AJSLP

Research Article

# Measuring Global Coherence in People With Aphasia During Unstructured Conversation

Marion C. Leaman[a] and Lisa A. Edmonds[a]

**Purpose:** Global coherence is an essential macrolinguistic discourse skill that speakers use to formulate discourse to convey meaning with maintenance to a topic. When global coherence is poor, the listener's ability to understand how the discourse makes sense as a whole is diminished. Measures exist to evaluate global coherence in people with aphasia during monologue tasks (e.g., picture description). The aim of the current research is to develop such a measure for unstructured conversation and to explore how global coherence is impacted by aphasia during conversation. A global coherence measure for conversation is required because markedly different cognitive and linguistic demands are made for production of different types of discourse. Thus, a structured monologue measure cannot be used with validity for unstructured conversation. To adequately evaluate global coherence during conversation, a measure specific to the demands of conversation is required.
**Method:** We adapted the 4-point Global Coherence Scale (Wright & Capilouto, 2012; Wright et al., 2013), a monologue-level measure of global coherence to conversation, resulting in the 4-point Global Coherence Scale in unstructured conversation (GCSconv). We conducted statistical evaluation of the reliability/stability of the 4-point GCSconv in 18 unstructured conversations held by nine people with aphasia. Utterances with low global coherence scores were classified following a recent methodology to describe how breakdown in these utterances contributed to diminished global coherence (Hazamy & Obermeyer, 2019).
**Results:** The 4-point GCSconv demonstrated excellent inter/intrarater reliability and test–retest stability. Nonspecific language and off-topic comments contributed most frequently to lowered global coherence.
**Conclusions:** Findings suggest the 4-point GCSconv may be a feasible and reliable measure of global coherence in conversation. This measure adds to a core of emerging reliable discourse measures for conversation. As such, it has potential to inform assessment and treatment of everyday conversation and to investigate the relationship of global coherence in structured monologue and unstructured conversation.
**Supplemental Material:** https://doi.org/10.23641/asha.12469187

In everyday life, speakers use discourse for a variety of functions, such as telling stories, explaining how to do something, and conversing with others. Speakers convey overall meaning by organizing and expressing discourse in a coherent way so as to be easily and efficiently understood by the listener. When this is achieved, the discourse is said to be globally coherent. When discourse does not demonstrate global coherence, listeners may be confused, and the speaker's purpose is not easily understood, even if individual utterances are intact from syntactic, semantic, and lexical standpoints. Furthermore, even when ideas are well organized, linguistic deficits such as lexical retrieval difficulty in people with aphasia (PWA) may interfere with ideas from being expressed in a way that is perceived as globally coherent by the listener. Global coherence capability in PWA is critical for understanding the communication challenges they may have in everyday situations, because everyday communication occurs through discourse.

Evaluation of communication skills in PWA that is limited to the word and sentence level lacks the capacity to shed light on crucial communication skills, such as global coherence, that are only present in discourse. Objective understanding and measurement of discourse-level communication skills are key in developing aphasia treatment that can impact daily communication needs. This is of the utmost importance because PWA and their families desire

[a]Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, NY

Correspondence to Marion C. Leaman: ml3833@tc.columbia.edu

Editor-in-Chief: Melissa Duff
Editor: Jacqueline Laures-Gore

Received September 17, 2019
Revision received January 6, 2020
Accepted February 13, 2020
https://doi.org/10.1044/2020_AJSLP-19-00104

**Publisher Note:** This article is part of the Special Issue: Select Papers From the 49th Clinical Aphasiology Conference.

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

rehabilitation that focuses on improving discourse skills, most especially for conversation (Mayer & Murray, 2003; Wallace et al., 2016). Furthermore, conversation has been identified as the most frequent communication activity in adults (Davidson et al., 2003) and supports friendship and quality of life in PWA (Davidson et al., 2008). In addition to being a primary need for daily communication, conversation is unique in that speakers often embed other types of discourse in conversation to explain ideas (expository discourse), tell stories (narrative discourse), and provide instructions (procedural discourse). Thus, assessment of an individual's ability to produce globally coherent conversation may include assessment of global coherence in additional discourse genres and does so in an ecologically valid manner, capturing real-world capabilities of the speaker in naturally occurring discourse contexts.

Global coherence has been investigated in structured, monologue-level discourse, but not as of yet in unstructured conversation. Thus, to guide assessment, intervention, and future research with the potential to contribute to the discourse needs of PWA, the current investigation aims to develop a reliable measure of global coherence tailored to unstructured conversation and to investigate how global coherence is impacted by aphasia in this discourse context.

Sherratt (2007) provides a hypothetical model of discourse production (adapted from Frederiksen et al., 1990) for understanding how speakers construct globally coherent discourse. Early stages of discourse production focus on generation of a needed overall structure for the type of discourse being produced (e.g., story structure for a narrative discourse, step-by-step structure for procedural discourse; see also van Dijk & Kintsch, 1983). During the middle stage of production, semantic propositions are generated and selected. This process is mediated by social, pragmatic, and contextual needs of the discourse conditions, during which speakers select and integrate information final production. A plan for global coherence is developed as the speaker selects structural and organizational parameters to frame the resulting discourse to meet the needs of the specific context (e.g., a conversation on a familiar subject with a close friend or presenting a complex speech to an audience). Lexical items and syntactic structures are assigned to the propositions in the next stage. In the final stage, motor production occurs. In summary, global coherence planning occurs in the early and mid-stages of discourse production, is then layered with lexical items and syntactic structures, and is then finally produced motorically. The final production or surface-level production that is received by the listener may thus be potentially impacted by microlinguistic (i.e., lexical, grammatical, syntactical) deficits or motor production deficits that occur in the late stages of discourse production.

### *Measures of Global Coherence in Structured Monologue*

Global coherence is typically analyzed through monologue-level discourse samples collected in response to picture stimuli (e.g., picture descriptions or wordless picture books), a researcher's request for a personal narrative on a specified topic (e.g., "tell me about a frightening experience"), or a request to explain a procedure (e.g., how to make a peanut butter and jelly sandwich). Measures either tally the rate of global coherence violations (Andreetta et al., 2012; Andreetta & Marini, 2015; Christiansen, 1995; Marini et al., 2011) or rate the degree of global coherence using a scale (Coelho & Flewellyn, 2003; Glosser & Deser, 1990; Hazamy & Obermeyer, 2019; Olness & Englebretson, 2011; Rogalski et al., 2010, 2019; Ulatowska, Doyel, et al., 1983; Ulatowska, Freedman-Stern, et al., 1983; Ulatowska et al., 1981, 2003; Van Leer & Turkstra, 1999; Wright & Capilouto, 2012; Wright et al., 2013).

Several measures evaluate the global coherence of each utterance in discourse as it relates to the overall specified topic or aim of the sample. The resulting global coherence score is calculated as the mean of the scores of the utterances in the sample (Coelho & Flewellyn, 2003; Glosser & Deser, 1990; Marini et al., 2011; Rogalski et al., 2010; Van Leer & Turkstra, 1999; Wright & Capilouto, 2012; Wright et al., 2013). This approach is based on definitions of global coherence indicating that each utterance within a sample can be interpreted individually in terms of its relative coherence to the topic of the whole sample (Agar & Hobbs, 1982; van Dijk, 1977). However, every utterance is not expected to maintain perfect global coherence, even in adults without communication disorders (Charolles, 1983; van Dijk, 1977). These measures focus on how global coherence is achieved through verbal discourse production, and utterances are rated for global coherence considering parameters such as vague language, tangentiality, lack of critical details, inclusion of personal information, or irrelevant details. Utterances with reduced global coherence scores may reflect surface-level microlinguistic deficits such as lexical retrieval and/or syntactic difficulty present in aphasia (Christiansen, 1995; Hazamy & Obermeyer, 2019; Wright & Capilouto, 2012) that could contribute to observations such as lack of critical details or use of vague language, in addition to underlying deficits in earlier stages of processing (Sherratt, 2007).

Another approach conceptualizes global coherence as it applies to a discourse sample *in its entirety* rather than at an utterance level (G. Brown & Yule, 1983; Charolles, 1983; Fillmore, 2015; Patry & Nespoulous, 1990; Ulatowska & Olness, 2000). This approach views the whole as greater than the sum of the parts. Thus, global coherence is not rated on an utterance-by-utterance basis but rather through consideration of how all of the utterances together create a unified whole (Rogalski et al., 2019; Ulatowska, Doyel, et al., 1983; Ulatowska, Freedman-Stern, et al., 1983; Ulatowska et al., 1981, 2003). Raters use scales to either make a single judgment regarding the relative degree of interconnectedness/clarity of the verbal production for the entire sample (Ulatowska et al., 2003) or respond to several questions about the sample aimed at capturing aspects of global coherence, such as the following: "Is the story unambiguous as to what each of the participants did?" or "Did the speaker stay focused on the topic?" (Rogalski

et al., 2019; Ulatowska, Doyel, et al., 1983; Ulatowska, Freedman-Stern, et al., 1983; Ulatowska et al., 1981). Olness and Englebretson (2011) describe a methodology that likewise measures global coherence at a discourse sample level using elicited personal narratives. However, their measure rates global coherence as produced through both verbal and nonverbal means and focuses on utterances that express an opinion or make a point. These utterances are identified and then integrated to create a paraphrase of the point of the discourse, which is evaluated for its global coherence.

Lastly, some researchers analyze the discourse sample exactly as spoken, including microlinguistic errors (i.e., rating the surface production; Andreetta et al., 2012; Andreetta & Marini, 2015; Coelho & Flewellyn, 2003; Glosser & Deser, 1990; Marini et al., 2011; Rogalski et al., 2010, 2019; Wright & Capilouto, 2012; Wright et al., 2013), whereas others remove the impact of microlinguistic errors by correcting or paraphrasing each utterance before analysis for global coherence (Christiansen, 1995; Olness & Englebretson, 2011). The former method may result in analysis that includes linguistic errors occurring at late stages of processing (e.g., lexical retrieval and syntactic errors) rather than strictly measuring global coherence capacity that takes place at an early level of processing (Sherratt, 2007). Meanwhile, the latter method requires rater modification of the sample.

Variation in measures of global coherence, severities/types of aphasia of participants, discourse sample types, and elicitation methods have resulted in varied findings regarding global coherence difficulties in PWA (for a review, see Ellis et al., 2016). In addition, methodological differences may contribute to differing findings regarding how global coherence may be impacted by microlinguistic errors. However, this relationship is likely complex as global coherence is neither guaranteed by intact microlinguistic skills nor prevented by impaired microlinguistic skills (Coates, 1995; Huber, 1990; Patry & Nespoulous, 1990; van Dijk, 1977). Some aphasia researchers suggest global coherence is unimpaired (Glosser & Deser, 1990) or is relatively intact in comparison to microlinguistic skills in PWA (Olness & Englebretson, 2011; Ulatowska, Doyel, et al., 1983; Ulatowska, Freedman-Stern, et al., 1983; Ulatowska et al., 1981, 2004), whereas others show a more direct association between microlinguistic deficits and global coherence difficulty (Andreetta et al., 2012; Andreetta & Marini, 2015; Hazamy & Obermeyer, 2019; Marini et al., 2011; Wright & Capilouto, 2012). In addition, differential recovery of microlinguistic and global coherence skills has been reported (Coelho & Flewellyn, 2003). Microlinguistic skills are also reported to contribute to maintenance of global coherence (Wright & Capilouto, 2012).

Hazamy and Obermeyer (2019) have recently developed a global coherence error analysis methodology that provides further insight into the factors contributing to reduced global coherence. They assigned utterances receiving low global coherence scores (Wright & Capilouto, 2012) to one of seven error categories that they defined as follows: *commentary* (speaker makes a comment on the discourse), *nonspecific* (vague language), *not complete* (incomplete utterance), *repeated* (repetition of utterance), *incorrect* (wrong information provided), *detail* (unimportant information), or *off-topic* (not related to the topic). They noted that category assignment could be related to various underlying deficits leading to reduced global coherence. For example, they remarked that linguistic deficits (e.g., lexical retrieval difficulty) often resulted in utterances classified as not complete, whereas compensation for linguistic deficits resulted in commentary utterances, and traditional coherence deficits often led to utterances classified as off-topic. They found that most errors in their sample of 31 PWA were classified to the *not complete* and *commentary* categories rather than to the *off-topic* category. Their findings suggest microlinguistic contributions to reduced global coherence (Hazamy & Obermeyer, 2019), consistent with Wright and Capilouto (2012).

Despite a growing body of global coherence literature in structured discourse, a lack of a reliable measure of global coherence in conversation prevents assessment and understanding of global coherence in real-world contexts for PWA. Given that various types of discourse have varying structural demands (e.g., story grammar in narrative discourse, step-by-step structure in procedural discourse), as well as varying cognitive and linguistic demands (Beeke et al., 2007; Boyle, 2011; Dipper et al., 2018; Kurland & Stokes, 2018; Linnik et al., 2016; Longacre, 1996; Olness et al., 2002; Pritchard et al., 2017; Stark, 2019; Wilkinson, 1995), communication abilities in highly structured, elicited monologues cannot be assumed to parallel communication abilities in conversational discourse. Likewise, production of narrative, procedural, or expository discourse in highly structured clinical environments produced in response to a request does not necessarily reflect how a speaker would produce similar discourse when embedded in spontaneous conversation. For example, in conversation, the speaker must manage an ever-evolving plan for production of coherent turns in response to the ongoing interaction (van Dijk & Kintsch, 1983), which is not required in the structured, elicited monologue-level discourse used in clinical and research settings. In order to thoroughly and accurately understand global coherence in conversation, a measure specifically designed to meet the demands of conversational discourse is required (Kurland & Stokes, 2018; Leaman & Edmonds, 2019a). Furthermore, interpreting how global coherence ability in structured monologue-level tasks relates to similar abilities in unstructured discourse contexts requires development of a measure that is adaptable to each of these communication environments.

### *The 4-Point Global Coherence Scale*

Wright and Capilouto (2012) developed a 4-point global coherence scale (4-point GCS) for use in structured monologue-level discourse. Their scale demonstrates good interrater reliability (IRR), stability, and validity (Wright et al., 2013). Construct validity was established through comparison to a 5-point scale developed by Van Leer and Turkstra (1999) based on Glosser and Deser's (1990) previous

5-point scale. In addition, a recent study reports preliminary evidence of external validity for the 4-point GCS (Rogalski et al., 2019). Wright et al. (2013) developed a 4-point scale rather than a 5-point scale because they hypothesized that a 5-point scale may have offered too many options for raters, based on evidence in the literature that all five levels were seldom used (Rogalski et al., 2010; Van Leer & Turkstra, 1999).

The 4-point GCS evaluates how well each utterance relates to the overall topic of the discourse. In summary (see Wright & Capilouto, 2012; Wright et al., 2013, for the detailed 4-point GCS), Level "4" indicates clear and specific mention of elements of the stimulus (e.g., actors, actions, objects) related to the main details of the stimulus and thus requires no inferencing by the listener; "3" indicates the utterance is related to the stimulus, but may include tangential information relevant to the main topic, and/or lacks substantive information requiring inferencing by the listener; "2" indicates a distant relationship to the stimulus and may be tangential, egocentric, or have a focus on a noncritical aspect of the stimulus; and "1" indicates the utterance is entirely unrelated to the stimulus and may be a comment on the stimulus item.

### *Measuring Unstructured Conversation*

Although unstructured conversation is by nature unpredictable, recent research has demonstrated that at least some measures of verbal microlinguistic skills may be statistically reliable across unstructured conversations, even when topics are allowed to vary naturally (Carragher et al., 2013, 2015; Leaman & Edmonds, 2019a, 2019b) or when conversation partners who are familiar with aphasia are allowed to vary (i.e., family members/friends and speech-language pathologists [SLPs]; Leaman & Edmonds, 2019a, 2019b). To analyze stability of linguistic production in conversation, strict adherence to a traditional test–retest context with static stimuli materials is not possible, because the very nature of everyday conversation entails spontaneity of topic and occurs across a variety of partners. Introducing control of the conversational topic would eliminate the ecological validity of this research, thus preventing the needed answers to meaningful questions regarding language production by PWA in real-world conversations. Given the importance of understanding the needs and linguistic production of PWA in everyday conversation and the lack of research in this area, Carragher et al. (2013, 2015) and Leaman and Edmonds (2019a, 2019b) used methodologies that investigate stability of measures while allowing conversations to unfold naturally as well as performed statistical analyses on the resulting language samples. Carragher et al. (2013, 2015) reported stability as a lack of variance in the measures across varying time points, whereas Leaman and Edmonds (2019a, 2019b) referred to the test–retest stability of their measures, with the view that the stimulus, unstructured conversation, was the same at both time points. Supporting this view, they demonstrated no significant difference in the resulting data for mean length of utterance (R. Brown, 1973), type–token ratio (Johnson, 1944), and vocabulary frequency (Brysbaert & New, 2009) and reported similar topics of conversation (entertainment, travel, family, food, etc.) across conversations (Leaman & Edmonds, 2019a, 2020).

In the current research, we investigate whether the 4-point GCS can be reliably applied to unstructured conversation in PWA if adapted for the dyadic and contextually based aspects of conversation (referred to as "4-point GCSconv"). In addition, we investigate the relationship between microlinguistic skills and global coherence in unstructured conversation by using Hazamy and Obermeyer's (2019) methodology to analyze the utterances receiving low global coherence scores. To further understand how this relationship may vary with limitations on expressive language, we investigate this question separately for the PWA demonstrating mild expressive deficits and for those with aphasia resulting in severe expressive deficits. The research questions for this study ask what is

RQ1. the interrater and intrarater reliability of the 4-point GCSconv;
RQ2. the test–retest stability of the 4-point GCSconv across two unstructured conversations; and
RQ3. the rate of low global coherence utterances (i.e., scores of 1 or 2) by error category for (a) the entire group, (b) participants with mild expressive language deficits, and (c) participants with moderate expressive language deficits.

## Method

This research was conducted with approval from the Institutional Review Board of Teachers College, Columbia University in New York City.

### *Participants*

Nine participants with varying types/severities of aphasia and self-selected home communication partners (Home-Ps) were recruited from local university and community aphasia clinics (see Table 1). In addition, five licensed SLPs unfamiliar to the PWA who were experienced in working with PWA volunteered as partners (SLP-Ps). Two different types of partners were selected because our first measurement study (Leaman & Edmonds, 2019a) investigated the extent to which linguistic skills of PWA in unstructured conversation with an unfamiliar SLP-P would parallel the linguistic skills demonstrated during unstructured conversation with a familiar Home-P. The current study does not have a focus on the varying communication partners but rather on the reliability of the 4-point GCSconv across two conversations. Thus, we have not asked research questions specific to the varying partner condition but find value in using one set of conversations for all of our preliminary research in measurement (Leaman & Edmonds, 2019a, 2019b).

The PWA met the following criteria: aphasia due to a single left cerebrovascular accident (CVA) at least 6 months prior; age range of 18–75 years; adequate hearing

**Table 1.** Demographic information and assessment results for people with aphasia and home partners.

| Participant | Age/gender (race) | WABR-AQ | Aphasia type | CLQT overall score (clock drawing) | ACOM or CAT-DQ: T score | Conv. time (min′sec″) No. of utt. for each conv. | Language background | Education | TPO | Home partner (age/gender) – Education – (MoCA score) |
|---|---|---|---|---|---|---|---|---|---|---|
| P1 | 42/M (Cauc.) | 78.0 | Mild anomia | 3.2/4 Mild impairment (clock: WNL) | ACOM: 69.81 | H: 6′00″[a]<br>H: 50<br>SLP: 8′32″<br>SLP: 78 | Standard American English (SAE); bilingual:[b] English primary, Hebrew secondary | College | 9;2 | (41/F) – College – (28/30) |
| P2 | 66/M (AA) | 79.3 | Mild anomia & mod. apraxia of speech | 3.6/4 WNL (clock: WNL) | ACOM: 56.57 | H: 10′30″<br>H: 49<br>SLP: 9′45″<br>SLP: 54 | SAE; monolingual | Graduate school | 3;7 | (61/F) – Grad. school – (28/30) |
| P3 | 60/F (Cauc.) | 38.9 | Severe Wernicke | 2.2/4 Mod. impairment (clock: WNL) | CAT-DQ: 50.00 | H: 9′18″<br>H: 172<br>SLP: 9′11″<br>SLP: 100 | SAE; monolingual | College | 2;0 | (60/M) – Assoc. degree – (26/30) |
| P4 | 65/F (AA) | 85.0 | Mild anomia/mild dysarthria | 3.6/4 WNL (clock: WNL) | ACOM: 45.04 | H: 8′15″<br>H: 92<br>SLP: 8′12″<br>SLP: 90 | SAE & African American Vernacular English (AAVE)[c]; monolingual | High school | 0;8 | (42/F) – College – 26/30 |
| P5 | 71/F (Cauc.) | 75.7 | Mod. transcortical motor | 2.8/4 Mild impairment (clock: WNL) | ACOM: 41.35 | H: 11′32″<br>H: 60<br>SLP: 11′53″<br>SLP: 51 | SAE; monolingual | High school | 10;4 | (68/F) – Grad. school – (29/30) |
| P6 | 64/F (Cauc.) | 92.4 | Mild anomia & mild apraxia of speech | 2.6/4 Mild impairment (clock: WNL) | CAT-DQ: 57.00 | H: 8′57″<br>H: 48<br>SLP: 8′30″<br>SLP: 44 | SAE; monolingual | College | 5;9 | (59/F) – Grad. school – 27/30 |
| P7 | 72/M (Cauc.) | 92.0 | Mild anomia | 3.8/4 WNL (clock: WNL) | CAT-DQ: 49.00 | H: 12′00″<br>H: 48<br>SLP: 9′37″<br>SLP: 77 | SAE; monolingual | Graduate school | 4;2 | (68/F) – Grad. school – (27/30) |
| P8 | 75/F (Cauc.) | 88.0 | Mild anomia | 3.8/4 WNL (clock: WNL) | CAT-DQ: 50.00 | H: 8′52″<br>H: 116<br>SLP: 8′03″<br>SLP: 148 | SAE; monolingual | Graduate school | 13;7 | (66/M) – College – (29/30) |
| P9 | 63/F (Cauc.) | 51.0 | Moderate Broca's & mod./sev. apraxia of speech | ——Did not complete—— | | H: 11′34″<br>H: 60<br>SLP: 11′00″<br>SLP: 84 | SAE; monolingual | Graduate school | 21;5 | (63/M) – Grad. school – (30/30) |

*Note.* WABR-AQ = Western Aphasia Battery–Revised, Aphasia Quotient (Kertesz, 2006); CLQT = Cognitive Linguistic Quick Test (Helm-Estabrooks, 2001); ACOM = Aphasia Communication Outcome Measure (Hula et al., 2015); CAT-DQ = Comprehensive Aphasia Test–Disability Questionnaire (Swinburn et al., 2005); T score, mean = 50, ± 1 *SD* = 16–84; Conv. time = duration of conversation; (min′sec″) = time in minutes, seconds; No. of utt. = number of utterances analyzed; TPO = time postonset; MoCA = Montreal Cognitive Assessment (Nasreddine et al., 2005); M = male; Cauc. = Caucasian; WNL = within normal limits; H = home partner conversation; SLP = speech-language pathologist conversation; F = female; AA = African American. [a]Shortened conversation due to partner 7-min monologue. [b]Hebrew learned first, and SAE learned before 4 years old; participant grew up and received all schooling in the United States, with SAE as primary language per self-report and Hebrew rarely used for many years. [c]AAVE used 14.1% of utterances with home partner and 10% with speech-language pathologist partner (Charity, 2008; Wolfram, 2004).

in at least one ear (per screening at 40 dB); adequate self-reported, corrected vision; no self-reported history of speech, language, learning, psychiatric, or cognitive deficits prior to the CVA; no neurological condition other than the CVA; and native monolingual speaker of English (except P1; see Table 1). Home-Ps met the same criteria (except no neurological history and no aphasia) and achieved a "within normal limits" score (≥ 26/30) on the Montreal Cognitive Assessment (Nasreddine et al., 2005). SLP-Ps were native speakers of American English. Speech, language, hearing, vision, and neurological status were assumed adequate as all SLP-Ps were employed professionally and made no reports otherwise. Data for P1–P8 were previously reported in two studies regarding measurement in conversation (Leaman & Edmonds, 2019a, 2019b) and, for P1–P9, in a study regarding topic initiation (Leaman & Edmonds, 2020). Data for P2 are reported in a case study (Obermeyer et al., 2020).

### *Procedure*

During an initial visit, consent forms were signed following a review with the first author. In addition, the Western Aphasia Battery–Revised (Kertesz, 2006), Cognitive Linguistic Quick Test (Helm-Estabrooks, 2001), and either the Aphasia Communication Outcome Measure (Hula et al., 2015) or the Disability Questionnaire of the Comprehensive Aphasia Test (Swinburn et al., 2005) were administered (see Table 1 for data). Following the initial visit, each participant engaged in two 15- to 20-min conversations, one with the Home-P and one with an SLP-P. The only exception was P9, for whom scheduling of partners required conversations to occur on the same date as the testing. In this case, the conversations were collected prior to testing to prevent testing fatigue or frustration from interfering with the conversational data.

Conversations with the SLP-Ps were collected in the lab (P1–P7) or in the participant's home (P8, P9). Conversations with the Home-Ps were collected in participants' homes, except for P4, P5, and P7 who requested to hold their conversations in a private room in the lab. Conversations in the lab were video- and audio-recorded, whereas those at home were video-recorded. Home-P conversations were typically collected first, because the first author gave the video camera and instructions to the PWA/Home-P during Visit 1 following testing. The SLP-P conversation occurred during Visit 2. In instances where conversation collection occurred in one location on one date due to participant request (P4, P5, P7, P8, P9), the Home-P conversation occurred first for consistency, except for P4. For P4, the SLP-P conversation was conducted during Visit 2, and it was then discovered that there had been a video camera malfunction for the conversation previously collected at home. Thus, a conversation between P4 and her Home-P was collected after the SLP-P conversation. The location of the conversations did not impact use of environmental props (e.g., photos, writing, objects). Five participants used no props in either conversation; two participants used objects in both locations; one used drawing/writing with the Home-P but not the SLP-P (both collected in the lab setting); and one used no props with the Home-P, but the SLP-P referred to family photos on the wall one time (both collected in the home setting). All data collection visits were completed within 3 weeks.

Each dyad was asked to have a casual conversation for approximately 15–20 min. Our aim is to collect data during social conversations that can be used to assess capabilities of the PWA and inform treatment while minimizing influence from the partner, yet without eliminating the partner's active participation in the interaction. To construct similar conditions across conditions consistent with the unstructured nature of everyday conversation, the first author informed participants that there was no specified agenda. No eliciting questions or topics were provided. All dyads were told by the first author that either partner could initiate the conversation and that the topic of conversation could shift freely as occurs in everyday conversation. They were informed they could use any methods of communication, including but not limited to verbal, gestural, and written/drawn modalities. Paper and pen were provided in the lab and were included with the video camera loaned to each dyad for the home conversations. Due to a desire to maximize the social aspect of the conversations, the first author did not use a script to explain procedures but rather shared the procedures informally with the PWA/Home-P at the time of recruitment, when informed consent forms were signed, and prior to each conversation. Participants could ask the first author any questions they had regarding the conversations.

At the time they volunteered, all PWA knew they would hold conversations with familiar and unfamiliar partners and were aware that this research was investigating conversation in PWA and their partners. The PWA knew that the SLP-Ps were SLPs who had volunteered for the study and that the conversation was not an interview or a therapy session. All participants stated they understood the casual/social nature of the conversation. The SLP-P had no role in explaining procedures, operating recording devices, providing paper and pen in the room, or performing any other function related to data collection, as this would potentially cast the SLP-P in an authoritarian role, which was not desired. All data collection tasks were conducted by the first author. None of the participants had knowledge of the analyses that would be conducted on the resulting conversational data.

We developed a written and verbal protocol (Leaman & Edmonds, 2019a) used to train the SLP-Ps to be interested and engaged communication partners as well as to eliminate therapeutic behaviors. The protocol trains SLP-Ps to use behavior similar to what they would typically use in social conversations with adults without aphasia whom they have not met before. Accordingly, SLP-Ps were instructed not to suggest communication strategies or alternative methods of communication to PWA, but to accept these modalities if spontaneously used by the PWA. The SLP-Ps were instructed to share contextually appropriate personal

stories and to avoid an abundance of questions, especially yes/no questions, which could cause an institutionalized interview format of interaction to ensue. As would be typical in meeting someone for the first time, SLP-Ps were instructed to allow adequate time for the PWA to respond and to minimize interruptions/overlapping turns. SLP-Ps were allowed to use nonverbal communication (e.g., gestures, writing) and aligning turns (e.g., "mhm," "uhhuh," "yeah") if those behaviors felt natural in context. Home-Ps did not undergo a similar training, because they were not clinicians who needed to learn to maximize social behavior while eliminating clinician behavior. The PWA/Home-Ps were instructed to have an everyday social conversation as they typically would at home. The Home-Ps were told they did not need to try to elicit any type of behavior or verbalization. Such a behavior occurred in only one instance, when a Home-P requested that the person with aphasia produce the word "succulent" instead of "cactus." This utterance was not analyzed, consistent with our protocol that anticipated such occurrences and preemptively indicated that such interactions would be eliminated from analysis.

### *Transcription*

The first author transcribed all 18 conversations, using a procedure based on the Analysis of Speech Unit, developed for conversation (Foster et al., 2000). This protocol segments and analyzes utterances at the independent clause level (with inclusion of subordinating clause(s)). Elliptical clauses are segmented and analyzed independently as well because they can be expanded from context to an independent level. Transcription began at the start of a new topic following the first 2 min of conversation, allowing the partners to acclimate to being filmed (Beeke et al., 2007; Goodwin, 1981), continuing for 8–12 min. The stopping point aligned with the end of a conversation topic. Transcription included both partners' use of gesture/use of objects, writing, drawing, verbal/nonverbal alignments (e.g., head nods, "mhm"), overlapping speech, revisions, abandoned utterances, and any other behavior that the transcribers felt warranted notation.

### *Determining Topics Within Conversations for Rating the 4-Point GCSconv*

The 4-point GCS (Wright & Capilouto, 2012; Wright et al., 2013) was developed for structured monologue in which the expectation for coherence is explicit because a specific topic for the discourse is provided by the examiner. The speaker knows the topic he or she is meant to expand upon, and raters know the topic to which each utterance is rated for degree of global coherence. However, in unstructured conversation, topics are not explicitly stated and topics shift as the conversation unfolds. Rating global coherence requires that a topic be established against which each utterance can be rated. Thus, in unstructured conversation, raters must first determine location and presence of the topic(s) in the sample before evaluating global coherence.

In a previous study, we developed a methodology for reliably determining topic locations and mechanisms of topic initiation in unstructured conversation (Leaman & Edmonds, 2020). The conversations analyzed for global coherence in the current study served as the language samples in the earlier study. Thus, topic location methodology was established, and topic analysis for these specific language samples had already occurred and had been conducted independent of any potential influence of the current investigation. At the time of our earlier study, we did not anticipate the current study regarding the 4-point GCSconv or its need for a methodology for determining topic location.

A brief overview of the topic initiation methodology is provided. Please see Leaman and Edmonds (2020) for additional procedures, examples, and scoring methods. First, topics are located according to the definition of Mentis and Prutting (1991, p. 585), as follows:

> "(a topic is) a clause or noun phrase that identified the question of immediate concern and that provided a global description of the content of a sequence of utterances. The topic of any sequence of discourse was required to describe what the speakers were talking about and identify the central concern being addressed…. Each utterance within any topic sequence, however, was required to express a concept or set of concepts that could be directly subsumed under the topic label."

Second, after topics are located, each topic-initiating utterance is analyzed to determine which of the following mechanism(s) the speaker used to achieve a new topic: (a) waiting for a topic to end, (b) using a disjunctive word or phrase to indicate a new topic (e.g., "by the way" or "and, so"), (c) using cohesion (i.e., relating the new topic to the old through linking of vocabulary), or (d) no overt method of topic initiation (referred to as "noncoherent"; Mentis & Prutting, 1991). Third, each utterance spoken by PWA within each topic is evaluated with the 4-point GCSconv (the scale is detailed below) for its degree of coherence to the established topic. The utterances that initiate each new topic in the conversation are not coded for global coherence, because their function is to initiate a new topic rather than to cohere to the old topic. The exception to this is when topics are initiated with the noncoherent mechanism; in which case, such a topic-initiating utterance is given the lowest rating (i.e., "1") for global coherence. This is because noncoherent utterances are neither coherent to the old topic nor explicitly constructed with use of a topic-initiating mechanism informing the listener that a new topic had begun. Fourth, the 4-point GCSconv score for each conversation sample is calculated by averaging all of the scores in the entire conversation, consistent with the calculation used for monologues.

### *Four-Point GCSconv Protocol: Adaptations*

Using the 4-point GCS for unstructured conversation requires several adaptations to account for key components

of conversation contributing to global coherence. First, the 4-point GCSconv includes the context of the conversation in a way not required by the 4-point GCS. Second, the 4-point GCSconv rates expressions of opinion differently than is done by the 4-point GCS. Finally, the 4-point GCSconv does not rate utterances that are commentaries on the discourse (e.g., "I forgot the word"), whereas the 4-point GCS in monologue rates such utterances at a Level 1 (lowest level). These adaptations and their rationales are discussed below (and summarized in the Appendix).

**Context in Conversation**

A primary area of consideration for rating global coherence in conversation is the conversational context. As an interaction between two people in a specific evolving situational context, conversation differs pragmatically from a clinical task with discrete instructions to produce a decontextualized monologue for the purpose of language evaluation. Conversational context includes but is not limited to the relationship between the speakers, their common background knowledge or lack thereof (G. Brown & Yule, 1983; Coates, 1995; Wilkes-Gibbs, 1995), and how each utterance fits within the sequential series of utterances produced by both speakers as the dialogue unfolds (Sacks et al., 1974). Each of these parameters informs the meaning being built between the speakers during unstructured conversation. Furthermore, rather than a single specified plan for production, as expected in structured dialogue (Agar & Hobbs, 1982; van Dijk, 1977), speakers shift their plans as the conversation continues (G. Brown & Yule, 1983; van Dijk & Kintsch, 1983). Thus, while the 4-point GCS can focus on verbal production while minimizing contextual factors in structured tasks, the 4-point GCSconv protocol must include the contextual factors critical to constructing and understanding meaning that is globally coherent in conversation.

In using the 4-point GCSconv, raters were trained to observe and integrate contextual evidence. Raters use their own understanding of the conversation as external observers, yet are also required to consider the behavior and reactions of the individuals within the conversation in guiding their ratings. This is key, because turns are constructed for those within the conversation, not for external raters (Clark & Carlson, 1982; Sacks et al., 1974). Similar methods are used in conversation analysis to draw conclusions about conversational interactions (Schegloff, 2007).

Although inclusion of contextual factors potentially introduces a degree of subjectivity to the GCSconv rating scale, conversation cannot be meaningfully rated without considering context. We addressed this potential variation in ratings through a comprehensive training period for the research assistants (RAs), with detailed discussions between raters regarding differences. Furthermore, any subjectivity present in the ratings is accounted for by the IRR procedure because adequate IRR demonstrates that scores are similar across raters (i.e., not idiosyncratic or subjective to a single rater).

**Evaluative Function in Conversation**

In elicited monologue-level language samples, the speaker is expected to conscribe the output to the requested task (e.g., describing a picture or event, explaining how to complete a task), without providing an opinion or personal stories related to the task. The purpose of discourse production in this context is to comply with an examiner's request. Consequently, Wright and colleagues (Wright & Capilouto, 2012; Wright et al., 2013) indicate that the 4-point GCS rates utterances at a Level 2 or Level 1 when the speaker includes opinions and personal stories related to the topic (described as "egocentric," "tangential," and "noncritical" by the 4-point GCS). As a measure developed for structured, elicited monologue, the 4-point GCS focuses on *referential* language (i.e., the speaker's expression of relevant factual information, including complete/accurate references) rather than *evaluative* language (the speaker's expression of opinion or reaction to the discourse content; Labov, 1972).

Unstructured conversation, however, requires both referential and evaluative communication. Conversation occurs within an interpersonal context and has an interpersonal purpose. Speakers converse to share factual information about an event and, oftentimes, more importantly to express why or how that event matters. Narratives embedded in conversations occur because they are "tellable" (Labov & Waletzky, 1967; Sacks, 1992); otherwise, speakers would not choose to tell them. In conversation, speakers use evaluation to express why they are telling what they are telling, by providing opinions and reactions to the concurrent referential content (Labov, 1972; Olness et al., 2010). Evaluative resources include lexical resources (e.g., use of repetition; vocabulary that intensifies such as "it was *so* wonderful"; and speech production resources, i.e., highlighting information through use of pitch, rate, and volume; see Olness et al., 2010, for a further discussion of evaluative resources in aphasia).

Lastly, utterances in which the speaker comments on the discourse itself (i.e., commentary utterances) are not rated for global coherence by the 4-point GCSconv. In monologue, such utterances receive a score of "1" per the 4-point GCS, because they do not maintain coherence to the monologue topic requested by the examiner. In conversational contexts, however, commentary utterances perform an interactional function that often serves to build coherence and thus should not be penalized. Although a commentary utterance such as "I can't say it" shifts the topic to a focus on a lexical retrieval error that has just occurred or an error that is anticipated, the person with aphasia is demonstrating awareness of the error and is responding by self-initiating a repair sequence that may be vital for the conversation to continue forward effectively. This verbal expression of the person with aphasia's insight to the communication difficulty allows the partner to explicitly know the person with aphasia is aware of the error. Furthermore, in some instances, the commentary utterance may be alerting the listener to an error of which he or she had been unaware (such as a paraphasic error, which could

be interpretable as reasonable in context). In addition, initiating repair on one's own error utterance (i.e., self-repair) is preferred to repairs that are initiated by the partner (i.e., other repair; Schegloff et al., 1977). Thus, when using a commentary utterance, the person with aphasia is effectively and appropriately using a typical conversational tool to repair the previous utterance, thereby restoring global coherence to the dialogue disrupted by microlinguistic difficulty. Furthermore, the commentary utterance is a way to invite assistance from the partner in achieving the repair, an important interactional and functional aspect of commentary utterances in conversation. As such, although commentary utterances momentarily reduce coherence to the topic, they often do so with the purpose to restore coherence rapidly and are thus coherence building rather than coherence diminishing.

These differences in the functions of monologue and unstructured dialogue require that commentary utterances on discourse performance or that express opinion or make a point through use of stories (i.e., evaluation) are not automatically rated at a lower level on the 4-point GCSconv than utterances that contain mostly or only referential content. Opinion within conversation that is otherwise coherent to the topic is rated a "4" on the GCSconv. We also added a guideline regarding repetition used as an evaluative resource (see below), which is not mentioned in Wright and colleagues' global coherence work (Wright & Capilouto, 2012; Wright et al., 2013) because it did not occur (H. H. Wright, personal communication, August 2019).

### *Four-Point GCSconv Protocol: Additions*

In developing the GCSconv, we developed several procedures to consistently account for interactional material that does not occur in monologues (see Appendix). Utterances that are interrupted/overlapped by the partner and are subsequently incomplete are not evaluated for global coherence. Clarification requests by PWA are not analyzed because they perform a necessary function for PWA to confirm or request understanding. Utterances that consist only of yes, no, or filler language such as "and so…," "anyways…," and "you know" are not analyzed. These types of utterances fill a social purpose to organize conversational interactions (Schiffrin, 1987; Simmons-Mackie & Damico, 1996) rather than either a referential or evaluative purpose, thus are not expected to demonstrate global coherence to the topic in the way investigated in this research. In addition, incomplete utterances are not analyzed, as it is often not possible to accurately determine the degree of global coherence of such utterances. Nonverbal turns in which the person with aphasia writes, points to written language, or spells words aloud or in the air/on the tabletop are treated as if they were produced verbally. Utterances containing pronouns used to refer to objects in the environment are not penalized if otherwise coherent (i.e., deictic utterances).

Lastly, the 4-point GCS does not specify coding procedures for exact repetition as it did not occur in the data of Wright and colleagues (H. H. Wright, personal communication, August 2019). However, several global coherence measures penalize repetition (Andreetta et al., 2012; Andreetta & Marini, 2015; Christiansen, 1995; Marini et al., 2011), thus we added a guideline for repetition because it occurred in many of our samples. When repetition is used as an evaluative resource to emphasize a point during conversation, it is coded without penalty if it is otherwise coherent. Likewise, when repetition is used to respond to a partner's question/comment, it is not penalized, with the utterance receiving the same score it did in its first iteration. When repetition is not used in these ways, it is not coded, as it neither adds to nor diminishes overall coherence of the sample.

### *Analysis of Utterances With Low Global Coherence*

We performed an analysis for utterances receiving a low global coherence score (1 or 2) to better understand how various global coherence errors contribute to reduced global coherence. We categorized utterances receiving scores of 1 or 2, consistent with the methodology of Hazamy and Obermeyer (2019). However, we did not use their categories "commentary," "not complete," and "repeated," as the 4-point GCSconv does not evaluate these types of utterances. We provide two additional clarifications to Hazamy and Obermeyer's definitions for each category as used in our work, noted in parentheses (examples in Table 2). The *nonspecific* category is used for utterances containing vague language, such as "thing" or "stuff" (or when not enough information was provided reducing the specificity of the utterance); the *incorrect* category reflects utterances that contained wrong information (significant enough to impair global coherence, including paraphasic errors that were not understandable in context); the *detail* category describes verbalizations in which the level of detail is excessive/irrelevant; and the *off-topic* category is defined as utterances with no decipherable connection to the topic. Topic-initiating utterances, which were noncoherent (i.e., abrupt new topic, as discussed earlier), received a score of "1" and were assigned to the off-topic category. This occurred for 13.2% of topic-initiating utterances (23/174; Leaman & Edmonds, 2020).

### *Reliability*

All reliability data for the 4-point GCSconv are reported in the Results section, as reliability of this measure is the focus of Research Question 1. Transcription reliability was conducted by three graduate students in speech-language pathology trained as RAs by the first author. Two RAs were also trained to apply the 4-point GCSconv. One RA was trained to code utterances for low global coherence categories (Hazamy & Obermeyer, 2019). In our earlier study regarding topic initiation (Leaman & Edmonds, 2020), one of the above RAs was trained to locate topics and code mechanisms of topic initiation. Training for transcription and coding included verbal/written instruction, practice,

**Table 2.** Categories for reduced global coherence with examples.[a]

| Category | Description | Examples of utterances | 4-point GCSconv | Conversational context |
|---|---|---|---|---|
| Nonspecific | Vague language (thing, stuff, etc.) | And Sue Sue Sue | 1 | (P5/SLP-P) Talking about her children; Sue is not one of her children, has not been mentioned in the conversation; listener has no context to understand this utterance. |
| | Paucity of language | The Chicago | 2 | (P9/SLP-P) Dyad talking about details of P9's neighborhood, her house and road, which are 1,000 miles away from Chicago. |
| | | I got the stuff out of here together | 1 | (P5/Home-P) Partner asked about a Christmas decoration, a village on the tabletop; P5 points around the room, unclear what she's referring to or even if she is referring to decorations. |
| | | It's a lot of stuff | 2 | (P1/SLP-P) Talking about an upcoming event (e.g., a wedding). |
| | | One and Ellie, Ell, Anna was going to go | 2 | (P5/SLP-P) P5 drew a picture of a video cassette, trying to explain something about a family video. The utterance contains her name and her partner's name, but unclear connection between the action P5 wants them to complete with the video. |
| Incorrect | Incorrect word(s) are extreme enough to reduce coherence **(in bold)** | And she **waves** and gets them | 2 | (P8/SLP-P) Trying to explain her daughter's job (owner of a restaurant, which SLP-P does not know); she has expressed "selling at a store" already, which is followed by this utterance. "Waves" is subsequently identified as a real word paraphasia, reducing the global coherence of this utterance to the topic of owning a restaurant. |
| | Nonfactual information | I **thub** telephone | 2 | (P9/Home-P) Partner has asked P9 to be sure to text him later that night to let him know that she has safely cooked her dinner. The neologism "thub" reduces the coherence to the topic because it is unclear what she intends to do with the phone or is asking about the phone; it is even possible in context that she is commenting that she lost her phone. |
| | | I know every **play** right now | 2 | (P3/SLP-P) The partners have just realized they both used to live in the same small town, and SLP mentions it is a farming town. This real word paraphasia reduces the global coherence because it seems that P3 may be talking about a theater in the town. |
| Off-topic | Not related to topic Could be sudden/ noncoherent topic change, resulting in GCSconv score of 1 *(in italics)* | *Daniel called me last night* | 1 | (P4/Home-P) Partners are talking about the advantages of irons versus steam cleaners. P4 abruptly changes topic with this utterance, which is about Daniel, her son, sending her tea and chocolate for Mother's Day. |
| | | *My mother have to come to stroo {in context "stroo" is understandable as "school"}* | 1 | (P6/SLP-P) Partners are talking about how many siblings P6 has. P6 abruptly changes topic with this utterance and begins telling a story about her mother getting lost when she would come to P6's elementary school when she was growing up. |
| | | Vermont | 1 | (P1/SLP-P) Partners talking about a CEO of a major company in a large city, which is far from Vermont. P1 says "Vermont," with no apparent connection to previous or subsequent conversation. |
| Detail | Extraneous detail not related to topic | I know that people are working | 2 | (P1/Home-P) The partner just stated she is going to look for work in a foreign country; P1 responds by telling her he has contacts there, either "friends or workers," then says he doesn't know the workers, followed by the comment in the example. |

*Note.* Incorrect words are in bold. Utterances designated as topic-initiating utterances are in italics. GC = global coherence; 4-point GCSconv = Global Coherence Scale in unstructured conversation; SLP-P = speech-language pathologist partner; Home-P = home partner.

[a]Hazamy & Obermeyer (2019).

and feedback from the first author. All transcription and coding discrepancies were resolved through collaborative review.

RAs selected a random location to begin reliability procedures for each conversation, continuing for at least 30% of each conversation, as determined by time (rather than number of utterances; for the 4-point GCSconv coding, the minimum was 33.33%). IRR began at the start of the conversation for seven transcripts and at a middle location for 10 transcripts, and one transcript was coded in its entirety by the RA, due to limited utterances (P9/Home-P). We chose to conduct IRR on at least 30% of 100% of transcripts rather than 100% of 30% of the transcripts (i.e., 6/18 full transcripts), to ensure that transcribers and raters demonstrated reliability across all profiles of PWA in the study. In addition, this procedure is more conservative because the fewer data points per transcript means that each discrepancy between raters represents a higher percentage of the sample in disagreement. Lastly, using 30% of 100% of the 18 samples for IRR yields 18 sets of data points (rather than only six) for analysis, allowing a more strongly powered intraclass correlation, which we used to evaluate IRR. To ensure the validity of this decision and our decision to use at least 30% of the time elapsed rather than 30% of the number of utterances per conversation as well as to compare methodologies, we also conducted IRR on 100% of six transcripts (three from Home-P conversations, three from SLP-P conversations; one from a person with severe aphasia, one from someone with moderate aphasia, four from participants with mild aphasia) and on at least 30% of the PWA utterances per transcript (rather than using 30% of the data per transcript as determined by time) for the 4-point GCSconv (see Supplemental Material S1). Differing methodologies resulted in negligible differences only.

Point-to-point reliability for transcription and low coherence categories was calculated using the following formula: total agreements/(total agreements + total disagreements) × 100. Transcription reliability included words, Analysis of Speech Unit segmentation, interruptions, and abandoned utterances. The average transcription IRR was 90% (range: 80%–96%). Coding reliability for reduced global coherence categorization was 83% and was calculated for all samples together because some conversations had three or fewer utterances rated with poor global coherence (i.e., 1 or 2). IRR was 90% for topic location and 93% for mechanism of topic initiation. These data are a subset of reliability data from a previous study (Leaman & Edmonds, 2020; please note because of the global nature of topic in discourse, topic location and coding reliability protocol requires IRR to be conducted on entire transcripts; the IRR data reported are for 6/18 of the conversations in this study, with three from the Home-P condition and three with the SLP-P).

Intrarater reliability was completed by the first author on at least 10% of the data for transcription and coding, consistent with Wright et al.'s procedures (2013) for the 4-point GCS. The above formula was used to make the following calculations: Average intrarater reliability for the first author for transcription was 95% (range: 92%–100%; conducted on 10% of every conversation); for low coherence categories, 100% (40/40; full sample due to few items); for topic location, 89% (for 23/26 topics in 2/18 full transcripts; range: 88%–90%); and for topic mechanism, 100% (for 24/24 mechanisms in 2/18 full transcripts). A minimum of 3 months elapsed between the ratings by the first author.

## Results

The data were complete, with normal distributions per the Shapiro–Wilk test, meeting the assumptions required for the analyses below. The 10% intrarater reliability data (RQ1) were an exception, displaying a nonparametric distribution; thus, analysis was via Spearman rho (ρ) procedure. IBM SPSS 26 was used for data analysis. Please see Table 1 for number of utterances and time for each conversation.

RQ1: IRR for the GCSconv was evaluated for two rater pairs using an interclass correlation with a two-way random effects model with absolute agreement and single rater measures with the following results: RA1/first author: $\alpha = .983$, $p = .000$, 95% CI [.921, .997]; RA2/first author: $\alpha = .932$, $p = .000$, 95% CI [.711, .986]. Intrarater reliability for the first author's ratings for the GCSconv was evaluated using Spearman ρ correlational procedure: $\rho = .985$, $p = .000$.

RQ2: Test–retest stability of the GCSconv was evaluated using an interclass correlation two-way random effects model with consistency agreement and single rater measures, demonstrating a significant correlation of the measure across the two conversational conditions: $\alpha = .905$, $p = .000$, 95% CI [.639, .978].

RQ3: Rates of global coherence errors by category for the group were as follows: Nonspecific = 59.4%, Off-topic = 26.3%, Incorrect = 9.8%, Detail = 1.5%; for the people with mild expressive deficits: Nonspecific = 22.6%, Off-topic = 48.4%, Incorrect = 16.1%, Detail = 6.5%; for the people with severe expressive deficits: Nonspecific = 60.7%, Off-topic = 12.5%, Incorrect = 5.4%, Detail = 0%.

## Discussion

The data presented for nine people with varying types and severities of aphasia demonstrate excellent IRR for the 4-point GCSconv measure, with excellent correlation between the first author and each of the two RAs (i.e., $\alpha \geq .87$; Koo & Li, 2016; Portney & Watkins, 2015). Intrarater reliability for the GCSconv was excellent as well. Likewise, the test–retest stability of the 4-point GCSconv was excellent, with a high significant correlation of performance across conversations adequate to demonstrate suitability of the measure to evaluate performance of individuals, although with some degree of caution given the wide confidence interval (Boyle, 2014; Fitzpatrick et al., 1998). These results are consistent with our previous findings regarding a number of microlinguistic measures and

for a measure of overall communicative success (Leaman & Edmonds, 2019a, 2019b). These findings extend the research base suggesting potential to reliably measure linguistic skills in unstructured conversation in PWA (Carragher et al., 2013, 2015; Leaman & Edmonds, 2019a, 2019b).

A possible concern in statistical analysis of unstructured conversation is that conversation is inherently variable in terms of subject matter. As discussed earlier, we have used the term "test–retest reliability" in our work, but another option to consider is "parallel forms reliability," used by Wright et al. (2013) to evaluate reliability of the 4-point GCS across two different wordless picture books, in which exact content also varies. Parallel forms reliability typically used evaluates the stability of scores derived from two different formal test battery forms with varied test items on each form, whereas test–retest reliability evaluates the same test on two different dates (Trochim, 2006). Both procedures use the same correlational statistics to establish the extent of reliability of the measures across two conditions. Neither definition of reliability is an exact fit meeting our needs to describe reliability of measures collected across two different dates in unstructured conversation. However, we use the term "test–retest reliability" because our methodology conceptualizes the conversation condition as being more closely aligned with being an example of the same test on two different dates rather than an example of a formalized test battery containing two different forms each with a list of test questions evaluating similar constructs. Additionally, evaluating reliability in conversation with varying topics and/or differing conversation partners potentially reduces the reliability of the measures rather than the opposite, as these contexts are less controlled. Thus, previously reported results such as those by Carragher et al. (2013, 2015) and Leaman and Edmonds (2019a, 2019b), along with the current findings, demonstrate reliability of measures despite methodology that did not control unstructured conversation. Even in the context of unstructured conversation, the adapted 4-point GCSconv maintained stability. This is important information for moving the field toward inclusion of natural communication contexts in assessment and intervention.

The analysis of utterances receiving global coherence scores of 1 and 2 (RQ3) demonstrates that most of the utterances receiving 4-point GCSconv scores of 1 and 2 occurred due to use of nonspecific language (59.4%) for the group. Utterances coded to this category either contained vague words such as "thing," "it," and "stuff" (indicating lexical retrieval difficulty), or simply did not contain an adequate amount of language for the listener to determine how the utterance was related to the topic. Utterances that were coded as off-topic (26.3%) were generally adequate in terms of microlinguistic production but either were overtly off-topic or demonstrated such paucity of language that the listener could make no connection to the ongoing topic. When no such connection could be made by the rater (including the rater's observation of the partner's behavior and response), the off-topic category was coded. Hazamy and Obermeyer (2019) reported the off-topic category occurred for a much lower percentage of their data (3.6%; calculated from published data) than in our findings. However, it may be that the off-topic category accounted for a higher percentage of the low global coherence utterances in our data at least in part because the 4-point GCSconv does not rate commentary and incomplete utterances, which were the two most frequent classifications of the monologue data evaluated by Hazamy and Obermeyer.

The incorrect category accounted for 9.8% of low coherence utterances. This category captures lexical retrieval errors produced as paraphasias that were powerful enough to erode the listener's ability to understand the coherence of the utterance. Utterances containing lexical retrieval errors were not automatically coded to the incorrect category; the error needed to diminish global coherence (e.g., "I thubbed telephone," stated by P9, with precise meaning of her intended verb ambiguous and unclear in context). The final category, detail (1.5%), was coded for only two utterances by one participant (P1) that provided excessive detail rated as ancillary to the topic. Each of these two utterances was syntactically intact, and each contained one semantic paraphasia.

The nonspecific, off-topic, and detail categories were characterized by utterances that may have contained microlinguistic errors, but often did not. Instead, a characteristic observed frequently was reduced or incomplete output (in nonfluent participants) or reduced informative output (in fluent participants), rather than frank errors. These findings are consistent with findings by Hazamy and Obermeyer (2019) who described fluent participants as demonstrating reduced informativeness and nonfluent participants as demonstrating incomplete utterances. Thus, for utterances assigned to these categories, so much information was vague or left unspoken that the listener had difficulty knowing how to effectively use the normal conversational process of inferencing to piece together how the utterances were coherent to the whole (G. Brown & Yule, 1983; van Dijk, 1977).

Although speakers are expected to reduce and eliminate unnecessary language in conversation (i.e., use of ellipsis is expected), they must produce enough language (or express ideas nonverbally) so as to allow the partner to easily and rapidly make inferences as to how utterances within a discourse connect to the topic (G. Brown & Yule, 1983; van Dijk, 1977). During early stages of discourse production, during which ideas are developed and organized coherently (i.e., global coherence planning occurs), the speaker must make careful selections about which information must be included and what may be left out (Sherratt, 2007; van Dijk, 1977), based on many contextual parameters (e.g., the nature of the relationship with the partner, assumed common knowledge between the speakers, events of the ongoing discourse, knowledge of the environment [props, objects, etc.]; Coates, 1995; Sherratt, 2007). Underlying difficulty with global coherence at this stage of production could result in the absence of elements considered crucial for the listener's comprehension of a globally

coherent discourse, such as seen in our samples. However, in PWA, reduced language could also occur secondary to later stages of discourse production when lexical, grammatical, and syntactic assignments are made. Rather than producing overt lexical, grammatical, and syntactic errors, PWA may simply produce less verbal output, with greater reliance on ellipsis (i.e., higher demand for inferencing by the listener). Consequently, in utterances characterized by reduced language, teasing out the exact location responsible for the breakdown of discourse production is challenging.

However, if paucity of verbal language is due solely to later stage discourse production challenges (e.g., microlinguistic challenges), it would be reasonable to expect that in order to ease the listener inferencing load, speakers would use alternative communication modalities and/or select maximally salient verbal language (and omit low-saliency verbalizations altogether) more often than the group results demonstrate. Given that the 4-point GCSconv scoring system includes consideration of nonverbal communication and context, utterances with reduced language are not automatically scored for low global coherence if they include nonverbal communication and/or use context to express the utterance's coherence to the topic. Consequently, utterances receiving low global coherence scores in this study were produced with insufficient nonverbal/contextual additions to support expression of global coherence. The data indicate that utterances coded as nonspecific, off-topic, and irrelevant appear to be manifestations of both microlinguistic difficulties and higher level difficulty with coherence.

We further examined the distribution of utterances across the categories of low global coherence for the participants with severe expressive deficits (P3, P5, P9) separately from those with milder expressive language difficulty. However, results for the mild group should be interpreted with caution, as the sample size of utterances scored with low global coherence was small (i.e., 31 utterances). The sample of utterances in the severe expressive language group was composed of 112 utterances. This analysis showed different patterns of errors, with the group with severe expressive deficits using utterances described as nonspecific (60.7%) far more often than those with milder expressive difficulty (22.6%). In addition, those with mild expressive deficits used utterances described as off-topic (48.4%) more often than those with severe expressive deficits (12.5%). These findings are somewhat intuitive, suggesting that people with milder expressive language may have adequate skills to produce utterances with greater specificity than those with severe expressive deficits (i.e., fewer utterances categorized as nonspecific), whereas those with more severe expressive deficits demonstrate a degree of lexical retrieval difficulty that may result in higher levels of nonspecific language including a general impoverishment of output that is captured by this category. Furthermore, with better access to verbal language, the people with mild aphasia had more facility to produce sentences with enough language so as to be recognized as off-topic, whereas the extremely limited language of the participants with severe expressive language deficits often resulted in utterances that were vague either because output prevented specification of relationships between lexical items (the nonfluent participants) or due to reduced informativeness (the fluent participants). In these situations, reduced content, due to either reduced number of words or reduced informativeness of words, resulted in assignment to the nonspecific category, not the off-topic category, as an utterance could not be assumed to be off-topic unless the language produced overtly demonstrated that this was the case.

Lastly, to further explore and understand patterns of global coherence, we conducted a post hoc analysis examining the performance of each individual (see Supplemental Material S2). Consistent with findings in monologue, global coherence in conversation is reduced as aphasia becomes more severe (see data for P3, P5, and P9). The data demonstrate test–retest score differences that varied only by an average of .23 points (range: .04–.50, $SD$ = .18), yet with a pattern where the GCSconv measure tends to be higher in the Home-P conversation in comparison to the SLP-P conversation. This occurred for all participants except one (P7, mild anomia; his difference between conversations was .04 points SLP-P > Home-P). A factor that may have contributed to this trend could be related to the Home-P having a more thorough knowledge of conversational topics, contexts, and/or persons discussed by the person with aphasia than the SLP-P. This difference in relationship status may have allowed for a more efficient and effective understanding of the person with aphasia's utterances by the Home-P in comparison to the SLP-P, even when degraded by microlinguistic errors, resulting in higher 4-point GCSconv scores for those utterances. For example, in some of the SLP-P conversations, the condition of unfamiliarity meant that the person with aphasia needed to explain the referential "who, what, where, when" aspects of stories more comprehensively than with the Home-P, who often had such background knowledge in common with the PWA. This is illustrated by an example produced by P8 (mild anomia) in which she produced utterances with reduced global coherence while trying to explain her family relationships (see Supplemental Material S3) to the SLP-P. The observed pattern of 4-point GCSconv scores appears to be (at least in part) related to the conversation conditions that vary on use of familiar versus unfamiliar partners, which introduced social, contextual, and interactional differences into the samples. However, importantly, while a pattern is detectable upon examination of the individual data, it must be noted that this pattern was not so pronounced as to prevent stability of the 4-point GCSconv across the two conditions.

## Limitations and Future Directions

The current research provides preliminary data supporting the potential to measure global coherence in unstructured conversation, a critical context for everyday interactions and a key focus for rehabilitation. This initial sample is fairly small, consisting predominantly of people with mild aphasia. The focus in this research has been on factors reducing global coherence, with limited attention to

the specifics of unique or compensatory coherence-building resources that may be used by some PWA. The current data indicate that some people with severe expressive language deficits may have difficulty maximizing contextual resources to build coherence, suggesting potential benefit from treatment addressing these skills. Future research with a larger, more diverse group of PWA and attention to how PWA build coherence within the context of limited verbal language will be beneficial in understanding global coherence in conversation and in further developing measures and interventions.

In addition, our current methodology for use of the 4-point GCSconv necessitates a process for locating and coding multiple topics per conversation, which requires added training and time for use of the scale. Investigation as to whether one or just a few highly salient (i.e., easily identifiable) topics arising in unstructured conversation could be used reliably in place of an entire conversational sample with multiple topic switches would be of value to streamline application of the 4-point GCSconv.

Lastly, utterance-by-utterance analysis of discourse samples is not practical for most clinical needs. Thus, future research is indicated investigating relationships between sample-level and utterance-level measures of global coherence. Such research has begun to emerge in the monologue literature for several discourse measures, including the 4-point GCS (Rogalski et al., 2019) and Nicholas and Brookshire's (1993) Correct Information Unit (Webster et al., 2018). If similar positive relationships are demonstrated between utterance- and sample-level measures of unstructured conversation, development of reliable and efficient sample-level measures could be possible, making discourse measurement more clinically feasible.

## Conclusions

This research demonstrates the feasibility of measuring global coherence in unstructured conversation using an utterance-level measure, the 4-point GCSconv, based on a similar monologue-level measure, the 4-point GCS (Wright & Capilouto, 2012; Wright et al., 2013). The findings demonstrate excellent inter- and intrarater reliability and test–retest stability for the measure as applied in unstructured conversation. Importantly, the GCSconv includes contextual aspects of communication crucial to establishing global coherence in unstructured conversation. It is a measure that includes surface-level production and nonverbal/contextual information. Thus, it is sensitive to microlinguistic errors extreme enough to negatively impact global coherence and is sensitive to nonverbal/contextual resources that PWA may use to build global coherence. PWA included in this sample demonstrated low global coherence most often due to use of nonspecific (vague) language (59.4% of data). Off-topic utterances accounted for only 26.3% of low global coherence utterances, and irrelevant and incorrect utterances interfered with global coherence infrequently. Error patterns differed for people with mild and severe expressive language deficits (respectively, 48.4% errors were off-topic; 60.7% were vague). Demonstrating the 4-point GCSconv as a reliable measure allows future research that can further explore the relationships between global coherence in unstructured conversation and structured discourse, relationships between sample-level and utterance-level measures, and that can be used to inform assessment and treatment of everyday conversation in PWA.

## Acknowledgments

Thank you to the participants in this study and to the Aphasia Rehabilitation and Bilingualism Lab, Teachers College, Columbia University. In addition, many thanks to research assistants Holly DiClemente, Olivia Kelly, Hannah Penner, and Trisha Vu.

## References

**Agar, M., & Hobbs, J. R.** (1982). Interpreting discourse: Coherence and the analysis of ethnographic interviews. *Discourse Processes, 5*(1), 1–32. https://doi.org/10.1080/01638538209544529

**Andreetta, S., Cantagallo, A., & Marini, A.** (2012). Narrative discourse in anomic aphasia. *Neuropsychologia, 50*(8), 1787–1793. https://doi.org/10.1016/j.neuropsychologia.2012.04.003

**Andreetta, S., & Marini, A.** (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. *Aphasiology, 29*(6), 705–723. https://doi.org/10.1080/02687038.2014.979394

**Beeke, S., Wilkinson, R., & Maxim, J.** (2007). Individual variation in agrammatism: A single case study of the influence of interaction. *International Journal of Language & Communication Disorders, 42*(6), 629–647. https://doi.org/10.1080/13682820601160087

**Boyle, M.** (2011). Discourse treatment for word retrieval impairment in aphasia: The story so far. *Aphasiology, 25*(11), 1308–1326. https://doi.org/10.1080/02687038.2011.596185

**Boyle, M.** (2014). Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research, 57*(3), 966–978. https://doi.org/10.1044/2014_JSLHR-L-13-0171

**Brown, R.** (1973). *A first language: The early stages*. Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

**Brown, G., & Yule, G.** (1983). *Discourse analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511805226

**Brysbaert, M., & New, B.** (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

**Carragher, M., Sage, K., & Conroy, P.** (2013). The effect of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation. *Neuropsychological Rehabilitation, 23*(6), 846–887. https://doi.org/10.1080/09602011.2013.832335

**Carragher, M., Sage, K., & Conroy, P.** (2015). Outcomes of treatment targeting syntax production in people with Broca's-type aphasia: Evidence from psycholinguistic assessment tasks and everyday conversation. *International Journal of Language & Communication Disorders, 50*(3), 322–336. https://doi.org/10.1111/1460-6984.12135

**Charity, A. H.** (2008). African American English: An overview. *SIG 14 Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations, 15*(2), 33–42. https://doi.org/10.1044/cds15.2.33

**Charolles, M.** (1983). Coherence as a principle in the interpretation of discourse. *Text, 3*(1), 71–97. https://doi.org/10.1515/text.1.1983.3.1.71

**Christiansen, J. A.** (1995). Coherence violations and propositional usage in the narratives of fluent aphasia. *Brain and Language, 51*(2), 291–317. https://doi.org/10.1006/brln.1995.1062

**Clark, H. H., & Carlson, T. B.** (1982). Hearers and speech acts. *Language, 58*(2), 332–373. https://doi.org/10.2307/414102

**Coates, J.** (1995). The negotiation of coherence in face-to-face interaction: Some examples from the extreme bounds. In M. A. Gernsbacher & T. Givon (Eds.), *Coherence in spontaneous text* (pp. 41–58). John Benjamins. https://doi.org/10.1075/tsl.31.03coa

**Coelho, C., & Flewellyn, L.** (2003). Longitudinal assessment of coherence in an adult with fluent aphasia: A follow-up study. *Aphasiology, 17*(2), 173–182. https://doi.org/10.1080/729255216

**Davidson, B., Howe, T., Worrall, L., Hickson, L., & Togher, L.** (2008). Social participation for older people with aphasia: The impact of communication disability on friendships. *Topics in Stroke Rehabilitation, 15*(4), 325–340. https://doi.org/10.1310/tsr1504-325

**Davidson, B., Worrall, L., & Hickson, L.** (2003). Identifying the communication activities of older people with aphasia: Evidence from naturalistic observation. *Aphasiology, 17*(3), 243–264. https://doi.org/10.1080/729255457

**Dipper, L., Pritchard, M., Walkden, E., & Cruice, M.** (2018). How do speakers with and without aphasia use syntax and semantics across two discourse genres. *Aphasiology, 32*(6), 720–738. https://doi.org/10.1080/02687038.2018.1447642

**Ellis, C., Henderson, A., Wright, H. H., & Rogalski, Y.** (2016). Global coherence during discourse production in adults: A review of the literature. *International Journal of Language & Communication Disorders, 51*(4), 359–367. https://doi.org/10.1111/1460-6984.12213

**Fillmore, C. J.** (2015). The future of semantics. In R. Austerlitz (Ed.), *The scope of American linguistics: The first golden anniversary of the Linguistic Society held at the University of Massachusetts, Amherst, on July 24–25, 1974* (pp. 135–157). De Gruyter Mouton.

**Fitzpatrick, R., Davey, C., Buxton, M., & Jones, D. R.** (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment, 2*(14), 1–74. https://doi.org/10.3310/hta2140

**Foster, P., Tonkyn, A., & Wigglesworth, G.** (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354–375. https://doi.org/10.1093/applin/21.3.354

**Frederiksen, C. H., Bracewell, R. J., Breuleux, A., & Renaud, A.** (1990). The cognitive representation and processing of discourse: Function and dysfunction. In Y. Joanette & H. H. Brownell (Eds.), *Discourse ability and brain damage: Theoretical and empirical perspectives*. Springer-Verlag. https://doi.org/10.1007/978-1-4612-3262-9_4

**Glosser, G., & Deser, T.** (1990). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language, 40*(1), 67–88. https://doi.org/10.1016/0093-934X(91)90117-J

**Goodwin, C.** (1981). *Conversational organization: Interaction between speakers and hearers*. Academic Press.

**Hazamy, A. A., & Obermeyer, J.** (2019). Evaluating informative content and global coherence in fluent and nonfluent aphasia. *International Journal of Language & Communication Disorders, 55*(1), 110–120. https://doi.org/10.1111/1460-6984.12507

**Helm-Estabrooks, N.** (2001). *Cognitive Linguistic Quick Test (CLQT)*. The Psychological Corporation.

**Huber, W.** (1990). Text comprehension and production in aphasia: Analysis in terms of micro- and macrostructure. In Y. Joanette & H. H. Brownell (Eds.), *Discourse ability and brain damage: Theoretical and empirical perspectives* (pp. 154–179). Springer-Verlag. https://doi.org/10.1007/978-1-4612-3262-9_7

**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

**Johnson, W.** (1944). Studies in language behavior: I. A program of research. *Psychological Monographs, 56*(2), 1–15. https://doi.org/10.1037/h0093508

**Kertesz, A.** (2006). *Western Aphasia Battery–Revised (WAB-R)*. The Psychological Corporation. https://doi.org/10.1037/t15168-000

**Koo, T. K., & Li, M. Y.** (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

**Kurland, J., & Stokes, P.** (2018). Let's talk real talk: An argument to include conversation in a D-COS for aphasia research with an acknowledgment of the challenges ahead. *Aphasiology, 32*(4), 475–478. https://doi.org/10.1080/02687038.2017.1398808

**Labov, W.** (1972). *Language in the inner city: Studies in the black English vernacular*. University of Pennsylvania Press.

**Labov, W., & Waletzky, J.** (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society* (pp. 12–44). University of Washington Press.

**Leaman, M. C., & Edmonds, L. A.** (2019a). Conversation in aphasia across communication partners: Exploring stability of microlinguistic measures and communicative success. *American Journal of Speech-Language Pathology, 28*(1S), 359–372. https://doi.org/10.1044/2018_AJSLP-17-0148

**Leaman, M. C., & Edmonds, L. A.** (2019b). Revisiting the Correct Information Unit: Measuring informativeness in unstructured conversations in people with aphasia. *American Journal of Speech-Language Pathology, 28*(3), 1099–1114. https://doi.org/10.1044/2019_AJSLP-18-0268

**Leaman, M. C., & Edmonds, L. A.** (2020). "By the way" … How people with aphasia and their communication partners initiate new topics of conversation. *American Journal of Speech-Language Pathology, 29*(1S), 375–392. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0198

**Linnik, A., Bastiaanse, R., & Höhle, B.** (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology, 30*(7), 765–800. https://doi.org/10.1080/02687038.2015.1113489

**Longacre, R. E.** (1996). *The grammar of discourse* (2nd ed., pp. 7–31). Springer Plenum Press. https://doi.org/10.1007/978-1-4899-0162-0

**Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S.** (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology, 25*(11), 1372–1392. https://doi.org/10.1080/02687038.2011.584690

**Mayer, J., & Murray, L.** (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology, 17*(5), 481–497. https://doi.org/10.1080/02687030344000148

**Mentis, M., & Prutting, C.** (1991). Analysis of topic as illustrated in a head-injured and normal adult. *Journal of Speech and*

*Hearing Research, 34*(3), 583–595. https://doi.org/10.1044/jshr.3403.583

**Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H.** (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

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

**Obermeyer, J., Leaman, M. C., & Edmonds, L. A.** (2020). Evaluating change in the conversation of a person with mild aphasia after Attentive Reading with Constrained Summarization–Written treatment. *American Journal of Speech-Language Pathology*. Advance online publication. https://doi.org/10.1044/2020_AJSLP-19-00078

**Olness, G. S., & Englebretson, E. F.** (2011). On the coherence of information highlighted by narrators with aphasia. *Aphasiology, 25*(6–7), 713–726. https://doi.org/10.1080/02687038.2010.537346

**Olness, G. S., Matteson, S. E., & Stewart, C. T.** (2010). "Let me tell you the point": How speakers with aphasia assign prominence to information in narratives. *Aphasiology, 24*(6–8), 697–708. https://doi.org/10.1080/02687030903438524

**Olness, G. S., Ulatowska, H. K., Wertz, R. T., Thompson, J. L., & Auther, L. L.** (2002). Discourse elicitation with pictorial stimuli in African Americans and Caucasians with and without aphasia. *Aphasiology, 16*(4–6), 623–633. https://doi.org/10.1080/02687030244000095

**Patry, R., & Nespoulous, J.-L.** (1990). Discourse analysis in linguistics: Historical and theoretical background. In Y. Joanette & H. H. Brownell (Eds.), *Discourse ability and brain damage: Theoretical and empirical perspectives* (pp. 3–28). Springer-Verlag. https://doi.org/10.1007/978-1-4612-3262-9_1

**Portney, L. G., & Watkins, M. P.** (2015). *Foundations of clinical research: Applications to practice* (3rd ed.). F. A. Davis Company.

**Pritchard, M., Hilari, K., Cocks, N., & Dipper, L.** (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders, 52*(6), 689–732. https://doi.org/10.1111/1460-6984.12318

**Rogalski, Y., Altmann, L. J. P., Plummer-D'Amato, P., Behrman, A. L., & Masriske, M.** (2010). Discourse coherence and cognition after stroke: A dual task study. *Journal of Communication Disorders, 43*(3), 212–224. https://doi.org/10.1016/j.jcomdis.2010.02.001

**Rogalski, Y., Key-DeLyria, S. E., Mucci, S., Wilson, J. P., & Altmann, L. J. P.** (2019). The relationship between trained ratings and untrained listeners' judgments of global coherence in extended monologues. *Aphasiology, 34*(2), 214–234. https://doi.org/10.1080/02687038.2019.1643002

**Sacks, H.** (1992). *Lectures on conversation*. Blackwell.

**Sacks, H., Schegloff, E. A., & Jefferson, G.** (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4–1), 696–735. https://doi.org/10.2307/412243

**Schegloff, E. A.** (2007). *Sequence organization in interaction: A primer in conversation analysis* (Vol. 1). Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

**Schegloff, E. A., Jefferson, G., & Sacks, H.** (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*(2), 361–382. https://doi.org/10.2307/413107

**Schiffrin, D.** (1987). *Discourse markers*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611841

**Sherratt, S.** (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology, 21*(3–4), 376–393. https://doi.org/10.1080/02687030600911435

**Simmons-Mackie, N. N., & Damico, J. S.** (1996). The contribution of discourse markers to communicative competence in aphasia. *American Journal of Speech-Language Pathology, 5*(1), 37–43. https://doi.org/10.1044/1058-0360.0501.37

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

**Swinburn, K., Porter, G., & Howard, D.** (2005). *Comprehensive Aphasia Test*. Psychology Press. https://doi.org/10.1037/t13733-000

**Trochim, W. M.** (2006). *The research methods knowledge base* (2nd ed.). http://www.socialresearchmethods.net/kb

**Ulatowska, H. K., Doyel, A. W., Stern, R. F., Haynes, S. M., & North, A. J.** (1983). Production of procedural discourse in aphasia. *Brain and Language, 18*(2), 315–341. https://doi.org/10.1016/0093-934X(83)90023-8

**Ulatowska, H. K., Freedman-Stern, R., Doyel, A. W., Macaluso-Haynes, S. M., & North, A. J.** (1983). Production of narrative discourse in aphasia. *Brain and Language, 19*(2), 317–334. https://doi.org/10.1016/0093-934X(83)90074-3

**Ulatowska, H. K., North, A. J., & Macaluso-Hayes, S.** (1981). Production of narrative discourse and procedural discourse in aphasia. *Brain and Language, 13*(2), 345–371. https://doi.org/10.1016/0093-934X(81)90100-0

**Ulatowska, H. K., & Olness, G. S.** (2000). Discourse revisited: Contributions of lexico-syntactic devices. *Brain and Language, 71*(1), 249–251. https://doi.org/10.1006/brln.1999.2261

**Ulatowska, H. K., Olness, G. S., Wertz, R., Samson, A., Keebler, M., & Goins, K.** (2003). Relationship between discourse and Western Aphasia Battery performance in African Americans with aphasia. *Aphasiology, 17*(5), 511–521. https://doi.org/10.1080/0268703034400102

**Ulatowska, H. K., Olness, G. S., & Williams, L. J.** (2004). Coherence of narratives in aphasia. *Brain and Language, 91*(1), 42–43. https://doi.org/10.1016/j.bandl.2004.06.024

**van Dijk, T. A.** (1977). *Text and context*. Longman.

**van Dijk, T. A., & Kintsch, W.** (1983). *Strategies of discourse comprehension*. Academic Press.

**Van Leer, E., & Turkstra, L.** (1999). The effect of elicitation task on discourse coherence and cohesion in adolescents with brain injury. *Journal of Communication Disorders, 32*(5), 327–349. https://doi.org/10.1016/S0021-9924(99)00008-8

**Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., Pak Hin Kong, A., Simmons-Mackie, N., Scarinci, N., & Gauvreau, C. A.** (2016). Which outcomes are most important to people with aphasia and their families? An international nominal group technique study framed within the ICF. *Disability and Rehabilitation, 39*(14), 1364–1379. https://doi.org/10.1080/09638288.2016.1194899

**Webster, J., Harrison, S., & Morris, J.** (2018). Communicative informativeness in aphasia. *Aphasiology, 32*(Suppl. 1), 243–245. https://doi.org/10.1080/02687038.2018.1485840

**Wilkes-Gibbs, D.** (1995). Coherence in collaboration: Some examples from conversation. In M. A. Gernsbacher & T. Givon (Eds.), *Coherence in spontaneous text* (pp. 239–267). John Benjamins. https://doi.org/10.1075/tsl.31.09wil

**Wilkinson, R.** (1995). Aphasia: Conversation analysis of a non-fluent aphasic person. In M. Perkins & S. Howard (Eds.), *Case studies in clinical linguistics* (pp. 272–292). Whurr Publishers.

**Wolfram, W.** (2004). The grammar of urban African American vernacular English. In B. Kormann, C. Upton, K. Burridge, E. W. Schneider, & R. Mesthrie (Eds.), *A handbook of varieties of English: Phonology, morphology and syntax* (pp. 111–132). De Gruyter Mouton.

**Wright, H. H., & Capilouto, G. J.** (2012). Considering a multilevel approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology, 26*(5), 656–672. https://doi.org/10.1080/02687038.2012.676855

**Wright, H. H., Capilouto, G. J., & Koutsoftas, A.** (2013). Evaluating measures of global coherence ability in stories in adults. *International Journal of Language & Communication Disorders, 48*(3), 249–256. https://doi.org/10.1111/1460-6984.12000

## Appendix

Four-Point GCSconv: Scale Summary, Adaptations, Additions to the 4-Point GCS[a]

| Scale summary |
|---|
| ▪ 4 - The utterance is overtly related to the topic, with needed mention of actors, actions, and/or objects of significant importance. |
| ▪ 3 - The utterance is related to the topic but may include tangential information. Key information may be missing, or irrelevant information may be provided requiring inferencing. |
| ▪ 2 - The utterance is remotely related to the topic and may include nonsignificant details. |
| ▪ 1 - The utterance is not related to the topic in any way. |
| **Adaptations** |
| Rater includes context, including, for example (list is not exhaustive):[b] |
| ▪ Partner's behaviors and verbal responses that suggest whether the PWA's turn is received as globally coherent by the participant in the conversation |
| ▪ Situational context (i.e., objects in the room, material from the conversations) |
| ▪ General world knowledge contexts, reasonably expected |
| ▪ Common knowledge developed in the course of the conversation |
| ▪ Each utterance within the sequence of the unfolding conversation |
| Not penalized |
| ▪ Opinions, elaborations, and personal stories if otherwise coherent (i.e., evaluative comments) |
| Not rated |
| ▪ Commentary utterances regarding the discourse (i.e., "I can't think of the word") |
| **Additions** |
| Rated |
| ▪ Turns using writing, pointing to writing, and oral spelling are treated as if they are verbal |
| ▪ Repetition when used to make a point (i.e., evaluative function) |
| ▪ Repetition when used to answer a question from partner (receives same score as previous iteration of the utterance) |
| Not penalized |
| ▪ Utterances containing pronouns referenced by objects in the environment (i.e., deictic usage), for example, if speaker says "I went there" while showing a postcard of Italy, if otherwise coherent |
| Not rated |
| ▪ Interrupted/overlapped utterances that are incomplete |
| ▪ Abandoned utterances |
| ▪ Requests for clarification by PWA |
| ▪ Single-word yes/no responses (including "mhm," "yeah," "ok," "alright," "nope," etc.) |
| ▪ Utterances made up only of filler language (such as "and so…," "but…," "you know," "anyways…," etc.) |

*Note.* Four-point GCSconv = 4-point Global Coherence Scale in unstructured conversation; PWA = people with aphasia.

[a]Wright & Capilouto (2012) and Wright et al. (2013). [b]Coates (1995), Wilkes-Gibbs (1995), and Sacks et al. (1974).