Aphasiology

ISSN: 0268-7038 (Print) 1464-5041 (Online) Journal homepage: www.tandfonline.com/journals/paph20

Effects of conversation group treatment for individuals with moderate to severe aphasia

Gayle DeDe, Gretchen Szabo, Francine Kohen, Sarah Vitale, Nicholas McCloskey, Yorghos Tripodis & Elizabeth Hoover

To cite this article: Gayle DeDe, Gretchen Szabo, Francine Kohen, Sarah Vitale, Nicholas McCloskey, Yorghos Tripodis & Elizabeth Hoover (19 Dec 2025): Effects of conversation group treatment for individuals with moderate to severe aphasia, Aphasiology, DOI: 10.1080/02687038.2025.2602510

To link to this article: https://doi.org/10.1080/02687038.2025.2602510

View supplementary material

Published online: 19 Dec 2025.

Submit your article to this journal

Article views: 31

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=paph20

APHASIOLOGY
https://doi.org/10.1080/02687038.2025.2602510

Effects of conversation group treatment for individuals with moderate to severe aphasia

Gayle DeDe[a], Gretchen Szabo[b], Francine Kohen[a], Sarah Vitale[c], Nicholas McCloskey[a], Yorghos Tripodis[d] and Elizabeth Hoover[c]

[a]Department of Communication Sciences and Disorders, Temple University, Philadelphia, USA; [b]Aphasia Communication Groups & Research, Adler Aphasia Center, Maywood, USA; [c]Department of Speech, Language, and Hearing Sciences, Boston University, Boston, USA; [d]Boston University School of Public Health, Boston University, Boston, USA

ABSTRACT

Background: Conversation treatment is effective for individuals with aphasia, but the optimal ingredients are poorly understood. This study examined factors related to severity of aphasia and characteristics of the group setting.

Aim: We examined whether individuals with moderate to severe aphasia (IwMSA) benefit from conversation treatment, and whether treatment effects differ based on group size (dyad vs. large group) or composition (mixed vs. homogenous aphasia severity).

Method: IwMSA (*n* = 91) were assigned to one of five conditions: delayed control, mixed or homogenous dyad, mixed or homogenous large group. Treatment was 60-minutes, twice weekly, for 10 weeks. Individualized goals were addressed within conversational group treatment. During each session, multi-modal supports were available and modeled. The primary outcome measure was the Aphasia Communication Outcome Measure (ACOM). Secondary outcome measures included the CADL-3, Comprehensive Aphasia Test (CAT), and discourse measures (e.g. Complete Utterances and Core Lexicon).

Results: We first examined overall effects of conversation treatment (treated versus control group). On the ACOM (primary outcome measure) and CADL-3 (secondary outcome measure), the treated condition showed significant improvement from pre to post treatment, and at maintenance. The controls did not show significant differences. The interaction of time and condition did not approach significance in any of the other secondary outcome measures, so only ACOM and CADL-3 were analyzed in follow up analyses. Regarding effects of group size (Large Group vs Dyad), the Large Group Condition showed significant changes for pre- to post-treatment and at maintenance on the ACOM. For the Dyad Condition, there was a significant change only from pre-treatment to maintenance. On the CADL-3, both the Large Group and Dyad conditions showed significant changes from pre to post treatment and at maintenance. Regarding effects of group composition (homogeneous vs mixed), individuals in the homogenous group

ARTICLE HISTORY
Received 7 June 2025
Accepted 5 December 2025

KEYWORDS
Conversation group treatment; randomized controlled trial, mod-severe aphasia

CONTACT Elizabeth Hoover ehoover@bu.edu Department of Speech, Language, and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, MA 02215, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/02687038.2025.2602510
© 2025 Informa UK Limited, trading as Taylor & Francis Group

demonstrated significant improvements from pre- to post-treatment and a maintenance on both the ACOM and CADL-3. Individuals in the mixed group showed no significant change over time
Conclusions: Participants in the treatment conditions showed significant changes on the ACOM and CADL-3, demonstrating that IwMSA benefit from conversation treatment. These findings support the hypotheses that IwMSA benefit from conversation treatment in both large groups and dyads, and that there may be additional benefits associated with more homogeneous (versus mixed) group environments.

Introduction

Approximately 25% of stroke survivors have persistent aphasia three months after stroke (Ali et al., 2015). Among these survivors, initial aphasia severity (e.g., global aphasia) is associated with poorer language recovery and has been shown to persist into chronic states. Individuals with severe aphasia also experience social isolation and significantly poorer health-related quality of life outcomes compared to less severely impaired peers (Hilari & Byng, 2009; Parr, 2007). From a health economics perspective, the burden of severe aphasia is likely to be greater than that of less severe aphasia (c.f., Ellis et al., 2012; Jacobs & Ellis, 2023).

Given these negative outcomes, it is critical to develop cost-effective treatments for individuals with severe aphasia. Early studies suggested that those with global aphasia showed minimal improvements in spoken language and suggested a focus on alternative means of communication (e.g., gestures). However, observational studies suggest that individuals with global aphasia show ongoing, slow improvements in all communication modalities (e.g., Smania et al., 2010). It may be that this population responds differently to treatment than those with mild to moderate profiles of aphasia severity.

There has been a growing interest in developing effective treatments for individuals with moderate to severe aphasia (IwMSA). Some of these approaches have focused on impairment-based treatments targeting specific linguistic functions (e.g., Edmonds & Babb, 2011; Knollman-Porter et al., 2018). Others have focused on therapeutic conversation (e.g., (Basso, 2009; Leaman & Edmonds, 2023), augmentative communication (Nicholas & Elliott, 1998; Steele et al., 1989), or functional, multimodal communication (e.g., Elman & Bernstein-Ellis, 1999; Hinckley, 2014; Hoover et al., 2020; Lyon & Sims, 1989; Roper et al., 2016). However, most of these studies are pilot studies or small in scale (e.g., ≤ 4 participants).

Conversation group treatment has the potential to address both communication and psychosocial needs of individuals with moderate-severe to severe aphasia. In their seminal study of group communication treatment, Elman and Bernstein-Ellis (1999) examined subgroups of participants with moderate to severe aphasia ($n = 12$) and mild-moderate aphasia. Both severity groups improved on the Western Aphasia Battery (WAB) Aphasia Quotient and Short-Porch Index of Communication Ability (SPICA). However, a two-way ANOVA revealed a treatment-by-severity effect on the Communicative Activities of Daily Living (CADL), such that the more severely impaired group showed greater changes on

the CADL compared to individuals with milder aphasia. This was attributed to potential ceiling effects in the less severely impaired group, who performed near ceiling on the CADL on pre-treatment testing.

The implementation of conversation group treatment is highly variable (Simmons-Mackie et al., 2014). The Rehabilitation Treatment Specification System (RTSS; Hart et al., 2019) is a method for describing mechanisms of action and critical ingredients of a treatment. Hoover et al. (2021) applied the RTSS to group conversation treatment, in an effort to identify key components of the treatment and hypothesize potential mechanisms of action driving behavioral change. Critical ingredients were grouped into three domains: communication practice (e.g., turn taking, initiation, multimodal communication), linguistic practice (comprehension of discourse, production of words and sentences), and group dynamics (psychosocial support and vicarious learning). However, these features of treatment cannot reasonably be disentangled. Thus, we have manipulated the relative amount of language/communication practice and group dynamics by varying group size. Our logic is that a large group of six to eight people with aphasia will offer fewer opportunities for expressive communication, but more opportunities for group dynamics. In contrast, a dyad of two people with aphasia would offer more opportunities to practice expressive communication, but fewer benefits of group dynamics. These assumptions have been borne out in our previous studies of conversation treatment (cf. DeDe et al., 2019; Sharkey et al., 2024).

Hoover et al. (2020) examined effects of conversation treatment in five participants with severe aphasia, who were extracted from a larger study of conversation treatment. Four of these participants completed dyadic conversation treatment (two people with aphasia and one clinician), and one participated in heterogenous large group treatment (six to eight people with mild to severe aphasia profiles). Hoover et al. did not report consistent effects of treatment in these individuals on a standardized language battery (Comprehensive Aphasia Test subsection scores), the Aphasia Communication Outcome Measure, or discourse measures; instead, individuals showed change on different measures. Hoover and colleagues posited several potential causes of their results, including the small number of participants, the selected measures not being sensitive to change, or that individuals with severe aphasia do not benefit from conversation treatment. Another issues was that only one of the participants with severe aphasia was randomly assigned to receive treatment in the large group format. Thus, additional research is needed to better understand the optimal parameters in which IwMSA might benefit from conversation treatment. The present study builds on our previous work (DeDe et al., 2019; Hoover et al., 2020, 2021, 2025) by (1) examining whether a larger sample of IwMSA are responsive to conversation treatment, and (2) examining effects of group size and group composition.

Group size: large group versus dyad

First, we investigated whether IwMSA are responsive to conversation treatment delivered in a dyad (two people with aphasia) or a large group (6–8 people with aphasia), as compared to a no-treatment control condition. In the context of conversation treatment, there are two competing hypotheses that may underly decisions about group size. According to the Treatment Dosage Hypothesis, linguistic and communication practice are critical ingredients of conversation treatment (cf. DeDe et al., 2019; Hoover et al.,

2025). In the context of conversation treatment, a practice trial may be operationalized as a conversational turn. In our work, the large group and dyad conditions offer the same number of treatment hours. However, dyads offer a greater number of practice trials than large groups (DeDe et al., 2019) because there are fewer people competing for conversational turns. Thus, dyads afford greater opportunities for linguistic and communication practice than large groups. According to the Group Dynamics Hypothesis, the psychosocial support and vicarious learning opportunities afforded by a larger group are critical ingredients of conversation treatment (cf. DeDe et al., 2019; Hoover et al., 2025). Large groups offer greater opportunities to learn from and receive support from peers with aphasia.

Our previous work suggested that people with mixed aphasia profiles benefit from both dyadic and large group conversation treatment (e.g., DeDe et al., 2019; Hoover et al., 2025). However, the relative benefits of large group and dyad conversation treatment may differ for IwMSA. For example, IwMSA may require greater dosage to show treatment effects. Suggestive evidence can be found in Hoover et al. (2015). Hoover investigated effects of Verb Network Strengthening Treatment (VNeST) when administered in individual and group formats. Secondary analyses examined when treatment effects emerged in severely impaired participants. Those with mild and moderate aphasia began to show an upward trend after two weeks of treatment, whereas those with severe aphasia only showed the beginning of an upward trend after six weeks of treatment. Similarly, Edmonds and Babb (2011) reported that individuals with moderate-severe aphasia require more weeks of VNeST and show reduced treatment effects, as compared to individuals with moderate aphasia. If treatment dosage is particularly important for IwMSA, then this population may benefit more from dyadic than large group treatment settings.

On the other hand, IwMSA have extensive psychosocial needs (cf. Parr, 2007). Larger group may better address the psychosocial needs of IwMSA, by decreasing social isolation and providing a wider range of opportunities for vicarious learning. Larger groups may increase communication confidence, resulting in IwMSA engaging in more conversations outside the clinic. On this account, large groups may be more beneficial to IwMSA than dyads.

Group composition: mixed vs. Homogenous group treatment

Another question under investigation was whether group composition plays a role in treatment outcomes. At present, clinicians are likely to construct groups based on pragmatic considerations (e.g., availability of participants) and clinical intuition about the relative benefits of mixed versus homogenous groups. We investigated whether mixed or relatively homogenous groups are associated with different treatment benefits for IwMSA.

Mixed groups may contain a wide range of aphasia profiles, ranging from global aphasia to very mild or even latent aphasia (sometimes known as not aphasic by WAB). These groups expose participants to a diversity of more and less complex models and a wide range of learning opportunities. There is evidence that being exposed to a greater variation of models or practice opportunities results in greater learning (e.g., Gómez, 2002; Goode et al., 2008; Lee et al., 1985; Maas et al., 2008). Mixed groups also offer

opportunities to help and be helped by individuals with more or less severe aphasia. This premise relates to the concept of "helper-therapy" in the group dynamics literature. Helper therapy is the idea that the act of providing support to other group members is therapeutic for the individual providing support (Toseland & Rivas, 2017). A related body of literature focuses on the benefits of positive emotions and having purpose on stroke outcomes (Ostir et al., 2008; Yu et al., 2015).

On the other hand, homogenous groups may offer greater cohesiveness than mixed groups due to a greater degree of shared experiences (Yalom & Leszcz, 2005). Homogenous groups can also be tailored to individual needs. For example, individuals with severe aphasia are likely to use more compensatory strategies and external supports as compared to more mildly impaired participants. Rate of speech within the conversation may also be faster among individuals with mild aphasia. As a result, severely impaired participants may be better able to interject conversational turns in homogenous than mixed groups. In cycles 1 and 2 of this research, which included all aphasia severities (severe through latent), clinicians noted that more severely impaired participants took disproportionately fewer conversational turns in large groups as compared to their more mildly impaired peers.

In clinical settings, pragmatic concerns may drive choices about group composition. Yet, to date there is little evidence to support these decisions, and it is important to know whether there are advantages to one option. For this reason, we compared IwMSA who were enrolled in mixed compared to homogenous large groups to determine whether outcomes differed.

To summarize, the present study reports data from a non-randomized (quasi-experimental) controlled group design trial of conversation treatment in IwMSA. We addressed three research questions (RQ). The primary question (RQ 1) was whether IwMSA show treatment benefits following conversation treatment (i.e., treated versus control conditions). The secondary questions examined whether there was evidence for differential effects of group size (RQ 2: dyad versus large group) or group composition (RQ 3: homogeneous versus mixed large groups) in measures that were sensitive to treatment effects in research question 1.

Methods

This project was approved by the institutional review board (IRB) at Boston University (BU), which served as a single IRB for all sites (Temple University-TU; Adler Aphasia Center-AC).

Participants

This study included participants with moderate to severe aphasia (IwMSA). This profile was defined as a maximum WAB-R AQ of 70; with severe aphasia defined as WAB AQ < 51 and Moderate as WAB AQ 51–75. The cut-off at 70 allows for a distinct separation between profiles with Mild and Latent aphasia profiles which are the focus of another manuscript (in submission). When the WAB-R AQ was not available (see below), the cut off was having a Total Number Correct of less than 17 on the Comprehensive Aphasia Test Object Naming Section (Swinburn et al., 2004). This CAT Naming score represented more than one standard deviation below the mean score of all previous participants in the study (mixed cohorts from 2017 and 2022) and thus defined this current cohort as having relatively worse anomia, which is a cardinal feature of all profiles of aphasia.

Recruitment and randomization

To run studies of group conversation treatment, all participants must be recruited before testing or treatment can begin. Thus, the study was run in three cycles. In each cycle, participants were randomly assigned to condition. However, the number of randomization conditions was not the same across all cycles in the present study. This decision reflected pragmatic concerns, due to the challenges associated with recruiting IwMSA.

In cycles one and two, a mixed group of people with aphasia was recruited. 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 age of 6), and d) diagnosed with aphasia based on clinical judgment and standardized test results (Western Aphasia Battery-Revised [WAB-R]; Kertesz, 2006 or Comprehensive Aphasia Test [CAT]; Swinburn et al., 2004). After meeting inclusion criteria, participants were randomly assigned to one of three conditions using a 1:1:1 allocation ratio: a large group (6–8 people with aphasia), dyad (2 people with aphasia), and a natural history control group. 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. Analyses of the complete data set are reported in Hoover et al. (2025). For the purposes of this study, we extracted participants who met the criteria for moderate-severe to severe aphasia as defined above. Across these two cycles, there were 11 IwMSA in the control condition,18 in the dyads, and 15 in the large group.

Only IwMSA were recruited in cycle 3. Otherwise, the same inclusion criteria were used as in cycles 1 and 2. Participants were screened using the Western Aphasia Battery. The target recruitment goal at each site was 12 to 16 IwMSA. The Boston University team contacted 30 people with aphasia. Ten of these were screened out on the phone; 20 were screened using the WAB-R, and 16 were enrolled. Temple University contacted 33 people with aphasia. Of these, 13 did not respond or were screened out on the phone. Twenty were screened using the WAB-R, and 16 were enrolled. At the Adler Aphasia Center, 48 people with aphasia were contacted, 20 completed the WAB-R, and 15 were enrolled. Figure 1 shows the total number of participants at each site and testing point.

In cycle 3, participants who met screening criteria were randomly assigned to only one of two conditions using a 1:1 allocation ratio: the large group or dyad treatment condition. A block size of 4 was used to roughly match the number of people in each of the two conditions. Otherwise, randomization procedures paralleled those in cycles 1 and 2.

Table 1 presents demographic data and sample size for each condition. 84 IwMSA completed all phases of the study (11 in the control condition, 39 in the dyads, 34 in the large group (14 mixed; 20 homogenous). Participants randomized to the dyad condition were paired based on pragmatic concerns (e.g., scheduling) and clinician judgment of compatibility following random assignment. The rationale was to approximate how participants might be paired in clinical settings.

At BU and TU, treatment for the large group and dyad conditions occurred in the summer. The natural history control groups (cycles 1 and 2) were offered treatment in the fall for ethical reasons, but post-treatment data were not collected. At the AC, treatment was in the fall and the natural history control group was offered treatment in the spring. Participants were asked to abstain from other communication treatment during the study period. Informed written consent was obtained from all participants.

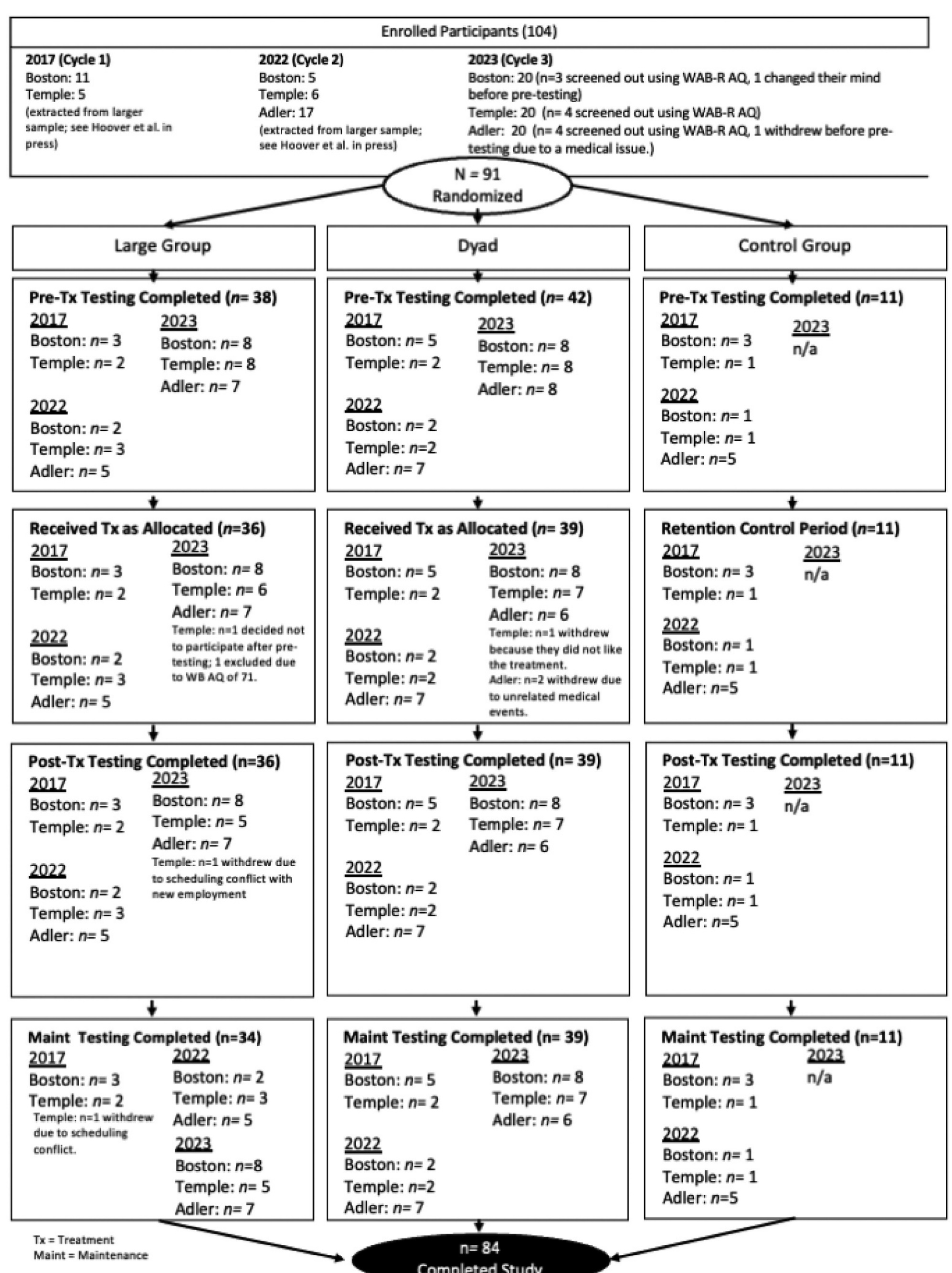

Figure 1. Recruitment and retention of participants across site and time.

Assessment Protocol

The testing protocol was administered at three time points: immediately prior to and following the experimental treatment and approximately 6-weeks post-treatment (maintenance). Given the number of participants, testing occurred within a two week period

Table 1. Demographic information & sample size. Values given are mean (standard deviation).

Measure		Large Group (*n* = 34)	Dyad (*n* = 39)	Control (*n* = 11)
Race/Ethnicity	Asian	*n* = 1	*n* = 2	*n* = 1
	Black or African- American	*n* = 6	*n* = 8	*n* = 3
	White	*n* = 27	*n* = 29	*n* = 7
Gender (self-reported)	Men	n = 22	n = 26	n = 10
	Women	n = 12	n = 13	n = 1
Age		61.77 (10.88; 34.7–77.9)	65.14 (13.80; 30.3–88.1)	65.53 (9.72; 49.6–76.8)
Years Education		14.38 (2.88; 10–21)	14.89 (2.57; 12–20)	16.27 (4.61; 8–20)
Time Post Onset		100.62 (97.43; 5.0 – 100.3)	88.86 (88.2; 5.0 – 420.7)	99.77 (69.74 49.9 – 207.8)
WAB-R AQ		38.99 (25.95; 5.0–67.1)	37.56 (23.75; 8.8 – 69.4)	30.89 (23.49; 31.9 – 65.3)
Naming Total Number Correct		8.17 (7.38; 0–15)	8.38 (8.07; 0–22)	7.45 (6.68; 0–20)

(e.g., within two weeks before the first treatment session). Assessors were blinded to treatment condition whenever possible. In addition, equipoise was maintained in all conversations with participants and clinicians. Testers were certified speech-language pathologists (SLPs) with extensive experience working with this population or graduate students in speech-language pathology supervised by the certified SLPs.

Treatment outcome measures were aligned with the constructs of a core outcome set for people with aphasia established by the collaboration of aphasia trialists (CATS; Wallace et al., 2019). The test battery included objective and patient-reported outcome measures across linguistic, communication and quality of life domains. Some measures were added after cycle one and thus, include a smaller sample size (see Table 2 for a complete list).

Table 2. Assessment battery.

Type	Measure	Domain	Dependent Variable
Primary Outcome Measure			
PROM	Aphasia Communication Outcome Measure (ACOM)	Self-reported communicative effectiveness	T-score
Secondary Outcome Measures			
Standardized Test (Linguistic)	Comprehensive Aphasia Test (CAT) *Language Section*	Naming, Repetition, Oral Reading, & Picture Description	Total Scores for each subsection
Standardized Test (Communication)	Communication Activities of Daily Living – 3rd Edition*	Functional communication	Raw Score (max = 100)
PROM	Assessment of Living with Aphasia (ALA) Wall Question*	Overall impact of aphasia on daily life.	Raw Score (Likert Scale)
PROM	MOS Social Support Scale*	Self-report of social support available to individuals.	Total Score
PROM	Communication Confidence Rating Scale for Aphasia (CCRSA)*	Self-reported communication confidence.	Total Score
Discourse (Linguistic)	CoreLex*	Key words produced in discourse	Number of CoreLex Words
Discourse (Linguistic)	Complete Utterances*	Relevance and grammaticality of discourse	Number of C-Units & Complete Utterances

Note: PROM = Patient Reported Outcome Measure; *Added after Cycle 1.

Primary Outcome Measure: The primary outcome measure was the *Aphasia Communication Outcome Measure* (ACOM; Hula et al., 2015). The ACOM is a psychometrically validated patient-reported outcome measure. Participants were asked to self-rate how effectively they communicate in a variety of functional situations such as "how effectively do you tell a story to your close friends and family" or "how effectively do you write a simple to do list". This measure was chosen to be sensitive to treatment changes across a wide range of aphasia severity levels and goals. In addition, this measure reflects the overall treatment aim of improved functional communication.

Secondary Outcome Measures: Six subtests from the *CAT* (spoken language comprehension, reading comprehension, naming, repetition, oral reading, and picture description) were included. 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. Auditory and reading comprehension are tested at the word and sentence levels. The naming subtest includes verbal fluency and confrontation naming for nouns and verbs. Repetition includes simple and complex words and non-words, as well as digit and sentence repetition. Oral reading includes content words, function words, and non-words. The picture description total score combines number of information carrying words with ratings of grammaticality, speech rate, and diversity of syntactic structures.

The remaining measures were collected only in cycles two and three. The *Communication Activities of Daily Living-3rd 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 CADL-3 was intended to provide a more objective measure of functional communication than the ACOM, which provides a self-reported measure of functional communication.

The Wall Question from the *Assessment of Living with Aphasia* (ALA: Wall Question; Kagan et al., 2018) is a simple questionnaire used to quantify the overall impact of aphasia on daily life. There is good reliability between the Wall Question and the entire ALA battery in determining the impact of aphasia on overall quality of life (Simmons-Mackie et al., 2007). This item evaluates how much communication ability acts as a barrier to participation in life activities. The *MOS Social Support Survey* (MOS; 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 for Aphasia* (CCRSA; Babbitt et al., 2011; Cherney & Babbit, 2011) is a psychometrically validated measure that evaluates self-reported communication confidence across a variety of communicative contexts.

Participants were also asked to produce four monologic picture descriptions (Cat Rescue, Birthday Scene, Refused Umbrella, and Broken Window; Nicholas & Brookshire, 1995) following AphasiaBank procedures (MacWhinney et al., 2011). Eleven participants (6 in the dyads and 5 in the large groups) produced 2 or fewer content words across all stimuli and time points; these participants were excluded from all analyses of the discourse samples. Picture descriptions were orthographically transcribed by trained research assistants and then checked by a second trained research assistant. A third research assistant independently transcribed 20% of the samples. 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 0.73, and 62% of samples had a score greater than 0.70. Inspection of transcripts with scores below .70 revealed deviations in spelling of non-words, fillers, and part word repetitions. Transcription reliability was relatively low due to the number of nonwords, pauses, and other fillers. Transcriptions were segmented into C-units and checked by trained research assistants. Disagreements noted during checks were reconciled between coders or via discussion with the PIs and supervising clinicians.

Discourse samples were coded for Complete Utterances and CoreLex after being de-identified for testing time, participant, and condition (cf. Dalton & Richardson, 2015; Edmonds et al., 2009). Complete Utterances are C-units that are relevant to the narrative (± rel) and contain a verb with all of its obligatory arguments (±SV). 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. Average reliability was 0.89 for Grammaticality (91.2% > 0.70), 0.84 for Relevance (91.2% > 0.70), and 0.88 for Complete Utterances (91.3% > 0.70). Data from all five picture description tasks – AphasiaBank stimuli plus the Grandpa scene from the CAT were averaged to increase reliability and stability of the measure (cf., Boyle, 2007).

Core Lexicon (CoreLex) is a reliable and valid measure of word production in discourse. Lexical items produced in the narrative were compared to lists of lexical items included in 50% or more of the normative samples. Published lists were available for three stimuli: Cat Rescue, Refused Umbrella, and Broken Window (Dalton & Richardson, 2015). CoreLex scores, including number of CoreLex words and CoreLex words per minute, were obtained using automated procedures (Dalton & Richardson, 2015).

In Cycle 3 only, an aphasia-friendly Treatment Impact Survey was also administered after treatment to understand the impact of the intervention for the participants and their care partners This survey followed guidelines proposed by Breitenstein et al. (2023) and included three questions related to the impact of the treatment on 1) communication 2) quality of life and 3) general well-being. Consistent with the guidelines proposed by Revicki et al. (2008), the prompts use a 5-point Likert scale ranging from −2= "much worse", − 1 = "slightly worse", 0= "no change", +1 = "slightly improved", +2= "much improved" to ask how the referenced areas have changed since the beginning of treatment.

Treatment

The treatment protocol for the dyad and large group conditions was identical at all sites and cycles. Treatment was offered twice weekly, in 60-minute sessions for 10 weeks (20 hours total). At BU and TU, the treatment was provided by trained graduate students under the supervision of licensed SLPs with extensive experience delivering conversation group treatment. Treatment at AC was delivered by licensed SLPs with previous experience in group treatment programming.

The conversation treatment protocol aligns with a socially-oriented approach and aims to foster naturalistic, dynamic conversations on functional and personally relevant topics (see Elman, 2007; Simmons-Mackie et al., 2014). One challenge in studying conversation treatment is that it is inherently unstructured and thus difficult to deliver consistently. We

therefore established several systems to increase consistency and reproducibility of the treatment between sites and cycles.

First, we developed five broad conversational topics, each with four associated subtopics (see Table 3). For example, personal history included family, hometowns, education/vocation, and hobbies. Treatment materials were developed for each subtopic and were shared across all sites via Microsoft™ Teams and One Drive. Materials included PowerPoint slides with visual cues and questions, conversation supports specific to scheduled topics (e.g., visual depictions of family roles/family members, list of music genres). In addition, general supports such as writing implements, maps, and a computer or tablet were available in all sessions. Participants were informed about the upcoming topic and were encouraged to bring materials (e.g., photos) to facilitate their engagement in the next conversation.

Each session began with a social question (e.g., *how is everyone today*?). The conversational topic was introduced using PowerPoint slides, allowing for both verbal and visual presentation. Conversation was facilitated with supportive communication and standard conversation treatment techniques (Elman, 2007; Kearns & Elman, 2008) (Archer et al., 2019, 2021; Simmons-Mackie et al., 2007).; Conversations evolved naturally. Clinicians did not attempt to direct the flow of the topic, except in rare situations to minimize harmful perseveration on a topic or idea. Although prepared materials were available to support conversation, clinicians were not required to use all questions or slides. Clinicians attempted to equalize opportunities for participants to take conversation turns and encouraged multi-modal communication. Clinicians explicitly modeled communication strategies such as gesturing, writing key words, pointing to conversation supports, and repeating key phrases to both illustrate the strategy and support auditory comprehension for all group members.

Prior to the start of treatment, each participant identified two individual communication goals to address during the treatment, in consultation with clinicians. These

Table 3. Conversation topics by week and session.

	Week	Session	Topic
Personal History	1	1	Stroke story/Family
		2	Where we grew up
	5	11	Education, occupation, vocation
		12	Hobbies/interests
Dining	2	3	Dining in/cooking
		4	Dining out
	6	13	Food that travels
		14	Holiday food
Travel	3	5	Transportation
		6	Local Sights
	7	15	Places in America
		16	Places worldwide
News/ Events	4	7	Political
		8	Sports
	8	17	Global/natural events
		18	People who inspire
Arts/Entertainment	9	9	Books/Films
		10	TV
	10	19	Art
		20	Music

goals were addressed by creating practice opportunities within conversations each session. For example, one participant's goal was to increase their participation in conversations by using multimodal communication (e.g., writing, gestures). In addition to the writing implements and multimodal communication models available to all participants, clinicians paired gesture and written words when posing questions to the participant, allowed extra time, and encouraged the use of alternative communication modalities (e.g., is it on your phone? Can you gesture it?). Another individual's goal was to demonstrate auditory comprehension by answering Yes/No questions. They were provided with external supports (yes/no icons) and comprehension support (e.g., key words, repetition) and asked Yes/No or other forced-choice questions to address this goal.

Treatment Fidelity

Clinicians received extensive training about the treatment protocol, including a 2-hour orientation meeting to review procedures and goals. Clinicians were provided with a treatment manual, which included information about conversation treatment, cueing hierarchies, and group facilitation techniques. Clinical supervisors at BU and TU observed all sessions and provided daily feedback to ensure correct implementation of the protocol. Co-PI DeDe provided training prior to the first treatment cycle at AC and observed the first week of sessions. Within cycle, clinicians provided treatment in both the large group and dyad conditions. Clinical supervisors were consistent across cycles, but student clinicians were different for each cycle. Trained observers used a checklist to record the number of facilitator models for multimodal communication strategies. Number of conversational turns for each participant and facilitator was recorded for 20% of the sessions. These data were regularly reviewed to ensure accurate implementation of the protocol.

Statistical Analysis Plan

Data were analyzed using profile analysis model (also known as multivariate ANOVA, or MANOVA) with an unstructured covariance matrix to account for the within subject correlation. It was estimated using maximum likelihood in SAS 9.4. The model imposes no structure on the mean change or the correlation, and tests hypotheses of differences in change between time points by condition. For all research questions, the independent variables were time (pre-treatment, post-treatment, maintenance) and condition. For effects of time, planned comparisons were pre- versus post-treatment and pre-treatment versus maintenance within condition using Tukey HSD comparisons. The dependent measures were performance on primary and secondary outcome measures.

Analyses proceeded as follows. We examined effects of treatment (RQ1) for all primary and secondary outcome measures. We then examined whether there were effects of group size (RQ2) or group composition (RQ3) for variables that showed evidence of treatment effects in RQ1. Main effects of time and interactions between time and condition were followed up with planned comparisons as described above. Significance was set at $p < .05$, and trends at $p < .10$. With respect to data interpretation, significant interactions between time and condition indicate that one condition demonstrated significantly greater changes across the three testing times (pre, post, maintenance) as compared to

the other condition in the analyses. Planned comparisons indicate whether one condition shows treatment changes, but not whether those changes are greater than another condition.

Results

Research Question 1: Do IwMSA show treatment benefits following conversation treatment (regardless of group size) as compared to a control group?

Primary outcome measure: ACOM

For research question 1, the independent variables were time (pre-treatment, post-treatment, maintenance) and condition (treated vs untreated, collapsing across group size). Figure 2 illustrates these results. The main effect of time was significant, $F(2, 81) = 8.87$, $p < .001$. The effect of condition was not significant, $F(1, 84) = 0.74$, $p = 0.39$. The interaction of time by condition was also significant, $F(2,81) = 3.56$, $p = 0.033$. This interaction indicates that the treated condition showed significantly greater changes than the control condition as a function of time. Specifically, the treated condition showed significant improvement from pre to post treatment (diff = 3.81, 95% confidence interval 0.76 - 6.86), $t = 3.64$, $p = .006$) and from pre-treatment to maintenance (diff = 5.23, 95% confidence interval 2.33–8.13), $t = 5.26$, $p < .001$). The control group (untreated) did not show significant change from pre- to post-treatment or pre-treatment to maintenance ($t_{absolute \leq}$ 1.36, p's $\geq$0.75).

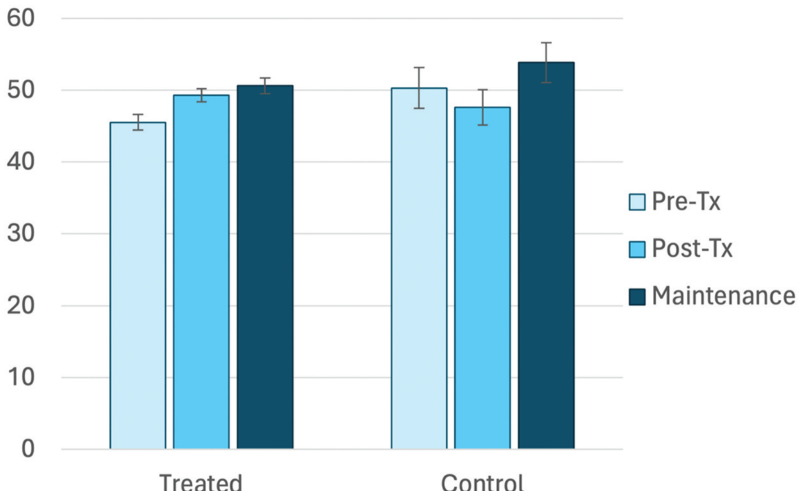

Figure 2. Mean ACOM scores for treated and control groups at pre-treatment, post-treatment, and maintenance. Note: Error bars represent ±1 standard error of the mean (SEM). Asterisks depict significant effects ($p < .05$) compared to pre-treatment in planned comparisons.

Secondary outcome measures

Functional Communication: On the CADL-3, the main effects of time and condition were not significant, F's $\leq$ 1.84, p's $\geq$0.17). Figure 3 illustrates these results. The interaction of time by condition reached the level of a trend, F(2,66) = 2.48, *p* = 0.09. The sample size for the CADL-3 was smaller than that for the ACOM (*n* = 68 vs 84 at Time 3), because the CADL-3 was only added in for cycles 2 and 3. Planned comparisons were examined to determine whether there was evidence of treatment benefit on this measure. The treated conditions showed significant improvement from pre to post treatment (diff = 4.91, 95% confidence interval 2.33–7.48), *t* = 5.59, *p* < .001) and from pre-treatment to maintenance (diff = 6.01, 95% confidence interval, 2.96 - 9.07, *t* = 5.77, *p* < .001). The control group did not show significant change from pre- to post-treatment or pre-treatment to maintenance ($t_{absolute} \leq 1$, *p* = 1).

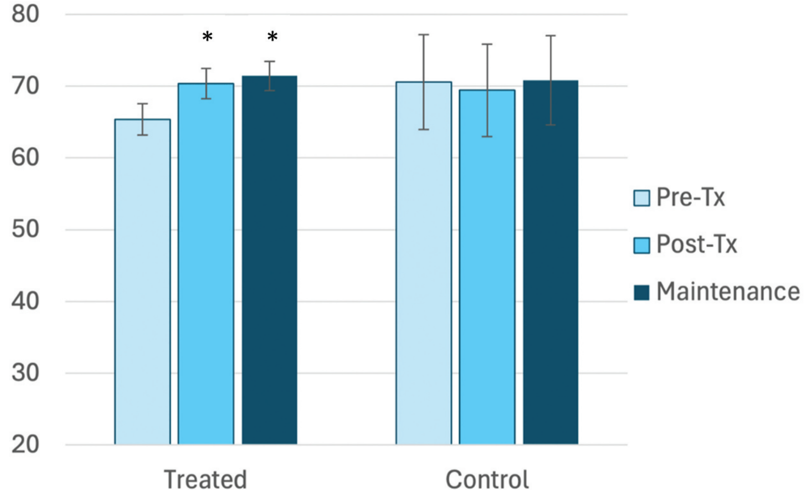

Figure 3. Mean CADL-3 scores for treated and control groups at pre-treatment, post-treatment, and Maintenance. Error bars represent ±1 standard error of the mean (SEM). Asterisks depict significant effects (*p* < .05) compared to pre-treatment in planned comparisons. In all analyses, *condition* refers to treated versus untreated conditions, and *time* refers to pre-treatment, post-treatment, and maintenance.

The interaction of time and condition did not approach significance in any of the other secondary outcome measures. Complete results are available in the supplementary data.

Treatment Impact Survey Results: Forty of the 48 participants from cycle 3 completed the Treatment Impact Surveys. Responses ranged from −2 to + 2 on a Likert Scale. Mean scores of 1.25 (sd 0.7), 1.28 (sd 0.7) and 1.23 (sd 0.8) were reported in response to questions about self-perceived changes in communication, quality of life and well-being, respectively. Performance was judged to be "somewhat worse" on 2/120 items (1.7%) and "unchanged" for 13/120 items (13%). In contrast, improvement was noted on approximately 88% of the items. Specifically, 55/120 items (45.8%) were scored as "somewhat improved" and 50/120 items (41.7%) were scored as "much improved". Twenty-two

Table 4. Mean (standard error; N) ACOM and CADL-3 scores for research questions 2 and 3.

		Pre-Tx	Post-Tx	Maintenance
RQ2: Effects of Group Size (Dyad vs Large Group Condition)				
ACOM	Dyad	47.22 (1.46; 42)	50.14 (1.27; 40)	52.48* (1.47; 39)
	Large Group	43.48 (1.60; 35)	48.32* (1.59; 34)	48.46* (1.59; 34)
CADL-3	Dyad	64.97 (3.05; 35)	69.02* (2.99; 32)	70.86* (2.90; 33)
	Large Group	65.90 (3.29; 30)	71.90* (3.22; 29)	72.11* (3.16; 29)
RQ3: Effects of Group Composition (Large Group Only)				
ACOM	Homogeneous	42.01 (2.12; 20)	48.17* (1.84; 20)	47.60* (1.98; 19)
	Mixed	45.43 (2.44; 15)	48.61 (2.15; 14)	49.66 (2.31; 14)
CADL-3	Homogeneous	65.50 (2.64; 20)	71.24* (2.57; 19)	72.03* (2.50; 19)
	Mixed	66.54 (3.31; 10)	68.77 (3.22; 9)	70.43 (3.10; 10)

*Indicates significant difference relative to pre-treatment testing in planned comparisons. See text for details.

of 48 care partners responded to a care partner version of the survey. Care partners reported no declines in performance since beginning treatment. The distribution of scores was as follows: “no change” – 7/66 items (10.6%), “somewhat improved” – 22/66 items (33.3%) and “much improved” – 37/66 items (56.1%).

Research questions 2 and 3 focused on whether there was an advantage for certain subgroups (RQ2: Dyad vs Large Group; RQ3: Homogeneous vs Mixed Groups). Because these analyses do not include a control group, only measures that showed evidence of treatment effects were evaluated. Thus, only results of the ACOM and CADL-3 were analyzed for research questions two and three. Table 4 presents the means and standard errors for Research Question 2 (Dyad vs. Large Group) and Research Question 3 (Homogeneous vs. Mixed Groups) for the ACOM and CADL-3.

Question 2: Do IwMSA show differential effects of treatment in large groups versus dyads?

Primary outcome measure: ACOM

The independent variables were time (pre-treatment, post-treatment, maintenance) and group size (dyad vs large group). The ACOM showed a significant effect of time, $F(2, 72) = 13.22$, $p < 0.001$. The interaction was not significant, $F(2, 72) = 0.99$, $p = 0.38$. Planned comparisons for the effect of time examined pre-treatment versus post-treatment and maintenance, separately by treatment condition. The Large Group showed significant changes for pre- versus post-treatment (diff = 4.88, 95% confidence interval, 0.50 – 9.26, $t = 3.26$, $p = 0.02$) and pre-treatment versus maintenance (diff = 5.24, 95% confidence interval, 0.92 – 9.56, $t = 3.55$, $p = 0.01$). For the Dyad Condition, there was no significant change from pre-treatment to post-treatment (diff = 2.93, 95% confidence interval: −1.17 – 7.03, $t = 2.09$, $p = 0.30$). There was a significant change from pre-treatment to maintenance (diff = 5.27, 95% confidence interval, 1.19 – 9.34, $t = 3.78$, $p = 0.004$).

Table 5. Average and range of cat Total Naming scores, WAB-R aphasia Quotient scores for mixed and homogeneous large groups (separately by site and cycle).

			# WAB-R Severity Ratings		
	CAT Total Naming Score	WAB-R AQ	Mild	Mod	Severe
Adler Aphasia Center					
Mixed (Cycle 2)	34.4 (0 - 86)	58.2 (14.7 - 93.1)	2	3	2
Homogeneous (Cycle 3)	18.9 (0–44)	51.2 (15.0–67.1)	0	5	3
Boston University					
Mixed (Cycle 1)	50.6 (5 - 72)	N/A			
Mixed (Cycle 2)	44 (10–78)	72.1 (40.3 - 93.7)	5	1	2
Homogeneous (Cycle 3)	19 (0–53)	45.8 (18.1 – 63.7)	0	4	4
Temple University					
Mixed (Cycle 1)	49.8 (22 - 74)	N/A			
Mixed (Cycle 2)	47.4 (14 - 81)	74.4 (48.9 - 95)	4	2	1
Homogeneous (Cycle 3)	23 (0 – 59)	35.5 (5 - 63.8)	0	4	4

Note that individuals in the mixed cycles who did not meet criteria for inclusion in the present study were not included in data analysis. This table illustrates the diversity of the mixed compared to homogeneous conditions. For WAB-R severity ratings, severe is WAB AQ <51, moderate is WAB AQ 51 - 75, and mild is >75.

Secondary outcome measure: CADL-3

The CADL-3 showed a significant effect of time, $F(2, 60) = 18.43$, $p < 0.001$. The interaction was not significant, $F(2, 60) = 0.64$, $p = 0.53$. Planned comparisons showed that both the Large Group and Dyad conditions showed significant changes from pre to post treatment (Large Group: diff = 5.88, 95% confidence interval: 1.95 – 9.72, $t = 4.42$, $p < 0.001$; Dyad: diff = 4.05, 95% confidence interval: 0.33 – 7.77, $t = 3.21$, $p = 0.02$). There were also significant changes from pre-treatment to maintenance treatment in both treated conditions (Large Group: diff = 6.14, 95% confidence interval: 1.53 – 10.74, $t = 3.92$, $p = 0.003$; Dyad: diff = 5.89, 95% confidence interval: 1.44 – 10.34, $t = 3.90$, $p = 0.003$).

Research Question 3: Do IwMSA show differential effects of conversation treatment in mixed versus homogenous large groups?

Table 5 presents the average and range of CAT Total Naming and WAB-R Aphasia Quotient scores for the mixed and homogeneous large groups for each site and cycle. These data demonstrate the range of aphasia severities in the two conditions.

ACOM (Primary Outcome Measure): The independent variables were time (pre-treatment, post-treatment, maintenance) and large group composition (mixed vs homogeneous). The main effect of time was significant, $F(2,31) = 10.48$, $p = 0.003$. In the homogeneous group, planned comparisons showed significant improvements from pre to post treatment (diff = 6.16, 95% confidence interval: 1.44 – 10.34, $t = 3.90$, $p = 0.003$), and from pre-treatment to maintenance (diff = 6.16, 95% confidence interval: 2.06 – 10.33, $t = 4.49$, $p = 0.001$). The mixed (heterogeneous) group did not show a significant change from pre to post treatment (diff = 3.19, 95% confidence interval: −1.83 – 8.21, $t = 1.93$, $p = 0.40$) or from pre-treatment to maintenance (diff = 4.23, 95% confidence interval: −1.49 – 9.95, $t = 2.24$, $p = 0.25$). Neither the main effect of group context nor the interaction between group context and time were significant (F's < 1, $p > .38$).

CADL-3 (Secondary Outcome Measure): The main effect of time was significant, $F(2,29) = 14.46$, $p < 0.001$. In the homogeneous group, planned comparisons showed significant improvements from pre to post treatment (diff = 6.71, 95% confidence interval: 3.27 – 10.16, $t = 5.97$, $p < 0.001$), and from pre-treatment to maintenance (diff = 6.55, 95%

confidence interval: 0.50 – 12.61, $t = 3.32$, $p = 0.03$). The mixed (heterogeneous) group did not show a significant change from pre to post treatment (diff = 4.21, 95% confidence interval: −1.04 – 9.44, $t = 2.46$, $p = 0.17$) or from pre-treatment to maintenance (diff = 5.30, 95% confidence interval: −3.31 – 13.91, $t = 1.89$, $p = 0.43$). The interactions between group context and time were not significant (F's < 1, $p > .49$).

Discussion

The present study investigated whether individuals with moderate to severe aphasia (IwMSA) benefit from conversation treatment. Based on the Aphasia Communication Outcome Measure (ACOM), the primary outcome measure, IwMSA showed significant benefits from conversation treatment. There was no evidence that this population showed improvements on linguistic measures of communication ability such as the Comprehensive Aphasia Test (CAT). The results were largely consistent with those found by Elman and Bernstein-Ellis (1999). In their study, the subgroup of individuals with moderate-severe aphasia showed significantly greater changes on their functional communication measure (CADL-2). With respect to linguistic change, their full group of people with mixed aphasia profiles showed significant change on the WAB AQ. It is not clear whether this change was similar in people with moderate-severe and mild-moderate aphasia. Regardless, taken together, the results of the two studies support the claim that IwMSA benefit from conversation treatment.

In addition to significant treatment effects on the ACOM, the present results also showed that the IwMSA who received treatment – but not the control group – significantly improved on the CADL-3, a standardized measure of functional communication. This result must be interpreted cautiously, since the effects were not significantly different in the treated compared to the control group. The CADL-3 was only added in Cycle 2, meaning that the sample size for the CADL-3 was smaller than the ACOM. This sample size difference could account for the lack of significant interaction. Regardless, the CADL-3 results suggest that self-reported changes in functional communication on the ACOM were reflected in observable changes on an objective measure. This is important in part because the participants in the present study had aphasia that ranged from moderate-severe to severe, which affected both expressive and receptive language. Receptive language impairments may have interfered with comprehension of items on the ACOM for some participants, which could threaten the validity of the measure. However, the fact that consistent changes were found on the CADL-3 support the validity of the ACOM results. Thus, when taken together, the ACOM and CADL-3 results suggest that conversation treatment was associated with real changes in functional communication ability.

The present study also investigated whether group size (RQ 2: dyad versus large group) or group composition (RQ 3: homogeneous versus mixed large groups) affected treatment outcomes. Although there were no significant effects of treatment condition and time, planned comparisons separately by condition revealed interesting patterns. Regarding group size, both the dyad and large group conditions showed changes on the CADL-3 from pre- to post-treatment and pre-treatment to maintenance. In contrast, the dyad condition showed changes on the ACOM from pre-treatment to maintenance only, but not from pre-treatment to post-treatment, whereas the large group showed significant improvement on the ACOM at both time points. Taken together, the present

results suggest that IwMSA in both the dyad and large group conditions showed similar changes in functional communication immediately after treatment, as evidenced by significant changes on the CADL-3. However, IwMSA in dyads may not have recognized and thus not self-reported those changes on the ACOM at post-treatment. That is, IwMSA in the dyads may have taken longer to adjust their own perception of their communication. If this interpretation is correct, both dyads and large groups are associated with improvements in functional communication and changes in how IwMSA perceive their own communication effectiveness, but those in dyads take a bit longer to self-report those changes.

Analysis of group size did not show a statistically significant benefit for homogeneous or mixed groups. In the planned comparisons, however, the homogeneous but not mixed group showed evidence of treatment benefits on both the CADL-3 and ACOM. This pattern of results provides some evidence in support of homogenous groups, possibly because they can be more tailored to the participants than a mixed group. Another potential benefit of the homogeneous condition is that those with more moderate or moderate-severe aphasia had the opportunity to be among the best communicators in the room, and those with very severe aphasia had more opportunities to learn from and receive psychosocial support from peers with more similar communication profiles.

It is worth noting that both the mixed and homogeneous groups included individuals with a range of aphasia severities. The WAB-R Aphasia Quotients (AQ) for the homogenous large group ranged from 5 to 67.1, whereas WAB-R AQs in the mixed groups ranged from 14.7 to 95. Thus, the homogeneous groups were homogeneous in that they excluded those with mild and latent aphasia. Overall, the results of the present study suggest that there may be advantages to stratifying aphasia groups such that those with very mild aphasia and IwMSA are assigned to separate groups. Further work is necessary to know if such a division would also benefit those with mild aphasia. However, we emphasize that there were no statistically significant differences between the mixed and homogeneous conditions, meaning that including people with different aphasia severities in the same large group is not inconsistent with the results of the present study.

The present study did not find evidence of treatment effects on other functional outcome measures, including the Wall Question for the ALA, the MOS Social Support Scale, or the CCRSA. These null results mirror our previous findings, which focused on effects of group size in a larger sample of people with mixed aphasia profiles (De De et al., 2019; Hoover et al., 2025). These measures may not be sensitive to the effects of goal-directed conversation treatment. It may also be that the Wall Question and CCRSA scores do not show enough variability to detect treatment changes. For example, calculating the CCRSA scores requires binning items into four response categories, which would reduce the sensitivity of the measure. However, a Treatment Impact Survey specific to this intervention revealed that this conversation treatment was impactful for participants with IwMSA. Eighty-eight percent of the IwMSA reported improvement on communication, quality of life and overall well-being. Although the response rate was lower for care partners of this cohort, the scores from partners also reflected a positive impact.

There were also no significant effects on the linguistic measures, including subsections of the CAT and discourse measures. In previous studies, treatment effects have consistently been observed on the CAT Naming section (De De et al., 2019; Hoover et al., 2025). Many participants in the present study had apraxia of speech, which limited their verbal

output. To compensate for effects of speech impairments, we developed protocols for participants to respond using writing after standardized administration of tests. These protocols did not appreciably change performance, so the non-standardized results were not reported here. Interested readers may contact the authors for additional information. Another possibility is that people with severe apraxia of speech or very severe aphasia did not benefit from treatment, introducing floor effects that masked changes on linguistic tasks such as naming. To address this possibility, we identified nine participants who performed at floor (score of 0) at two of three time points and removed their data from the analysis of the CAT Naming Section. Though this reduced power, no significant or near-significant effects emerged. Thus, IwMSA may not show changes in discrete linguistic tasks following goal-directed conversation treatment. This does not mean that IwMSA cannot show improvement on discrete linguistic tasks, but that conversation treatment may not be the best approach to improve performance on word retrieval for this population. Instead, they may require more intensive, focused practice than is possible in group treatment.

Limitations and future directions

Data collection for the present study was challenging. On average, we enrolled one participant for every three we contacted in Cycle 3. This was in part due to inclusion criteria but also reflected challenges in contacting and scheduling participants. Though many participants were excluded based on WAB-R AQ scores greater than 70, enrollment was also complicated by participants' reliance on family or friends for communication and transportation, as well as more extensive physical impairments than less impaired peers with aphasia. One strategy to facilitate recruitment and retention would be additional funds to support transportation, as it would reduce the burden on care partners. We also encountered challenges in administering standardized tests and coding discourse samples. During standardized tests, impairments in auditory and written comprehension made it difficult to evaluate whether participants fully understood test items. This may have introduced noise in the data, making it more difficult to observe statistically reliable results. Some of the measures in our test battery would not typically be administered to someone with severe aphasia, but were included here because these are part of a larger randomized controlled trial. For example, someone with severe aphasia might not typically be asked to orally read complex words or complete multiple discourse production tasks. To minimize frustration and maintain consistency across sites, we developed extensive protocols for when and how to modify test administration for this population. Regarding discourse samples, transcription of narratives was complicated by differences in how research staff perceived non-words and stereotypical utterances, resulting in relatively low reliability. Finally, performance on many measures showed floor effects across time due to severely impaired spoken production (e.g., discourse tasks), which made it difficult to elicit or detect any treatment changes at the group level.

These challenges account for many of the limitations in the present study. One issue is that a fully randomized design was not possible due to the recruitment challenges. Participants were randomly assigned to condition within cycle, but the likelihood of being assigned to a given condition was 33% in cycles 1 and 2 (control, large group, or dyad) and 50% in cycle 3 (large group or dyad only, no control

group). An implication is that the control group was much smaller than the treatment condition. For the ACOM, there were 11 untreated controls at Time 3 compared to 73 treated participants. The smaller number of controls may have contributed to additional variability in that group or made it harder to detect changes across time in this condition. Ideally, future work would be powered for both primary and secondary outcome measures and would contain balanced sample sizes across condition. Nonetheless, the consistency of effects, relative magnitude of effect sizes, and distribution of confidence intervals support the claim that treated IwMSA showed benefits of conversation treatment compared to the control conditions.

An unanswered question is the extent to which apraxia of speech contributed to observed results. We did not formally test for apraxia of speech. In many cases, it was difficult to distinguish apraxia of speech from aphasia due to the severity of impairment. All participants had significant aphasia as evidenced by significant difficulty writing words and comprehension deficits. It was difficult to disentangle aphasia from apraxia in participants with limited verbal output. Thus, it may be that concomitant speech impairments made it difficult to detect changes on linguistic measures such as naming, suggesting that the results may underestimate the benefits of this treatment for IwMSA.

Another issue was the wide range of aphasia profiles contained within the study. Ideally, it would be possible to compare outcomes for individuals stratified by narrower severity bands. Practically, however, this is unlikely in a study of group treatment because of the need to recruit all participants at once. Future studies may be able to examine effects of aphasia severity as a continuous variable using latent variable path models or machine learning approaches. Another direction for future work would be to examine longer term follow up. In the present study, longer term follow up was not practically possible because many of the participants wanted to join other activities within the aphasia centers after the study period. Asking participants to refrain from treatment for a longer period would not be ethical, and effects observed after other group treatments would not necessarily reflect the treatment protocol.

Conclusion

Given these challenges, it is all the more noteworthy that significant treatment effects were observed in the present study. These results highlight the importance of including IwMSA in research studies and demonstrate that this population can improve their functional communication following goal-directed conversation treatment. From a clinical perspective, best practice may be for IwMSA to be enrolled in conversation treatment groups that do not include those with very mild or mild-moderate aphasia. However, where this is not possible for pragmatic reasons, mixed groups are also of benefit for IwMSA.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Institute on Deafness and Other Communication Disorders grants [R21 DC015859] and [R01 DC018781] to Elizabeth Hoover and Gayle DeDe.

References

Ali, M., Lyden, P., Brady, M., & VISTA Collaboration. (2015). Aphasia and Dysarthria in acute stroke: Recovery and functional outcome. *International Journal of Stroke*, *10*(3), 400–406.

Archer, B., Azios, J. H., Gulick, N., & Tetnowski, J. (2021). Facilitating participation in conversation groups for aphasia. *Aphasiology*, *35*(6), 764–782. https://doi.org/10.1080/02687038.2020.1812030

Archer, B., Azios, J. H., Tetnowski, J., Damico, J., Freer, J. C., Schmadeke, S., & Christou-Franklin, E. (2019). Key wording practices in three aphasia conversation groups: A preliminary study. *Aphasiology*, *33*(10), 1248–1269. https://doi.org/10.1080/02687038.2019.1630596

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

Basso, A. (2009). “Natural” conversation: A treatment for severe aphasia. *Aphasiology*, *24*(4), 466–479. https://doi.org/10.1080/02687030802714165

Boyle, M. (2007). Discourse analysis with individuals with aphasia: The use of AphasiaBank. *Aphasiology*, *21*(4), 351–364. https://doi.org/10.1080/02687030600797172

Breitenstein, C., Hilari, K., Menahemi-Falkov, M., Rose, M. L., Wallace, S. J., Brady, M. C., Hillis, A. E., Kiran, S., Szaflarski, J. P., Tippett, D. C., Visch-Brink, E., & Willmes, K. (2023). Operationalising treatment success in aphasia rehabilitation. *Aphasiology*, *37*(11), 1693–1732. https://doi.org/10.1080/02687038.2021.2016594

Cherney, L. R., & Babbit, E. (2011). *Communication confidence rating scale in aphasia (CCRSA)*. Rehabilitation Institute of Chicago. https://doi.org/10.1310/tsr1804-352

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology*, *24*(4), S923–S938. https://doi.org/10.1044/2015_AJSLP-14-0161

De De, 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., & Babb, M. (2011). Effect of verb network strengthening treatment in moderate-to-severe aphasia. *American Journal of Speech-Language Pathology*, *20*(2), 131–145. https://doi.org/10.1044/1058-0360(2011/10-0036)

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., 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. (2007). 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., & Bernstein-Ellis, E. (1999). 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

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476

Goode, M. K., Geraci, L., & Roediger, H. L. (2008). Superiority of variable to repeated practice in transfer on anagram solution. *Psychonomic Bulletin & Review*, *15*(3), 662–666. https://doi.org/10.3758/pbr.15.3.662

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

Hilari, K., & Byng, S. (2009). Health-related quality of life in people with severe aphasia. *International Journal of Language & Communication Disorders*, *44*(2), 193–205. https://doi.org/10.1080/13682820802008820

Hinckley, J. (2014). Facilitating life participation in severe aphasia with limited treatment time. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, *24*(3), 89–99. https://doi.org/10.1044/nnsld24.3.89

Holland, A. L., Fromm, D., & Wozniak, L. (2018). *Communication activities of daily* (living-3rd ed.). Pro-Ed.

Hoover, E., De De, 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

Hoover, E. L., Caplan, D., Waters, G., & Budson, A. (2015). Effects of impairment-based individual and socially-oriented group therapies on verb production in Aphasia. *Aphasiology*, *29*(7), 781–798. https://doi.org/10.1080/02687038.2015.1052682

Hoover, E., McFee, A., & De De, G. (2020). Efficacy of group conversation intervention in individuals with severe profiles of aphasia. *Seminars in Speech and Language*, *41*(1), 71–82. https://doi.org/10.1055/s-0039-3400991

Hoover, E., Szabo, G., Kohen, F., Vitale, S., McCloskey, N., Maas, E., Kulkarni, V., & De DE, G. (2025). 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. *American Journal of Speech Language Pathology*, *34*(3), 3, 12–3–1218. https://doi.org/10.1044/2025_AJSLP-24-00279

Hula, W. D., Doyle, P. J., Stone, C. A., Hula, S. N. A., Kellough, S., Wambaugh, J. L., Schumacher, J. G., St. Jacque, A., & Ross, K. B. (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., Rowland, A., Huijbregts, M., Shumway, E., & Victor, J. C. (2018). *Assessment for living with aphasia (ALA)* (2nd ed.). Aphasia Institute.

Kearns, K., & Elman, R. (2008). Group therapy for aphasia: Theoretical and practical considerations. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (4th ed. pp. 316–337). Lippincott, Williams & Wilkins.

Kertesz, A. (2006). *Western aphasia battery (Revised)*. PsychCorp.

Knollman-Porter, K., Dietz, A., & Dahlem, K. (2018). Intensive auditory comprehension treatment for severe aphasia: A feasibility study. *American Journal of Speech-Language Pathology*, *27*(3), 936–949. https://doi.org/10.1044/2018_AJSLP-17-0117

Leaman, M. C., & Edmonds, L. A. (2023). Analyzing language in the picnic scene picture and in conversation: The type of discourse sample we choose influences findings in people with aphasia. *American Journal of Speech-Language Pathology*, *32*(4), 1413–1430. https://doi.org/10.1044/2023_AJSLP-22-00279

Lee, T. D., Magill, R. A., & Weeks, D. J. (1985). Influence of practice schedule on testing schema theory predictions in adults. *Journal of Motor Behavior*, *17*(3), 283–299. https://doi.org/10.1080/00222895.1985.10735350

Lyon, J. G., & Sims, E. (1989). Drawing: Its use as a communicative aid with aphasic and normal adults. In T. E. Prescott (Ed.), *Clinical aphasiology* (Vol. 18, pp. 339–356). College-Hill Press.

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, *17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

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

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech, Language, and Hearing Research*, *38*(1), 145–156. https://doi.org/10.1044/jshr.3801.145

Nicholas, M., & Elliott, S. (1998). *C-speak aphasia: A communication system for adults with aphasia*. Mayer-Johnson Co.

Ostir, G. V., Berges, I. M., Ottenbacher, M. E., Clow, A., & Ottenbacher, K. J. (2008). Associations between positive emotion and recovery of functional status following stroke. *Psychosomatic Medicine*, *70*(4), 404–409. https://doi.org/10.1097/PSY.0b013e31816fd7d0

Parr, S. (2007). Living with severe aphasia: Tracking social exclusion. *Aphasiology*, *21*(1), 98–123. https://doi.org/10.1080/02687030600798337

Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, *61*(2), 102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012

Roper, A., Marshall, J., & Wilson, S. (2016). Benefits and limitations of computer gesture therapy for the rehabilitation of severe aphasia. *Frontiers in Human Neuroscience*, *10*, Article 595. https://doi.org/10.3389/fnhum.2016.00595

Sharkey, C., De De, G., & Hoover, E. (2024). Quantification of group dynamics in conversation treatment for individuals with aphasia. Poster Presented at the International Aphasia Rehabilitation Conference, Brisbane, Australia.

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., 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. https://doi.org/10.1097/00011363-200701000-00003

Simmons-Mackie, N., Kagan, A., & Hogan, N. (2007). Measures of communicative success: The development and evaluation of the assessment for living with aphasia (ALA). *Aphasiology*, *21*(6), 722–733. https://doi.org/10.1080/02687030600989689

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

Smania, N., Gandolfi, M., Aglioti, S. M., Girardi, P., Fiaschi, A., & Girardi, F. (2010). How long is the recovery of global aphasia? Twenty-five years of follow-up in a patient with left hemisphere stroke. *Neurorehabilitation and Neural Repair*, *24*(9), 871–875. https://doi.org/10.1177/1545968310368962

Steele, R. D., Weinrich, M., Wertz, R. T., Kleczewska, M. K., & Carlson, G. S. (1989). Computer-based visual communication in aphasia. *Neuropsychologia*, *27*(4), 409–426. https://doi.org/10.1016/0028-3932(89)90048-1

Swinburn, K., Howard, D., & Porter, G. (2004). *Cat: Comprehensive Aphasia Test*. Psychology Press.

Toseland, R. W., & Rivas, R. F. (2017). *An introduction to group work practice*. Allyn & Bacon.

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., & Collaboration of Aphasia Trialists. (2019). A core outcome set for aphasia treatment research: An international standardised minimum set of outcomes. *International Journal of Stroke*, *14*(4), 330–337. https://doi.org/10.1177/1747493019831250

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

Yu, L., Boyle, P. A., Wilson, R. S., Levine, S. R., Schneider, J. A., & Bennett, D. A. (2015). Purpose in life and cerebral infarcts in community-dwelling older people. *Stroke*, *46*(4), 1071–1076. https://doi.org/10.1161/STROKEAHA.114.008010