

Old Dominion University

## ODU Digital Commons

---

Communication Disorders & Special Education  
Faculty Publications

Communication Disorders & Special Education

---

2022

# Conversation as an Outcome of Aphasia Treatment: A Systematic Scoping Review

Jamie H. Azios

Brent Archer

Nina Simmons-Mackie

Anastasia Raymer

*Old Dominion University, sraymer@odu.edu*

Marcella Carragher

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.odu.edu/cdse\\_pubs](https://digitalcommons.odu.edu/cdse_pubs)



Part of the [Speech Pathology and Audiology Commons](#)

---

### Original Publication Citation

Azios, J. H., Archer, B., Simmons-Mackie, N., Raymer, A., Carragher, M., Shashikanth, S., & Gulick, E. (2022). Conversation as an outcome of aphasia treatment: A systematic scoping review. *American Journal of Speech-Language Pathology*, 31(6), 2920-2942. [https://doi.org/10.1044/2022\\_ajslp-22-00011](https://doi.org/10.1044/2022_ajslp-22-00011)

This Article is brought to you for free and open access by the Communication Disorders & Special Education at ODU Digital Commons. It has been accepted for inclusion in Communication Disorders & Special Education Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact [digitalcommons@odu.edu](mailto:digitalcommons@odu.edu).

---

**Authors**

Jamie H. Azios, Brent Archer, Nina Simmons-Mackie, Anastasia Raymer, Marcella Carragher, Shriya Shashikanth, and Eleanor Gulick

## Review Article

# Conversation as an Outcome of Aphasia Treatment: A Systematic Scoping Review

Jamie H. Azios,<sup>a</sup> Brent Archer,<sup>b</sup> Nina Simmons-Mackie,<sup>c</sup> Anastasia Raymer,<sup>d</sup> Marcella Carragher,<sup>e,f</sup> Shriya Shashikanth,<sup>a</sup> and Eleanor Gulick<sup>b</sup>

<sup>a</sup>Department of Communicative Disorders, University of Louisiana at Lafayette <sup>b</sup>Department of Communication Sciences and Disorders, Bowling Green State University, OH <sup>c</sup>Department of Health & Human Sciences, Southeastern Louisiana University, Hammond <sup>d</sup>Department of Communication Disorders & Special Education, Old Dominion University, Norfolk, VA <sup>e</sup>School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Victoria, Australia <sup>f</sup>Centre for Research Excellence in Aphasia Recovery and Rehabilitation, La Trobe University, Melbourne, Victoria, Australia

## ARTICLE INFO

## Article History:

Received January 15, 2022

Revision received May 24, 2022

Accepted August 4, 2022

Editor-in-Chief: Katherine C. Hustad

Editor: Sarah Elizabeth Wallace

[https://doi.org/10.1044/2022\\_AJSLP-22-00011](https://doi.org/10.1044/2022_AJSLP-22-00011)

## ABSTRACT

**Purpose:** Conversation-focused speech-language pathology services are a top priority for people living with aphasia, but little is known about how researchers measure conversation as an outcome of treatment. This scoping review was undertaken to systematically review the evidence regarding the measurement of conversation in aphasia studies and to identify current practices and existing gaps.

**Method:** A systematic literature search was conducted for studies published between January 1995 and September 2019 in multiple electronic databases. Covidence software was used to manage search results, study selection, and data charting processes. Data were extracted from each study and then collated and organized to elucidate the breadth of approaches, tools, or procedures oriented to measuring conversation as an outcome and identify gaps in the existing literature.

**Results:** The systematic search of the literature resulted in 1,244 studies. A total of 64 studies met inclusion criteria and were included in the review. The review summarizes the various tools and procedures used to measure conversation as an outcome of aphasia intervention, including variations in data collection and analysis procedures. The review also evaluates the quality of conversation measures in terms of psychometric properties and informal measures of validity. There was a total of 211 measures used across the 64 studies.

**Conclusions:** While there was no clear measure that was objectively superior, several measures show promise and warrant future exploration. Some of the orientations, conceptualizations, and procedures we have presented can be seen as options that might be included in a future conversation-focused core outcome set.

**Supplemental Material:** <https://doi.org/10.23641/asha.21514062>

Conversation is an integral part of the human experience. At its simplest, conversation is defined as at least two people taking turns, listening to one another and producing relevant utterances (Goodwin & Heritage, 1990). Conversation allows us to transmit information to other people, to coordinate our actions and resources in pursuit of complex goals (Goodwin, 2000), and to build the social

worlds we inhabit (Schegloff, 2006). Given the importance, ubiquity, and complexity of casual, talk-based interactions in everyday life, any restriction to full participation in conversation has the potential for devastating impacts. People with communication disabilities such as aphasia and their close others experience challenges participating in conversation, leading to negative consequences for relationships (Beeke et al., 2007), mental health (Cruice et al., 2003), and work opportunities (Morris et al., 2011). Unsurprisingly, improving conversation abilities is a top priority discussed by people living with aphasia (Wallace et al., 2017; Worrall et al., 2011).

Correspondence to Jamie H. Azios: [jamie.azios@louisiana.edu](mailto:jamie.azios@louisiana.edu). **Disclosure:** The authors have declared that no competing financial or non-financial interests existed at the time of publication.

To address the needs of people and families living with aphasia, researchers and clinicians have developed a range of treatments that explicitly seek to improve conversational skills. Treatments include those that are designed to promote meaningful participation by people with aphasia in everyday conversations and are informed by a broad array of theories and frameworks; for example, conversation analysis (CA), the Life Participation Approach to Aphasia (LPAA) and models of counseling (for details, see Simmons-Mackie et al., 2014). Treatments may also involve techniques designed to improve the underlying word-retrieval impairment through didactic or structured practice on a given set of stimuli (e.g., del Toro et al., 2008). The ultimate goal of many intervention studies is to improve everyday communication. Therefore, regardless of the underlying theoretical model that drives therapeutic structure or technique, outcomes can and do reflect the fundamental goal of improving the ability to participate in conversational exchanges.

Investigations of the efficacy and effectiveness of therapies aimed at improving conversation employ a wide range of outcome measures. The dimensions of the World Health Organization International Classification of Function model can be used to classify assessment procedures that appear in the literature. Some measures focus on the person with aphasia's individual *impairment* and count units such as words (e.g., Boles, 1997), morphemes (e.g., Goldberg et al., 2012), or "correct information units" (words that are intelligible, accurate, and relevant to the spoken context; e.g., del Toro et al., 2008) produced by the person with aphasia. Other procedures assess the extent to which the person with aphasia and partners are able to engage in the *activity* of having a conversation. For example, the Conversation Analysis Profile for People with Aphasia (CAPPA; Whitworth et al., 1997) or the Profile of Word Errors and Retrieval in Speech (POWERS; Herbert et al., 2013) both examine how turns produced by people with aphasia and interlocutors form part of and contribute to conversation. Similarly, Kagan et al. (2001, 2004) introduced, validated, and assayed measures that quantify how well people with aphasia and their conversation partners participate in conversation. Finally, some instruments identify the life domains and communicative settings in which people with aphasia are able to meaningfully *participate*, such as the Communication Activities of Daily Living (Holland et al., 2018) or the American Speech-Language-Hearing Association Functional Assessment of Communication Skills for Adults (Fratalli et al., 2017).

Conversation-focused measures are also used in nontreatment studies where variables such as conversation partner familiarity or discourse genre are manipulated and the effects on language or communication are monitored. For example, Leaman and Edmonds (2019b) investigated

the reliability and stability of seven linguistic measures in unstructured conversation.

## Challenges Related to Measuring Conversation-Focused Outcomes

In a recent survey of current practices of clinicians and researchers using spoken discourse, conversation with a clinician and/or family member was found to be a commonly collected sample of spoken discourse, second only to picture description (Stark et al., 2021). Yet, despite the widespread collection of conversation data, no consensus exists on which measures are most suitable for assessing conversations involving people with aphasia. Researchers use a variety of procedures for gathering spoken discourse data, such as semistructured interviews, video retell tasks, casual conversation, and interview guides (Bryant et al., 2016; Stark et al., 2021). Reviews of discourse measures more broadly have identified 536 unique structural measures (Bryant et al., 2016) and 58 information-focused measures (Pritchard et al., 2017).

This overabundance of approaches creates significant problems for researchers and clinicians. It is difficult for researchers to reproduce, compare, and interpret findings when they conceptualize and measure outcomes in a wide variety of ways. Meta-analyses of treatment studies are especially difficult without sufficient uniformity across study designs, methods, and measures. Speech-language pathologists who work with people with aphasia may find it extremely challenging to select appropriate and feasible measures from the hundreds of options featured in the literature. Moreover, lack of access to tools and resources, inadequate training in discourse collection, and the labor-intensive process of transcription and coding procedures (Kurland & Stokes, 2018; MacWhinney, 2014; Stark et al., 2021) may discourage clinicians from providing conversation-based services and measuring conversation outcomes.

It is timely to document the range of assessment practices related to conversation. While other areas of assessment in aphasia have moved toward a core outcome set (COS; Wallace et al., 2019), the overabundance of discourse measures and lack of psychometric data means it is not currently feasible to achieve this in conversation or other types of discourse (see Dietz & Boyle, 2018). An overview of the research landscape could suggest theoretical frameworks on which a future conversation-COS might be built, furnish insights concerning the relative validity of different measures so that comparisons between intervention approaches can be made, and provide practical guidance regarding how to gather and analyze conversation data. Pritchard et al. (2017, 2018) reviewed the psychometric properties of informational, structural, and cohesive measures in relation to monologic data; Doedens and Meteyard (2020) reviewed measures of functional,

real-world communication; Bryant et al. (2016) reviewed studies that provided descriptive reports or counts of lexical, grammatical, and semantic structures within samples; and a number of surveys have been conducted to identify clinician and researchers' collection and analysis of spoken discourse (Bryant et al., 2017; Cruice et al., 2020; Stark et al., 2021). The current scoping review will add to the literature by focusing on the specific discourse genre of conversation. This is warranted due to the priority placed on conversation by people with aphasia and their close others (e.g., Wallace et al., 2017) as well as the routine nature of clinicians collecting conversation samples as a part of assessment or treatment (Stark et al., 2021). This line of research could ultimately inform consensus positions about methods for assessing conversation outcomes.

## Aim

This scoping review will present a current synopsis of the tools and procedures used to assess conversation within the field of aphasiology. Eligible studies will include those that investigate the psychometric properties of a measure designed to assess conversation gains as well as studies that investigate the impact of a treatment on conversation abilities. The current scoping review will complement the review by Bryant et al. (2016) by focusing on particular issues related to using conversation as an outcome of aphasia intervention.

## Method

This scoping review was guided by the methodological framework suggested by Arksey and O'Malley (2005). Additionally, guidance and standards published by the Joanna Briggs Institute (JBI; Peters et al., 2020) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (Tricco et al., 2018) checklist were followed to increase methodological rigor. The scoping review methodology was chosen to investigate the range and nature of measurement of conversation, identify existing practices and gaps, and determine the feasibility and relevance of a full systematic review (Arksey & O'Malley, 2005). A review protocol was developed by the research team prior to initiating the study; however, this was not registered or published. A complete list of procedures in this protocol may be requested from the corresponding author, but we have provided as many details as possible throughout this section.

## Purpose of Scoping Review

The purpose of the review is to scope the nature and extent of the evidence base regarding the tools or procedures

used to capture treatment effects in conversation. The review will address the following research questions.

1. How are conversation outcome measures in aphasia studies designed, described, and reported?
  - a. What are the characteristics of studies (e.g., time of publication, study design, intervention focus) that use conversation as an outcome measure?
  - b. What are the range of practices employed for collecting and analyzing conversation data as a means of capturing treatment effects?
  - c. How are conversation outcome measures conceptualized and what specific features of conversation have been measured?
  - d. What characteristics of conversation demonstrate change as a result of intervention?
2. How do researchers address the quality of tools or procedures to measure conversation as an outcome of aphasia treatment?
  - a. What factors related to reliability and validity are considered in the selection of a specific measure?
  - b. How are conversation data collected and analyzed with regards to producing credible tools or procedures?

## Identifying Relevant Studies

A systematic literature search was conducted for studies published between January 1995 and September 2019 in the following electronic databases: CINAHL, PsychINFO, MEDLINE, PubMed, and ComDisDome. Databases were searched from 1995 because of the movement in the literature in the mid to late 1990s toward a focus on conversation as an important aspect of aphasia assessment and treatment (e.g., Kagan, 1998) and as a medium for the analysis of aphasic talk (e.g., Damico et al., 1999). The search terms *aphasia* and *conversation* were purposefully selected to capture the wide range of tools and procedures used to measure change in conversation. Hand searches were performed in addition to the systematic literature search. This included reviewing the references of (a) included studies during the data extraction stage and (b) published reviews covering similar topics (e.g., Simmons-Mackie et al., 2010).

## Study Selection

Studies were included in this scoping review if they reported original published research, were written in English, peer-reviewed, and used a tool or procedure to

elicit and analyze aspects of conversation as a means of measuring treatment effects. Eligible studies included treatment studies that employed at least one conversation-focused measure and nontreatment studies (e.g., development or refinement of conversation-based measures for use in future treatment studies). For treatment studies, no restrictions were imposed regarding the treatment approach used (e.g., impairment-focused, functional, social, or LPAA). While assessing the quality of the literature is not typically a goal of scoping reviews (Grant & Booth, 2009), we considered it important to chart key features related to reliability and validity of outcomes (as well as other practices related to quality) to document the range of factors that may be considered when determining the suitability of a conversational outcome measure. Study participants presented with aphasia of any etiology. Reviews and conference papers were excluded.

For the purposes of this review, conversation was operationally defined as either a structured or semistructured dialogue between two or more participants. Semistructured dialogue included conversations where topics were predetermined or where a priori procedures were followed in relation to speaker roles. For example, some studies asked participants to speak about a given topic (e.g., divorce; Croteau et al., 2007) or gave partners some general guidelines prior to the conversation (e.g., avoid consistent asking of questions; Leaman & Edmonds, 2019b). We considered conversation as any interaction involving at least two people in which participants alternate in the roles of speakers and listeners (Hutchby & Wooffitt, 2008). Crucially, in contrast to other forms of interaction (rituals, meetings with predefined agendas), conversations are generally unplanned in that many of the joint actions undertaken are locally negotiated (Schegloff, 2007). Studies were excluded if the tool or procedures measured aspects unrelated to conversation or if conversation was only a small part of the tool or procedure (e.g., Communication Confidence Rating Scale for Aphasia; Babbitt & Cherney, 2010). Studies were excluded if only the conversation partner's behaviors were measured.

Studies that described conversation (dis)abilities without quantification of behaviors or assessment of treatment efficacy or effectiveness were also excluded; all the studies excluded for this reason employed CA as a means of describing various aspects of aphasic conversation (e.g., repair, turn-taking). While these studies collected and analyzed naturally occurring conversations between a speaker with aphasia and a primary partner (usually a spouse), the goal of these CA studies was not to investigate the effects of treatment. Rather, these investigations sought to describe how naturally occurring interactions unfold. Therefore, unless the aim of the study included describing response to an intervention, these descriptive CA studies were excluded.

We used Covidence software (Version 1.0; Covidence Systematic Review Software, 2020) to manage search results, study selection and data charting processes. Covidence is cloud-based software that allows users to import all citations and work within a shared system to screen references and complete data extraction. Covidence keeps a full record of all decisions made during the review process and allows users to export data to common statistics packages. Covidence automatically identified and excluded all duplicate references. As suggested by JBI (Peters et al., 2020), abstract and full-text screening were performed by two or more independent reviewers. Initially, articles were screened by applying the inclusion and exclusion criteria to the title and abstract. The first and sixth authors reviewed all abstracts at this stage and studies moved to the full-text screening stage if both authors agreed on inclusion. At the full-text screening stage, the first and sixth authors validated the inclusion criteria by reviewing the full article. At both stages, disagreements were resolved by a third reviewer (second author) making the final decision.

## Data Charting

Data were extracted from each study according to a data charting template, which was discussed, created, and revised by the research team. All members of the research team reviewed the template by charting data from two separate studies and revisions were made based upon suggestions. The template and accompanying definitions of data items extracted were modified iteratively as the research team continued the data charting process. This process helped to ensure that the final data extraction items were consistent with the goals of the review. The data charting template was uploaded to Covidence. For each included study, two reviewers from the research team extracted data independently of each other. No reviewers were assigned studies in which they were an author. After both reviews on a study were complete, the study moved into the "consensus" stage where the first reviewer assigned to the study compared judgments for each item charted by both reviewers. When discrepancies occurred on a particular item, the first reviewer examined the study again, reviewed the judgments made by herself and the second reviewer, and selected the judgment that was most accurate and descriptive.

## Collating, Summarizing, and Reporting Results

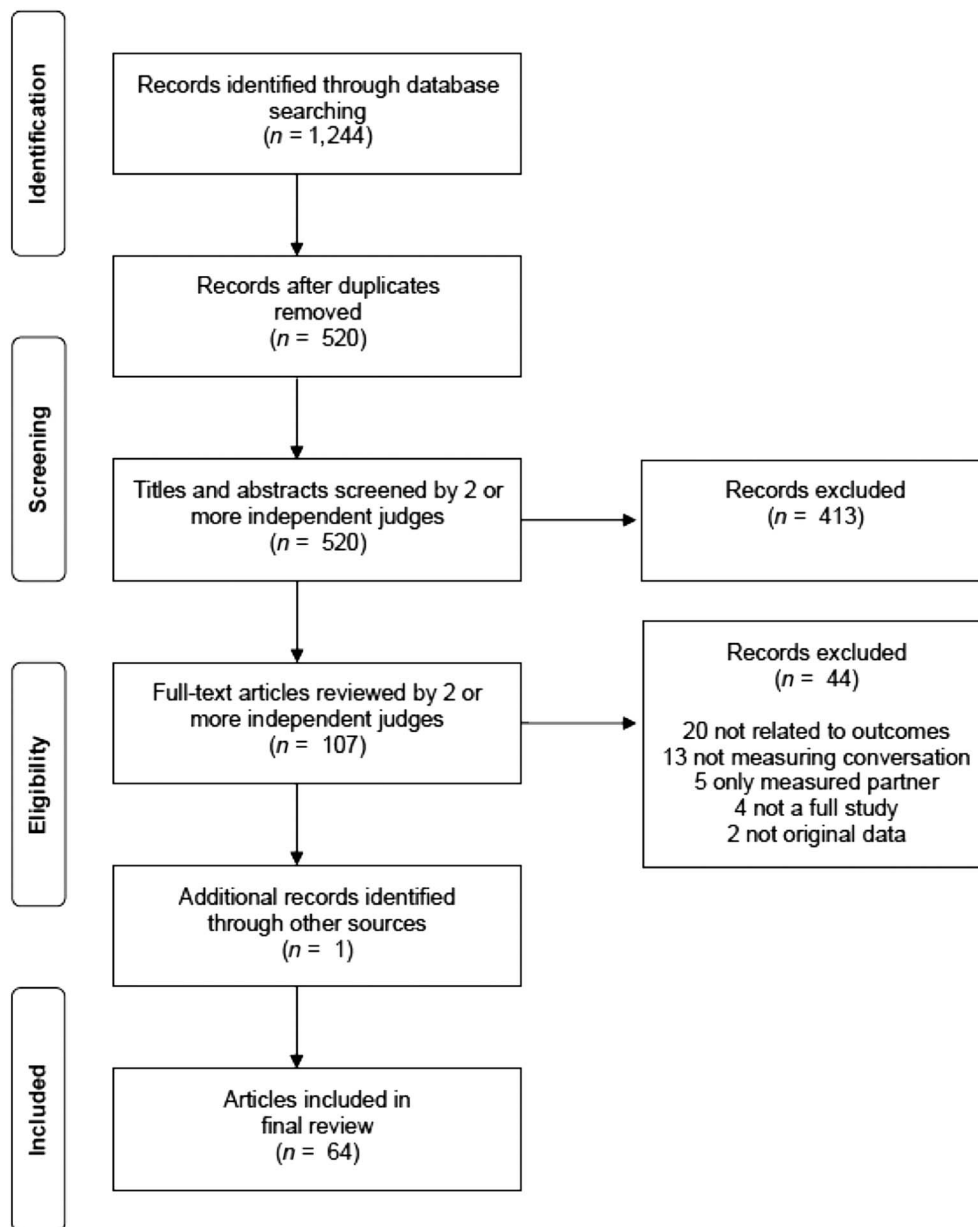
The first step in collating and summarizing the results included gathering data on items represented by a basic numerical analysis such as the total number and design of studies (designs categorized according to Simmons-Mackie et al., 2010), year of publication, intervention type, and total number of measures. This part of the analysis helped

to elucidate the breadth of intervention approaches and tools or procedures oriented to measuring conversation as an outcome and to identify gaps in the existing literature. The second step was to thematically organize the extracted data so that key elements contributing to the design and implementation of a tool or procedure measuring conversation could be aptly described. Conceptual categories were derived from the research aims and included information on processes for collecting and analyzing samples and information related to the quality of the tool or procedure (concepts including, but not limited to, reliability and validity).

## Results

Search strategy results are presented in Figure 1. The systematic search of the literature resulted in 1,244 studies. After removing duplicates and conducting the abstract screening and full-text review, 63 total studies met the inclusion criteria. One additional study was identified through searching references of included studies. A total of 64 studies are included in this review (see Supplemental Material S1 for a list of all included studies).

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart.



## Publication Year

Of relevant studies published between 1995 and 2019, interest in using conversation as an outcome of aphasia intervention has peaked in the last decade, with a majority of studies ( $n = 38$ ) published from 2010–2019 (see Figure 2).

## Participant Characteristics

Data from 611 participants were reported across the 64 studies: 372 males and 239 females, ranging in age from 23 to 89. The number of participants in each study ranged from 1 to 240. Most studies ( $n = 56$ ) described participants as having chronic aphasia (i.e., greater than 6 months post onset). Participants' etiologies of aphasia included stroke ( $n = 58$ ), a dementia-producing condition resulting in primary progressive aphasia ( $n = 2$ ), and tumor ( $n = 1$ ). Three studies did not include information on etiology. Type of aphasia was reported in 55 studies. Aphasia severity was reported in 50 studies. Level of education was reported in 27 studies.

## Study Characteristics

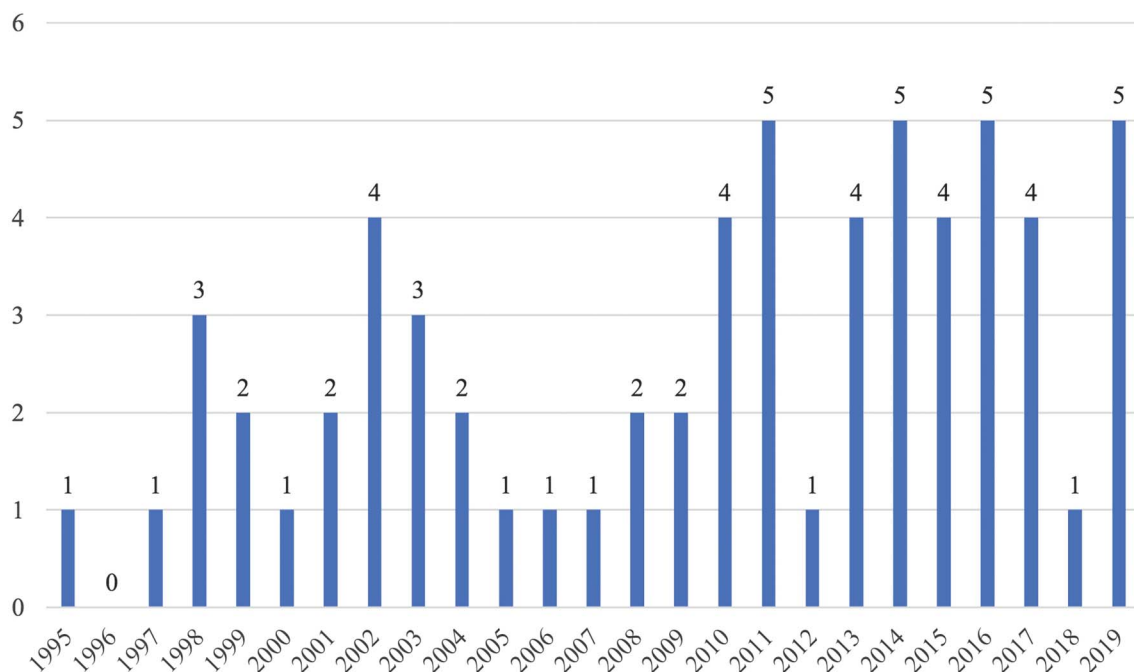
Out of the 64 studies included, 53 were characterized as intervention studies and 11 were studies that examined the psychometric properties of a tool or process intended to capture conversation as an outcome measure. Study

designs were classified using procedures from Simmons-Mackie et al. (2010): the yield included case studies ( $n = 26$ ), single-subject design studies ( $n = 18$ ), group studies ( $n = 8$ ), and qualitative studies ( $n = 1$ ). In keeping with Simmons-Mackie et al. (2010), case studies included those without experimental control and were descriptive in nature. Therefore, case studies including a pre–post comparison using qualitative methods such as CA are included in the case study count. Qualitative studies included those that conformed to interpretive traditions such as ethnography or grounded theory (Simmons-Mackie et al., 2010). Of the group studies, four did not utilize a control group, three were considered randomized controlled trials, and one was considered a quasi-randomized trial. Of the 11 studies examining psychometric properties, four studies examined an aspect of reliability, one examined an aspect of validity, five examined aspects of both reliability and validity, and one study investigated the potential impact of conversation duration on perceptual ratings of conversation.

## Focus of Intervention

A variety of therapy approaches was investigated across the 53 intervention studies. These intervention approaches were classified using the framework proposed by Worrall et al. (2017). As a result, intervention approaches included social approaches ( $n = 28$ ), cognitive neuropsychological (including psycholinguistic) approaches ( $n = 11$ ), pragmatic approaches ( $n = 10$ ), stimulation-based

**Figure 2.** Number of published studies by year.





approaches ( $n = 3$ ), and neurolinguistics ( $n = 1$ ). Social approaches included interventions that defined aphasia as a communication disability that was exacerbated by external environmental factors, whereas pragmatic approaches defined aphasia as difficulty engaging in everyday communication activities (Worrall et al., 2017). Therefore, an intervention such as communication partner training was classified as a social approach while a treatment like Promoting Aphasics Communicative Effectiveness was categorized as a pragmatic approach. Mode of treatment delivered differed across studies, including group studies ( $n = 5$ ), individual studies ( $n = 45$ ), and mixed group and individual studies ( $n = 3$ ).

Goals of the intervention were classified according to the Framework for Living with Aphasia (A-FROM; Kagan et al., 2008) and included reducing aphasia severity ( $n = 16$ ), improving participation in life situations ( $n = 11$ ), and improving the communication environment ( $n = 7$ ). The remainder of studies ( $n = 19$ ) included a combination of goals with all but one of these studies classified as conversation partner training studies; that is, where the communication and language environment and participation in life situations were both targeted by working on outcomes related to the person with aphasia and a communication partner. None of the studies solely focused intervention at the A-FROM dimension of personal attitudes, emotions, and identity; however, the early family-oriented intervention described by Blom Johansson et al. (2013) did have sessions which focused on both the emotional consequences of aphasia and the communication environment of the person with aphasia.

## Data Collection

To better understand procedures related to the collection of conversation samples, data were charted across all 64 studies on the sample environment, conversation partner, elicitation task, and length of sample. Data samples were collected in the participant's home ( $n = 19$ ), a clinical setting ( $n = 17$ ), or a combination of home and clinic ( $n = 4$ ). Surprisingly, 38% ( $n = 24$ ) of studies did not report the location where the conversation sample was collected.

The conversation partner was consistent across a majority of the 64 studies ( $n = 49$ ). That is, the same conversation partner was used across all samples in the study. The consistent partner included a spouse or significant other ( $n = 14$ ), the researcher or a clinician ( $n = 13$ ), a familiar partner ( $n = 11$ ), or an unfamiliar partner ( $n = 11$ ). The remainder of studies ( $n = 15$ ) either employed a mix of familiar and unfamiliar conversation partners ( $n = 11$ ) or did not report the relationships between the participant with aphasia and the conversation partner ( $n = 4$ ).

Elicitation tasks included both semistructured ( $n = 24$ ) and unstructured ( $n = 36$ ) samples; elicitation procedure

was unreported by a few studies ( $n = 4$ ). Semistructured conversation samples included procedures such as providing a set of predetermined topics, discussing a previously watched television show or news segment, providing artifacts such as a newspaper clipping, and giving specific instructions to the partner regarding conversation practices (e.g., asking questions). For example, Croteau et al. (2007) elected to use a semistructured interview situation in which participants were instructed to talk about their opinions on specific topics (e.g., divorce, technology). In contrast, unstructured samples did not impose predetermined topics and participant dyads were typically given instructions to talk as they normally would. Often, unstructured samples were collected in the absence of the researcher; participants were instructed to set up a camera in a place and during a time of day that they usually conversed.

Conversation varied both in terms of length of the sample collected as well as how much of the sample was used for analysis. Most studies did not analyze the full data sample collected, with the length of the sample varying widely including  $> 20$  min of the sample used for analysis ( $n = 10$  studies), 15–20 min ( $n = 8$ ), 10–15 min ( $n = 16$ ), 5–10 min ( $n = 15$ ), and  $< 5$  min ( $n = 1$ ). One study used the number of conversation turns rather than minutes to measure sample length. Twenty percent ( $n = 13$ ) did not report the length of the sample used for analysis. Most studies did not provide a rationale for the length of sample selected for analysis.

## Data Analysis

Data analysis procedures included preparing the sample for analysis (including transcription) and methods for operationalizing and coding behaviors of interest in conversation. While some data analysis processes described here inherently impact the quality of the tool or procedure (e.g., transcription examples, transcriptionist training), we have elected to report those processes here for ease in interpretation.

### Preparing Sample for Analysis

Many studies ( $n = 46$ ) used transcription as a means of preparing the sample and assessing the component of conversation being measured. For those studies that did not employ transcription, conversation data were analyzed via observation of video samples ( $n = 11$ ) or audio samples ( $n = 1$ ), or a combination of video and real-time ratings ( $n = 1$ ), while a few studies ( $n = 4$ ) did not report procedures for assessing conversation. One study did not use researcher/examiner-led assessment, instead employing a self-report measure as the only conversation-based outcome.

*Transcription.* There were several aspects of the transcription process that varied across the 46 studies that employed transcription. First, the type of transcription

differed, with three main types identified as verbatim transcription plus nonverbal communication using CA conventions ( $n = 12$ ), verbatim plus nonverbal communication without CA conventions ( $n = 8$ ), and verbatim without nonverbal communication ( $n = 9$ ). Eleven studies did not report the type of transcription that was done.

A few studies ( $n = 6$ ) included a second stage of the analysis process which involved segmenting or processing transcriptions prior to coding and analyzing behaviors of interest. For example, Leaman and Edmonds (2019a, 2019b) detailed a process by which transcripts were segmented into utterances using the Analysis of Speech Unit (Foster et al., 2000), a modification of the T-unit. This procedure formed the basis for the analysis so that clauses could be separated and scored independently, regardless of whether they were produced within the same turn.

Of the 46 studies using transcription, half ( $n = 23$ ) published examples of the transcripts in the body of the article or in the appendix. Transcribers included research assistants (including students;  $n = 14$ ) and the authors of the study ( $n = 6$ ). However, the majority of studies ( $n = 26$ ) did not report who completed the transcription. Only 13 studies reported that the transcriber underwent training prior to or during the course of the study. Similarly, only 18 studies reported procedures for monitoring and ensuring the reliability of transcription.

### Operationalizing and Coding Behaviors

Almost half of the studies ( $n = 28$ ) used a theoretical or pre-existing framework to code and analyze behaviors of interest and to measure the effects of therapy. Examples included previously developed and published measures such as T-units, conversational repair, turns, communication success, quantitative production analysis, correct information units (CIUs), mean length of utterance (MLU), or scales such as the Measure of Participation in Conversation (MPC) and POWERS that were based upon well-established theories of communication. In contrast, some studies ( $n = 11$ ) developed bespoke measures, created through a data-driven approach in which the researchers used participant data to define conversation outcomes. Data-driven techniques were most often used in partner training studies that aimed to reduce barriers and/or increase use of facilitative strategies by persons with aphasia and their partners. One example of a data-driven approach to coding and analysis is provided by Barnes and Nickels (2018): Baseline conversation samples were analyzed to highlight potential therapy targets. These therapy target behaviors were then described, operationalized, and served as outcome measures for the intervention. A few studies ( $n = 5$ ) used a combination of the two approaches to define and measure behaviors. In almost one third of the included studies ( $n = 20$ ), procedures for operationalizing conversation-based indices were not

reported or were not clear enough to determine whether a data-driven or pre-existing framework was employed.

### Quality

Items were charted to report on the quality of the tools or procedures for measuring conversation: person assessing outcomes including training and blinding, formal measures of reliability and validity, and informal measures of validity.

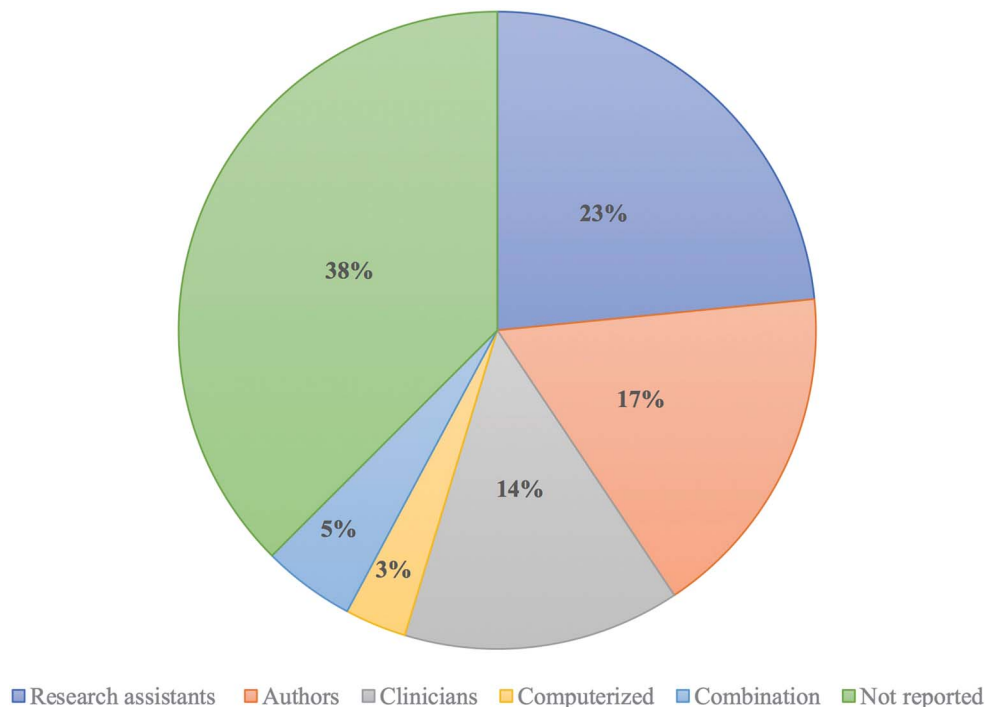
### Assessor

For the purposes of categorizing assessors, research assistants included students if studies designated them as research assistants. Similarly, student clinicians were designated as “clinicians” if they were the treating therapist in the study. Assessors included a range of individuals (see Figure 3). Thirty eight percent of studies ( $n = 24$ ) did not report the identity of the assessor. In terms of assessor training, less than half of the studies ( $n = 31$ ) reported details of how assessors were trained. Blinding assessors to some aspect of the study (e.g., sample date, type of treatment) was not commonly reported: some studies ( $n = 23$ ) reported using blinded raters, whereas the remainder ( $n = 41$ ) either did not use or did not report these details.

### Formal Measures of Reliability and Validity

Sixty-four percent of all studies ( $n = 42$ ) employed at least one measure of reliability such as consensus reliability ( $n = 6$ ), interrater reliability ( $n = 24$ ), or both intra- and interrater reliability ( $n = 12$ ). Methods for obtaining reliability included the use of multiple raters where consensus on each item must be agreed upon by a set portion of the raters, point-to-point reliability, and statistical analysis (e.g., *Cohen's Kappa*; Kazdin, 1982). Some studies bundled all measures and calculated one reliability score for several conversation-based measures. Studies that did not bundle measures and quantified intra- and/or interrater reliability for a specific measure ( $n = 26$ ) can be found in Table 1. The table is stratified by categories. Traditional linguistics refers to measures based upon structural aspects of language and formal units of speech. Conversation analysis refers to measures based upon sociolinguistics and ethnomethodology. Rating scales include measures that used ordinal or interval scales with predetermined categories to assess a given construct. Other includes a range of quantifiable measures from various theoretical orientations besides traditional linguistics or CA. The table also includes methods for evaluating reliability and whether the measure achieved acceptable reliability. While there is no formally recognized minimal level of agreement for conversation-based measures, a benchmark of  $> .80$  has been used (Nunnally & Bernstein,

Figure 3. Person assessing outcome.



1994; Pritchard et al., 2017; Streiner & Norman, 2000). Therefore, this benchmark was applied to studies assessing aspects of reliability. However, it should be noted that Best et al. (2016) question the appropriateness of applying a threshold of 80% agreement, particularly for interactional features of conversation.

For many studies that assessed intra- or interrater reliability, scores were reported as moderate to high. For example, moderate reliability scores were reported by Hickey et al. (2004) when raters coded for the comprehensibility of utterances (.63 to .65). Best et al. (2016) reported reliability at 69% for number of conversation facilitators and 64% for conversation barriers. In their study of the effect of pictographic books on communicative interaction, Ho et al. (2005) found that frequency of variables impacted interrater reliability agreement. That is, all frequently occurring dependent variables achieved high interrater reliability but infrequent variables (i.e., negative affect and communication breakdowns) achieved lower levels of agreement. The only other study that reported low reliability scores was Oelschlaeger and Thorne (1999). In this study, interrater reliability for %CIU was never greater than 55% and intrarater reliability for %CIU was 73%.

A few studies ( $n = 5$ ) formally evaluated test-retest reliability as a way of capturing the stability of a given measure. For example, Leaman and Edmonds (2019a) investigated the stability of CIUs in conversation and found significant correlations between conversations with

a familiar home partner and an unfamiliar speech-language pathologist partner. Ramsberger and Rende (2002) explored the stability of their procedure for measuring transactional success in conversation. While the measure did not achieve the authors' criteria for statistical significance, there was a moderately high correlation ( $r_s = .80$ ;  $p_s > .01$ ) between conversations after watching the same four *I Love Lucy* episodes 6 months apart. Based upon the measure developed by Ramsberger and Rende (2002), Leaman and Edmonds (2019b) created a measure of "communication success" (p. 365) and tested the stability of the measure in unstructured conversations. In this study, the authors found strong, significant correlations for communicative success between two, separate unstructured conversation conditions. They also tested six linguistic variables for stability and found significant correlations for three of the six variables: relevance, grammaticality, and other lexical errors (e.g., paraphasias, neologisms, and vague language). Other measures that were found to be stable across time included speech rate (Oelschlaeger & Thorne, 1999) and several measures reported in Herbert et al. (2008) that included nouns, content words, and substantive turns as a proportion of speech units, turns, and substantive turns.

Most studies did not formally evaluate validity. For studies that did, evaluation included construct validity ( $n = 2$ , focused on convergent and discriminant validity), concurrent validity ( $n = 5$ ), and social validity ( $n = 3$ ).

**Table 1.** Inter- and intrarater reliability scores.

| Category                              | Measure                                      | Frequency            | Interrater reliability |   | Intrarater reliability   |   |
|---------------------------------------|--|----------------------|------------------------|---|--------------------------|---|
|                                       |  |                      | Evaluation             | > .80   | Evaluation               | • .80                                   |
| Traditional Linguistics               | CIUs   | 5 studies            | • ICC                  | • Yes <sup>1,2,3,4</sup><br>• No <sup>5</sup> | • % agreement            | • Yes <sup>4</sup><br>• No <sup>5</sup> |
|                                       | Utterances                                   | 2 studies            | • % agreement          | • Yes <sup>4,6</sup>                          | • % agreement            | • Yes <sup>4</sup>                      |
|                                       | Incoherent utterances                        | 1 study              | • % agreement          | • Yes <sup>6</sup>                            | —                        | —                                       |
|                                       | Questions                                    | 1 study              | • % agreement          | • Yes <sup>6</sup>                            | —                        | —                                       |
|                                       | Trained words                                | 1 study              | • Pearson correlation  | • Yes <sup>7</sup>                            | —                        | —                                       |
|                                       | Grammatical morphemes                        | 1 study              | • Pearson correlation  | • Yes <sup>7</sup>                            | —                        | —                                       |
|                                       | Disfluencies                                 | 1 study              | • Pearson correlation  | • Yes <sup>7</sup>                            | —                        | —                                       |
|                                       | Content words                                | 2 studies            | • % agreement          | • Yes <sup>8</sup><br>• No <sup>9</sup>       | • % agreement            | • No <sup>9</sup>                       |
|                                       | Nouns  | 2 studies            | • % agreement          | • Yes <sup>9,10</sup>                         | • % agreement            | • Yes <sup>9</sup>                      |
|                                       | Verbs  | 1 study              | • % agreement          | • Yes <sup>10</sup>                           | —                        | —                                       |
|                                       | Lexical retrieval                            | 1 study              | • % agreement          | • Yes <sup>9</sup>                            | • % agreement            | • Yes <sup>9</sup>                      |
|                                       | # of words                                   | 1 study              | • % agreement          | • Yes <sup>5</sup>                            | • % agreement            | • Yes <sup>5</sup>                      |
|                                       | MLU  | 1 study              | • % agreement          | • Yes <sup>4</sup>                            | • % agreement            | • Yes <sup>4</sup>                      |
|                                       | Grammatically complete utterances            | 1 study              | • % agreement          | • No <sup>4</sup>                             | • % agreement            | • Yes <sup>4</sup>                      |
|                                       | Conversation Analysis                        | Conversation repairs | 3 studies              | • % agreement                                 | • Yes <sup>6,12,13</sup> | • % agreement                           |
| Prop                                  |  | 1 study              | • ICC                  | • Yes <sup>14</sup>                           | —                        | —                                       |
| Gesture                               |  | 1 study              | • ICC                  | • Yes <sup>14</sup>                           | —                        | —                                       |
| Write                                 |  | 1 study              | • ICC                  | • Yes <sup>14</sup>                           | —                        | —                                       |
| Touch                                 |  | 1 study              | • ICC                  | • Yes <sup>14</sup>                           | —                        | —                                       |
| Idiosyncratic behavior                |  | 1 study              | • ICC                  | • Yes <sup>14</sup>                           | —                        | —                                       |
| Occurrence of word finding difficulty |  | 1 study              | • Cohen's kappa        | • Yes <sup>15</sup>                           | —                        | —                                       |
| Conversation facilitators             |  | 2 studies            | • % agreement          | • Yes <sup>16</sup><br>• No <sup>17</sup>     | • % agreement            | • Yes <sup>16</sup>                     |
| Conversation barriers                 |  | 1 study              | • % agreement          | • No <sup>17</sup>                            | —                        | —                                       |
| Turn types                            |  | 1 study              | • % agreement          | • Yes <sup>9</sup>                            | • % agreement            | • Yes <sup>9</sup>                      |
| Substantive turns                     |  | 1 study              | • % agreement          | • Yes <sup>9</sup>                            | • % agreement            | • Yes <sup>9</sup>                      |
| # of turns                            |  | 1 study              | • % agreement          | • Yes <sup>18</sup>                           | • % agreement            | • Yes <sup>18</sup>                     |
| Topic initiation                      |  | 1 study              | • % agreement          | • No <sup>18</sup>                            | • % agreement            | • Yes <sup>18</sup>                     |
| Breakdowns                            |  | 1 study              | • % agreement          | • No <sup>18</sup>                            | • % agreement            | • No <sup>18</sup>                      |
| Pointing                              |  | 1 study              | • % agreement          | • Yes <sup>18</sup>                           | • % agreement            | • Yes <sup>18</sup>                     |
| Opportunity strategy use              | 1 study                                      | • % agreement        | • Yes <sup>16</sup>    | • % agreement                                 | • Yes <sup>16</sup>      |   |
| Rating Scales                         | MPC Interaction                              | 3 studies            | • ICC                  | • Yes <sup>21,22</sup><br>• No <sup>20</sup>  | • ICC                    | • Yes <sup>22</sup>                     |
|                                       | MPC Transaction                              | 3 studies            | • ICC                  | • Yes <sup>20,22</sup><br>• No <sup>21</sup>  | • ICC                    | • Yes <sup>22</sup>                     |
|                                       | Comprehensiveness rating                     | 1 study              | • % agreement          | • No <sup>23</sup>                            | —                        | —                                       |
| Other                                 | MIC  | 1 study              | • ICC                  | • No <sup>24</sup>                            | • ICC                    | • No <sup>24</sup>                      |
|                                       | Propositional idea density                   | 1 study              | • % agreement          | • Yes <sup>25</sup>                           | —                        | —                                       |
| Other                                 | Speaking rate                                | 1 study              | • Pearson correlation  | • Yes <sup>7</sup>                            | —                        | —                                       |
|                                       | Comprehensibility/utterance                  | 1 study              | • Cohen's kappa        | • No <sup>26</sup>                            | —                        | —                                       |
|                                       | # of main concepts                           | 1 study              | • % agreement          | • Yes <sup>19</sup>                           | —                        | —                                       |
|                                       | Conversation initiations (speech act theory) | 3 studies            | • % agreement          | • Yes <sup>8,11,18</sup>                      | • % agreement            | • Yes <sup>11,18</sup>                  |
|                                       | Conceptual complexity                        | 1 study              | • % agreement          | • No <sup>8</sup>                             | —                        | —                                       |
|                                       | Main idea analysis                           | 1 study              | • % agreement          | • Yes <sup>23</sup>                           | —                        | —                                       |
|                                       | % successful communication exchanges         | 1 study              | • % agreement          | • Yes <sup>11</sup>                           | • % agreement            | • Yes <sup>11</sup>                     |
|                                       | Speech units                                 | 1 study              | • % agreement          | • Yes <sup>9</sup>                            | • % agreement            | • Yes <sup>9</sup>                      |
|                                       | No response                                  | 1 study              | • % agreement          | • Yes <sup>18</sup>                           | • % agreement            | • Yes <sup>18</sup>                     |
|                                       | Negative affect                              | 1 study              | • % agreement          | • Yes <sup>18</sup>                           | • % agreement            | • No <sup>18</sup>                      |

Note. Em dashes indicate no studies evaluated. CIU = correct information units; MLU = mean length of utterance; MPC = Measure of Participation in Conversation; MIC = Measure of Interaction in Conversation; <sup>1</sup>Doyle et al. (1995); <sup>2</sup>Savage & Donovan (2017); <sup>3</sup>Leaman & Edmonds (2019a); <sup>4</sup>Murray & Ray (2001); <sup>5</sup>Oeschlaeger & Thome (1999); <sup>6</sup>Boles (2000); <sup>7</sup>Goldberg et al. (2012); <sup>8</sup>Hux et al. (2010); <sup>9</sup>Herbert et al. (2008); <sup>10</sup>Rose et al. (2016); <sup>11</sup>Garrett & Huth (2002); <sup>12</sup>Boles (1997); <sup>13</sup>Boles (1998); <sup>14</sup>Cunningham & Ward (2003); <sup>15</sup>Lavoie et al. (2019); <sup>16</sup>Lustig & Tompkins (2002); <sup>17</sup>Best et al. (2016); <sup>18</sup>Ho et al. (2005); <sup>19</sup>Hopper et al. (2002); <sup>20</sup>Kagan et al. (2001); <sup>21</sup>Kagan et al. (2004); <sup>22</sup>Muò et al. (2019); <sup>23</sup>Ramsberger & Rende (2002); <sup>24</sup>Eriksson et al. (2016); <sup>25</sup>Bryant et al. (2013); <sup>26</sup>Hickey et al. (2004).

Both Kagan et al. (2004) and Ramsberger and Rende (2002) investigated construct validity (i.e., convergent and discriminant validity) of conversation-based measures. Social validity measures were used as a measure of clinical significance that enabled researchers to formally evaluate outsiders' perspective on the effectiveness of a treatment. As pointed out by Lustig and Tompkins (2002), social validity judgments "can provide a valuable perspective on treatment effectiveness and relevance across a variety of communicative contexts" (p. 512). In all three studies where social validity was measured, undergraduate and/or graduate students were asked to rate items such as comfort level, amount of accurate information, effectiveness of communication strategies, equity of turn-taking, topic maintenance, difficulty in communicating information, communication satisfaction, and comprehensibility.

When formal validity was reported, it was usually adequate. Exceptions include one measure reported by Bryant et al. (2013) and three measures reported by Herbert et al. (2008). In Bryant et al. (2013), type-token ratio (TTR) was found to be significantly higher in participants with aphasia when compared with control participants, which was an unexpected finding. In Herbert et al. (2008), relationships between picture naming and word-retrieval in conversation were examined to better understand the validity of the picture-naming task. While several measures were correlated across tasks (e.g., nouns as a proportion of turns), no significant correlation was found for nouns as a proportion of speech units, content words as a proportion of speech units, and substantive turns as a proportion of total turns.

Of the 64 studies, some studies ( $n = 17$ ) used a scale that was validated in a previous study (e.g., MPC, CAPP, POWERS). Other studies ( $n = 15$ ) discussed procedures related to ecological validity including purposefully collecting an unstructured conversation that was recorded in the home environment, ensuring the researcher was not present, discarding the first 5 min of the recording, and using conversation-based indices that were participant-driven and based upon initial assessment of the dyad. Several studies ( $n = 28$ ) did not explicitly report validity measures; however, it should be noted that the purpose of five of these studies was to capture reliability on a conversation-based tool or procedure rather than formal validity.

### Informal Measures of Validity

There were various ways that researchers attempted to exercise a certain amount of experimental control so that outcomes related to conversation could be interpreted as valid. We considered these procedures as a type of validity, even though psychometric definitions of validity were not used. Of the 64 studies, over half were concerned with controlling for the variability in conversation data by

keeping time consistent across samples ( $n = 36$ ) or keeping number of turns consistent across samples ( $n = 4$ ). Four of these studies also incorporated procedures to account for the types of turns that may contribute to differences across samples and used proportional (rather than raw) data as the final measure. For example, Herbert et al. (2013) discussed the importance of measuring the number of nouns per substantive turn (i.e., turn that contains at least one content word; Herbert et al., 2008) instead of total number of nouns since the latter measure may be more vulnerable to changes across conversation samples. Additionally, 11 studies used a consistent part of the sample (e.g., middle 5 min) to ensure similarity across samples.

Another common method included exercising control over the content of the conversation ( $n = 26$  studies). In these studies, researchers kept topics consistent across samples by offering structured topics or by using prompts. One method developed initially by Ramsberger and Rende (2002) included having participants view *I Love Lucy* episodes and conversing about the main points in the episode in a conversation after viewing. Similar procedures were also developed in which participants watched news segments or other video clips and then had a conversation after the viewing (e.g., Hopper et al., 2002).

A less common procedure ( $n = 7$ ) incorporated explicit instructions delivered to partners prior to collecting conversation samples. These studies either gave specific instructions on ways to converse with the person with aphasia (e.g., avoid asking too many questions) or exercised consistency in the type of materials or supports used in conversations across the data set.

A minority of studies ( $n = 2$ ) examined properties of baseline data to ensure stability of a conversation measure or procedure prior to conducting analysis of treatment effects. Carragher et al. (2013) investigated the stability of verb retrieval for each participant to ensure a lack of variance in the baseline conversation data before making statistical comparisons to posttherapy conversations. Similarly, Rayner and Marshall (2003) investigated the mean ratings of the MPC from baseline conversations before making comparisons to posttreatment ratings.

Only seven of the 64 studies did not describe any type of control across the conversation samples. Importantly, of the 53 intervention studies, less than half engaged in repeated sampling across all phases of the study. While there was a great range in the number of samples collected across baseline and posttreatment phases (from one to eight), only 22 studies collected two or more conversation samples across both baseline and posttreatment phases. That is, the vast majority of the intervention studies using a conversation measure as an outcome only collected one sample at baseline and/or at posttreatment to determine response to intervention.

## Procedure and Tools Identified as Outcomes

### Type and Nature of Measure

Most of the 64 studies ( $n = 52$ ) used or evaluated objective measures in which the phenomenon of interest was collected and scored by an assessor. A minority of studies ( $n = 11$ ) used a combination of objective and self- or caregiver-report measures. As described above, one study employed a self-report measure as the only conversation-based outcome. The CAPPa was the most commonly used tool for self- or caregiver-report.

In terms of level of analysis, several studies ( $n = 41$ ) detailed measures intended to capture change at the micro or local level of analysis (e.g., word errors, use of effective strategy). A few studies ( $n = 9$ ) described macro or global aspects of conversation (e.g., general participation, engagement). Some studies ( $n = 14$ ) used a combination of micro and macro measures.

Of the 53 intervention studies, further information was charted to describe the range and nature of outcomes including the name or description of every conversation measure, whether the measure demonstrated positive change, and if the measure was a primary or secondary outcome of intervention. Criteria for positive change included that the study described a statistically significant change in the measure for at least one participant. When significance testing was not appropriate (e.g., qualitative studies) or not reported, determination of positive change was made based upon explanations provided by the authors of each study and their interpretations of data.

There was a total of 211 measures used across the 64 studies. Almost half ( $n = 100$ ) of the measures were unique (89 unique measures in intervention studies, 11 unique measures in psychometric studies). See Table 2 for a list of all unique measures in the 53 intervention studies and whether the study reported a positive change in the measure as a result of the intervention.

Outcome measures that had slightly different names but followed the same procedures were included under a single label. Several studies also had slight variations in the way outcomes were ultimately calculated even if they measured the same construct. For example, some studies calculated the total number of words while others calculated words per utterance, words per minutes, or percentage of words in the sample. There were not consistent definitions or procedures reported in each study when these differences occurred. Therefore, outcomes measuring the same construct (e.g., “words”) regardless of denominator (e.g., words/utterance) were also included under a single label. As shown in Table 2, most measures demonstrated change in at least one participant across one or more studies.

Figure 4 below represents a hierarchy of the most widely used outcome measures identified in the scoping

review. The innermost ring of the diagram lists the domain into which measures were classified, the intermediate ring lists the specific measures identified during the review, and the outermost ring provides a representation of the proportion of studies in which the given measure indicated significant change. For all three rings, segment size is proportional to the number of measures associated with a category and its subcategories. In the outermost ring, studies in which a given measure indicated a significant change in the index of interest are represented by green segments while studies in which a given measure did not indicate significant change are represented by red segments. It is important to note that it was not a goal of this scoping review to compare levels of evidence for interventions. Therefore, the figure demonstrates change in a given measure rather than change due to a specific treatment.

As shown in Figure 4, several measures were used across multiple studies. In the traditional linguistics domain, the most common measures included words, trained or targeted words, verbs, utterances, CIUs, MLUs, and TTR. None of these measures showed consistent change as a result of the interventions. For example, trained words were examined in six separate studies. In half of these studies, participants’ number of trained words or percentage of trained words in conversation did not significantly change. Likewise, MLU and TTR were each measured in four separate intervention studies and more often than not, these measures did not change in post-treatment conversation samples. CIUs were calculated in five studies to measure posttreatment gains, with only three of these studies demonstrating significant change. There was no procedure for counting and analyzing a traditional linguistics measure that was clearly superior; that is, whether studies calculated measures per minute, as a total number of a sample, or as percentage of a sample, measures did not consistently show change in participants.

Regarding measures within the CA domain, traditional CA (i.e., qualitative ethnomethodology CA) and gestures were by far the most frequent measures calculated. While significance testing is not appropriate for traditional CA, descriptions and explanations provided by the authors of each study were used to determine whether positive change occurred. These studies all presented detailed, positive changes to turn construction, strategy use, and repair, as a result of the treatment under investigation. Gestures, however, only demonstrated positive change in one study. Four studies measuring facilitative, nonfacilitative, or neutral gestures as an outcome of treatment did not show significant change. Likewise, total number of turns appears less likely to exhibit positive change. Repair appears to be a mechanism that is particularly sensitive to posttreatment change as self-repair and

**Table 2.** Conversation outcomes and reported change, stratified by framework.

| Measure name                      | Studies with significant change  | Studies with no change   |
|-----------------------------------|--|--|
| <b>Traditional linguistics</b>    |  |  |
| Words                             | words/minute, Boles (1997)<br>words/utterance, Boles (1997)<br>frequency of words, Boles (1997)<br>contribution of words, Boles (1998)<br>words/utterance, Boles (1998)<br>total words, Croot et al. (2015)<br>different words, Savage & Donovan (2017)    | % words, Fox et al. (2009)<br>total words, Rose et al. (2016)<br>words/minute, Rose et al. (2016)  |
| Correct information units         | total CIUs, Avent et al. (2009)<br>total CIUs, Boo & Rose (2011)<br>CIUs/min, Savage & Donovan (2017)  | total CIUs, Savage et al. (2014)*<br>% CIUs, Murray & Ray (2001)   |
| Trained or targeted words         | total target words, Carragher et al. (2013)<br>total target words, Mason et al. (2011)<br>% target words, Ulmer et al. (2017)  | total target words, Croot et al. (2015)<br>% target words, Goldberg et al. (2012)<br>total target words, Palmer et al. (2019)  |
| Verbs                             | mean verb retrieval, Carragher et al. (2013)<br>verbs/total utterances, Carragher et al. (2015)<br>verbs/content words, Carragher et al. (2015)<br>total verbs, del Toro et al. (2008)<br>substantive verbs, Rose et al. (2016)<br>Savage & Donovan (2017) | verbs/total words, Carragher et al. (2013)   |
| Mean length of utterance          |  | del Toro et al. (2008)<br>Murray & Ray (2001)<br>Ulmer et al. (2017)   |
| Utterances                        | utterances/minute, Boles (1997)<br>frequency of utterances, Boles (1997)<br>total utterances, Boles (2000)<br>utterances new information, del Toro et al. (2008)   | % utterances, Murray & Ray (2001)<br>% utterances, Garrett & Huth (2002)   |
| Type-token ratio                  | Avent et al. (2009)<br>Savage & Donovan (2017)   | Ulmer et al. (2017)<br>del Toro et al. (2008)  |
| Questions                         | Boles (2000)   | Fox et al. (2009)  |
| Nouns                             | Rose et al. (2016)   | del Toro et al. (2008)   |
| Content words                     | Croot et al. (2015)<br>Hux et al. (2010)   | —  |
| Closed-class words                | Croot et al. (2015)  | —  |
| Grammatical morphemes             | Goldberg et al. (2012)   | —  |
| Disfluencies                      | —  | Goldberg et al. (2012)   |
| Grammatical complete utterances   | Murray & Ray (2001)  | —  |
| % non-sentential utterance        | Savage & Donovan (2017)  | —  |
| % simple utterance                | Savage & Donovan (2017)  | —  |
| % complex utterances              | Savage & Donovan (2017)  | —  |
| elliptical utterances             | —  | del Toro et al. (2008)   |
| incoherent utterances             | Boles (2000)   | —  |
| # targets abandoned               | Lustig & Tompkins (2002)   | —  |
| Turn-taking – facilitative        | Savage et al. (2014)   | —  |
| Turn-taking – inhibitory          | Savage et al. (2014)   | —  |
| Fillers                           | —  | Croot et al. (2015)  |
| “yes” and “no”                    | —  | Croot et al. (2015)  |
| grammatical sentences             | del Toro et al. (2008)   | —  |
| minimal sentences                 | del Toro et al. (2008)   | —  |
| one-word responses                |  | del Toro et al. (2008)   |
| <b>Conversation analysis</b>      |  |  |
| Traditional Conversation Analysis | Beckley et al. (2013)<br>Beeke et al. (2015)<br>Beeke et al. (2014)<br>Beeke et al. (2011)<br>Cunningham & Ward (2003)<br>Wilkinson et al. (1998)<br>Wilkinson et al. (2010)<br>Wilkinson et al. (2011)<br>Tuomenoksa et al. (2016)                        | —  |
| Gesture                           | + gesture, Damico et al. (2015)  | + gesture, Beeke et al. (2015)<br>+ gesture, Beeke et al. (2014)<br>+ gesture, Boles & Lewis (2003)<br>neutral gesture, Boles & Lewis (2003)<br>– gesture, Boles & Lewis (2003)<br>+ gesture, Cunningham & Ward (2003) |

(table continues)

**Table 2.** (Continued).

| Measure name  | Studies with significant change  | Studies with no change                    |
|---|--|---|
| Profile of Word Errors and Retrieval in Speech        | Greenwood et al. (2010)<br>Mason et al. (2011)<br>Best et al. (2011)<br>Herbert et al. (2014)  | Woolf et al. (2016)                       |
| Total no. of turns                                    | Brock et al. (2017)<br>Hux et al. (2010)   | Ho et al. (2005)<br>Booth & Swabey (1999) |
| Self-repair   | Boles (1997)<br>Boles (1998)<br>Boles (2000)<br>Damico et al. (2015)   | —   |
| Topic initiations                                     | Barnes & Nickels (2018)<br>Fox et al. (2009)<br>Ho et al. (2005)<br>Waller et al. (1998)   | —   |
| Conversation Analysis Profile for People with Aphasia | Beckley et al. (2013)<br>Ross et al. (2006)  | Booth & Swabey (1999)                     |
| Repair length   | No. of turns in repair, Booth & Swabey (1999)<br>average length, Booth & Swabey (1999)   | total length, Fox et al. (2009)           |
| Writing   | Beeke et al. (2015)<br>Beeke et al. (2014)   | Cunningham & Ward (2003)                  |
| Occurrence of word finding difficulty                 | Damico et al. (2015)<br>Lavoie et al. (2019)   | —   |
| Conversation facilitators                             | Best et al. (2016)<br>Lustig & Tompkins (2002)   | —   |
| Keyword   | Beeke et al. (2015)  | Beeke et al. (2014)                       |
| Other-repair  | —  | Boles (1997)                              |
| Repair sequences                                      | Fox et al. (2009)  | —   |
| Trouble sources                                       | Booth & Swabey (1999)  | —   |
| Breakdowns  | —  | Ho et al. (2005)                          |
| Simple turns  | —  | Barnes & Nickels (2018)                   |
| Multimodal turns                                      | —  | Barnes & Nickels (2018)                   |
| Topic expansion                                       | Waller et al. (1998)   | —   |
| Request for assistance                                | —  | Barnes & Nickels (2018)                   |
| Change in modalities                                  | —  | Barnes & Nickels (2018)                   |
| Mime  | Beeke et al. (2015)  | —   |
| Drawing   | —  | Cunningham & Ward (2003)                  |
| Touch   | —  | Cunningham & Ward (2003)                  |
| Prop  | —  | Cunningham & Ward (2003)                  |
| Pointing  | Ho et al. (2005)   | —   |
| Idiosyncratic non-verbals                             | —  | Cunningham & Ward (2003)                  |
| Opportunity strategy use                              | —  | Lustig & Tompkins (2002)*                 |
| Conversation barriers                                 | Best et al. (2016)   | —   |
| Ineffective recall strategy                           | Damico et al. (2015)   | —   |
| Sequential association                                | Hickey et al. (2004)   | —   |
| <b>Rating scales</b>                                  |  |   |
| Measure of Participation in Conversation              | Blom Johansson et al. (2013)<br>Finch et al. (2017)<br>Fox et al. (2009)<br>Sorin-Peters & Patterson (2014)<br>Kagan et al. (2001)<br>Rayner & Marshall (2003) | —   |
| Likert-scale perception of conversation               | Boles & Lewis (2003)<br>Hux et al. (2010)  | —   |
| Self-rating of conversation                           | Fox et al. (2009)  | —   |
| Estimation of conversation skills                     | Blom Johansson et al. (2013)   | —   |
| Therapy Outcome Measure                               | —  | Palmer et al. (2019)                      |
| Measure of Interaction in Conversation                | —  | Eriksson et al. (2016)                    |
| La Trobe Communication Questionnaire                  | Boo & Rose (2011)  | —   |

(table continues)



Table 2. (Continued).

| Measure name  | Studies with significant change  | Studies with no change |
|---|--|------------------------|
| <b>Other qualitative analyses</b>                             |  |                        |
| Content analysis  | Blom Johansson et al. (2013)   | —                      |
| Structural context analysis                                   | Carragher et al. (2013)  | —                      |
| Analysis of interaction                                       | Sorin-Peters (2004)  | —                      |
| Analysis of transaction                                       | Sorin-Peters (2004)  | —                      |
| Discourse analysis  | Booth & Swabey (1999)  | —                      |
| <b>Other quantitative analyses</b>                            |  |                        |
| Conversation initiations and/or responses (speech act theory) | Garrett & Huth (2002)<br>Ulmer et al. (2017)<br>Ulmer et al. (2017)<br>Hux et al. (2010) | Ho et al. (2005)       |
| Conversation time   | Brock et al. (2017)<br>Savage & Donovan (2017)   | Garrett & Huth (2002)  |
| Speaking rate   | Boles (1998)<br>Goldberg et al. (2012)   | Garrett & Huth (2002)  |
| Conceptual complexity   | Brock et al. (2017)<br>Hux et al. (2010)   | —                      |
| Instances of frustration                                      | Brock et al. (2017)  | —                      |
| Units   | Brock et al. (2017)  | —                      |
| # of main concepts  | Hopper et al. (2002)   | —                      |
| Propositional idea density                                    | Savage & Donovan (2017)  | —                      |
| Metalinguistic comments to indicate trouble                   | Fox et al. (2009)  | —                      |
| % successful exchanges  | Garrett & Huth (2002)  | —                      |
| Comprehensibility/utterance                                   | Hickey et al. (2004)   | —                      |
| Negative affect   | Ho et al. (2005)   | —                      |
| No response   | Ho et al. (2005)   | —                      |
| Phonologically related response                               | Mason et al. (2011)  | —                      |
| Navigational errors   | Brock et al. (2017)  | —                      |
| % revised words   | Ulmer et al. (2017)  | —                      |
| Responses   | Waller et al. (1998)   | —                      |
| Confirmations   | Waller et al. (1998)   | —                      |
| Qualitative production analysis                               | Boo & Rose (2011)  | —                      |

Note. Em dashes indicate no studies reported. \* = not reported; + = facilitative; - = nonfacilitative.

repair length were found to significantly change on six out of seven occurrences.

The MPC was the most commonly used rating scale in the 53 intervention studies ( $n = 6$ ) and demonstrated posttreatment change in every study. There were only six other rating scales used to measure response to intervention; three of these were formalized measures and three appeared to be bespoke scales developed by the authors for individual studies. Two of these bespoke measures were self-reported perceptions of conversation that described the ease or success of information transfer in conversation. While significance testing was not possible due to small sample studies, both studies using these measures discussed self-reported changes as a result of the intervention.

Beyond the domains of traditional linguistics and CA, there were several other measures described as “other quantitative measures.” The most common of these was described according to speech act theory and was operationalized as conversation initiations and/or responses.

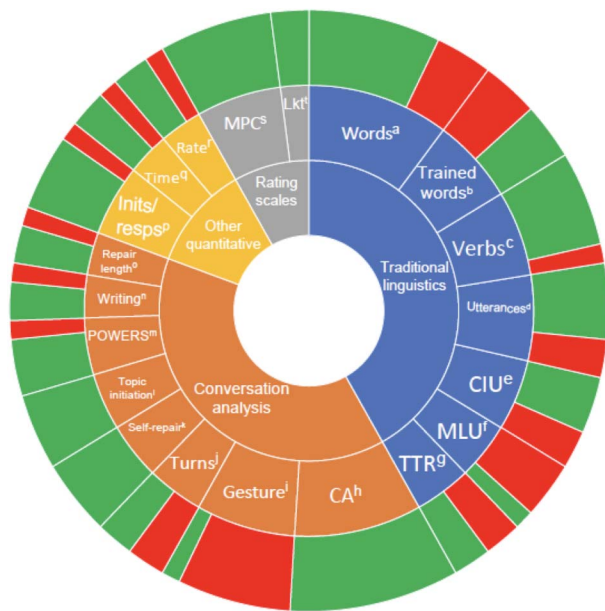
Conversation initiations and/or responses showed significant change on four out of five occurrences. Conversation time and speaking rate were both measured 3 times and posttreatment change was reported in two of the three studies for both measures.

In terms of whether the measure was a primary or secondary outcome measure, 90% of intervention studies described some measure of conversation as the primary measure of change. Only two studies did not use some aspect of conversation as a primary outcome measure and three studies did not report whether the conversation measure was a primary or secondary measure. A majority of studies ( $n = 38$ ) collected an additional measure separate from conversation as a means of quantifying response to intervention. Only 13 intervention studies used conversation as the sole outcome measure.

### Philosophical Orientation

Across the studies, measurement of aspects of conversation was motivated by different philosophical or

**Figure 4.** Most widely used outcome measures and reported change. The innermost ring includes domain for each measure, the intermediate ring includes specific measures identified in review, and the outermost ring denotes change in green or lack of change in red; a: Number of words; b: Number of trained (target) words; c: Number of verbs; d: Number of utterances; e: Number of CIUs (correct information units); f: MLU (mean length of utterance); g: TTR (type-token ratio); h: Qualitative CA (conversation analysis) measures; i: Number of gestures; j: Number of turns; k: Number of self-repairs; l: Number of topics initiated; m: POWERS (Profile of Word Retrieval and Errors); n: Number of times participants used writing; o: Length of repair; p: Number of initiations and responses (based on speech act theory analytical framework); q: Length of conversation; r: Speaking rate; s: MPC (Measure of Participation in Conversation); t: Likert-scale of global perception of conversational ability.



theoretical orientations. These included ethnomethodology (i.e., CA;  $n = 26$ ), structural linguistics ( $n = 17$ ), pragmatic or social theories ( $n = 12$ ), and functional or applied linguistics ( $n = 6$ ). Two studies used a combination of theoretical approaches. One study did not include enough information about the measure to determine the philosophical orientation.

The decision on what specific feature of conversation to measure was often based upon the philosophical underpinning of the treatment under study or the theoretical background that gave rise to tool development. For example, some intervention studies ( $n = 11$ ) included investigations of partner training programs with theoretical roots in CA. In all of these studies except for one, phenomena of interest were defined, measured, and explained according to CA frameworks. Some of these studies used traditional CA, but others used quantitative procedures where definitions were based upon CA philosophies. Along similar lines, conversation partner therapies that focused on improving transaction and interaction (e.g., Kagan et al.,

2001) were often evaluated with a scale intended to measure these constructs (i.e., MPC; Kagan et al., 2004).

## Discussion

There has been increasing interest in measuring conversation as an outcome of aphasia intervention over the last several years (Stark et al., 2021). This growth reflects the trends in discourse-related research in general, with recent increases in the frequency of discourse analysis (Bryant et al., 2016), number of discourse-related outcome measures (Bryant et al., 2013; Pritchard et al., 2017; Pritchard et al., 2018), qualitative methods in aphasia research (Simmons-Mackie & Lynch, 2013), and conversation-based treatments (Simmons-Mackie et al., 2014). Research expansion may reflect a growing awareness of the desires and needs of people with aphasia and their families. Like previous reviews in outcomes related to discourse (e.g., Bryant et al., 2016), there was great heterogeneity in types of measures and procedures for collecting and analyzing samples. Most studies elected to use multiple measures so that a variety of conversation variables could be assessed across samples. Despite the volume of measures (211 total across 64 studies) several measures were not operationalized in a way that could be replicated across studies and/or were not psychometrically profiled, a finding consistent with discourse measures reviewed by Pritchard et al. (2017).

## Selecting a Measure

While there is currently no “gold standard” and further investigation is needed, three measures are promising. Firstly, the MPC appears to be sensitive to postintervention change, but its interrater reliability and stability have not been unequivocally demonstrated. Another promising measure is POWERS, which has been investigated across several studies and undergone considerable psychometric development. It is theoretically grounded in CA and authors have completed several studies to better understand which type of proportional data is most stable and valid across a range of variables. With strong validity data, more data is needed to assess rater reliability and the stability of POWERS. Conversational repair, whether measured as self-repair or as repair length, and conversation initiations have also demonstrated adequate reliability (albeit using percentage agreement which may be a less robust measure) and appear to be sensitive to postintervention changes; however, neither have been assessed for stability. Rater reliability and stability of measures must be established to determine what treatments are most effective, replicate findings, and interpret treatment effects across participants.

Other measures may be reliable but are less likely to change as a result of intervention. Several conversation strategies used by people with aphasia (often referred to as conversation facilitators) have been well defined in the literature and achieve acceptable levels of reliability (e.g., Cunningham & Ward, 2003), but do not consistently evince improvement after treatment. Many of these conversation facilitators are multimodal strategies such as writing, drawing, gesture, and so forth. It could simply be that multimodal strategies are less likely to improve. More likely, the lack of change is more nuanced and could be due to the context-dependent nature of these strategies and the variability of their use when compared with verbal strategies (e.g., lexical self-repair). Research investigating mechanisms of change in conversation therapy have also revealed that behavior change is a complex process that involves not only an increased awareness of the strategy, but also strengthening a speaker's reasons to do something different in conversation and structuring efforts to make changes in context (Johnson et al., 2017). Alternatively, lack of change could be a product of how researchers are defining outcomes across studies and decisions on whether to analyze multimodal strategies as discrete or combined.

An ideal measure would be reliable within and across raters, achieve stability across time, be relevant and meaningful to people with aphasia, and feasible to administer in clinical settings. Perhaps equally important is a measure's suitability for assessing elements of conversation likely to change as a result of intervention. While measures such as the MPC, POWERS, and conversational repair are most promising, they are not likely to demonstrate change across a range of interventions. For example, if a conversation partner intervention is focused on changing discrete behaviors (e.g., reducing "test questions" asked by a spouse), elements of conversation described in all three of these measures may not change. The underlying philosophical orientation and goals of the treatment will necessarily influence decisions about what elements of conversation should be assessed (Simmons-Mackie et al., 2018).

Finally, there are a number of measures that warrant further assessment before researchers consider selecting them as an outcome of aphasia intervention. Several measures did not consistently show positive change after treatment in participants with aphasia: trained words, CIUs, MLU, TTR, gesture, and number of turns. Of these measures, CIUs in conversation did not achieve adequate intra- or interrater reliability in one study (Oelschlaeger & Thorne, 1999), but did achieve high test-retest stability in one study (Leaman & Edmonds, 2019a). Interrater reliability on trained words, MLU, gesture, and number of turns was only captured once despite being used as an outcome measure in 20 different studies. Reliability of

TTR was not reported but the limitations of this measure are well known (Hess et al., 1986).

## Moving Toward Consensus

A COS has been established in aphasia but there is no consensus on outcomes in discourse (Wallace et al., 2019). If as a field we intend to work toward a COS in conversation, it requires an international collaborative effort. While there is no single outcome measure in conversation that will likely be appropriate for every treatment study or clinical setting, the field would benefit from clarifying and establishing consensus on a number of foundational methodological issues. International discussions through a network such as the Collaboration of Aphasia Trialists could begin immediately with the following priorities.

First, international consensus is needed to define the most common and most promising measures. Several outcomes and procedures were vague or contradictory across studies in the current review (e.g., definitions for gesture, conversation initiations, topic). These outcomes and procedures should be discussed, defined, and made available for future studies.

Second, international consensus should focus on how best to count and analyze conversation data. Herbert et al. (2008) proposed that proportional, rather than raw, data is a more accurate way to measure performance. Furthermore, findings from the same study suggest that denominators in conversation measures play a key role in determining the validity of a given measure. For example, when Herbert et al. (2008) used a speech unit as a denominator for measuring linguistic variables, there was no relationship between conversation and picture naming. However, when the denominator was turns, there was a significant relationship between word retrieval in conversation and picture naming. This suggests that these two methods of quantifying the "same" variable (e.g., nouns, content words) produce contrasting results when the denominator is different. Very few studies ( $n = 4$ ) in the current review opted to use turns as a denominator for calculating proportional data, although this trend does appear to be changing in recent studies (e.g., see Tetnowski et al., 2021). Most studies in this review kept time consistent across samples or calculated proportional data by segmenting transcripts into utterances or clauses (e.g., Leaman & Edmonds, 2019a). If proportional data matters, these findings raise questions about whether researchers are measuring the same constructs if procedures for segmenting transcripts, coding, and analyzing are not uniform across studies.

A third point for decision-making pertains to benchmarks for the number of samples collected across

intervention studies. It is well established that there is inherent variability in conversation data (Perkins et al., 1999), although some studies in the current review (e.g., Ramsberger & Rende, 2002) and also more recent studies (e.g., Leaman & Edmonds, 2021) have investigated stability of conversation measures. To account for variability, Best et al. (2016) suggest that multiple samples be collected in intervention studies to ensure that conversation patterns are stable and unrelated to variables such as topic or fatigue. Currently, there are no guidelines that indicate a specific number of conversation samples to be collected across phases of a treatment, but it is generally agreed that study designs include multiple baseline sampling to account for variability. The exact number or range of samples required is unknown. However, 58% of studies in this review collected one sample at each time point, which may be problematic given the inherent variability of conversation data. An international effort toward establishing guidelines for an optimal number of samples required for treatment studies is of vital importance.

It is also important to note that all of the measures identified are based on the assumption that increased frequency is a marker of success or effectiveness, leading to improved communication. However, a more nuanced, functional approach may be needed to assess how the behavior is used and if it is used successfully in conversation. This may be particularly true for “facilitative” strategies that have been described in the literature such as gesture, writing, circumlocution, and so on, which could be inherent, pre-existing strategies for a person with aphasia that could be exploited in therapy (Johnson et al., 2017; Simmons-Mackie & Damico, 1997). More recent studies refining conversation outcomes have accounted for both frequency and success of a given behavior by examining the partner’s turn after the strategy use to evaluate success (Azios et al., 2021). As suggested by Beeke et al. (2011), it is difficult to separate target outcomes (i.e., behaviors of a person with aphasia) from a partner’s turn due to the sequential nature of turn taking in conversation. Measures that account for success, in addition to frequency, are of utmost importance and using “next-turn proof” (Hutchby & Wooffitt, 2008) or similar concepts could be a type of informal validity that would improve stability and soundness of measures.

While a conversation COS is a long-term goal, a more reachable, near-term objective may be to gain consensus on a set of measures based upon the current review and others (e.g., Bryant et al., 2016) and prioritize these for psychometric evaluation and use in clinical settings. Measures that best demonstrate aspects of behavior that conversation is built upon and are reliable, valid, stable, and clinically applicable could then be discussed for inclusion into the current aphasia COS (Wallace et al., 2019).

## Limitations and Future Directions

This review did not capture any information relevant to stakeholders’ opinions on outcomes related to conversation. Our field still needs to know what measures matter most to people with aphasia, their significant others, and to clinicians. Importantly, many of the measures described in this review are labor- and time-intensive and require extensive training, which calls into question the clinical utility of many of the outcome measures reviewed. Stark et al. (2021) have highlighted important differences in the perceptions of researchers versus clinicians using discourse-based outcomes, including the number of samples collected to measure treatment effects. Surveying all stakeholder groups and incorporating these findings into recommendations for conversation outcome measures should be an immediate priority.

A second limitation is that our research team was limited to native English speakers. As a result, we could not include studies in languages other than English. The cultural and linguistic differences in the mechanics of conversation across languages limit the extrapolation of findings to other languages or cultures. A recent review of communication partner training of Chinese-speaking persons with aphasia revealed key differences in the training of health care professionals delivering the partner training and a range of outcome measures that were not appropriately adapted or translated to a Chinese version (Kong et al., 2021). However, some aspects of conversation may be similar across cultures and languages. CA-based studies have demonstrated that certain elements of conversation, such as overlapping talk and silence between turns, are strikingly similar across a variety of languages (Stivers et al., 2009). Regardless, descriptions and reports of outcomes in conversation across languages other than English are an important area for future investigations.

This review considered whether change in conversation occurred as a result of intervention, but did not provide data on how researchers approach the issue of effect sizes. Several experimental studies did not calculate effect sizes, making it impossible to determine how much change occurred due to the intervention, whereas others reported “significant” changes when small effect sizes were calculated. Future research should examine a subset of studies in the current review (e.g., single-subject experimental designs) to provide an overview of effect size practices. The most widely used methods for calculating treatment effects (e.g., *Tau U*) have been shown to be susceptible to Type I errors due to autocorrelation (Archer et al., 2019). Additionally, little is known about the minimal significant difference that should be captured for specific outcomes in conversation. What is statistically significant may not be clinically significant, and vice versa. A systematic review

focused on effect sizes in conversation-based studies could provide initial insights about more defensible, autocorrelation immune statistics and about the best, most clinically meaningful way to interpret these statistics.

## Conclusions

The 64 studies in the current review demonstrate a high degree of variability. The number of participants who provided data, the goals of intervention, the procedures used to elicit samples, the length of conversation that investigators analyzed, transcription practices and measures used and other characteristics varied considerably across the publications we analyzed. Clearly, the body of literature that focuses on conversation-based assessments and treatments for people with aphasia features a very wide range of approaches to gathering, processing, and interpreting recordings of conversations.

As regards the quality of tools used in conversation-focused studies, our findings indicate that most researchers include reliability procedures within study designs. Unfortunately, validity is rarely addressed in a rigorous, replicable way, though many studies did include more informal elements designed to increase this psychometric property.

This review provides an overview of the measures used by aphasiologists who focus on conversation as an outcome, on the procedures researchers use when implementing these measures, and the quality of these measurement procedures. By cataloguing the ways in which researchers deal with conversation, we have provided information that can inform the development of a conversation-specific COS. Some of the orientations, conceptualizations, and procedures we have presented can be seen as options that might be included in a future COS. In other instances, the approaches detailed (e.g., informal validity procedures identified, procedures for ensuring stability) could be refined and expanded upon before being disseminated to other researchers and clinicians. We hope the comprehensive analysis we undertook will function as the first step in a journey that leads to a uniform, robust and clinically relevant conversation-specific, aphasia COS.

## Acknowledgments

The authors would like to acknowledge Jack S. Damico for his helpful suggestions, comments, and feedback during the conceptualization stage of this project.

## References

\*Asterisk indicates reference was included in the scoping review. For the full list of studies included, see Supplemental Material S1.

- Archer, B., Azios, J. H., Müller, N., & Macatangay, L. (2019). Effect sizes in single-case aphasia studies: A comparative, autocorrelation-oriented analysis. *Journal of Speech, Language, and Hearing Research, 62*(7), 2473–2482. [https://doi.org/10.1044/2019\\_JSLHR-L-18-0186](https://doi.org/10.1044/2019_JSLHR-L-18-0186)
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- \*Avent, J., Patterson, J., Lu, A., & Small, K. (2009). Reciprocal scaffolding treatment: A person with aphasia as clinical teacher. *Aphasiology, 23*(1), 110–119. <https://doi.org/10.1080/02687030802240211>
- Azios, J. H., Archer, B., & Lee, J. B. (2021). Detecting behavioural change in conversation: Procedures and preliminary data. *Aphasiology, 35*(7), 961–983. <https://doi.org/10.1080/02687038.2020.1812031>
- Babbitt, E. M., & Cherney, L. R. (2010). Communication confidence in persons with aphasia. *Topics in Stroke Rehabilitation, 17*(3), 214–223. <https://doi.org/10.1310/tsr1703-214>
- \*Barnes, S., & Nickels, L. (2018). Interaction-focussed therapy for aphasia: Effects on communication and quality of life. *International Journal of Speech-Language Pathology, 20*(5), 528–540. <https://doi.org/10.1080/17549507.2017.1329851>
- \*Beckley, F., Best, W., Johnson, F., Edwards, S., Maxim, J., & Beeke, S. (2013). Conversation therapy for agrammatism: Exploring the therapeutic process of engagement and learning by a person with aphasia. *International Journal of Language & Communication Disorders, 48*(2), 220–239. <https://doi.org/10.1111/j.1460-6984.2012.00204.x>
- \*Beeke, S., Maxim, J., Best, W., & Cooper, F. (2011). Redesigning therapy for agrammatism: Initial findings from the ongoing evaluation of a conversation-based intervention study. *Journal of Neurolinguistics, 24*(2), 222–236. <https://doi.org/10.1016/j.jneuroling.2010.03.002>
- Beeke, S., Maxim, J., & Wilkinson, R. (2007). Using conversation analysis to assess and treat people with aphasia. *Seminars in Speech and Language, 28*(2), 136–147. <https://doi.org/10.1055/s-2007-970571>
- \*Best, W., Grassly, J., Greenwood, A., Herbert, R., Hickin, J., & Howard, D. (2011). A controlled study of changes in conversation following aphasia therapy for anomia. *Disability and Rehabilitation, 33*(3), 229–242. <https://doi.org/10.3109/09638288.2010.534230>
- \*Beeke, S., Beckley, F., Johnson, F., Heilemann, C., Edwards, S., Maxim, J., & Best, W. (2015). Conversation focused aphasia therapy: Investigating the adoption of strategies by people with agrammatism. *Aphasiology, 29*(3), 355–377. <https://doi.org/10.1080/02687038.2014.881459>
- \*Beeke, S., Johnson, F., Beckley, F., Heilemann, C., Edwards, S., Maxim, J., & Best, W. (2014). Enabling better conversations between a man with aphasia and his conversation partner: Incorporating writing into turn taking. *Research on Language and Social Interaction, 47*(3), 292–305. <https://doi.org/10.1080/08351813.2014.925667>
- \*Best, W., Maxim, J., Heilemann, C., Beckley, F., Johnson, F., Edwards, S. I., Howard, D., & Beeke, S. (2016). Conversation therapy with people with aphasia and conversation partners using video feedback: A group and case series investigation of changes in interaction. *Frontiers in Human Neuroscience, 10*(562), 562. <https://doi.org/10.3389/fnhum.2016.00562>
- \*Blom Johansson, M., Carlsson, M., Östberg, P., & Sonnander, K. (2013). A multiple-case study of a family-oriented

- intervention practice in the early rehabilitation phase of persons with aphasia. *Aphasiology*, 27(2), 201–226. <https://doi.org/10.1080/02687038.2012.744808>
- \*Boles, L. (1997). Conversation analysis as a dependent measure in communication therapy with four individuals with aphasia. *Asia Pacific Journal of Speech, Language and Hearing*, 2(1), 43–61. <https://doi.org/10.1179/136132897805577468>
- \*Boles, L. (1998). Conversational discourse analysis as a method for evaluating progress in aphasia: A case report. *Journal of Communication Disorders*, 31(3), 261–274. [https://doi.org/10.1016/S0021-9924\(98\)00005-7](https://doi.org/10.1016/S0021-9924(98)00005-7)
- \*Boles, L. (2000). Aphasia therapy in a bilingual speaker: Treatment in language one, with spousal support in language two. *Asia Pacific Journal of Speech, Language and Hearing*, 5(2), 137–142. <https://doi.org/10.1179/136132800805577013>
- \*Boles, L., & Lewis, M. (2003). Working with couples: Solution focused aphasia therapy. *Asia Pacific Journal of Speech, Language and Hearing*, 8(3), 153–159. <https://doi.org/10.1179/136132803805576110>
- \*Boo, M., & Rose, M. L. (2011). The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology*, 25(2), 154–175. <https://doi.org/10.1080/02687031003743789>
- \*Booth, S., & Swabey, D. (1999). Group training in communication skills for carers of adults with aphasia. *International Journal of Language & Communication Disorders*, 34(3), 291–309. <https://doi.org/10.1080/136828299247423>
- \*Brock, K., Koul, R., Corwin, M., & Schlosser, R. (2017). A comparison of visual scene and grid displays for people with chronic aphasia: A pilot study to improve communication using AAC. *Aphasiology*, 31(11), 1282–1306. <https://doi.org/10.1080/02687038.2016.1274874>
- Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, 30(7), 489–518. <https://doi.org/10.3109/02699206.2016.1145740>
- Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, 31(10), 1105–1126. <https://doi.org/10.1080/02687038.2016.1239013>
- \*Bryant, L., Spencer, E., Ferguson, A., Craig, H., Colyvas, K., & Worrall, L. (2013). Propositional idea density in aphasic discourse. *Aphasiology*, 27(8), 992–1009. <https://doi.org/10.1080/02687038.2013.803514>
- \*Carragher, M., Sage, K., & Conroy, P. (2013). The effects 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). Preliminary analysis from a novel treatment targeting the exchange of new information within storytelling for people with nonfluent aphasia and their partners. *Aphasiology*, 29(11), 1384–1408. <https://doi.org/10.1080/02687038.2014.988110>
- Covidence Systematic Review Software. (2020). *Better systematic review management*. <http://www.covidence.org>
- \*Croot, K., Taylor, C., Abel, S., Jones, K., Krein, L., Hameister, I., Ruggero, L., & Nickels, L. (2015). Measuring gains in connected speech following treatment for word retrieval: A study with two participants with primary progressive aphasia. *Aphasiology*, 29(11), 1265–1288. <https://doi.org/10.1080/02687038.2014.975181>
- \*Croteau, C., Le Dorze, G., & Baril, G. (2007). Development of a procedure to evaluate the contributions of persons with aphasia and their spouses in an interview situation. *Aphasiology*, 21(6–8), 791–801. <https://doi.org/10.1080/02687030701192398>
- Cruice, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., & Dipper, L. (2020). UK speech and language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. *International Journal of Language & Communication Disorders*, 55(3), 417–442. <https://doi.org/10.1111/1460-6984.12528>
- Cruice, M., Worrall, L., Hickson, L., & Murison, R. (2003). Finding a focus for quality of life with aphasia: Social and emotional health, and psychological well-being. *Aphasiology*, 17(4), 333–353. <https://doi.org/10.1080/02687030244000707>
- \*Cunningham, R., & Ward, C. (2003). Evaluation of a training programme to facilitate conversation between people with aphasia and their partners. *Aphasiology*, 17(8), 687–707. <https://doi.org/10.1080/02687030344000184>
- Damico, J. S., Simmons-Mackie, N., Oelschlaeger, M., Elman, R., & Armstrong, E. (1999). Qualitative methods in aphasia research: Basic issues. *Aphasiology*, 13(9–11), 651–665. <https://doi.org/10.1080/026870399401768>
- \*Damico, J., Tetzowski, J., Lynch, K., Hartwell, J., Weill, C., Heels, J., & Simmons-Mackie, N. (2015). Facilitating authentic conversation: An intervention employing principles of constructivism and conversation analysis. *Aphasiology*, 29(3), 400–421. <https://doi.org/10.1080/02687038.2014.945388>
- \*del Toro, C. M., Altmann, L. J., Raymer, A. M., Leon, S., Blonder, L. X., & Gonzalez Rothi, L. J. (2008). Changes in aphasic discourse after contrasting treatments for anomia. *Aphasiology*, 22(7–8), 881–892. <https://doi.org/10.1080/02687030701844204>
- Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: Have we reached the tipping point? *Aphasiology*, 32(4), 459–464. <https://doi.org/10.1080/02687038.2017.1398803>
- Doedens, W. J., & Meteyard, L. (2020). Measures of functional, real-world communication for aphasia: A critical review. *Aphasiology*, 34(4), 492–514. <https://doi.org/10.1080/02687038.2019.1702848>
- \*Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology*, 4(4), 130–134. <https://doi.org/10.1044/1058-0360.0404.130>
- \*Eriksson, K., Hartelius, L., & Saldert, C. (2016). On the diverse outcome of communication partner training of significant others of people with aphasia: An experimental study of six cases. *International Journal of Language & Communication Disorders*, 51(4), 402–414. <https://doi.org/10.1111/1460-6984.12216>
- \*Finch, E., Cameron, A., Fleming, J., Lethlean, J., Hudson, K., & McPhail, S. (2017). Does communication partner training improve the conversation skills of speech-language pathology students when interacting with people with aphasia. *Journal of Communication Disorders*, 68, 1–9. <https://doi.org/10.1016/j.jcomdis.2017.05.004>
- 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>
- \*Fox, S., Armstrong, E., & Boles, L. (2009). Conversational treatment in mild aphasia: A case study. *Aphasiology*, 23(7–8), 951–964. <https://doi.org/10.1080/02687030802669526>
- Frattali, C., Thompson, C., Holland, A., Wohl, C., Wenck, C., Slater, S., & Paul, D. (2017). *American Speech-Language-Hearing Association Functional Assessment of Communication Skills for Adults (ASHA FACS)*. American Speech-Language-Hearing Association.

- \*Garrett, K. L., & Huth, C. (2002). The impact of graphic contextual information and instruction on the conversational behaviours of a person with severe aphasia. *Aphasiology*, 16(4–6), 523–536. <https://doi.org/10.1080/02687030244000149>
- Goldberg, S., Haley, K. L., & Jacks, A. (2012). Script training and generalization for people with aphasia. *American Journal of Speech-Language Pathology*, 21(3), 222–238. [https://doi.org/10.1044/1058-0360\(2012/11-0056\)](https://doi.org/10.1044/1058-0360(2012/11-0056))
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10), 1489–1522. [https://doi.org/10.1016/S0378-2166\(99\)00096-X](https://doi.org/10.1016/S0378-2166(99)00096-X)
- Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology*, 19(1), 283–307. <https://doi.org/10.1146/annurev.an.19.100190.001435>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health information & libraries journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- \*Greenwood, A., Grassly, J., Hickin, J., & Best, W. (2010). Phonological and orthographic cueing therapy: A case of generalised improvement. *Aphasiology*, 24(9), 991–1016. <https://doi.org/10.1080/02687030903168220>
- \*Herbert, R., Best, W., Hickin, J., Howard, D., & Osborne, F. (2013). *Powers: Profile of word errors and retrieval in speech: An assessment tool for use with people with communication impairment*. CQUniversity.
- \*Herbert, R., Gregory, E., & Best, W. (2014). Syntactic versus lexical therapy for anomia in acquired aphasia: Differential effects on narrative and conversation. *International Journal of Language & Communication Disorders*, 49(2), 162–173. <https://doi.org/10.1111/1460-6984.12054>
- \*Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, 22(2), 184–203. <https://doi.org/10.1080/02687030701262613>
- Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research*, 29(1), 129–134. <https://doi.org/10.1044/jshr.2901.129>
- \*Hickey, E., Bourgeois, M., & Olswang, L. (2004). Effects of training volunteers to converse with nursing home residents with aphasia. *Aphasiology*, 18(5–7), 625–637. <https://doi.org/10.1080/02687030444000093>
- \*Ho, K. M., Weiss, S. J., Garrett, K. L., & Lloyd, L. L. (2005). The effect of remnant and pictographic books on the communicative interaction of individuals with global aphasia. *Augmentative and Alternative Communication*, 21(3), 218–232. <https://doi.org/10.1080/07434610400016694>
- Holland, A., Frattali, C., & Fromm, D. (2018). *Communicative activities of daily living* (3rd ed.). Pro-Ed
- \*Hopper, T., Holland, A., & Rewega, M. (2002). Conversational coaching: Treatment outcomes and future directions. *Aphasiology*, 16(7), 745–761. <https://doi.org/10.1080/02687030244000059>
- Hutchby, I., & Wooffitt, R. (2008). Conversation analysis. Polity.
- \*Hux, K., Buechter, M., Wallace, S., & Weissling, K. (2010). Using visual scene displays to create a shared communication space for a person with aphasia. *Aphasiology*, 24(5), 643–660. <https://doi.org/10.1080/02687030902869299>
- Johnson, F. M., Best, W., Beckley, F. C., Maxim, J., & Beeke, S. (2017). Identifying mechanisms of change in a conversation therapy for aphasia using behaviour change theory and qualitative methods. *International Journal of Language & Communication Disorders*, 52(3), 374–387. <https://doi.org/10.1111/1460-6984.12279>
- Kagan, A. (1998). Supported conversation for adults with aphasia: Methods and resources for training conversation partners. *Aphasiology*, 12(9), 816–830. <https://doi.org/10.1080/02687039808249575>
- \*Kagan, A., Black, S. E., Duchan, J. F., Simmons-Mackie, N., & Square, P. (2001). Training volunteers as conversation partners using “Supported Conversation for Adults With Aphasia” (SCA): A controlled trial. *Journal of Speech, Language, and Hearing Research*, 44(3), 624–638. [https://doi.org/10.1044/1092-4388\(2001/051\)](https://doi.org/10.1044/1092-4388(2001/051))
- Kagan, A., Simmons-Mackie, N., Rowland, A., Huijbregts, M., Shumway, E., McEwen, S., & Sharp, S. (2008). Counting what counts: A framework for capturing real-life outcomes of aphasia intervention. *Aphasiology*, 22(3), 258–280. <https://doi.org/10.1080/02687030701282595>
- \*Kagan, A., Winckel, J., Black, S., Felson Duchan, J., Simmons-Mackie, N., & Square, P. (2004). A set of observational measures for rating support and participation in conversation between adults with aphasia and their conversation partners. *Topics in Stroke Rehabilitation*, 11(1), 67–83. <https://doi.org/10.1310/CL3V-A94A-DE5C-CVBE>
- Kazdin, A. E. (1982). The token economy: A decade later. *Journal of Applied Behavior Analysis*, 15(3), 431–445. <https://doi.org/10.1901/jaba.1982.15-431>
- Kong, A. P. H., Chan, K. P. Y., & Jagoe, C. (2021). Systematic review of training communication partners of Chinese-speaking persons With aphasia. *Archives of Rehabilitation Research and Clinical Translation*, 3(4), 100152. <https://doi.org/10.1016/j.arrct.2021.100152>
- 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>
- \*Lavoie, M., Bier, N., & Macoir, J. (2019). Efficacy of a self-administered treatment using a smart tablet to improve functional vocabulary in post-stroke aphasia: A case-series study. *International Journal of Language & Communication Disorders*, 54(2), 249–264. <https://doi.org/10.1111/1460-6984.12439>
- \*Leaman, M. C., & Edmonds, L. A. (2019a). 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](https://doi.org/10.1044/2019_AJSLP-18-0268)
- \*Leaman, M. C., & Edmonds, L. A. (2019b). 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](https://doi.org/10.1044/2018_AJSLP-17-0148)
- Leaman, M. C., & Edmonds, L. A. (2021). Measuring global coherence in people with aphasia during unstructured conversation. *American Journal of Speech-Language Pathology*, 30(1S), 359–375. [https://doi.org/10.1044/2020\\_AJSLP-19-00104](https://doi.org/10.1044/2020_AJSLP-19-00104)
- \*Lustig, A. P., & Tompkins, C. A. (2002). A written communication strategy for a speaker with aphasia and apraxia of speech: Treatment outcomes and social validity. *Aphasiology*, 16(4–6), 507–521. <https://doi.org/10.1080/02687030244000211>
- MacWhinney, B. (2014). Challenges facing COS development for aphasia. *Aphasiology*, 28(11), 1393–1395. <https://doi.org/10.1080/02687038.2014.930263>
- \*Mason, C., Nickels, L., McDonald, B., Moses, M., Makin, K., & Taylor, C. (2011). Treatment of word retrieval impairments in aphasia: Evaluation of a self-administered home programme using personally chosen words. *Aphasiology*, 25(2), 245–268. <https://doi.org/10.1080/02687038.2010.489258>

- Morris, J., Franklin, S., Menger, F., & GD (2011). Returning to work with aphasia: A case study. *Aphasiology*, 25(8), 890–907. <https://doi.org/10.1080/02687038.2010.549568>
- \*Muò, R., Rinaudo, M., Rabino, G., Massari, E., Schindler, A., Steni, P., & Iacomussi, T. (2019). Validation of the Italian version of the Kagan scales for people with aphasia and their conversation partners. *Aphasiology*, 33(3), 352–371. <https://doi.org/10.1080/02687038.2018.1482402>
- \*Murray, L. L., & Ray, A. H. (2001). A comparison of relaxation training and syntax stimulation for chronic nonfluent aphasia. *Journal of Communication Disorders*, 34(1–2), 87–113. [https://doi.org/10.1016/S0021-9924\(00\)00043-5](https://doi.org/10.1016/S0021-9924(00)00043-5)
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.
- \*Oelschlaeger, M. L., & Thorne, J. C. (1999). Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. *Journal of Speech, Language, and Hearing Research*, 42(3), 636–648. <https://doi.org/10.1044/jslhr.4203.636>
- \*Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., Latimer, N., Julious, S., Cross, E., Alshreef, A., Harrison, M., Bradley, E., Witts, H., & Chater, T. (2019). Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): A multicentre, single-blinded, randomised controlled trial. *The Lancet Neurology*, 18(9), 821–833. [https://doi.org/10.1016/S1474-4422\(19\)30192-9](https://doi.org/10.1016/S1474-4422(19)30192-9)
- Perkins, L., Crisp, J., & Walshaw, D. (1999). Exploring conversation analysis as an assessment tool for aphasia: The issue of reliability. *Aphasiology*, 13(4–5), 259–281. <https://doi.org/10.1080/026870399402091>
- Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11: Scoping Reviews (2020 version). In E. Aromataris & Z. Munn (Eds.), *JBI Manual for evidence synthesis*. <https://doi.org/10.46658/JBIMES-20-12>
- 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>
- Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2018). Psychometric properties of discourse measures in aphasia: Acceptability, reliability, and validity. *International Journal of Language & Communication Disorders*, 53(6), 1078–1093. <https://doi.org/10.1111/1460-6984.12420>
- \*Ramsberger, G., & Rende, B. (2002). Measuring transactional success in the conversation of people with aphasia. *Aphasiology*, 16(3), 337–353. <https://doi.org/10.1080/02687040143000636>
- \*Rayner, H., & Marshall, J. (2003). Training volunteers as conversation partners for people with aphasia. *International Journal of Language & Communication Disorders*, 38(2), 149–164. <https://doi.org/10.1080/1368282021000060308>
- \*Rose, M. L., Mok, Z., Carragher, M., Kathagen, S., & Attard, M. (2016). Comparing multi-modality and constraint-induced treatment for aphasia: A preliminary investigation of generalisation to discourse. *Aphasiology*, 30(6), 678–698. <https://doi.org/10.1080/02687038.2015.1100706>
- \*Ross, A., Winslow, I., Marchant, P., & Brumfitt, S. (2006). Evaluation of communication, life participation and psychological well-being in chronic aphasia: The influence of group intervention. *Aphasiology*, 20(5), 427–448. <https://doi.org/10.1080/02687030500532786>
- \*Savage, M. C., & Donovan, N. J. (2017). Comparing linguistic complexity and efficiency in conversations from stimulation and conversation therapy in aphasia. *International Journal of Language & Communication Disorders*, 52(1), 21–29. <https://doi.org/10.1111/1460-6984.12252>
- \*Savage, M. C., Donovan, N. J., & Hoffman, P. R. (2014). Preliminary results from conversation therapy in two cases of aphasia. *Aphasiology*, 28(5), 616–636. <https://doi.org/10.1080/02687038.2013.843153>
- Schegloff, E. A. (2006). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In N. Enfield & S. Levinson (Eds.), *The roots of human sociality* (pp. 70–96). Routledge. <https://doi.org/10.4324/9781003135517-4>
- Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511791208>
- Simmons-Mackie, N., Ahlsen, E., & Jensen, L. R. (2018). Overview of special issue on complexity, alignment, and enrichment in communication partner training for aphasia. *Aphasiology*, 32(10), 1123–1134. <https://doi.org/10.1080/02687038.2018.1453044>
- Simmons-Mackie, N., & Damico, J. S. (1997). Reformulating the definition of compensatory strategies in aphasia. *Aphasiology*, 11(8), 761–781. <https://doi.org/10.1080/02687039708250455>
- Simmons-Mackie, N., & Lynch, K. E. (2013). Qualitative research in aphasia: A review of the literature. *Aphasiology*, 27(11), 1281–1301. <https://doi.org/10.1080/02687038.2013.818098>
- Simmons-Mackie, N., Raymer, A., Armstrong, E., Holland, A., & Cherney, L. R. (2010). Communication partner training in aphasia: A systematic review. *Archives of Physical Medicine and Rehabilitation*, 91(12), 1814–1837. <https://doi.org/10.1016/j.apmr.2010.08.026>
- Simmons-Mackie, N., Savage, M. C., & Worrall, L. (2014). Conversation therapy for aphasia: A qualitative review of the literature. *International Journal of Language & Communication Disorders*, 49(5), 511–526. <https://doi.org/10.1111/1460-6984.12097>
- \*Sorin-Peters, R. (2004). The evaluation of a learner-centred training programme for spouses of adults with chronic aphasia using qualitative case study methodology. *Aphasiology*, 18(10), 951–975. <https://doi.org/10.1080/02687030444000453>
- \*Sorin-Peters, R., & Patterson, R. (2014). The implementation of a learner-centred conversation training programme for spouses of adults with aphasia in a community setting. *Aphasiology*, 28(6), 731–749. <https://doi.org/10.1080/02687038.2014.891094>
- Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., Ramage, A. E., & Roberts, A. C. (2021). Spoken Discourse Assessment and Analysis in Aphasia: An International Survey of Current Practices. *Journal of Speech, Language, and Hearing Research*, 64(11), 4366–4389. [https://doi.org/10.1044/2021\\_JSLHR-20-00708](https://doi.org/10.1044/2021_JSLHR-20-00708)
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K. E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Streiner, D., & Norman, R. (2000). *Health measurement scales: A practical guide to their development and use*. Oxford University.
- Tetnowski, J. T., Tetnowski, J. A., & Damico, J. S. (2021). Patterns of conversation trouble source and repair as indices of improved conversation in aphasia: A multiple-case study using conversation analysis. *American Journal of Speech-Language Pathology*, 30(1S), 326–343. [https://doi.org/10.1044/2020\\_AJSLP-19-00100](https://doi.org/10.1044/2020_AJSLP-19-00100)



- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. <https://doi.org/10.7326/M18-0850>
- \*Tuomenoksa, A., Pajo, K., & Klippi, A. (2016). Collaborative participation in aphasic conversation before and after intensive language-action therapy. *Clinical Linguistics & Phonetics*, *30*(10), 749–769. <https://doi.org/10.1080/02699206.2016.1220621>
- \*Ulmer, E., Hux, K., Brown, J., Nelms, T., & Reeder, C. (2017). Using self-captured photographs to support the expressive communication of people with aphasia. *Aphasiology*, *31*(10), 1183–1204. <https://doi.org/10.1080/02687038.2016.1274872>
- Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A.-C., Marshall, J., . . . Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*, *14*(2), 180–185. <https://doi.org/10.1177/1747493018806200>
- Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., Hin Kong, A. P., Simmons-Mackie, N., Scarinci, N., & Gauvreau, C. A. (2017). 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>
- \*Waller, A., Dennis, F., Brodie, J., & Cairns, A. Y. (1998). Evaluating the use of TalksBac, a predictive communication device for nonfluent adults with aphasia. *International Journal of Language & Communication Disorders*, *33*(1), 45–70. <https://doi.org/10.1080/136828298247929>
- Whitworth, A., Perkins, L., & Lesser, R. (1997). *Conversation Analysis Profile for People with Aphasia (CAPPA)*. Whurr Publishers.
- \*Wilkinson, R., Bryan, K., Lock, S., Bayley, K., Maxim, J., Bruce, C., Edmundson, A., & Moir, D. (1998). Therapy using conversation analysis: Helping couples adapt to aphasia in conversation. *International Journal of Language & Communication Disorders*, *33*(S1), 144–149. <https://doi.org/10.3109/13682829809179412>
- \*Wilkinson, R., Bryan, K., Lock, S., & Sage, K. (2010). Implementing and evaluating aphasia therapy targeted at couples' conversations: A single case study. *Aphasiology*, *24*(6–8), 869–886. <https://doi.org/10.1080/02687030903501958>
- \*Wilkinson, R., Lock, S., Bryan, K., & Sage, K. (2011). Interaction-focused intervention for acquired language disorders: Facilitating mutual adaptation in couples where one partner has aphasia. *International Journal of Speech-Language Pathology*, *13*(1), 74–87. <https://doi.org/10.3109/17549507.2011.551140>
- \*Woolf, C., Caute, A., Haigh, Z., Galliers, J., Wilson, S., Kessie, A., Hirani, S., Hegarty, B., & Marshall, J. (2016). A comparison of remote therapy, face to face therapy and an attention control intervention for people with aphasia: A quasi-randomised controlled feasibility study. *Clinical Rehabilitation*, *30*(4), 359–373. <https://doi.org/10.1177/0269215515582074>
- Worrall, L., Sherratt, S., & Papathanasiou, I. (2017). Therapy approaches in aphasia. In I. Papathanasiou & P. Coppens (Eds.), *Aphasia and related neurogenic communication disorders* (2nd ed., pp. 109–127). Jones & Bartlett Learning.
- Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology*, *25*(3), 309–322. <https://doi.org/10.1080/02687038.2010.508530>