatments of all kinds. Using the RTSS terminology, we have identified three key ingredients for conversation treatment: group dynamics, communication practice, and language practice (see Figure 1). Investigating the relative impact of these ingredients on outcomes is critical to understand the mechanisms of action underlying this intervention and, hopefully, to build a strong research base that will increase the accessibility of conversation group treatment more broadly (i.e., via reimbursed health care models).

Group dynamics is based on the psychodynamic literature and proposes several beneficial features of groups (Elman, 2007b). Group treatment brings together individuals with similar interests or issues and provides an opportunity for group members to share resources and gain support (Elman, 2007b; Yalom & Leszcz, 2005). In addition, group membership offers psychosocial support, which may, in turn, relieve social isolation (Elman, 2007a). Groups of PwA engender vicarious and interpersonal learning, such as observing others with aphasia who use strategies (e.g., gesturing or writing) to communicate successfully (Elman, 2007b; Holmes & Kivilighan, 2000; Luterman, 1991). Communication practice refers to the opportunity to *use* multimodal forms of communication such as writing, gesturing, and technology to communicate thoughts and ideas. Language practice refers to opportunities to process auditory input, produce key words, and so on.

We have argued that modifications in conversation treatment delivery, such as the number of people in the group, have the potential to alter the impact of intervention ingredients and, thus, differentially affect treatment outcomes (DeDe et al., 2019; Hoover et al., 2021). In a traditional group of six to eight participants, one might expect a relatively larger influence of the group dynamics ingredient on outcome measures given that there are more members with shared lived experiences who can model communication strategies, offer social support, and so on. However, large groups offer fewer opportunities for linguistic and communication practice and increased competition for speaking time (cf. DeDe et al., 2019). In contrast, dyads (two PwA) offer more opportunities for language and communication practice but only one other

Figure 1. Hypothesized pathway by which modifications in group size influence ingredients and mechanisms of action in conversation treatment. The shading indicates the relative contribution of each ingredient. The large group condition maximizes the benefit of group dynamics over linguistic practice, and vice versa for the dyads. Both groups are hypothesized to offer opportunities for communication practice. PwA = people with aphasia; CAT = Comprehensive Aphasia Test; ACOM = Aphasia Communication Outcome Measure; ALA = Assessment for Living with Aphasia.



individual with whom to share experiences or observe for vicarious learning (DeDe et al., 2019; Hoover et al., 2021). Thus, we hope to examine the effects of group dynamics and language practice by comparing patterns of outcomes for conversation treatment delivered in large groups compared to dyads.

The results of our preliminary work showed that both treatment groups (dyads and large group) showed significant improvements on selected standardized tests of language and patient-reported measures in contrast to the control group, which showed no changes. In addition, our analyses suggested that there are differences in outcomes related to the size of the group. The dyads improved on more discrete linguistic tasks: repetition, two different word retrieval tasks, and number of complete utterances produced in five narrative samples. The large group also improved on one naming measure, in addition to changes on more functional communication measures: a patient-reported outcome measure of functional communication, the standardized analysis of the Comprehensive Aphasia Test (CAT) picture description task, and the number and relevancy of utterances produced across three narratives. The sample size in the preliminary study was underpowered for large-scale analyses; thus, results were based on nonparametric statistics. In addition, there were fewer measures of functional communication. While the overall efficacy for conversation treatment is building, there remains a critical need for additional research to refine outcome measures and to systematically explore the relative contributions of the critical ingredients of this intervention.

The goals of this Phase 1b randomized controlled clinical trial were to examine the efficacy of the conversation treatment and investigate the optimal group size for treatment delivery with an increased sample size and additional secondary outcome measures. The primary outcome measure was the Aphasia Communication Outcome Measure (ACOM; Hula et al., 2015). Our research questions were the following:

1. Is conversation treatment associated with self-reported and directly measured changes in linguistic ability, communication ability, and psychosocial health for PwA? For this research question, we examine whether the treatment conditions (large group and dyad) show significantly greater improvements on the primary and secondary outcome measures as compared to the control condition.
2. Does the pattern of change on outcome measures differ for large groups compared to dyads? For this research question, we examine whether the treatment conditions (large group and dyad) show different patterns of effects across the outcome measures. Based on our preliminary studies (DeDe et al., 2019; Hoover et al., 2021), we hypothesized that the dyad condition would show treatment changes on discrete linguistic measures (e.g., repetition, oral reading), whereas the large group condition would show more changes on functional communication measures such as the ACOM (primary outcome measure) and narratives.

Method

This project was approved by the institutional review board (IRB) at Boston University (BU; Protocol No. 4341E), which served as a single IRB for all sites.

Participants

One hundred seventeen PwA were recruited in this multicenter, unblinded, parallel group RCT Phase 1b study with a concurrent natural history control group and balanced randomization.[1] To complete a RCT of group treatment, all participants must be recruited prior to beginning the treatment. Thus, the study was run in two cycles and at three sites, with a recruitment goal of 24 PwA in each cycle (n = 8 each in large group, dyadic, and control conditions).[2] Cycle 1 was recruited and completed testing in 2017, and Cycle 2 was recruited and completed testing in 2022. Figure 2 shows the total number of participants at each site and in each cycle. A total of 117 PwA were enrolled, and 104 PwA completed this study.

Inclusion criteria included (a) at least 18 years of age, (b) at least 5 months post-onset of aphasia, (c) native English speaker (learned before the age of 6 years), and (d) diagnosed with aphasia based on clinical judgment and standardized test results (WAB–Revised [WAB-R]; Kertesz, 2007;[3] or CAT; Swinburn et al., 2004). Table 1 presents basic descriptive data about the participants.

Allocation to condition was concealed in opaque, sealed envelopes until participants completed eligibility testing and enrolled in the study. Participants were randomly assigned to one of three conditions using a 1:1:1 allocation ratio. The randomization sequence (block size = 6) was determined by a blinded statistician who provided a password-protected Excel sheet to a study researcher. This block size was implemented so that if recruitment targets were not met, there would be roughly equal numbers in each of the three conditions. Dyads were paired based on pragmatic concerns (e.g., scheduling) and clinician judgment of compatibility following random assignment to maximize the clinical relevance of this study. Participants assigned to the control condition were asked to abstain from other communication treatments during the experimental period and were offered large group conversation treatment after the maintenance testing period had concluded. This group served as a delayed/no-treatment “control group” as they did not receive any intervention during the data collection phase of this study.

At BU and Temple University (TU), recruitment occurred in spring and treatment for the large group and dyad conditions occurred in summer. A community partner, the Adler Aphasia Center (Adler), joined the study for Cycle 2. Recruitment at Adler took place in summer, with the dyad and large group treatments occurring in fall. Participants

[1]Data from a subset of the participants reported here (n = 41) were previously reported in DeDe et al. (2019).
[2]In 2022, institutional precautions related to the coronavirus disease 2019 pandemic limited the recruitment target to 18 at Temple University.
[3]Western Aphasia Battery–Revised was only administered in Cycle 2.

Figure 2. Flowchart of participant recruitment and retention. Tx = treatment; Maint = maintenance.

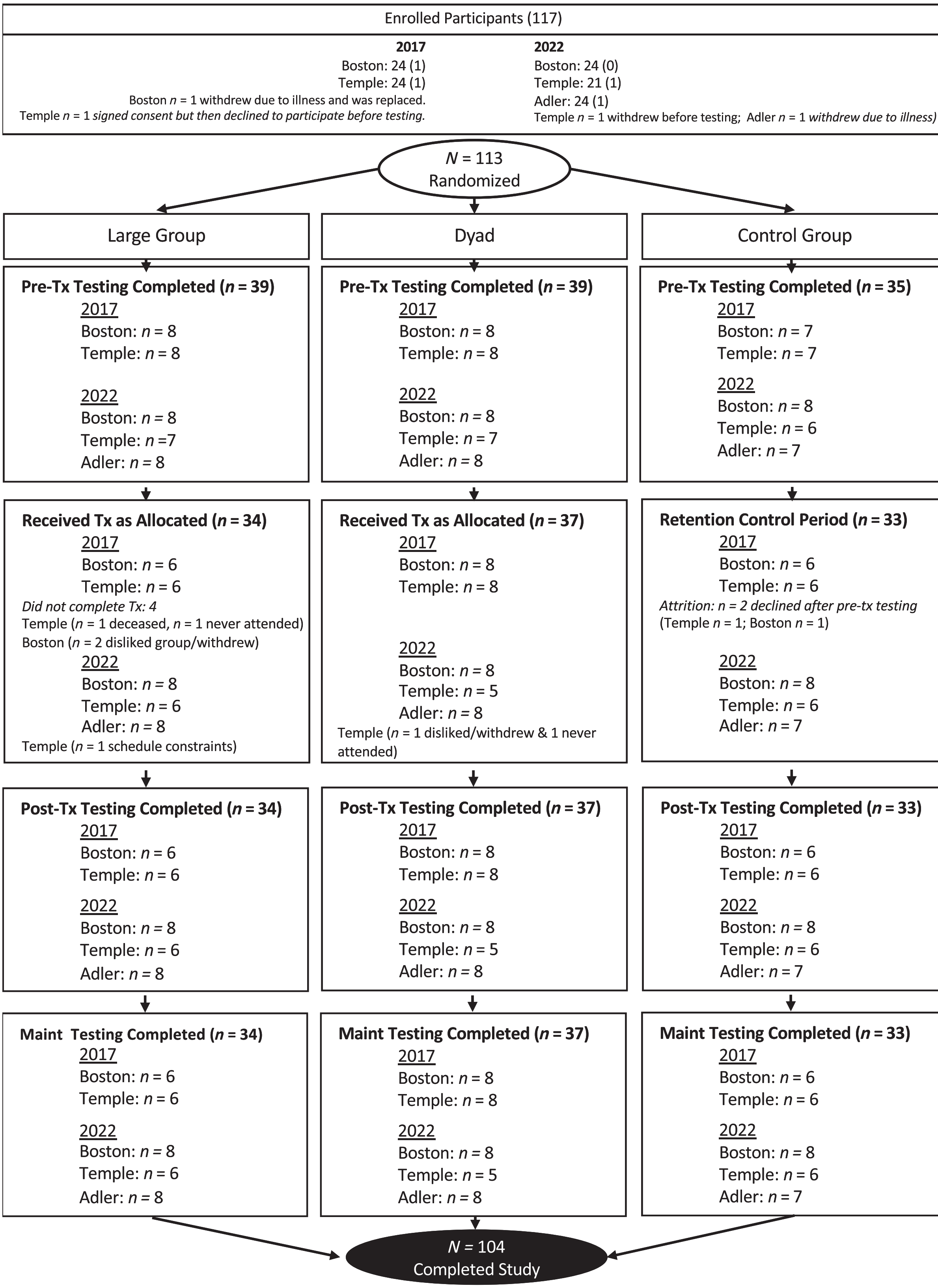


Table 1. Mean demographic data.

Condition	Age (years)	Education (years)	TPO (months)	WAB-AQ	Gender	Race/ethnicity
Control (*SD*)	66.6 (12.7)	15.6 (3.1)	94 (60.4)	55.7 (21.9)	Male: 25, female: 10	3 A; 8 B/AA; 1 M/O; 23 W
Dyad (*SD*)	61.9 (12.5)	16.2 (3.0)	76 (66.9)	76.5 (17.34)	Male: 21, female: 17	1 A; 11 B/AA; 1 M/O; 25 W
Large group (*SD*)	59.0 (9.0)	14.1 (2.9)	79 (67.1)	73.3 (22.22)	Male: 20, female: 18	11 B/AA; 27 W

Note. TPO = time post onset; WAB-AQ = Western Aphasia Battery–Aphasia Quotient; A = Asian; B/AA = Black/African American; M/O = mixed race/other; W = White.

were asked to abstain from other communication treatments during the study period. Participants in the control condition were offered conversation treatment after the study (in fall at BU and TU and in winter at Adler). Informed written consent was obtained from all participants.

Assessment Protocol

The testing protocol was administered at three time points: immediately prior to and following the experimental treatment and 6 weeks posttreatment (maintenance). It was not possible to blind participants or treating clinicians to the condition. However, whenever possible, assessors were blinded to the treatment condition. To mitigate the effects of bias, we maintained equipoise between conditions, never presenting one condition as likely to be more effective to participants or clinicians. Examiners were licensed speech-language pathologists (SLPs) with extensive experience in aphasia or graduate students in speech-language pathology under the supervision of licensed SLPs. These testing sessions were completed before any treatment was provided to the delayed treatment group, allowing these participants to serve as a concurrent, natural history control group.

Six standardized measures were selected as outcomes in this study. These measures align with the constructs of a core outcome set for PwA established by the Collaboration of Aphasia Trialists (Wallace et al., 2019) and represent both objective and patient-reported outcome measures across linguistic, communication, and quality of life domains. Table 2 provides a list of these measures with accompanying domains. Importantly, the final three measures listed in the table were added in the second cycle of the study and, thus, were administered to a smaller sample size.

Primary Outcome Measure

The primary outcome measure was the ACOM (Hula et al., 2015). The ACOM is a psychometrically

Table 2. Assessment battery of formal measures.

Type	Measure	Domain	Dependent variable
Primary outcome measure			
PROM	Aphasia Communication Outcome Measure	Self-reported communicative effectiveness	*T* score
Secondary outcome measures			
Standardized test (linguistic)	Comprehensive Aphasia Test language section	Comprehension of auditory and written language, naming, repetition, oral reading word, and picture description	Total scores for each subsection
Standardized test (linguistic)	Northwestern Assessment of Verbs and Sentences: Verb Naming Test	Verb naming	Total number correct
Standardized test (communication)	Communication Activities of Daily Living–Third Edition[a]	Functional communication via role-play	Raw score
PROM	Assessment for Living with Aphasia: Wall Question[a]	Likert scale ranking of the overall impact of aphasia on daily life	Raw score
PROM	MOSS Social Support Scale[a]	Self-report of social support available to individuals	Total score
PROM	Communication Confidence Rating Scale for Aphasia[a]	Self-reported communication confidence	Total score
Discourse (linguistic)	CoreLex[a]	Key words produced in discourse	Number of CoreLex words
Discourse (linguistic)	Complete utterances[a]	Relevance and grammaticality of discourse	Number and percent of c-units and of grammatical, relevant, and complete c-units

Note. PROM = patient-reported outcome measure; MOS = medical outcomes study.
[a]Administered only in the second cycle.

validated patient-reported outcome measure that reflects the overarching aim of conversation treatment, which is a self-reported communication ability in functional settings. Participants are asked to rate "the effectiveness" of their communication on multimodal functional tasks, such as "How effectively do you explain how to do something?" or "How effectively do you make your wants and needs known?" This was chosen as the primary outcome measure because it reflects the overall purpose of conversation treatment (i.e., improved functional communication) and might reasonably be expected to reflect change for people with a wide range of aphasia severity profiles.

Secondary Outcome Measures

Language measures included the Verb Naming Test (VNT; Cho-Reyes & Thompson, 2012) and four subtests from the CAT (naming, repetition, oral reading, and picture description). The CAT is a comprehensive psycholinguistic test battery whose scores account for delayed or self-corrected responses and requests for repetition, increasing the sensitivity of the measure. The naming subtest includes verbal fluency and confrontation naming for nouns and verbs. Repetition includes words and nonwords, as well as digit and sentence repetition. Oral reading includes content words, function words, and nonwords. The picture description total score combines a number of information carrying words with ratings of grammaticality, speech rate, and diversity of syntactic structures.

The CADL–Third Edition (CADL-3; Holland et al., 2018) is a standardized measure of functional communication that uses role-played scenarios such as shopping, driving, and visiting a doctor. The Wall Question from the Assessment for Living with Aphasia (ALA: Wall Question; Kagan et al., 2018) is a single survey question used to quantify the overall impact of aphasia on daily life. The Wall Question is one item on the larger ALA battery. Kagan et al. found good reliability between the Wall Question and the entire ALA battery in determining the impact of aphasia on the overall quality of life. This single item was used as an index of how much communication ability acts as a barrier to participation in life activities. The Medical Outcome Study Social Support Survey (Sherbourne & Stewart, 1991) is a 20-item instrument on a 5-point response scale, designed to evaluate participants' perception of the social support available to them. The Communication Confidence Rating Scale (Babbitt et al., 2011; Cherney & Babbit, 2011) is a psychometrically validated measure that evaluates self-reported communication confidence across a variety of communicative contexts.

Monologic picture descriptions were analyzed using core lexicon (Dalton et al., 2020) and complete utterance (Edmonds et al., 2009) metrics. Participants produced four monologic picture descriptions (cat rescue, birthday scene, refused umbrella, and broken window; Nicholas & Brookshire, 1993) following Aphasia Bank procedures (MacWhinney et al., 2011). Picture descriptions for these stimuli, as well as the picture description from the CAT, were orthographically transcribed and checked by a second trained research assistant. In addition, 20% of the samples were randomly selected to be independently transcribed. Interrater transcription reliability was evaluated using normalized similarity scores based on Levenshtein distance (1 − Levenshtein distance/maximum transcription length in characters). The average similarity score was 0.88, and 92% of the samples had a score greater than 0.70. Inspection of the transcripts with scores below 0.70 revealed deviations in spelling of nonwords, fillers, and part word repetitions. Transcriptions were segmented into C-units and checked by trained research assistants; any disagreements noted during checks were reconciled between coders or via discussion with the principal investigators and supervising clinicians.

Samples were deidentified with respect to the testing time, participant, and condition prior to being coded for complete utterances (cf. Edmonds et al., 2009) and core lexicon (Dalton et al., 2020). Complete utterances are C-units that are both grammatical and relevant to the narrative. Grammaticality is based on whether C-units contain verbs plus their obligatory arguments. Relevance is based on whether the C-units are relevant to the topic of the discourse. Complete utterance codes were checked by a second rater, and issues were resolved by consensus. Then, 20% of the samples were independently coded to evaluate interrater reliability. The average reliability was 0.97 for the number of C-units (98.3% > 0.70), 0.91 for grammaticality (93.1% > 0.70), 0.84 for relevance (91.9% > 0.70), and 0.87 for complete utterances (89.1% > 0.70). A core lexicon analysis examines the specific lexical items used to tell a story and allows for the comparison of items to published normative samples. This method was selected to serve as a reliable and valid measure of word retrieval in discourse. Analyses were run for the three stimuli with published lists (Dalton et al., 2020) using automated procedures (Dalton et al., 2022). Reliability between hand and automated core lexicon scores were computed for approximately 30% of the samples. Intraclass correlations were calculated using single-rater one-way random effects models. Intraclass correlations (ICC) were calculated using single-rater one-way random effects models. The average ICC was .998, $F(160) = 580.1$, $p < .0001$.

Treatment

Treatment sessions were 1 hr, twice weekly for 10 weeks (20 hr in total). Treatment at BU and TU was provided by trained graduate students supervised by licensed SLPs with previous experience in delivering conversation group treatment. Treatment at Adler was provided by licensed SLPs with previous experience in group treatments.

The treatment protocol followed a socially oriented approach in which participants were encouraged to engage in meaningful, authentic conversations surrounding functional topics (see Elman, 2007a; Simmons-Mackie et al., 2014). Although conversation treatment is inherently unstructured, we developed systems to maximize consistency and reproducibility of treatment delivery across sites and cycles. We generated five broad conversational topics with four functional subtopics (see Figure 3).

For each subtopic, we developed materials including PowerPoint slides with visual cues and questions, as well as conversational supports. For all sessions, general support materials such as writing implements, maps, and a computer or tablet to be used as a digital librarian were available. We also generated conversation supports specific to scheduled topics (e.g., visual depictions of family roles/family members). Materials were shared across all sites via Microsoft Teams and OneDrive applications.

Sessions began with social questions (e.g., "How is everyone today?"). Next, the topic was introduced, and conversation was facilitated following established guidelines of conversation treatment (Elman, 2007a; Kearns & Elman, 2008). Clinicians allowed the conversation to evolve naturally and did not attempt to direct the flow of the topic, except for rare situations such as diverting conversations from perseverative topics. Clinicians attempted to provide equal opportunities for all PwA in the group to contribute conversation turns. To normalize multimodal communication, clinicians modeled communication strategies such as gesturing, writing key words, and repeating key phrases. These strategies also helped support auditory comprehension for all group members.

Each participant identified individual goals prior to initiating treatment. These goals were addressed within the sessions by creating practice opportunities within the conversation. For example, one PwA wanted to contribute more turns and produce more accurate key words in a conversation. This individual was provided with personalized visual supports (alphabet board), picture/photo prompts, extra time to contribute turns, and support such as word recasts or requests for repetition to increase the number of communication attempts.

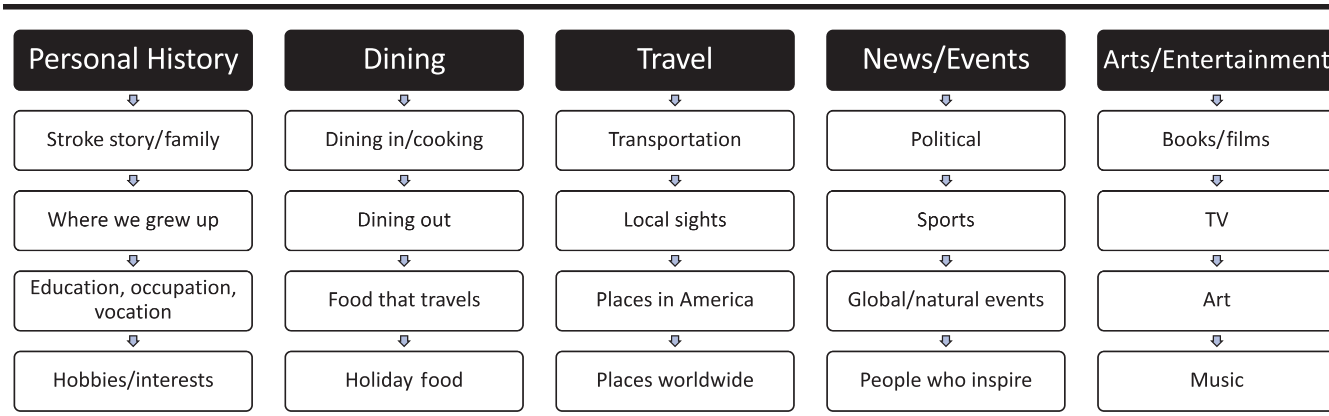
Treatment Fidelity

All clinicians received extensive training about the treatment protocol, which included a 2-hr orientation meeting to review procedures and goals. All clinicians also received a treatment manual with information about the conversation treatment, cueing hierarchies, and group facilitation techniques. At BU and TU, supervisors observed all treatment sessions and provided daily feedback to ensure the correct implementation of the protocol. At Adler, co-PI DeDe provided training prior to treatment and observed the first week of sessions. Researchers at all three sites met weekly to discuss any questions about the treatment delivery. All facilitators provided treatment in both experimental conditions within each cycle (large group and dyad). Finally, trained observers used a checklist to record variables related to treatment, such as the number of facilitator models for multimodal communication strategies and the number of conversation turns for each PwA and facilitator. These data were regularly reviewed to ensure accurate implementation of the protocol. No significant differences or deviations from the treatment plan were observed or reported between sites.

Statistical Analysis Plan

Data cleaning and primary data analysis were completed by a statistician blinded to the condition using R (Version 4.2.1; R Core Team, 2021) and the lmer,

Figure 3. Conversation topics.



lmertest, and emmeans packages. Examination of residuals did not reveal systematic departures from linear model assumptions of normality and heteroscedasticity. Therefore, no data transformations were implemented.

Modified intention-to-treat analyses were implemented, in which participants with missing data were excluded. Given the relatively small sample size for each condition and generally random causes of attrition, participants with missing data were excluded rather than using data imputation. The planned analyses were omnibus linear mixed-effects models, which were fit to the data using random-effect maximum likelihoods. Data were treatment-coded, and reference factors are in italics. The fixed effects were condition (large group, dyad, *control*), time (*pretreatment*, posttreatment, maintenance), and the interaction of condition by time. The model included a random intercept for participants and testing site (BU, TU, Adler) as a control variable. Models did not converge when random slopes were included.

Planned analyses of outcome measures were followed up with post hoc analyses that examined treatment effects separately for the two experimental conditions (large group vs. control and dyad vs. control). Effects that were significant in the planned, omnibus analysis were not reanalyzed. The same statistical methods were used except that the treatment groups (large group, dyad) served as the reference factors. This allowed us to ask (a) whether the treatment group showed an effect of time (pre- vs. posttreatment and pretreatment vs. maintenance) and (b) whether that change was significantly different in the treatment versus the control conditions (interactions of time by condition; Brehm & Alday, 2022).

Results

Table 3 provides the mean scores and standard deviations for each condition at each time point. Here, only significant effects of treatment or interactions between time and condition are reported. Full model results are available in Supplemental Materials S1–S12.

Planned Analyses

Primary outcome measure. Supplemental Material S1 presents full results of the mixed-effect models for the ACOM. For the large group condition at posttreatment, there was a trend toward an interaction between condition and time, ß = 3.39, $SE = 1.82$, $t(207.27) = 1.86$, $p = .06$. No other effects approached significance.

Secondary outcome measures. There were significant effects of treatment on the CAT naming total score (see Supplemental Material S2). From pre- to posttreatment, there was a significant interaction with condition for the large group compared to the control, ß = 6.37, $SE = 1.87$, $t(203.8) = 3.41$, $p < .001$, and the dyads compared to controls, ß = 4.26, $SE = 1.81$, $t(203.34) = 2.36$, $p = .02$. Compared to controls, the large group maintained a significant difference on CAT naming from pretreatment to maintenance, ß = 4.27, $SE = 1.87$, $t(203.81) = 2.29$, $p = .02$. The dyads' maintenance data showed a nonsignificant improvement compared to controls, ß = 3.20, $SE = 1.83$, $t(203.40) = 1.75$, $p = .08$. There were also main effects of the testing site, reflecting that participants at one site (Adler) presented with more severe aphasia compared to the participants at other sites; the testing site did not interact with any other factors.

The dyads showed additional effects of treatment on the CAT repetition total score. Compared to controls, the dyads showed a significantly greater improvement from pretreatment to posttreatment testing, ß = 5.02, $SE = 1.57$, $t(203.47) = 3.20$, $p < .01$. At maintenance (6 weeks posttreatment), the dyads showed a marginally significant effect compared to controls, ß = 3.12, $SE = 1.58$, $t(203.54) = 1.97$, $p = .05$. As Supplemental Material S2 shows, there was no Group × Time interaction for the large group condition in repetition performance. There were no additional effects of treatment in the omnibus analyses (see Supplemental Materials S1–S4).

Post Hoc Analyses: Large Group Versus Control Condition

Primary Outcome Measure

Supplemental Material S5 presents model results. For the ACOM, the main effects of time in the large group were significant for both pre- versus posttreatment, ß = 3.56, $SE = 1.11$, $t(133.8) = 3.20$, $p = .01$, and pretreatment versus maintenance, ß = 4.75, $SE = 1.12$, $t(132.8) = 4.26$, $p < .001$. The interaction of Condition × Time was significant at posttreatment, ß = −3.34, $SE = 1.58$, $t(133.24) = -2.16$, $p = .03$. The treated group showed significantly larger improvements on the ACOM from pre- to posttreatment than the control condition. The change was not significant from pretreatment to maintenance.

Secondary Outcome Measures

CAT picture description. The main effects of time in the large group were significant for both pre- versus posttreatment, ß = 2.58, $SE = 1.22$, $t(131.93) = 2.12$, $p = .04$, and pretreatment versus maintenance, ß = 5.52, $SE = 1.22$, $t(131.91) = 4.28$, $p < .001$. The main effects of location were also significant, reflecting a more severe group of participants at Adler.

Table 3. Results of standardized testing and narrative analysis for all conditions and time points.

Measure (max points)	Control			Dyad			Large Group		
	Pre Tx	Post Tx	6 wks	Pre Tx	Post Tx	6 wks	Pre Tx	Post Tx	6 wks
ACOM (70)	53.4 (8.4)	53.8 (8.6)	55.8 (8.6)	51.3 (11.2)	52.8 (10.0)	53.7 (9.9)	49.0 (9.7)	52.6* (9.5)	53.8 (10.0)
Comprehensive Aphasia Test (CAT) subsection scores									
Naming (94)	52.3 (27.5)	51.6 (27.8)	51.5 (24.7)	42.5 (28.6)	45.7* (29.1)	43.7 (29.7)	44.9 (24.7)	50.5* (27.5)	47.7 (27.6)
Repetition (74)	50.5 (19.0)	49.8 (19.4)	51.1 (18.7)	41.3 (21.0)	45.8* (21.4)	44.0 (22.2)	47.0 (21.1)	47.5 (21.6)	45.8 (22.0)
Oral Reading (70)	43.8 (22.7)	44.6 (20.7)	43.1 (21.9)	35.4 (24.6)	36.4 (25.1)	36.9 (25.3)	41.5 (23.8)	41.6 (24.5)	40.5 (24.4)
Pict Desc (n/a)	19.1 (13.7)	19.5 (13.1)	20.5 (15.2)	16.1 (16.1)	17.7 (15.2)	16.1 (19.6)	14.2 (13.7)	17.5 (14.5)	19.0 (15.0)
VNT (22)	13.7 (7.0)	13.1 (7.0)	15.2 (6.8)	11.3 (8.0)	12.6 (8.1)	11.6 (8.3)	13.8 (6.8)	14.3 (6.6)	14.1 (6.9)
CADL-3 (100)	84.0 (14.1)	82.9 (13.5)	85.2 (12.9)	76.2 (20.2)	78.3 (20.4)	79.3 (19.7)	79.6 (17.0)	82.9 (14.9)	83.5 (14.8)
CCRSA (40)	31.7 (5.0)	32.3 (4.3)	33.4 (4.2)	29.9 (5.7)	30.9 (5.2)	31.6 (5.2)	27.9 (6.7)	27.9 (4.3)	29.5 (5.5)
MOS-SSS (100)	71.5 (0.8)	72.9 (0.8)	74.5 (0.8)	74.8 (0.8)	77.0 (0.6)	79.3 (0.6)	68.2 (0.9)	68.2 (0.9)	71.7 (0.8)
ALA: Wall Q (4)	2.9 (0.9)	2.9 (0.8)	2.9 (1.0)	2.8 (0.9)	2.8 (1.0)	2.9 (0.8)	2.5 (1.0)	2.6 (0.8)	2.614 (0.8)
Complete utterances (CU)									
# C-units	7.9 (3.5)	8.2 (3.8)	8.0 (3.4)	7.8 (2.3)	7.9 (2.1)	7.7 (1.4)	9.2 (5.1)	9.2 (3.8)	9.2 (4.0)
# SV	2.2 (2.3)	1.3 (1.9)	2.3 (2.3)	2.2 (1.6)	1.6 (2.2)	2.2 (1.9)	2.1 (2.2)	1.7 (1.9)	1.9 (1.9)
% SV	64.0 (39.4)	81.8 (36.8)	70.4 (39.7)	63.8 (26.6)	78.0 (29.4)	72.2 (28.8)	64.3 (26.9)	76.8 (28.6)	73.1 (29.8)
# Rel	3.8 (2.6)	3.8 (2.4)	3.7 (2.5)	3.3 (2.9)	3.0 (2.4)	3.2 (2.8)	5.0 (5.2)	4.6 (4.1)	4.5 (3.5)
% Rel	50.2 (32.0)	48.4 (30.3)	50.8 (34.3)	58.6 (35.2)	61.9 (30.0)	58.4 (32.4)	48.4 (32.1)	50.4 (33.0)	49.6 (33.3)
Avg CU	3.5 (3.5)	3.6 (3.6)	3.7 (3.7)	4.2 (4.2)	4.3 (4.3)	3.8 (3.8)	3.5 (3.5)	3.6 (3.6)	4.0 (4.0)
% CU	39.7 (35.1)	36.9 (33.8)	40.9 (34.9)	54.1 (37.6)	52.6 (37.9)	49.7* (37.3)	36.8 (34.3)	38.5 (35.0)	39.2 (34.6)
CoreLex words (31)	14.7 (7.1)	14.4 (7.6)	14.3 (6.4)	12.7 (8.4)	12.3 (8.8)	12.8 (8.9)	13.2 (7.3)	13.7 (7.3)	14.3 (7.8)

Note. Data are group means and (standard deviations). Pre Tx = pretreatment; Post Tx = posttreatment; wks = weeks; ACOM = Adaptive Aphasia Communication Outcome Measure; Pict desc = picture description; VNT = Verb Naming Test; CADL-3 = Communication Activities of Daily Living–Third Edition; CCRSA = Communication Confidence Rating Scale for Aphasia; MOS-SSS = Medical Outcome Study Social Support Scale; ALA: Wall Q = Assessment for Living with Aphasia: Wall Question; SV = grammatical utterances; Rel = relevant utterances; Avg = average; CoreLex words = number of core lexicon words.

*$p \leq .05$ in secondary analyses.

CADL-3. The main effect of time was significant from pretreatment to maintenance, ß = 3.91, *SE* = 1.25, *t*(80.97) = 3.13, *p* = .01. Because of the treatment coding, this statistic indicates that the large group showed a significant improvement from pretreatment to maintenance but that it was not reliably greater than the change seen in the control group (Brehm & Alday, 2022).

Discourse measures. In the analysis of monologic narratives, the number of complete utterances showed a significant main effect of time from pretreatment to maintenance, ß = 0.5, *SE* = 0.23, *t*(83.01) = 2.36, *p* = .02. The interaction between condition and time was not significant. In the core lexicon analysis, the large group condition produced more words from the core lexicon list at maintenance than pretreatment, ß = 1.02, *SE* = 0.52, *t*(80.96) = 1.97, *p* = .05. The interaction between condition and time was not significant.

The remaining standardized test scores and discourse measures showed no main effects of testing time or significant interactions between testing time and condition (see Supplemental Materials S5–S8), other than ones that were also observed in the omnibus analysis.

Post Hoc Analyses: Dyad Versus Control Condition

Primary Outcome Measure

For the ACOM, there was no main effect or significant interaction between condition and time for the dyads compared to controls (see Supplemental Material S9).

Secondary Outcome Measures

VNT. The main effect of time was significant from pre- to posttreatment, ß = 1.43, *SE* = 0.47, *t*(138.46) = 3.04, *p* < .01. The interaction was not significant, suggesting that the magnitude of the effect was not significantly greater in the dyadic than control condition.

CADL-3. The main effect of time was significant from pretreatment to maintenance, ß = 3.14, *SE* = 1.31, *t*(82.0) = 2.39, *p* = .02. The interaction was not significant, suggesting that the magnitude of the effect was not significantly greater in the dyadic than control condition.

Discourse measures. In the analysis of monologic narratives, the percentage of complete utterances showed a significant main effect of time from pretreatment to maintenance, ß = −4.46, *SE* = 2.01, $t(82.00) = -2.22$, $p = .03$. The interaction between condition and time was also significant from pretreatment to maintenance, ß = 5.69, *SE* = 2.81, $t(82.00) = 2.02$, $p = .05$. The direction of the effects indicates that the dyadic condition achieved lower percent complete utterances at maintenance than pretreatment, whereas the control condition did not show a significant change from pretreatment to maintenance. The remaining standardized test scores and discourse measures showed no main effects of testing time or significant interactions between testing time and condition (see Supplemental Materials S9–S12).

Discussion

This Phase 1b study investigated the effects of conversation treatment on patient-reported and directly measured changes in linguistic ability, communication ability, and psychosocial health for PwA. In addition, we investigated how difference in the group size (dyad compared to large group) influences treatment outcomes. Overall, our findings reveal significant change (improvement and treatment effect) on outcome measures in both experimental treatment conditions (dyad and large group) but *not* in the no-treatment control condition. The primary outcome measure showed significant changes only in the large group condition. Furthermore, the results point to a complex association between group size and treatment outcomes.

Primary Outcome Measure

Effect of Treatment on Participant-Perceived Communicative Effectiveness

Our primary outcome measure was the ACOM, a patient-reported outcome measure that reflects how effectively the PwA feel they communicate in different settings. The omnibus analysis of the ACOM showed a null result, with neither the treatment group nor the control group showing a statistically significantly greater improvement. This statistic offers no support for the efficacy of conversation treatment on this primary measure. However, secondary analyses showed that the large group condition demonstrated significantly greater improvements on the ACOM than the control group following treatment, indicating a treatment effect. This effect did not persist at maintenance. No significant treatment effect was observed in the dyads posttreatment or at maintenance, although they also showed a positive trajectory from pre- to posttreatment. In this way, we demonstrate that conversation treatment produces significant changes in the self-reported communication ability when delivered in a large group format.

One critical finding from the RELEASE (RELEASE Collaborators, 2022) systematic review was that a dosage of 2–4 hr of weekly speech-language therapy results in the greatest clinical gains in the overall language and functional communication. While group aphasia interventions were included in this review, the proportion of group interventions was relatively small (6.4%) and based on individual participant data (Williams et al., 2022). It is, therefore, unknown whether these recommendations apply to dyadic and/or group conversation treatments. Nonetheless, it may be that a greater treatment dosage (i.e., hours of treatment) is required to show maintenance of effects.

Secondary Outcome Measures

Effects of Conversation Treatment on Linguistic Ability

The selected subtests from the CAT and the VNT serve as the linguistic measures from our assessment protocol. In the planned analysis, significant changes on these measures were observed *only* in the large group and dyad treatment conditions. On the CAT naming subtest, both treatment groups showed significant effects immediately after treatment relative to the large group. The large group condition, but not the dyad condition, maintained the treatment effects for 6 weeks posttreatment. The dyad condition also showed significant effects of treatment on the CAT repetition total score. In the post hoc analyses, there was additional evidence of change on the VNT in the dyad condition. The dyad condition showed a significant improvement from pre- to posttreatment. Although the interaction was not significant, inspection of the means reveals that the control condition showed poorer performance from pre- to posttreatment on the VNT. This trend suggests that the effect was largely driven by the dyad condition. Together, these results support that conversation treatment is associated with linguistic change for PwA on some language-specific measures. Critically, these results indicate that treatment at the level of discourse can generalize to discrete linguistic tasks such as naming and repetition.

Effects of Conversation Treatment on Functional Communication

The planned analyses did not show any significant effects on the discourse measures or on the CADL-3, which served as a standardized measure of functional communication. Post hoc analyses, run separately for each treatment condition, showed statistically significant changes on discourse measures for the large group analysis only. Both treatment conditions showed significant improvement from pretreatment to maintenance on the CADL-3, but the interaction with condition was not significant. As with the VNT, inspection of the means suggests that the effects were largely driven by the treatment conditions. Change scores on the CADL-3 were 1.2 for the controls, 3.1 for the dyad condition, and 3.9 for the large group. The null interaction may reflect the sample size; the CADL-3 was added in the second study cycle and, thus, has a smaller sample size. Another reason might be ceiling effects among some of the participants. The CADL-3 allows credit for communication across multiple modalities and relies largely on gist versus complexity. Thus, it is sensitive to the overall communication for individuals with more moderate and severe profiles of aphasia. It may be that within our smaller sample size, some of whom had milder aphasia profiles, there was a ceiling effect with this measure pre- to posttreatment. A third possibility is that participants with more severe aphasia may need more time to realize treatment effects. Further research is needed to determine which, if any, of these account for the observed results.

Effects of Conversation Treatment on Measures of Psychosocial Health

Three measures in the battery addressed aspects of quality of life or general well-being. The results provide no evidence that conversation treatment directly impacts psychosocial health on these measures. There are several possible accounts for these results. One possibility is that the findings reflect a smaller sample size, because the psychosocial health measures were only added in the second cycle. Another possibility is that our data do not reflect the gains one might expect from a first-time exposure to group treatments as most of the recruited participants had some previous exposure to aphasia groups. Thus, they may have already experienced the psychosocial benefits of a group prior to the study. The lack of findings may also reflect the treatment dosage or reduced sensitivity of our measures to treatment change. Previous studies that demonstrated changes in psychosocial health used qualitative interviewing methods following a substantially larger treatment dosage (e.g., Elman & Bernstein-Ellis, 1999b; Van der Gaag et al., 2005; Vickers, 2010). Finally, changes in the quality of life may take more time to achieve. Along these lines, Worrall et al. (2024) recently showed that only selected domains from the ALA were sensitive to change following an intensive aphasia treatment, and only at a later time point (i.e., not immediately posttreatment). A longitudinal evaluation of participation in an aphasia center reported that the ALA showed significant change following 1 year of participation, which was maintained at Year 2 (Edmonds & Morgan, 2022). Further research is needed to examine these possibilities.

Influence of Group Size on Conversation Treatment Outcomes

The second research question was whether group size influences the benefits of conversation treatment for PwA. Consistent with our previous work (DeDe et al., 2019), the study found a different pattern of effects of conversation treatment delivered in a large group versus a dyad. However, there was no evidence that one treatment condition was statistically better than the other across the board, with advantages depending on the outcome measure.

Both conditions showed positive changes on naming after conversation treatment. In addition, large groups show superior benefits on a measure of functional communication, the ACOM, as well as on some discourse measures. In contrast, the dyad condition showed additional changes on a measure of repetition and verb naming. These results largely mirror the findings from our earlier studies (DeDe et al., 2019; Hoover et al., 2021), where results similarly pointed to a complex relationship between group size and outcome measures. In both our earlier studies and the present work, a few treatment effects persisted at maintenance. These results may suggest the need for a greater treatment dosage or duration. The only other RCT of the communication group treatment comprised a significantly larger treatment dosage, 80 hr over 16 weeks compared to the 20 hr over 10 weeks delivered in this study (Elman & Bernstein-Ellis, 1999a).

One important distinction between our earlier work and the present study is that previously, data were analyzed using nonparametric statistics within condition rather than directly asking whether improvements were statistically greater in the treatment versus control conditions. While this larger study provided the opportunity for stronger statistical analyses, there remain challenges. Compared to DeDe et al. (2019), the relatively modest results likely reflect the use of statistical models that included both treatment and control conditions. Regardless, the differing statistical outcomes are consistent with the hypothesis that manipulating group size alters the relative influence of psychodynamics and language and communication practice. Greater change on ACOM reinforces the group dynamics hypothesis, suggesting that increased opportunities for vicarious learning, clinician models, communication practice, and psychosocial support are associated with change in self-reported functional communication. Dyads show superior benefits on an additional-task specific standardized test (i.e., repetition), which is consistent with the hypothesis that greater benefits result from a larger number of communication trials (i.e., dosage hypothesis).

Consistent with our earlier work, both treatment conditions showed significant improvement on the CAT naming total score. This finding is not surprising given the heterogeneity in profiles and severity across the participants in the study. We purposely enrolled PwA with any profile of aphasia. Anomia, or naming difficulty, is widely established as the cardinal feature of aphasia and, as such, the only symptom shared by all participants regardless of aphasia severity. Measures of naming, therefore, would provide some opportunity for change and minimize the likelihood of a ceiling effect for the entire cohort. Additionally, the complexity account (Elman, 2011; Thompson et al., 2003), which suggests that practicing a more complex task will generalize to a simpler task, might help explain this generalization. Word retrieval in conversation is considered more complex than single-word retrieval in a confrontation naming task; thus, we see generalization of naming in conversation to single-word naming tests.

Limitations

This Phase 1b study is a relatively large multisite randomized controlled clinical trial with a heterogenous cohort of individuals with aphasia. It represents both a replication and extension of an earlier RCT (DeDe et al., 2019). Nonetheless, the sample size within each condition is modest relative to some other RCTs (McGill et al., 2020). Furthermore, measures added in the second cycle had a smaller sample size than those included across the entire study.

It is important to acknowledge possible sources of bias in our data. Given the nature of our treatment, it was not possible to blind treating clinicians or participants to condition. Whenever possible, the researchers administering the assessments were blinded to condition, but resources did not permit this for all participants and participants may have indicated their treatment condition to the examiner. It is possible that this concern introduced bias in favor of experimental conditions to the study. This concern was somewhat mitigated by the use of equipoise during all communications. Another possible source of bias was attrition, which was less than 10%. After randomization, three participants withdrew because they did not like the treatment, one due to unrelated illness, one for scheduling reasons, and four for unknown reasons. Because of the relatively modest sample size, we did not impute missing data, instead opting for a modified intention-to-treat analysis in which missing data are omitted.

The current study enrolled individuals with mixed profiles of aphasia, which introduced variability in terms of individual treatment goals and which measures might show effects of treatment. We made this decision to reflect our current clinical practice of including a wide range of aphasia types in conversation treatment groups. However, it is unknown how group variability influences outcomes. Given this variability, it is encouraging that positive treatment effects were detected. Future work will be targeted at further explicating the optimal parameters for conversation treatment, prior to a larger scale efficacy study.

Conclusions

In summary, this study represents the largest and most rigorous examination of conversation group treatment of aphasia to date. Our first research question was whether conversation treatment for PwA is associated with self-reported and directly measured changes in linguistic ability, communication ability, and psychosocial health. The results offer a qualified yes in answer to this question. Our results provide evidence of significant change (improvement and treatment effect) on outcome measures in both experimental treatment conditions (dyad and large group) but *not* in the natural history (no-treatment) control condition. The caveat is that the primary outcome measure did not show significant changes in the dyad condition. The second question was whether the pattern of change on outcome measures differed for large groups compared to dyads. Although this question was not analyzed statistically, the pattern of results suggests that manipulation of group size affects treatment outcomes. Thus, this study also sheds light on how the critical ingredients in conversation treatment may affect outcomes. Together, we argue that the results of this study offer support for a *cost-effective* treatment option, in comparison to traditional individual (1:1) treatment, for individuals living with aphasia across the continuum of care.

Data Availability Statement

Data are not publicly available at this time. Data will be provided to researchers upon reasonable request to the corresponding author.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grants R21DC015859 and R01DC018781, awarded to Elizabeth Hoover and Gayle DeDe. The authors extend their deepest gratitude to the participants with aphasia and their care partners, without whom this work would not be possible. We also thank the small army of students and volunteers at Boston University, Temple University, and Adler Aphasia Center who were integral to the completion of this project.

References

Archer, B., Tetnowski, J. C., Feer, S., Schmadeke, E., & Christou-Franklin, E. (2019). Topic selection sequences in aphasia conversation groups. *Aphasiology, 32*(4), 394–416. https://doi.org/10.1080/02687038.2017.1413705

Babbitt, E. M., Heinemann, A. W., Semik, P., & Cherney, L. R. (2011). Psychometric properties of the Communication Confidence Rating Scale for Aphasia (CCRSA): Phase 2. *Aphasiology, 25*(6–7), 727–735. https://doi.org/10.1080/02687038.2010.537347

Boyle, M., Akers, C. M., Cavanaugh, R., Hula, W. D., Swiderski, A. M., & Elman, E. J. (2023). Changes in discourse informativeness and efficiency following communication-based group treatment for chronic aphasia. *Aphasiology, 37*(3), 563–597. https://doi.org/10.1080/02687038.2022.2032586

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language, 125,* Article 104334. https://doi.org/10.1016/j.jml.2022.104334

Cavanaugh, R., Kravetz, C., Jarold, L., Quique, Y., Turner, R., & Evans, W. S. (2021). Is there a research–practice dosage gap in aphasia rehabilitation? *American Journal of Speech-Language Pathology, 30*(5), 2115–2129. https://doi.org/10.1044/2021_AJSLP-20-00257

Cavanaugh, R., Olivieri-Mui, B., & Smith, L. (2023, November). *Treatment dosage in aphasia rehabilitation: From research efficacy to real-world evidence* [Platform presentation]. American Speech-Language-Hearing Association Annual Convention, Boston, MA.

Cherney, L. R., & Babbit, E. (2011). *Communication Confidence Rating Scale for Aphasia (CCRSA)*. Rehabilitation Institute of Chicago. https://doi.org/10.1037/t33851-000

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Cruice, M., Worrall, L., Hickson, L., & Murison, R. (2003). Finding a focus for quality of life with aphasia: Social and emotional health, and psychological well-being. *Aphasiology, 17*(4), 333–353. https://doi.org/10.1080/02687030244000707

Dalton, S. G. H., Kim, H., Richardson, J. D., & Wright, H. H. (2020). A compendium of core lexicon checklists. *Seminars in Speech and Language, 41*(1), 45–60. https://doi.org/10.1055/s-0039-3400972

Dalton, S. G., Stark, B. C., Fromm, D., Apple, K., MacWhinney, B., Rensch, A., & Rowedder, M. (2022). Validation of an automated procedure for calculating core lexicon from transcripts. *Journal of Speech, Language, and Hearing Research, 65*(8), 2996–3003. https://doi.org/10.1044/2022_JSLHR-21-00473

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, Language, and Hearing Research, 62*(5), 1437–1451. https://doi.org/10.1044/2019_JSLHR-L-18-0404

Edmonds, L. A., & Morgan, J. (2022). Two-year longitudinal evaluation of community aphasia center participation on linguistic functional communication and quality of life measures across people with a range of aphasia presentations. *American Journal of Speech-Language Pathology, 31*(5S), 2378–2394. https://doi.org/10.1044/2022_AJSLP-21-00308

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology, 23*(3), 402–424. https://doi.org/10.1080/02687030802291339

Ellis, C., & Peach, R. K. (2017). Racial-ethnic differences in word fluency and auditory comprehension among persons with poststroke aphasia. *Archives of Physical Medicine and Rehabilitation, 98*(4), 681–686. https://doi.org/10.1016/j.apmr.2016.10.010

Ellis, C., Simpson, A. N., Bonilha, H., Mauldin, P. D., & Simpson, K. N. (2012). The one-year attributable cost of poststroke aphasia. *Stroke, 43*(5), 1429–1431. https://doi.org/10.1161/STROKEAHA.111.647339

Elman, R. J. (Ed.). (2007a). *Group treatment of neurogenic communication disorders: The expert clinician's approach* (2nd ed.). Plural.

Elman, R. J. (2007b). The importance of aphasia group treatment for rebuilding community and health. *Topics in Language Disorders, 27*(4), 300–308. https://doi.org/10.1097/01.TLD.0000299884.31864.99

Elman, R. J. (2011). Starting an aphasia center? *Seminars in Speech and Language, 32*(3), 268–272. https://doi.org/10.1055/s-0031-1286180

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*(2), 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. https://doi.org/10.1055/s-2008-1064009

Flowers, H. L., Skoretz, S. A., Silver, F. L., Rochon, E., Fang, J., Flamand-Roze, C., & Martino, R. (2016). Poststroke aphasia frequency, recovery, and outcomes: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation, 97*(12), 2188–2201.e8. https://doi.org/10.1016/j.apmr.2016.03.006

Gialanella, B., Bertolinelli, M., Lissi, M., & Prometti, P. (2011). Predicting outcome after stroke: The role of aphasia. *Disability and Rehabilitation, 33*(2), 122–129. https://doi.org/10.3109/09638288.2010.488712

Guo, Y. E., Togher, L., & Power, E. (2014). Speech pathology services for people with aphasia: What is the current practice in Singapore? *Disability and Rehabilitation, 36*(8), 691–704. https://doi.org/10.3109/09638288.2013.804597

Hart, T., Dijkers, M. P., Whyte, J., Turkstra, L. S., Zanca, J. M., Packel, A., van Stan, J. H., Ferraro, M., & Chen, C. (2019). A theory-driven system for the specification of rehabilitation treatments. *Archives of Physical Medicine and Rehabilitation, 100*(1), 172–180. https://doi.org/10.1016/j.apmr.2018.09.109

Holland, A. L., Fromm, D., & Wozniak, L. (2018). *Communication Activities of Daily Living–Third Edition (CADL-3)*. Pro-Ed.

Hilari, K., Needle, J. J., & Harrison, K. L. (2012). What are the important factors in health-related quality of life for people with aphasia? A systematic review. *Archives of Physical Medicine and Rehabilitation, 93*(1), S86–S95. https://doi.org/10.1016/j.apmr.2011.05.028

Holmes, S. E., & Kivlighan, D. M., Jr. (2000). Comparison of therapeutic factors in group and individual treatment processes. *Journal of Counseling Psychology, 47*(4), 478–484. https://doi.org/10.1037/0022-0167.47.4.478

Hoover, E., DeDe, G., & Maas, E. (2021). A randomized controlled trial of the effects of group conversation treatment on monologic discourse in aphasia. *Journal of Speech, Language, and Hearing Research, 64*(12), 4861–4875. https://doi.org/10.1044/2021_JSLHR-21-00023

Hula, W. D., Doyle, P. J., Stone, C. A., Austermann Hula, S. N., Kellough, S., Wambaugh, J. L., Ross, K. B., Schumacher, J. G., & St. Jacque, A. (2015). The Aphasia Communication Outcome Measure (ACOM): Dimensionality, item bank calibration, and initial validation. *Journal of Speech, Language, and Hearing Research, 58*(3), 906–919. https://doi.org/10.1044/2015_JSLHR-L-14-0235

Jacobs, M., & Ellis, C. (2023). Aphasianomics: Estimating the economic burden of poststroke aphasia in the United States. *Aphasiology, 37*(1), 25–38. https://doi.org/10.1080/02687038.2021.1985426

Kagan, A., Simmons-Mackie, N., & Victor, J. (2018). *Assessment for Living with Aphasia toolkit (ALA)–Second Edition*. Aphasia Institute.

Kearns, K., & Elman, R. J. (2008). Group therapy for aphasia: Theoretical and practical considerations. In R. Chapey (Ed.). *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 376–400). Lippincott Williams & Wilkins.

Kertesz, A. (2007). *Western Aphasia Battery–Revised*. PsychCorp.

Koleck, M., Gana, K., Lucot, C., Darrigrand, B., Mazaux, J. M., & Glize, B. (2017). Quality of life in aphasic patients 1 year after a first stroke. *Quality of Life Research, 26*(1), 45–54. https://doi.org/10.1007/s11136-016-1361-z

Kong, A. P., & Tse, C. W. (2018). Clinician survey on speech pathology services for people with aphasia in Hong Kong. *Clinical Archives of Communication Disorders, 3*(3), 201–212. https://doi.org/10.21849/cacd.2018.00409

Lam, J. M., & Wodchis, W. P. (2010). The relationship of 60 disease diagnoses and 15 conditions to preference-based health-related quality of life in Ontario hospital-based long-term care residents. *Medical Care Journal, 48*(4), 380–387. https://doi.org/10.1097/MLR.0b013e3181ca2647

Luterman, D. (1991). *Counseling the communicatively disordered and their families*. Pro-Ed.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Mayo, N. E., Anderson, S., Barclay, R., Cameron, J. I., Desrosiers, J., Eng, J. J., Huijbregts, M., Kagan, A., MacKay-Lyons, M., Moriello, C., Richards, C. L., Salbach, N. M., Scott, S. C., Teasell, R., & Bayley, M. (2015). Getting on with the rest of your life following stroke: A randomized trial of a complex intervention aimed at enhancing life participation post stroke. *Clinical Rehabilitation, 29*(12), 1198–211. https://doi.org/10.1177/0269215514565396

McGill, K., Sackley, C. M., Godwin, J., McGarry, J., & Brady, M. C. (2020). A systematic review of the efficiency of recruitment to stroke rehabilitation randomised controlled trials. *Trials, 21,* Article 68. https://doi.org/10.1186/s13063-019-3991-2

National Quality Forum. (2016). *Measuring performance*. https://www.qualityforum.org/Measuring_Performance/Measuring_Performance.aspx

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators. (2022). Dosage, intensity, and frequency of language therapy for aphasia: A systematic review–based, individual participant data network meta-analysis. *Stroke, 53*(3), 956–967. https://doi.org/10.1161/STROKEAHA.121.035216

Rose, M., Ferguson, A., Power, E., Togher, L., & Worrall, L. (2014). Aphasia rehabilitation in Australia: Current practices, challenges and future directions. *International Journal of Speech-Language Pathology, 16*(2), 169–180. https://doi.org/10.3109/17549507.2013.794474

Shadden, B. B. (2005). Aphasia as identity theft: Theory and practice. *Aphasiology, 19*(3–5), 211–223. https://doi.org/10.1080/02687930444000697

Sherbourne, C. D., & Stewart, A. L. (1991). The MOS social support survey. *Social Science & Medicine, 32*(6), 705–714. https://doi.org/10.1016/0277-9536(91)90150-B

Simmons-Mackie, N., & Cherney, L. R. (2018). Aphasia in North America: Highlights of a white paper. *Archives of Physical Medicine and Rehabilitation, 99*(10), Article E117. https://doi.org/10.1016/j.apmr.2018.07.417

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

Swinburn, K., Howard, D., & Porter, G. (2004). *CAT: Comprehensive Aphasia Test*. Psychology Press. https://doi.org/10.1037/t13733-000

Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE). *Journal of Speech, Language, and Hearing Research, 46*(3), 591–607. https://doi.org/10.1044/1092-4388(2003/047)

Van Der Gaag, A., Smith, L., Davis, S., Moss, B., Cornelius, V., Laing, S., & Mowles, C. (2005). Therapy and support services for people with long-term stroke and aphasia and their relatives: A six-month follow-up study. *Clinical Rehabilitation, 19*(4), 372–380. https://doi.org/10.1191/0269215505cr785oa

Van Stan, J. H., Dijkers, M. P., Whyte, J., Hart, T., Turkstra, L. S., Zanca, J. M., & Chen, C. (2019). The rehabilitation treatment specification system: Implications for improvements in research design, reporting, replication, and synthesis. *Archives of Physical Medicine and Rehabilitation, 100*(1), 146–155. https://doi.org/10.1016/j.apmr.2018.09.112

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. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L .R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A. C., Marshall, J., . . . Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke, 14*(2), 180–185. https://doi.org/10.1177/1747493018806200

Williams, L. R., Myzoon, A., VandenBerg, K., Williams, L. J., Abo, M., Becker, F., Bowen, A., Brandenburg, C., Breitenstein, C., Bruehl, S., Copland, D. A., Cranfill, T. B., Di Pietro-Bachmann, M., Enderby, P., Fillingham, J., Galli, F. L., Gandolfi, M., Glize, B., Godecke, E., . . . The RELEASE Collaborators. (2022). Utilising a systematic review-based approach to create a database of individual participant data for meta- and network meta-analyses: The RELEASE database of aphasia after stroke. *Aphasiology, 36*(4), 513–533. https://doi.org/10.1080/02687038.2021.1897081

Worrall, L. E., Rodriguez, A. D., Dignam, J. K., Hill, A. J., Khan, Asaduzzaman., O'Connor, D. A., Pattie, M., Ward, E. C., McKinnon, E., Allan, R., & Copland, D. A. (2024). A modified intensive, comprehensive aphasia program (mICAP) has better reported outcomes than usual care: A randomized controlled trial. *Aphasiology,* 1–20. https://doi.org/10.1080/02687038.2024.2424493

Yalom, I. D., & Leszcz, M. (2005). *Theory and practice of group psychotherapy*. Basic Books.

Zanca, J. M., Turkstra, L. S., Chen, C., Packel, A., Ferraro, M., Hart, T., Van Stan, J. H., Whyte, J., & Dijkers, M. P. (2019). Advancing rehabilitation practice through improved specification of interventions. *Archives of Physical Medicine and Rehabilitation, 100*(1), 164–171. https://doi.org/10.1016/j.apmr.2018.09.110