Research Article

Spoken Discourse Assessment and Analysis in Aphasia: An International Survey of Current Practices

Brielle C. Stark,[a] [iD] Manaswita Dutta,[b] [iD] Laura L. Murray,[c] Davida Fromm,[d] [iD] Lucy Bryant,[e] Tyson G. Harmon,[f] [iD] Amy E. Ramage,[g] [iD] and Angela C. Roberts[h]

Purpose: Spoken discourse analysis is commonly employed in the assessment and treatment of people living with aphasia, yet there is no standardization in assessment, analysis, or reporting procedures, thereby precluding comparison/meta-analyses of data and hindering replication of findings. An important first step is to identify current practices in collecting and analyzing spoken discourse in aphasia. Thus, this study surveyed current practices, with the goal of working toward standardizing spoken discourse assessment first in research settings with subsequent implementation into clinical settings.
Method: A mixed-methods (quantitative and qualitative) survey was publicized to researchers and clinicians around the globe who have collected and/or analyzed spoken discourse data in aphasia. The survey data were collected between September and November 2019.
Results: Of the 201 individuals who consented to participate, 189 completed all mandatory questions in the survey (with fewer completing nonmandatory response questions). The majority of respondents reported barriers to utilizing discourse including transcription, coding, and analysis. The most common barrier was time (e.g., lack of time). Respondents also indicated that there was a lack of, and a need for, psychometric properties and normative data for spoken discourse use in the assessment and treatment of persons with aphasia. Quantitative and qualitative results are described in detail.
Conclusions: The current survey study evaluated spoken discourse methods in aphasia across research and clinical settings. Findings from this study will be used to guide development of process standardization in spoken discourse and for the creation of a psychometric and normative property database.
Supplemental Material: https://doi.org/10.23641/asha. 16639510

[a]Department of Speech, Language and Hearing Sciences and Program in Neuroscience, Cognitive Science Program, Indiana University, Bloomington, IN
[b]Department of Communication Disorders and Sciences, Rush University, Chicago, IL
[c]School of Communication Sciences and Disorders, Western University, London, CA
[d]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA
[e]Graduate School of Health, University of Technology Sydney, New South Wales, Australia
[f]Department of Communication Disorders, Brigham Young University, Provo, UT
[g]Department of Communication Sciences and Disorders, University of New Hampshire, Durham
[h]Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL

Correspondence to Brielle C. Stark: bcstark@iu.edu

Editor-in-Chief: Stephen M. Camarata
Editor: Julius Fridriksson

Discourse, or language production beyond the level of the sentence, is a fundamental component of communication. Indeed, impairments in spoken discourse have been shown to negatively affect individuals' social communicative competence and quality of life (Galski et al., 1998; Sim et al., 2013). As an expressive language measure, spoken discourse has good ecological validity and can be an important naturalistic language outcome measure (Davidson et al., 2003; Doedens & Meteyard, 2020; Linnik et al., 2016; Osborne et al., 1998). Accordingly, the evaluation of spoken discourse has gained widespread recognition as an important component of clinical aphasia assessment, treatment, and research (Brady et al., 2016; Bryant et al., 2016).

Spoken discourse provides a wealth of information about microstructural (e.g., linguistic elements such as syntax, lexical–semantics), macrostructural (e.g., communicative–linguistic elements such as cohesion, coherence, and main concepts; Armstrong, 2000; Cahana-Amitay & Jenkins, 2018; Whitworth et al., 2015), and interactional (e.g., turn-taking, topic maintenance, repair; Beeke et al., 2007; Tetnowski et al.,

2021; Wilkinson, 2014) properties of language and communication. Conducting multilevel analyses of language is difficult, if not impossible, using tasks such as confrontation naming or isolated sentence production. To collect spoken discourse samples, both structured and semistructured stimuli are frequently used with a variety of elicitation stimuli/tasks, including single picture or picture sequence descriptions, retelling a story or important life event, and topic-directed interviews (Brookshire & Nicholas, 1994; Bryant et al., 2016; Cherney et al., 1998). It is important to note that spoken discourse is proposed to be at least partially stimulus and task dependent, suggesting that micro- and macrostructural elements of language for a given individual may differ across stimuli (e.g., single picture, sequential pictures) and tasks (e.g., picture description, storytelling; Alyahya et al., 2020; Doyle et al., 1995; Fergadiotis & Wright, 2011; Fergadiotis et al., 2011; Stark, 2019; Ulatowska et al., 1981; Wright & Capilouto, 2009). To assess language constructed interactionally between two or more individuals, as well as pragmatic language abilities, clinicians and researchers commonly turn to conversations elicited with a clinician and/or familiar communication partner, which can be analyzed using methods (e.g., Conversation Analysis) and rating scales (e.g., Prutting & Kirchner, 1987) developed specifically for interactional tasks (Armstrong, 2000; Beeke et al., 2007; Damico et al., 1999; Ulatowska et al., 1992). It is therefore becoming increasingly clear that collecting language samples using multiple types of discourse genres can yield a comprehensive understanding of a speaker's language performance (Roberts & Orange, 2013; Shadden, 1998; Shadden et al., 1991; Stark, 2019; Ulatowska et al., 1981).

Methodological Issues Hampering Spoken Discourse Evidence

Despite its relevance for understanding the functional and pragmatic communication abilities of persons with aphasia and its potential to serve as a primary and important outcome measure, spoken discourse outcomes have been excluded from core outcome sets in aphasia (an agreed, minimum set of outcomes for treatment work; Brady et al., 2016; Wallace et al., 2019). A key reason why outcome sets currently exclude spoken discourse outcomes is due to a lack of standardization across data collection, analysis, and reporting. For the purposes of this study, we defined reporting as the explicit documentation and communication of information regarding how samples were recorded, interrater reliability, and other measures essential to replication, confidence, and reproducibility of the data. Another limitation in using discourse measures in outcome sets is that studies often report insufficient details around how language samples were collected and analyzed, which creates replication and reproducibility challenges. With few exceptions (e.g., Leaman & Edmonds, 2019; Roberts & Post, 2018), researchers have rarely (a) reported on how the raters responsible for transcribing, coding, and/or analyzing language samples were trained or (b) made rater training documents freely available (e.g., see Thompson, n.d.). Furthermore, in addition

to a large number of spoken discourse measures being reported in the aphasia literature, only infrequently (e.g., Boyle, 2014, 2015; Brookshire & Nicholas, 1994; Pritchard et al., 2018) have researchers intentionally studied the psychometric properties of specific discourse measures or developed robust normative data (Dietz & Boyle, 2018; Pritchard et al., 2017; Wallace et al., 2018). These issues preclude comparison/meta-analyses of discourse data and hinder the replication and reproducibility of findings, in both research and clinical settings. As such, there needs to be a concerted effort to standardize assessment, analysis, and reporting procedures in the field. Establishing and transparently reporting psychometric properties of discourse are necessary to instill confidence in users (e.g., clinicians, researchers) that the task and outcomes are reliable and valid for meaningful decision making.

Given the persistent heterogeneity in spoken discourse measurement and analysis procedures, their clinical utility in aphasia rehabilitation remains limited. Recent surveys have shown that although speech-language pathologists (SLPs) recognize the importance of targeting discourse in aphasia assessment and treatment, a large majority of them experience a variety of barriers in efficiently implementing spoken discourse analysis in their clinical practice. For instance, Bryant et al. (2017) surveyed 123 clinicians to better understand the contemporary uses of discourse analysis in clinical settings. Data were collected from five English-speaking countries with the majority of respondents being from Australia. The survey asked the SLPs about their discourse analysis practices in terms of frequency of use, collection, transcription, and analysis methods; perceptions and attitudes of SLPs regarding the use of discourse analysis in clinic were also queried. Over 85% of their study respondents indicated that they used a variety of methods to examine discourse productions of persons with aphasia, but perceptual, judgment-based approaches were most commonly used. The respondents noted that time to transcribe and analyze discourse samples was the most significant barrier, followed by other factors such as lack of adequate training and access to tools.

A more recent survey by Cruice et al. (2020) revealed similar findings amongst UK clinicians ($N = 211$). In addition to questions on participant demographics and procedures used to collect and analyze discourse samples, Cruice et al. also included questions about clinical feasibility and SLPs' capability, confidence, motivation, and opportunity to use discourse analysis in their practice. The authors found that only 30% of the clinicians who possessed appropriate knowledge and had good workplace support were *frequently* using discourse analysis. Most clinicians analyzed language samples in real time and were confident in making clinical judgments about the language abilities of persons with aphasia. While most respondents were open to implementing discourse analysis in their practice, they reported several barriers impeding the use including time constraints, lack of training, access to resources, aphasia severity, and uncertainty regarding selecting appropriate discourse outcome measures.

The aforementioned surveys provide important information about how discourse analysis is being used clinically by SLPs involved in aphasia rehabilitation (Bryant et al., 2017; Cruice et al., 2020). The studies highlight heterogeneity of the procedures used and identify significant barriers to the clinical use of discourse analysis such as lack of time, training, and resources. Given that both surveys focused primarily on SLPs working clinically, it is imperative to extend this enquiry to both researchers and clinicians working with persons with aphasia across the globe in different settings who also use discourse analysis in their work. Focusing on both groups is important because the goals and methods for discourse analysis, depending on the project/clinical need, may differ between research and clinical settings. For example, more granular, multilevel, hand-annotated analyses often required to characterize discourse impairments in a clinical cohort for a research study may not be feasible in clinical practice. Identifying similar and unique barriers across a range of professional roles and settings would help identify a coherent and comprehensive set of procedures to mediate barriers to using spoken discourse analysis in the assessment, treatment, and research of aphasia. Furthermore, it is critical to probe deeper into the specific methods currently being used by researchers and clinicians in terms of data collection, transcription, coding, and analysis procedures, as well as if and how they are determining the psychometric properties of discourse outcome measures. A more in-depth quantitative and qualitative analysis will identify the sources of methodological heterogeneity across different settings and form the basis toward establishing standards for discourse analysis to improve its transparency, replicability, and clinical utility.

Spoken Discourse Reporting Guidelines

To ensure that language science and, indeed, aphasiology continue to advance, we need to continuously examine principles and practices within our research and clinical communities, especially in light of failures to reproduce, replicate, or generalize findings in related behavioral disciplines (Ioannidis, 2005; Open Science Collaboration, 2015). The ability to replicate, reproduce, and interpret studies depends on the transparency and consistency of the reporting. There have been efforts across related fields to promote reproducibility and quality of evidence (e.g., Moher et al., 2010; Nichols et al., 2017), but there is a specific need to bring awareness to reporting within spoken discourse in aphasia, given the considerable lack of consistency and evidence in the existing research literature. There has been a recent advance in reporting standards relevant for aphasia, as seen in the ongoing DESCRIBE study (Establishing Standards for Reporting Participant Characteristics in Aphasia Research), aiming to gain consensus on reporting recommendations for participant characteristics in aphasia research studies (https://www.aphasiatrials.org). Relatedly, the Research Outcome Measurement in Aphasia consensus statement has recommended a set of core outcome measures to be used for aphasia intervention research to reduce heterogeneity and facilitate transparency, replicability, and reporting of meaningful outcomes (Wallace et al., 2019). Spoken discourse in aphasia comes with unique considerations for data collection, analysis, and dissemination that are not otherwise found in guidelines currently available for clinical trials or health studies. Examples of such considerations include the following: information about raters (i.e., those who transcribe, code, analyze, and/or interpret discourse data); availability of coding guidelines; use of transcription or coding methodology; and detailed information regarding the language proficiency and fluency of persons with aphasia (e.g., pre-aphasia bilingual status and proficiency).

Like the Committee on Best Practice in Data Analysis and Sharing, created by the Organization for Human Brain Mapping (Nichols et al., 2017), the objective of creating and maintaining reporting standards for spoken discourse in aphasia is to identify practices that maximize analytical stability and generalizability of study findings. Given the growth in spoken discourse research in aphasia across the past few decades (highlighted well in Bryant et al., 2016), we propose that the creation of reporting standards will "(a) encourage replication of studies, (b) ensure consistent reporting across studies, (c) recommend appropriate statistical modeling, thereby ensuring the most appropriate statistical inferences, and (d) overall, contribute to a more homogeneous, rigorous and standardized process by which spoken discourse research is evaluated and ultimately disseminated for clinical use" (p. 6; Stark et al., 2020). Reporting standards (and indeed, adherence to set standards) will enable meta-analytic consolidation of evidence and, more importantly, will have downstream, direct clinical implications by improving practices for collecting, analyzing, and accurately interpreting initial presenting status as well as changes in spoken discourse outcomes in aphasia.

Accordingly, to address the major gaps in the existing literature and improve the state of research in spoken discourse in aphasia, the FOQUSAphasia (FOstering the QUality of Spoken discourse in Aphasia) working group was created (Stark et al., 2020; http://www.foqusaphasia.com). This working group is composed of researchers, clinicians, and other stakeholders, including persons with aphasia and their caregivers. FOQUSAphasia has a relatively flat design that includes a steering committee who oversees and interfaces with the task forces and initiatives as well as the various stakeholders. Two of the task forces (i.e., "Best Practices" and "Methodology & Data Quality") focus on research initiatives, each with its own aim. For example, the "Best Practices" task force focuses on the creation of reporting standards, whereas the "Methodology & Data Quality" task force aims to create a shared, test–retest database across multiple sites.

Goals of This Study

The current survey was completed as part of the Best Practices Task Force within the FOQUSAphasia working group (Stark et al., 2020). The broad goal of the study was to survey current researchers and clinicians as a first step

toward creating recommendations for field-wide standards in methods, analysis, and reporting of spoken discourse outcomes, as has been done across other related disciplines (Nichols et al., 2017; Simmons-Mackie et al., 2017; Wallace et al., 2019). We used a mixed-methods survey to examine the current practices in spoken discourse collection, analysis, and interpretation undertaken by researchers and clinicians involved in aphasia assessment and rehabilitation across the globe. Albeit a prior survey provided a foundation to understand discourse use in clinical practice in aphasia (Bryant et al., 2017), the current survey focused on extending prior findings to a broader (research and clinical) audience to work toward standardization of discourse reporting in both clinical and research practice. That is, we created an extended survey to collect more detail regarding the current methods used for spoken discourse data collection, analysis, and interpretation in both research and clinical contexts. Such information is needed as a first step in creating guidelines because it can not only contribute further empirical rationale for the need for such guidelines (e.g., documenting heterogeneity in the methods used) but also inform our understanding of barriers that must be considered when developing such guidelines. Accordingly, the specific aims of our survey study were to (a) target concepts previously evaluated in prior related surveys (e.g., Bryant et al., 2017; Cruice et al., 2020), thus expanding on previous surveys in a different sample of respondents (i.e., those working in research and/or clinical settings), and (b) extend findings related specifically to data collection and analysis, with an emphasis on the psychometric properties of spoken discourse outcomes. To do so, we focused on the following research questions.

1. What are the current practices employed by clinicians and researchers using spoken discourse in their work relative to discourse sample collection, analysis approaches, and consideration of data reliability and validity?

2. What are the barriers faced by clinicians and researchers in using spoken discourse in their work relative to discourse sample collection, analysis, and data reliability and validity?

Method

We conducted a descriptive study of spoken discourse practice among researchers and clinicians; below, we outline related methodological details. We report our methods and results in accordance with the Checklist for Reporting Results of Internet E-Surveys (Eysenbach, 2004).

Participants

Participants self-selected to participate in the survey by responding to the following item: "You are being asked to participate in a research study. We are interested in understanding more about the methodology and analysis of spoken discourse in aphasia. We ask that you self-select to participate in this survey if you have in the past collected or are currently collecting and/or analyzing discourse data in speakers with aphasia, whether in a research or clinical capacity (or both). If you have not worked on discourse in aphasia, we ask that you do not continue on to the survey." A total of 201 participants consented to participate in this study, and their demographics are expanded upon in the results.

Survey Design and Procedure

The survey design and data collection were completed using REDCap, a secure web-based data management application (Harris et al., 2009, 2019). In the first iteration, the first and second authors (B. C. S. and M. D.) composed a set of items, built the survey, and designed the order of the items within the survey. The third, fourth, and fifth authors (L. L. M., D. F., and L. B.) reviewed the initial draft and suggested revisions to the survey questions as well as the order of items. Authors who contributed to designing the survey are certified SLPs who have worked in clinical settings with the aphasia population (M. D., L. L. M., D. F., L. B.) and who have conducted prior survey studies (L. L. M., L. B.; e.g., Bryant et al., 2017; Salis et al., 2018) and/or prior aphasia research focused on or utilizing spoken discourse assessment (B. C. S, M. D., L. L. M., D. F., L. B.; e.g., Fromm et al., 2017; Stark, 2019). Following this, the survey was piloted by the remainder of the authors along with select researchers and clinicians at Indiana University ($N = 5$ pilot participants, who were SLPs or communication sciences and disorders researchers reflecting the survey's target audience). Based on feedback received during piloting, some questions were revised and rechecked by the first five authors prior to distribution of the survey (via distribution methods discussed above). The presentation of survey questions was kept the same for all respondents (i.e., questions were not presented in a random order). Note that safety measures were used in REDCap to prevent the same respondent taking the survey on multiple occasions. Mobile compatibility settings were also used to enable potential respondents to fill out the survey on a variety of devices (e.g., desktop, laptop, tablet, phone).

The survey consisted of quantitative ("core" questions) and qualitative ("follow-up" or "clarification" questions as well as open-ended questions) items distributed across five sections.

1. Demographic information of participants (seven quantitative; seven qualitative items), discussing respondent's geographic location, age, years of working with persons with aphasia, education, and role;

2. Spoken discourse measurement in aphasia (eight quantitative; eight qualitative items), determining the extent to which respondents measured spoken discourse in aphasia, their reasons for doing so, and their barriers to discourse collection, transcription, analysis, and interpretation;

3. Data collection procedures (nine quantitative, 14 qualitative items), evaluating specific discourse data collection procedures, which expanded upon items asked in section 2 (Spoken discourse measurement in aphasia);

4. Data analysis procedures (30 quantitative, 15 qualitative items), examining transcription, coding, and analysis of spoken discourse data, including information regarding reliability analyses; and,

5. Psychometric properties and normative data (13 quantitative, 15 qualitative items), understanding common practices in and opinions about psychometric properties of discourse-derived outcomes (e.g., validity, stability, reliability) and the potential for normative data in discourse work.

Each section of the survey began with a short explanation of questions to be answered within that section. Question formats included multiple-choice, fill-in-the-blank, yes/no, and open-ended response options. Note that some questions allowed for multiple answers (e.g., one could be both an SLP and a researcher). While most items followed a forced-response format, answering all questions was not mandatory. Consequently, while $N = 189$ completed the survey, some questions were answered by fewer than 189 respondents; we note these instances in our Results section. The survey employed branching logic, such that for several items, respondents were shown an additional question if they answered in a certain way (e.g., if "yes" a follow-up question appeared). The number of items per page varied, as might be expected given that we created five sections of questions (described above); the maximum number of questions per page was 11. In total, the survey was 23 pages in length. A completeness check was not instituted by REDCap or the survey authors. There were 1986 page views from respondents (and potential respondents), including creating, updating, and viewing the survey on REDCap. See Supplemental Material S1 for the full survey.

Distribution of the Survey

Ethical approval for the study was obtained from the Indiana University Institutional Review Board. We employed convenience sampling for this open survey. A standard invitation e-mail message containing the survey link was distributed to several professional groups including the American Speech-Language-Hearing Association's Special Interest Group 2: Neurogenic Communication Disorders, Clinical Aphasiology Conference attendees, Speech Pathology Australia, and The Tavistock Trust for Aphasia. The survey was also publicized widely online and via social media (e.g., Facebook, Twitter, Google Groups, lab webpages). In that way, persons who took part in the survey or who saw the survey advertised were able to forward the survey to relevant parties ("snowballing"). The survey was protected using the Google reCAPTCHA feature to protect the survey from automated software programs (e.g., "bots").

Respondents were given the text, "I understand that this survey is assessing current methods and analysis techniques used to understand spoken discourse abilities in aphasia. By participating in this survey, I am currently or was involved in spoken discourse assessment in aphasia in my work or research setting," and then asked to click a button that said, "I consent to participate in this survey." The only way to advance into the survey was to select this option. Upon clicking "I consent to participate in this survey," REDCap assigned each respondent with a unique identifier.

Potential respondents were told in the informed consent information section that the survey would take 30–40 min to complete and that it could be completed in more than one sitting. To continue completing the survey at a future time, they entered their e-mail address, and the survey generated a password for reentry and access to their saved survey responses at a later date. The e-mail addresses of the respondents were not saved by REDCap (i.e., the authors of this study could not see these e-mail addresses). Respondents were allowed to change answers to their questions at any time during the survey period. Respondents were requested to click "Complete" when they had fully completed the survey, or REDCap would automatically select this option if all quantitative and qualitative sections had been completed. No identifying information was collected from participants during the completion of the survey. The survey was distributed in English from September to November 2019 and then closed for response analysis. No incentives were given as a part of this survey.

REDCap does not currently utilize an IP check to identify potential duplicate entries from the same user. Instead, we probed the log file to identify multiple entries, which were flagged if exact duplicates were identified in Section 1 of the survey (i.e., demographics, specifically age, gender, country, highest earned degree, "How would you describe yourself?" and "How many years have you been involved in aphasia assessment or rehabilitation?").

Data Analysis

After the survey was closed, responses were downloaded from REDCap in PDF and Microsoft Excel formats. Descriptive statistics were used to analyze responses to demographic questions as well as to quantify response frequencies to quantitative questions. Cross-tabulation analyses were also employed to investigate differences in categorical items by demographic categories. Responses to quantitative questions were entered for analysis into the Statistical Package for Social Sciences (SPSS 26; IBM Corp, 2019), while open-ended questions (e.g., qualitative) were exported to Excel for qualitative analysis.

Qualitative analysis was completed on 35 open-ended survey questions that had a response rate of 20% or higher ($Mdn = 73\%$, range: 21%–100%). This criterion excluded five open-ended survey questions from the analysis (see Supplemental Materials S2 and S3 for response rates by item) that had extremely low response rates (i.e., high nonresponse bias). Using a thematic analysis approach, informed by Braun and Clarke (2006), participant responses were coded independently by the final three authors (T.G. H., A. E. R.,

A. C. R.) using a five-step iterative process. First, all three coders familiarized themselves with the data set by reading through the open-ended participant responses while reflecting on the data and taking notes. Second, the three coders independently assigned possible codes to the responses using an inductive coding approach. Third, the three coders met to collaboratively discuss their independently assigned codes and to draft and revise a codebook. During this process, coders iteratively convened group discussion followed by independent review of a portion of the data until the codebook was deemed appropriate for the dataset (i.e., three times). Fourth, the coders returned to the data and independently (i.e., masked to the other raters' data) recoded responses based on the final codebook (see Tables 2–4 for themes and codes used). Coding consensus was defined as responses where at least two of three raters independently assigned the same code to a survey response. Discrepancies, responses for which coding consensus was not achieved, occurred on 40 (5.56%) survey responses. In the final step of the analysis, these discrepancies were discussed by the three raters collectively until coding consensus agreement was achieved. The qualitative coders were blinded to the quantitative survey responses and to the quantitative question prompts while coding and extracting themes from the open-ended responses to minimize extracting themes that would have been biased by the summary quantitative question statistics.

Results

Responses

Whereas a total of 201 participants consented to participate in the study, 12 respondents did not proceed to complete the survey after providing consent. Of the 189 respondents who completed the demographic section, 110 individuals completed all questions (i.e., 110/201 = 58%); as a reminder, not all questions were mandatory, and this feature was likely the cause of fewer responses across some questions. We considered all responses as contributing valuable information regardless of whether they came from participants who completed all survey questions; therefore, we did not reject entire surveys for noncompleters. Accordingly, in the following sections, we report the number of respondents for each data point in parentheses. We also highlight response rates for quantitative and qualitative questions in Supplemental Materials S2 and S3. Results reported below adhere to the order of the survey. Because our qualitative thematic analysis was approached holistically, qualitative themes and categories are mentioned in connection with their related quantitative results. We divide the results section into the five survey sections discussed in the Methods. Portions of Sections 2–5 address the research questions.

Survey Section 1: Demographic Information of Participants

We report data for the 189 participants who completed demographic questions. The descriptive statistics

results are presented in Table 1. The respondents were geographically dispersed, with more than half located in the United States. The majority of respondents identified as SLPs, with a large proportion identifying as researchers (note that respondents could identify as having more than one affiliation, so respondents who checked "SLP" could also check "researcher" as a response option). The majority of respondents were aged 26–40 years and identified as female. A variety of education backgrounds were represented. Some of the most common work settings at which respondents practiced or collected and analyzed spoken discourse data included a rehabilitation setting, acute care, hospital-based outpatient clinic, and university research lab or clinic. Respondents demonstrated a wide range of years working with persons with aphasia.

Survey Section 2: Spoken Discourse Measurement in Aphasia

This section examined the extent to which respondents measured spoken discourse in aphasia, their reasons for doing so, and their barriers to discourse collection, transcription, analysis, and interpretation. In terms of the frequency of discourse data collection and/or analysis ($N = 165$), a majority of respondents indicated that they *always* (33.3%) or *usually* (33.9%) collected and/or analyzed spoken discourse samples from persons with aphasia. The most common reasons for collecting spoken discourse data were to gain information regarding aphasia symptoms for clinical intervention purposes (72.1%), as an outcome measure for aphasia treatment in clinical practice (53.9%), and/or for research (31.5%). There was no significant association between years working with persons with aphasia and how often respondents collected spoken discourse, $\chi^2 [df = 140, N = 187] = 141.57$, $p = .45$, or between age of respondents and how often respondents collected spoken discourse, $(\chi^2 [df = 12, N = 189] = 4.36, p = .98)$. Furthermore, there was a nonsignificant association between primary work setting and how often respondents collected spoken discourse $(\chi^2 [df = 32, N = 189] = 46.12, p = .051)$, though university and hospital settings were the most common settings to '*always*' collect discourse data.

The most commonly collected spoken discourse genre was a description of a single picture (e.g., Cookie Theft picture from the *Boston Diagnostic Aphasia Examination* [BDAE], Goodglass et al., 2000) (89.1%), followed by a conversation with a clinician and/or family member (70.9%; $N = 165$). Other typically collected genres included a personal recount (e.g., important life event, 67.9%), procedural narrative (e.g., how to make a peanut butter and jelly sandwich, 57%), and an interview (biographical or otherwise, conducted by a clinician, 51.5%). To collect spoken discourse data ($N = 163$), respondents reported relying most on standardized aphasia assessment tools that included discourse generation tasks (e.g., *Western Aphasia Battery-Revised,* Kertesz, 2007; 74.8%), but some also endorsed using protocols such as the Nicholas and Brookshire protocol (31.3%, Nicholas & Brookshire, 1993), the AphasiaBank

Table 1. Demographic information of respondents.

Demographic information	Responses	Respondents (N)
Locations	USA (55%)	189
	United Kingdom (7.4%)	
	Australia (19.6%)	
	New Zealand (0.5%)	
	Canada (6.9%)	
	Other (10.6%)	
Roles	Researcher (43.4%)	189
(could select more than one)	Academic/teacher (22.2%)	
	Speech-language pathologist (81%)	
	Student (9%)	
	Other (3.7%)	
Age	< 25 years of age (4.2%)	189
years	26–40 years of age (47.6%)	
	41–55 years of age (30.7%)	
	> 55 years of age (17.5%)	
Gender	Female (93.1%)	189
	Male (5.8%)	
	Other (0.5%)	
Terminal degree	Bachelor's (15.9%)	189
	Master's (51.3%)	
	PhD (22.8%)	
	Postdoctoral (5.3%)	
	Clinical doctorate (1.1%)	
	Other (3.7%)	
Main area of data collection	Acute care (8%)	188
	Rehabilitation (23.4%)	
	Community health (6.9%)	
	Long-term care facility (3.2%)	
	Private practice (5.3%)	
	Hospital-based outpatient clinic (16%)	
	University research lab or clinic (33%)	
	Other (4.3%)	
Years of working with people with aphasia	M = 14.17 (SD = 10.45), range 1–45	187

protocol (15.3%, MacWhinney et al., 2011), or a self-developed protocol (20.2%). Qualitative responses elaborated on the quantitative findings discussed above, with respondents expanding on specific practices related to conversation elicitation (including both unscripted conversation interactions collected through naturalistic tasks and formal scripted exchanges elicited using interview guides, questions, and barrier-style tasks; these practices also included single-partner and group conversations), retellings or recounts (recounted content from videos, wordless picture books, and current events), and narrative descriptions from visual information (single, sequence, or wordless books; see Table 2, theme 1).

In terms of the number of samples (i.e., discourse tasks) typically collected and/or analyzed per individual with aphasia, respondents (N = 147) most often collected one to two samples (41.5%) or three to four samples (38.1%), with relatively few collecting five to six samples (11.6%) or more than six samples (8.8%). Additionally, respondents mentioned that, ideally, they would like to collect three to four samples (41.5%) or five to six samples (24.5%); few cited one to two samples (17.7%) or more than six samples (16.3%) as an ideal number.

To indicate the typical barriers to discourse sample collection, transcription, analysis, and interpretation, respondents could choose more than one barrier per section (e.g., collection, transcription) (see Figure 1). The most commonly selected barriers to discourse collection (N = 162) included lack of access to tools and resources (e.g., computer hardware/software, recording equipment) (34.6%), inadequate training in discourse collection (25.9%), and insufficient skills and/or knowledge in discourse collection (19.1%). Notably, 29.6% indicated no barriers to discourse collection. Regarding discourse transcription (N = 162), a major barrier was lack of time/time constraints (80.2%), with only 9.9% reporting no barriers. For discourse analysis (N = 161), respondents overwhelmingly endorsed lack of time/time constraints as a major barrier (75.8%), with only 6.8% indicating no barriers. Finally, the major barriers to discourse analysis interpretation (N = 161) included time constraints (50.9%), lack of training (33.5%), and lack of skills and knowledge (26.7%). A small proportion of respondents (20.5%) cited no barriers to discourse analysis interpretation.

Open-ended question responses provided further and clarifying information regarding perceived barriers to use of discourse assessment in aphasia (see Table 3). Responses were grouped by the following themes: resource-related barriers, clinician/researcher–related barriers, patient/participant–related barriers, and measurement-related barriers. Within resource-related barriers, different aspects of time were cited as issues (e.g., lacking "time to devote to self-training," "it takes a long time to train [others]," general feeling of lack of time), as well as limits in personnel and environment.

Table 2. Number of respondents reporting/endorsing use of specific discourse procedures in their clinical and/or research practice, organized by theme and subtheme.

Themes	Narrative response summary	Number of respondents
Theme 1: Elicitation methods		
1.1 Conversation or dialogue	Both unscripted conversation interactions collected through naturalistic tasks and formal scripted exchanges elicited using interview guides, questions, and barrier-style tasks; single-partner and group conversations	14
1.2 Story retelling/recounts	Recounting content from videos, wordless picture books, and current events	6
1.3 Narrative generation from pictures (e.g., single, composite, picture book)	Describing pictured scenes or "expository" discourse	5
Theme 2: Orthographic transcription		
2.1 No transcription	Do not routinely transcribe discourse because they perceive having sufficient ability to detect features of interest online or use perceptual rating scales that are scored during production	13
2.2 Partial transcription	Orthographically transcribing part of the sample verbatim or noting and transcribing errors only	9
2.3 Full transcription	Orthographically transcribing the entire sample verbatim	9
Theme 3: Audio recording		
3.1 No audio recording	"Samples aren't audio recorded" because they are "transcribed online" or scored/rated online during production.	8
3.2 Audio recording	Samples are routinely audio recorded for later transcription/rating.	8
Theme 4: Training individuals involved in discourse analysis		
4.1 Published protocols/annotation systems	Protocols such as those found in research articles, on websites, or as software tutorials used to train others to help with discourse annotation and analysis	16
4.2 Self-created protocols	Protocols developed in-house used to train others	12
Theme 5: Analysis approaches		
5.1 Granular language form and content	Analysis of language form and content (e.g., words, sentences, main concepts, CIUs [correct information units], target words, lexical diversity), syntax (e.g., parts of speech, syntactic complexity, phrase structure, predicate argument structure), errors (e.g., word-finding difficulties, paraphasia, morphosyntactic errors), fluency (e.g., speech rate), and macro-linguistic structure (e.g., coherence, cohesion, story grammar).	41
5.2 Global language form and content	Formal rating scales from standardized tests—primarily the Western Aphasia Battery; self-developed informal ratings to reflect "overall judgments of grammaticality" or "broad error patterns"	21
5.3 Granular pragmatic	Conversation analysis and analysis of conversation using formal approaches and behavior quantification instruments	13
5.4 Global functional	Rating scales of communicative effectiveness, comprehensibility, and conversation ability by expert and naïve listeners and also self-/conversation partner ratings	10
5.5 Global pragmatic	Informally or formally rating "conversational features" such as "turn-taking"	3
5.6 Global motoric	Rating scales that accounted for aspects of motor speech or judgments of intelligibility	2
Theme 6: People involved in discourse analysis		
6.1 Analyzed and collected by the same person (clinician or researcher)	Either worked alone or did not have access to trained personnel to support discourse analysis and reliability procedures	36
6.2 Students/trainees	Supported by graduate students or coursework that students were required to complete	23
6.3 Colleagues	Supported by colleagues or collaborators	7
6.4 Research staff	Supported by research assistants or lab managers	7

Responses that highlighted clinician/researcher–related barriers included perceived misalignment with high priority outcomes (e.g., "[not] relevant to dissect a client's verbal output—I prefer to look at the big picture [and their] primary goals"), lack of training or knowledge, a belief that discourse analysis was not related to their job position (e.g., "not my job") and/or that discourse analysis was not pursued because of a lack of professional interest, and historic or current practice patterns. Barriers were also endorsed by respondents in relation to the patient/participant, including the severity of impairment (e.g., most severe language impairments, especially those with concomitant motor speech disorders, may preclude usefulness or meaningfulness of discourse ["task will be too challenging for the client if more severe," "too little understandable speech to warrant an informative analysis"]) and the perceived burden on the patient/participant of collecting discourse samples. Finally, a number of measurement-related barriers were endorsed by respondents, including a general lack of psychometric data and shared processes specific to discourse (e.g., lack of standard practice, lack of psychometric data, variability in discourse collection and analysis methods ["[discourse measures ultimately] depend on the individuals doing the assessments and how they are trained"]), lack of linguistic

Figure 1. Barriers to data collection, transcription, analysis, and interpretation. Respondents could select more than one barrier. * = No response option for "data collection." ^ = No response option for "data interpretation." N/A = not applicable.
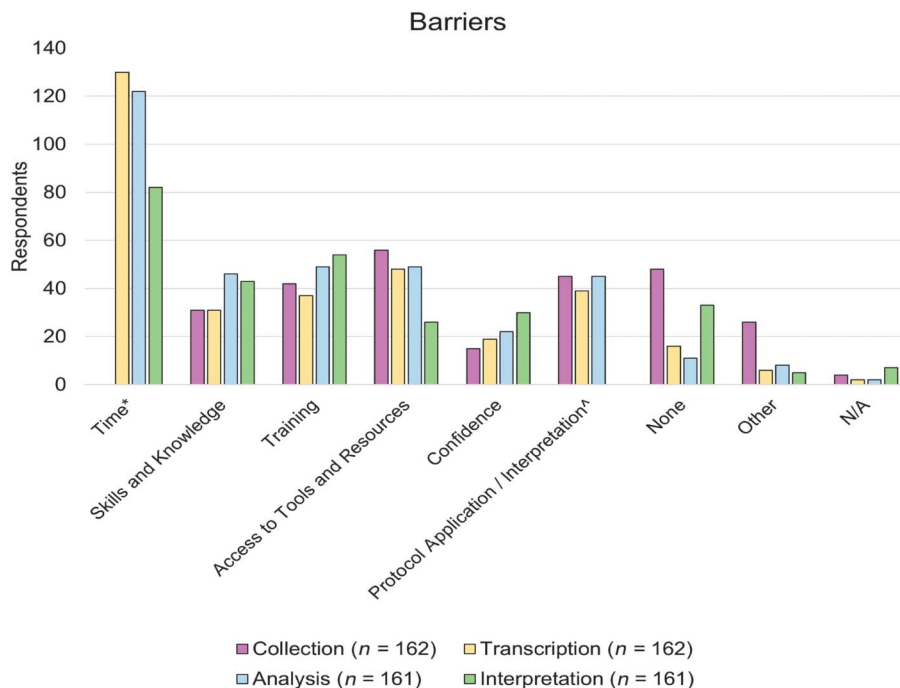
and culturally specific discourse methods/data (e.g., lack of protocols, normative data, and other psychometric properties in languages other than English and in cultures where monolingual speakers are not the majority), and lack of or difficulty finding empirical evidence.

Survey Section 3: Data Collection Procedures

This section included questions regarding specific discourse data collection procedures, expanding upon items in Section 2. In terms of the typical methods employed to collect spoken discourse data ($N = 147$), respondents indicated collecting samples in a quiet room (74.8%), in a hospital or rehab facility room with typical daily distractions (e.g., background noise; 48.3%), or at a participant's home (36.1%); few employed a sound booth (1.4%). Furthermore, a majority of respondents reported recording the spoken discourse data (77.6%). Individuals who indicated that they recorded discourse samples ($N = 114$) reported that, to collect this information, they used an audio recorder (58.8%), a video recorder (41.2%), a cellphone (31.6%), or a tablet (22.8%). Few used a laptop (with a webcam, 12.3%; sound only, 5.3%) or an external microphone (12.3%). Those respondents who indicated that they did not record spoken discourse data ($N = 33$) mentioned that they transcribed in real time (60.6%) and/or analyzed in real time without transcription (62.6%). Those who responded that they do not record samples and also answered open-ended questions reported feeling that transcription was not necessary or essential to their analysis because (a) they had sufficient

ability to detect features of interest online without a need to review the sample later; (b) they used perceptual rating scales that were scored during production or; (c) the desire to record only production errors could be accomplished through observations made online, only owing to the rater's skill or the low frequency of errors (see Table 2).

Respondents reported that a typical length of a recorded discourse sample ($N = 147$) was 1–3 min (24%), 3–5 min (19%), greater than 5 min (14.9%), or that the time varied by discourse type (17.7%). Few respondents indicated a length of less than 1 min (4.1%).

Survey Section 4: Data Analysis Procedures

This section asked about transcription, coding, and analysis of spoken discourse data, including information regarding reliability analyses. Regarding the steps undertaken once spoken discourse data are collected ($N = 145$), respondents reported listening to the recorded samples (68.3%), transcribing samples verbatim (71.7%), coding transcripts (48.3%), performing detailed analysis of transcripts (24.8%), making a perceptual judgment-based analysis (58.6%), making a clinical judgment of language ability (69.7%), and/or implementing other steps such as obtaining blinded listener ratings or conducting further pragmatic analysis (6.2%; see Figure 2).

Survey Subsection 4.a. Transcription

In terms of the frequency of transcriptions, respondents ($N = 144$) indicated *always* (33.3%), *usually* (31.3%),

Table 3. Number of respondents reporting/endorsing specific barriers to using discourse analysis in clinical and/or research practice, organized by theme and subtheme.

Themes	Narrative response summary	Number of respondents
Theme 1: Resource-related barriers		
1.1 Time	Perceived lack of time to collect and analyze discourse (e.g., "I don't have the time"), train oneself (e.g., "time to devote to self-training"), and train others (e.g., "it takes a very long time to train [others]")	152
1.2 Personnel	No, or limited, staff to assist with collecting or analyzing discourse data	23
1.3 Environment	Workplace, technology, and financial barriers including (a) no process or protocol in place for collecting or analyzing discourse data, (b) "unexpected" or "early" patient discharge or transfer, (c) no access to or knowledge of software used to process discourse, or (d) lack of equipment for high-quality recordings and worries about Health Insurance Portability and Accountability Act compliance relating to audio recordings and transcription processes	21
Theme 2: Clinician/researcher-related barriers		
2.1 Perceived misalignment with clinical and/or research priorities	Respondents perceived that discourse does not capture high priority outcomes for their clients and incorporating discourse in practice would not alter treatment goals and plans; discourse data are not always relevant for research questions or necessary for publication.	66
2.2 Training/knowledge	Lack of skills or knowledge to analyze discourse data; need for specific training in discourse collection and analysis.	56
2.3 Not related to job position or lack of professional interest	"Not my job"; perceived as not part of respondent's professional responsibility or disinterest in collecting, analyzing, or using discourse in practice	11
2.4 Historic or current practice pattern	Not the pattern of practice in the setting in which the respondent works or not part of their usual practice "habit"	9
Theme 3: Patient/participant-related barriers		
3.1 Severity of impairment	More severe language impairments, particularly with the co-occurrence of motor speech disorders (i.e., apraxia of speech, dysarthria), either made obtaining discourse data more difficult or less meaningful	9
3.2 Burden on patient/participant	Asking patients/participants to generate representative language samples multiple times might place too much of a burden on them.	2
Theme 4: Measurement related barriers		
4.1 Psychometric properties (lack of or problems with)	Lack of standard practice and psychometric data relative to discourse tasks and measures. Variability in types of discourse collected, elicitation techniques, analysis approaches or outcome measures used, and who administers the task makes discourse analysis "very messy."	19
4.3 Lack of linguistic and culturally specific discourse methods/data	Lack of "protocols," "normative data," and other "psychometric properties" for discourse in languages such as "French," "Dutch," and "Turkish"	10
4.4 Lack of/difficult to find empirical evidence	Insufficient research evidence to support discourse use in assessment or as an outcome measure; or evidence is hard to synthesize because of its disparate nature and reporting. No central access to discourse normative data.	4

sometimes (18.8%), *rarely* (11.8%), or *never* (3.5%) undertaking sample transcriptions. When asked if samples were typically personally transcribed (i.e., by the person collecting the discourse data), respondents ($N = 144$) answered as follows: *always* (41%), *usually* (11.1%), *sometimes* (18.1%), *rarely* (18.8%), or *never* (6.9%). Respondents who indicated that they did not always personally transcribe samples ($N = 79$) reported that the following personnel were involved in transcriptions: a graduate student volunteer [unpaid] (31.6%), a clinician/SLP (30.4%), a paid graduate-level research assistant (29.1%), a paid research assistant or lab manager [not a student] (26.6%), other (24.1%), a paid undergraduate research assistant (19%), a PhD student whose work involves the data collected (16.5%), an undergraduate student volunteer [unpaid] (24.1%), and/or a researcher with a PhD in a related field (8.9%). Open responses elaborated on these data (see Table 2, theme 2). Respondents who did not transcribe data cited that they perceived themselves as having sufficient ability to detect features of interest online (i.e., in real time) or to use perceptual rating scales. Those who completed partial transcription described orthographically transcribing part of the sample verbatim or noting and transcribing errors only.

Survey Subsection 4.b. Coding

The preceding transcription items were followed by questions about coding the spoken discourse samples. Respondents ($N = 144$) indicated that samples were coded *always* (27.1%), *usually* (14.6%), *sometimes* (22.2%), *rarely* (13.2%), or *never* (21.5%).

Respondents ($N = 143$) indicated that they *always* (29.4%), *usually* (12.6%), *sometimes* (21%), *rarely* (9.8%), or *never* (14%) personally coded the samples. Furthermore, the personnel reported ($N = 81$) being involved in coding included a graduate student volunteer [unpaid] (32.1%), a clinician/SLP (24.7%), a paid graduate-level research assistant (32.1%), a paid research assistant or lab manager

Figure 2. Follow-up steps respondents reported taking after collecting a spoken discourse sample. Respondents could select more than one option. N/A = not applicable.

After collecting a sample, what steps do you complete? ($N = 145$)
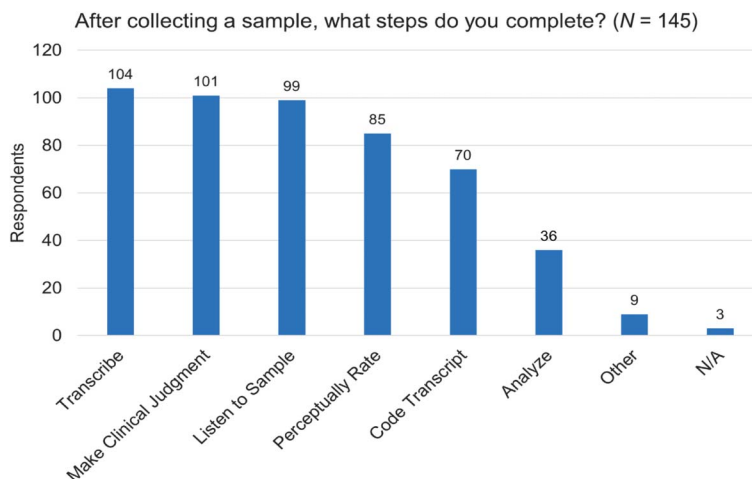
[not a student] (22.2%), other (25.9%), a paid undergraduate research assistant (13.6%), a PhD student whose work involves the data collected (19.8%), an undergraduate student volunteer [unpaid] (19.8%), and/or a researcher with a PhD in a related field (11.1%). Notably, of those respondents providing explanations or open responses, two respondents indicated not knowing what the term "coding" meant in reference to discourse analysis.

Survey Subsections 4.c and 4.d. Analysis

In terms of the frequency of data analysis, respondents ($N = 139$) indicated analyzing samples *always* (51.1%), *usually* (28.1%), *sometimes* (10.8%), *rarely* (6.5%), or *never* (3.6%). Furthermore, participants ($N = 138$) reported that they *always* (55.8%), *usually* (18.8%), *sometimes* (10.9%), *rarely* (8.7%), or *never* (5.8%) personally analyzed the samples. If the respondents did not themselves code the discourse samples ($N = 56$), the following personnel were involved in the analysis: graduate student volunteer [unpaid] (35.7%), a clinician/SLP (21.4%), a paid graduate-level research assistant (32.1%), a paid research assistant or lab manager [not a student] (25%), other (19.6%), a paid undergraduate research assistant (10.7%), a PhD student whose work involves the data collected (23.2%), an undergraduate student volunteer [unpaid] (21.4%), and/or a researcher with a PhD in a related field (21.4%). Of those who responded to open-ended questions regarding who was involved in analysis, most indicated that they either worked alone or did not have access to trained personnel to support discourse analysis and reliability procedures. For example, one respondent replied, "I work in the real world and do it all myself." In addition, some respondents mentioned receiving help from students, colleagues, required coursework, research assistants, and lab managers (see Table 2, theme 2).

To better understand the common practices in data analysis, we asked respondents *how* discourse samples were typically analyzed ($N = 133$). A majority indicated that

they used clinical judgment (69.9%), employed standardized aphasia assessment ratings/scoring (62.4%), and/or used manual coding and analysis (e.g., main concept analysis; 48.9%). Fewer respondents utilized computerized systems, such as Systematic Analysis of Language Transcripts (Miller & Chapman, 1983; 14.3%), Computerized Language Analysis (MacWhinney, 2000; 11.3%), Praat (Boersma & Van Heuven, 2001; 3.8%), or Computer Analysis of Speech for Psychological Research (Brown et al., 2007; Covington et al., 2007; 0.8%). Six percent indicated that they did not use a specific protocol, and 11.3% indicated "other" methods such as blinded listener ratings, pragmatic analysis protocol, or spontaneous speech analysis. In general, when expanding on their responses, respondents reported using both granular (e.g., specific coding of discourse features) and global (e.g., overall rating or singular judgment score) analyses relating to language form and content, pragmatics, and functional as well as motor speech measures (see Table 2, theme 5).

As shown in Figure 3 ($N = 122$), the most commonly extracted discourse outcome measures provided information about fluency (64.8%), informational content (65.6%), paraphasias/word retrieval errors (66.4%), conversational behaviors (62.3%), and grammatical errors (63.1%), with less than half of the respondents evaluating functional or macrostructural information (e.g., story grammar, cohesion) (40.2%). On average, respondents stated that they extracted 3.33 ($SD = 3.46$) outcome measures from spoken discourse.

Section 5: Psychometric Properties and Normative Data

This section asked about the psychometric properties of discourse-derived outcomes (e.g., validity, stability, reliability).

Raters

As noted earlier, different personnel were reported as being involved in the transcription, coding, and analyzing

Figure 3. Typically analyzed outcome measures. Respondents could select more than one option.
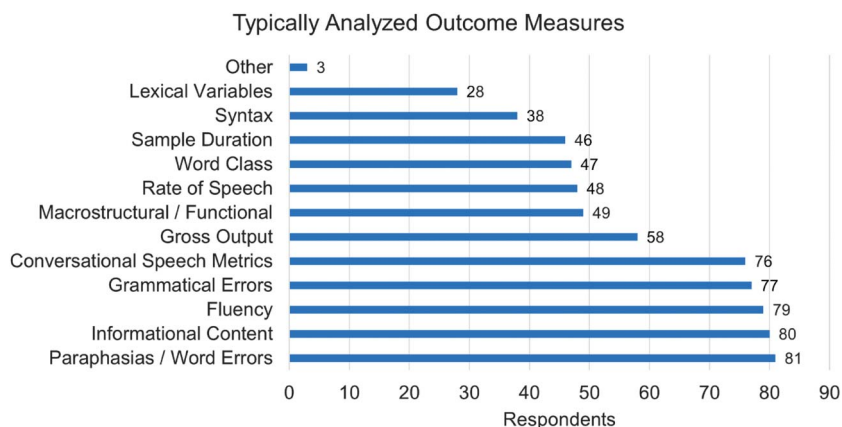
process. Accordingly, this section asked more detailed questions about raters and their training. Respondents ($N = 123$) were split as to whether they personally trained raters: 39% did, 47.2% did not, and 13.8% taught in conjunction with others (e.g., collaborative training of raters between multiple study personnel). Notably, we want to acknowledge a limitation on this result. This question made the assumption that other parties are involved in rating, which may not be the case. Therefore, it may be the case that some respondents who do all of the discourse work on their own were forced into an erroneous response option (e.g., "no"); we therefore urge readers to interpret these results in this context.

Furthermore, a majority of respondents indicated that they did not follow any specific training protocol for transcription, coding, and/or analysis (81.3%), with only a minority following a specific protocol (18.7%). Seventeen respondents reported using published protocols (such as those found in research articles, on webpages, or as software tutorials) to train others to help with discourse analysis, while 12 indicated using self-created protocols (see Table 2, theme 4). For respondents indicating the use of a protocol ($N = 23$), we asked if that protocol was freely available and easily accessible online; 39.1% said yes and 60.9% said no. In terms of those involved in transcribing, coding, and/or analysis of discourse data ($N = 121$), the most common educational background of these individuals (of which respondents could select more than one option) was speech-language pathology (91.7%) followed by linguistics (33.9%).

Decisions about discourse analysis

Due to the considerable downstream effects that utterance delineation has on many spoken discourse outcome measures (e.g., mean length of utterance, syntactic variables), we inquired how utterance boundaries were determined when transcribing (multiple answers allowed). We found that respondents ($N = 122$) used a variety of methods including "both syntactic and pause/intonation indicators" (45.9%), "syntactic indicators" (33.6%), "pauses" (30.3%), or "full ideas" (see Figure 4).

Regarding the rationale for selecting discourse outcome measures, a majority of respondents stated that they chose outcome measures because they were used in publications describing a similar therapy/assessment program (45.5%) or they had training/experience in using these measures (45.5%; $N = 121$; see Figure 5).

Psychometric properties: Rater reliability

The majority of respondents indicated that they did not usually collect data about rater agreement (53.4%), but some respondents ($N = 116$) indicated collecting rater agreement during transcription (29.3%), coding (37.1%), and analysis (38.8%). If reliability was checked, the personnel completing the reliability analysis were most commonly the respondents themselves (40.9%), a clinician/SLP (19.1%), or a paid graduate-level research assistant (18.3%; $N = 115$).

We then asked what discourse measures were examined for rater reliability (multiple answer), with the most common being all outcome measures of interest (29.6%), followed by total words (or tokens; 17.4%), only some outcome measures of interest (16.5%), total utterances (14.8%), other (3.5%), or not applicable (51.3%; $N = 115$). To quantify rater agreement, respondents ($N = 115$) frequently used percent agreement (29.6%) or intraclass correlation coefficient (17.4%), with less-used metrics being correlation coefficient (13%) or other (5.2%) (51.3% answered N/A and 9.6% answered "none").
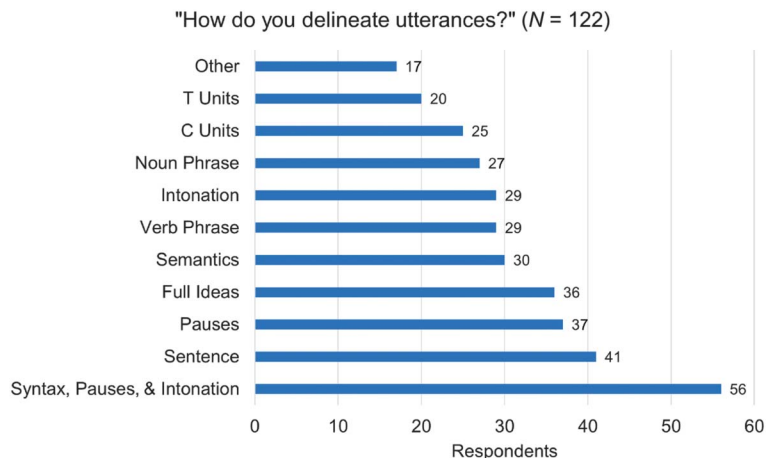
Psychometric properties: Test–retest data

Respondents ($N = 110$) typically never (36.4%) or sometimes (40%) collected test–retest data for spoken discourse samples (often, 6.4%; usually, 10.9%; always, 10%).

Psychometric properties of outcome measures

A majority of respondents indicated that, in general, there was inadequate availability of psychometric data (81.8%, $N = 111$) and normative data (51.8%, $N = 110$) for spoken discourse outcome measures. Respondents ($N = 110$) further stated that they would find a database of psychometric

Figure 4. Utterances were delineated in a variety of ways by respondents. Respondents could select more than one option.

"How do you delineate utterances?" (*N* = 122)

Category	Respondents
Other	17
T Units	20
C Units	25
Noun Phrase	27
Intonation	29
Verb Phrase	29
Semantics	30
Full Ideas	36
Pauses	37
Sentence	41
Syntax, Pauses, & Intonation	56

properties and/or normative data of discourse outcomes useful (93.6%).

We then asked if respondents looked for psychometric information of discourse outcome measures (e.g., reliability, validity, stability, acceptability), 39.6% said "no," 6.4%; usually "yes," and 27% said "sometimes" (*N* = 116). Respondents (*N* = 110) cited that major barriers to collecting psychometric data included time (82.7%), knowledge and training (60.9%), funds (46.4%), personnel (42.7%), and other (10%). Respondents were also asked if they looked for normative data for discourse outcome measures: 30% said "no," 32.7% said "yes," and 37.3% said "sometimes" (*N* = 110). Finally, participants had the opportunity to share their insights into ways to facilitate the collection, analysis, and publication of spoken discourse data in aphasia (*N* = 29).

Qualitative data indicated that respondents saw the potential value and usefulness of standardized discourse measures for comparison and interpretation, to ensure best practice, as a meaningful outcome measure, and for reimbursement purposes (see Table 4). Specifically, psychometric properties of discourse data were thought to be important

for comparing and interpreting discourse measures across individuals and approaches (e.g., "Without adequate psychometric properties described, interpretation of results is problematic, and clinical application of measures will be limited") and were thought to be useful for expressing "stable," "reliable, valid and sensitive" measures that are considered "best practice." Additionally, respondents described valuing discourse measures related to clients' goals, posttherapy change, and those that could be applied to "real life" and highlighted the importance of psychometrically sound measures for determining clinical outcomes and "gauging treatment effects." Furthermore, open-ended responses spoke to important issues in psychometric data collection and use: that it is often not part of the practice or what is done at the setting (e.g., "It's not current practice at our facility for people to even collect discourse samples, so I'm not sure how I'd recruit someone to assess my reliability") as well as that they acknowledge that psychometric data are important (e.g., "We retest over time to test for treatment effects. We rely on the published reliability of the measure for the test–retest stability").

Figure 5. Respondents indicated which discourse outcome measures they extracted based on a variety of factors. Respondents could select more than one option.
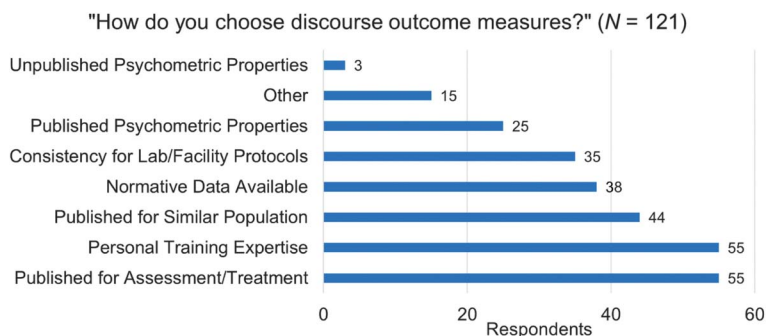

"How do you choose discourse outcome measures?" (*N* = 121)

Category	Respondents
Unpublished Psychometric Properties	3
Other	15
Published Psychometric Properties	25
Consistency for Lab/Facility Protocols	35
Normative Data Available	38
Published for Similar Population	44
Personal Training Expertise	55
Published for Assessment/Treatment	55

Table 4. Number of respondents reporting/endorsing the value of standardized spoken discourse measures in clinical and/or research practice, organized by theme and subtheme.

Themes	Narrative response summary	Number of respondents
Theme 1: Comparison and interpretation	Psychometric properties of discourse data are important for comparing and interpreting discourse measures across individuals and approaches.	45
Theme 2: Best practice	Spoken discourse can be useful for expressing "stable," "reliable, valid and sensitive" measures that are considered "best practice."	25
Theme 3: Clinical or research outcomes	Discourse measures related to clients' goals, post-therapy change, and those that could be applied to "real life" are important for determining clinical outcomes and "gauging treatment effects."	17
Theme 4: Reimbursement	Can be helpful for reimbursement of services	1

Comparison of Responses from Clinical and Research Settings

To compare the extent to which survey responses were driven by primary data collection site (e.g., clinical vs. research), we separated respondents into two groups based on what they selected as their primary setting for collecting discourse. Those collecting discourse at a university-affiliated clinic were considered "research" ($N = 62$), whereas all other primary settings (e.g., acute) were considered "clinical" ($N = 118$). There was a single respondent who did not answer this question, and eight who selected "other." We did not classify those that selected "other" into either group.

When comparing the clinical and research groups (Table 5), age of respondents by setting was not significantly different ($p = .11$), nor was the frequency with which discourse was collected ($p = .25$). Respondents from clinical settings had overall more years of working with persons with aphasia ($M = 17.58$ years) than respondents from research settings ($M = 12.66$ years; $p = .004$).

Primary differences in clinical and research settings were highlighted in the barriers endorsed by each group. Notably, clinical respondents endorsed a higher total number of barriers for each step of discourse analysis (collection, $p = .0003$; transcription, $p = .0006$; analysis, $p = .002$; interpretation, $p = .00005$). They did not significantly endorse a different number of barriers regarding psychometric data collection, compared with the research group ($p = .26$). Close analysis of specific barriers within each step of discourse analysis elucidated which barriers were more often endorsed by the clinical group. For example, under discourse collection, inadequate training and access to tools and resources were barriers that were more frequently endorsed by the clinical group. This was also the case for the negative response "no barriers" (thus meaning there were barriers), suggesting that persons in the clinical group were more likely to experience barriers during discourse collection. Similar patterns were found for transcribing discourse data (specifically, clinical group was more likely to select barriers, and these were inadequate training and access to tools and resources), analyzing discourse data (specifically, clinical group was more likely to select barriers, and these barriers were inadequate training and access to tools and resources), and interpreting discourse data (specifically,

clinical group was more likely to select barriers, and these barriers were time, access to tools and, resources, and, in general, a lower frequency of checking the "no barrier" box). Putting this together with the larger findings of the survey, it is interesting that "time" is only found to be a significantly greater barrier in clinical settings during interpretation of data, whereas respondents from both research and clinical settings are not significantly different in selecting "time" as a barrier during collection, transcription, and analysis.

We then evaluated differences in clinical and research groups regarding the discourse data itself. The clinical group tended to collect fewer samples than the research group (specifically, either one to two samples or three to four samples; $p = .001$). Notably though, the groups did not demonstrate a significant difference in the number of ideal discourse samples collected ($p = .18$), in that both groups preferred to collect more samples. There was a significant difference between respondents who recorded (e.g., audiotaped or videoed) discourse data, in that respondents from the research setting recorded discourse more often ($p = .00001$). Of those who did not record discourse data, there was not a significant difference between groups for whether they transcribed live ($p = .048$) or analyzed live ($p = .095$; note that corrected p value for significance for this comparison was $p < .025$, defined using Bonferroni correction). In general, persons from research settings tended to transcribe ($p = .003$) and code ($p = .00003$) discourse data more often as a part of their work, but the groups did not significantly differ on how often they analyze discourse ($p = .12$). This likely reflects a difference in the choice to transcribe and code rather than to perceptually analyze the discourse.

Respondents from the research setting were more likely to collect psychometric information about the discourse ($p = .00001$) and were more likely to seek out psychometric properties for discourse outcomes ($p = .013$). However, there was not a significant difference in whether a group sought out normative data for discourse outcomes ($p = .05$). We did not identify a significant difference between groups regarding the opinion that there is adequate psychometric data ($p = .74$) or normative data ($p = .51$) for discourse available. Both groups cited that they would be likely to use a normative and/or psychometric properties database if one were made available (i.e., no significant difference in groups, $p = .40$).

Table 5. A comparison of core survey responses (e.g., barriers) between clinical and research settings. Respondents were asked to select which primary setting they largely collected discourse.

Variable	Categories	Clinical (N = 118) M(SD) or frequency	Research (N = 62) M(SD) or frequency	Statistics
Age	Less than 25 years	4	4	Kruskal–Wallis χ^2 = 5.98, df = 3, p = .11
	26–40 years	64	23	
	41–55 years	35	21	
	More than 55 years	15	14	
Years in aphasia	Continuous	17.58(11.35)	M = 12.66(9.66)	W = 2660, p = .004*
How often do you collect discourse?	Always	28	23	χ^2 = 4.07, df = 3, p = .25
	Usually	34	22	
	Sometimes	27	11	
	Rarely	11	3	
Total barriers to collecting discourse data	Sum of categorical selections (yes/no)	1.54(1.15)	1.27(0.75)	Kruskal–Wallis χ^2 = 23.36, df = 5, p = .00028*
Insufficient skills	Yes/No	27/91	4/58	χ^2 = 6.59, df = 1, p = .01
Inadequate training	Yes/No	37/81	5/57	χ^2 = 11.056, df = 1, p = .00088*
Tool & resource access	Yes/No	47/71	9/53	χ^2 = 11, df = 1, p = .00091*
Confidence	Yes/No	11/107	4/58	χ^2 = 0.14, df = 1, p = .71
Protocol interpretation	Yes/No	28/90	17/45	χ^2 = 0.13, df = 1, p = .72
No barriers	Yes/No	14/104	30/32	χ^2 = 27.41, df = 1, p = 1.645e-07*
Total barriers to transcribing discourse data	Sum of categorical selections (yes/no)	1.93(1.53)	1.52(0.94)	Kruskal–Wallis χ^2 = 23.54, df = 6, p = .00064*
Time	Yes/No	80/38	47/15	χ^2 = 0.899, df = 1, p = .34
Insufficient skills	Yes/No	27/91	4/58	χ^2 = 6.59, df = 1, p = .01027
Inadequate training	Yes/No	32/86	5/57	χ^2 = 7.91, df = 1, p = .004924*
Tool & resource access	Yes/No	41/77	7/55	χ^2 = 10.27, df = 1, p = .001355*
Confidence	Yes/No	15/103	4/58	χ^2 = 1.09, df = 1, p = .2966
Protocol interpretation	Yes/No	24/94	15/47	χ^2 = 0.17, df = 1, p = .6847
No barriers	Yes/No	4/114	9/53	Fisher's exact test, p = .01201
Total barriers to analyzing discourse data	Sum of categorical selections (yes/no)	2.08(1.61)	1.63(1.12)	Kruskal–Wallis χ^2 = 20.38, df = 6, p = .0024*
Time	Yes/No	77/41	43/19	χ^2 = 0.151, df = 1, p = .697
Insufficient skills	Yes/No	35/83	10/52	χ^2 = 3.28, df = 1, p = .07
Inadequate training	Yes/No	41/77	8/54	χ^2 = 8.72, df = 1, p = .0032*
Tool & resource access	Yes/No	42/76	7/55	χ^2 = 10.92, df = 1, p = .00095*
Confidence	Yes/No	15/103	7/55	χ^2 = 0.0014, df = 1, p = .97
Protocol interpretation	Yes/No	26/92	16/46	χ^2 = 0.15, df = 1, p = .72
No barriers	Yes/No	3/115	7/55	Fisher's exact test, p = .034
Total barriers to interpreting discourse analysis	Sum of categorical selections (yes/no)	1.68(1.36)	1.23(0.76)	Kruskal–Wallis χ^2 = 27.23, df = 5, p = 5.154e-05*
Time	Yes/No	64/54	16/46	χ^2 = 12.18, df = 1, p = .00048*
Insufficient skills	Yes/No	35/83	8/54	χ^2 = 5.39, df = 1, p = .02
Inadequate training	Yes/No	43/75	10/52	χ^2 = 7.12, df = 1, p = .0076*
Tool & resource access	Yes/No	23/95	3/59	χ^2 = 5.93, df = 1, p = .015
Confidence	Yes/No	20/98	10/52	χ^2 = 4.3672e-30, df = 1, p > .99
No barriers	Yes/No	7/111	24/38	χ^2 = 28.37, df = 1, p = 9.999e-08*
Total barriers to psychometric data collection	Sum of categorical selections (yes/no)	1.29(1.43)	1.74(1.45)	Kruskal–Wallis χ^2 = 5.27, df = 4, p = .26
Time		53/65	36/26	χ^2 = 2.31, df = 1, p = .13
Funds		25/93	25/37	χ^2 = 6.495, df = 1, p = .01*
Personnel		25/93	21/41	χ^2 = 2.80, df = 1, p = .09
Knowledge/training		41/77	24/38	χ^2 = 0.13, df = 1, p = .72

(*table continues*)

Table 5. (Continued).

Variable	Categories	Clinical (N = 118) M(SD) or frequency	Research (N = 62) M(SD) or frequency	Statistics
Number of typical discourse samples collected	1–2 samples	45	14	Kruskal–Wallis χ^2 = 15.52, df = 3, p = .0014*
	3–4 samples	33	20	
	5–6 samples	6	11	
	> 6 samples	4	9	
Number of ideal discourse samples collected				Kruskal–Wallis χ^2 = 3.48, df = 2, p = .176
Do you record the discourse data (e.g., audio, visual)?	Yes/No	56/32	54/0	χ^2 = 23.31, df = 1, p = 1.38e-06*
If you do not record, do you transcribe live?	Yes/No	19/34	0/8	Fisher's exact test, p = .049
If you do not record, do you analyze live?	Yes/No	20/42	0/8	Fisher's exact test, p = .095
How often discourse samples are transcribed	Always	20	26	Kruskal–Wallis χ^2 = 18.19, df = 5, p = .0027*
	Usually	26	17	
	Sometimes	17	9	
	Rarely	17	2	
	Never	3	0	
	NA	2	0	
How often discourse samples are coded (e.g., phonetic coding, error marking)	Always	13	24	Kruskal–Wallis χ^2 = 33.37, df = 5, p = 3.173e-06*
	Usually	14	7	
	Sometimes	13	18	
	Rarely	16	2	
	Never	27	3	
	NA	2	0	
How often collected discourse samples are analyzed	Always	40	29	Kruskal–Wallis χ^2 = 7.33, df = 4, p = 0.12
	Usually	23	13	
	Sometimes	6	9	
	Rarely	7	2	
	Never	5	0	
How often respondents collect psychometric information	Always	1	6	Kruskal–Wallis χ^2 = 33.001, df = 4, p = 1.194e-06*
	Usually	9	17	
	Sometimes	7	11	
	Rarely	8	6	
	Never	42	5	
Do you seek out psychometric properties for discourse outcomes?	Yes	17	20	χ^2 = 8.69, df = 2, p value = .01297*
	Sometimes	16	14	
	No	31	9	

(table continues)

Table 5. *(Continued).*

Variable	Categories	Clinical (N = 118) M(SD) or frequency	Research (N = 62) M(SD) or frequency	Statistics
Do you believe there is adequate psychometric data available for discourse?	Yes/No	12/52	6/36	$\chi^2 = 0.11$, $df = 1$, $p = .74$
Do you seek out normative data for discourse outcomes?	Yes	17	18	$\chi^2 = 5.94$, $df = 2$, $p = .05$
	Sometimes	23	17	
	No	24	7	
Do you believe there is adequate normative data available for discourse?	Yes	3	1	Fisher's exact test, $p = .51$
	Sometimes	27	14	
	No	30	26	
	I don't believe normative data are required	4	1	
Would you use a normative and/or psychometric data database?	Yes/No	59/5	41/1	Fisher's exact test, $p = .399$

Note. If respondents selected "university-based setting or clinic," they were assigned to the "research" setting. If respondents selected any other response, they were assigned to the "clinic" setting. Note that respondents who did not respond ($n = 1$) or who checked "other" ($n = 8$) were not assigned to either group. W = Wilcoxon signed-ranks test statistic.

*Significant after Bonferroni correction. In terms of barriers to collection and interpretation, significant p was $p < .0083$; in terms of barriers to transcription and analysis, significant p was $p < .0071$; in terms of psychometric data collection, significant $p < .0125$; and "if you do not record…," significant $p < .025$. In some cases, due to sample size (respondents < 5), a Fisher's exact test was used.

Discussion

As part of the FOQUSAphasia working group, this study surveyed researchers and clinicians regarding their contemporary practices in spoken discourse assessment in aphasia, with the goal of working toward guiding and establishing standardization procedures first in research settings with subsequent implementation in clinical settings. Briefly, this survey identified considerable heterogeneity in the methods used to collect, analyze, and interpret discourse findings.

Demographics of Participants

An international sample of clinicians and researchers involved in aphasia assessment and rehabilitation across many geographical locations around the world participated in this survey study. Respondents were also heterogeneous regarding their place of practice/data collection, years working in aphasia, age, gender, and terminal professional degree. This sample boasts a unique, more diverse demographic makeup of respondents compared with prior research studies (e.g., geographic span of respondents: Bryant et al., 2017, was primarily focused on Australia and Cruice et al., 2020, was restricted to the United Kingdom), indicating that the participants and responses provide an extension from these previous studies, rather than a replication.

Spoken Discourse Data Collection Procedures

We found that respondents working in university or hospital (i.e., acute care, rehabilitation, hospital-based outpatient clinic) settings were most likely to report always collecting discourse information. However, this finding may reflect a selection bias as respondents who worked in settings outside of universities or hospitals (e.g., community health, long-term care, private practice) were underrepresented in our survey sample. Confirming the findings of Bryant et al. (2017), our survey respondents reported frequently relying on single picture descriptions to elicit spoken discourse samples from persons with aphasia. Stimuli from standardized aphasia assessments (e.g., BDAE and Western Aphasia Battery) and well-established protocols (e.g., Nicholas & Brookshire 1993, AphasiaBank from MacWhinney et al., 2011) were most used during discourse collection. Despite long-standing recommendations to collect and analyze discourse data from multiple genres to obtain a holistic and stable understanding of language use across variable communicative contexts (Armstrong, 2000; Brookshire & Nicholas, 1994; Olness, 2006; Stark, 2019; Stark & Fukuyama, 2020), our survey respondents (41.5%) most commonly collected one to two samples from a given persons with aphasia, with relatively few respondents collecting more than four samples per person. These findings differ from Bryant et al. (2016) who reported use of multiple genres and topics to elicit discourse samples in the aphasia research studies they reviewed.

Interestingly, over 70% of our total sample reported collecting conversation samples with a clinician and/or family member of persons with aphasia. This is a promising finding as it indicates that by sampling discourse within functional communicative interactions, SLPs and researchers are looking beyond impairment-level changes and are also focused on capturing and improving participation, confidence, and quality of life for persons with aphasia (Boyle, 2020; Kagan et al., 2008). However, only a limited number of respondents reported using formal conversation analysis tools, which examine the interactional level. Because this survey was not optimized to query the value and utility of conversational analysis, further investigation of conversational samples and analysis is warranted.

Expanding further on common discourse collection methods, our survey found that respondents typically collected discourse in a variety of environments, ranging from a quiet room to participants' homes. The endorsement by respondents of a variety of environments may reflect differences in work settings, patient needs, and/or research protocols. Additionally, the length of recorded discourse samples ranged between 1 and 5 min, with the time varying based on the type of discourse task. In contrast to Bryant et al. (2016) and Cruice et al. (2020) who noted that only 38.7% (total $N = 123$) and 16%–33% ($N = 211$) of their respondents recorded audio or video samples, respectively, we found that approximately 78% of our total sample recorded discourse data, primarily using audio recordings. Of those who did not record the samples, around 60% frequently transcribed the samples in real time or analyzed the discourse output without transcribing. This latter finding reflects preferences reported by Cruice et al. (2020) who found that 69% and 36% of their respondents favored transcribing in real time or analyzing in real time without transcription, respectively. While real-time analysis may be an efficient means of data analysis, there is little research comparing the accuracy of online versus offline transcription approaches. Qualitatively, some respondents indicated that recording and transcribing spoken discourse samples was "unnecessary" because they were able to evaluate these data and make performance judgments concurrently while collecting samples. For example, one respondent wrote, "I may make hash marks or take other notes of errors or successes versus full transcription." Although we were not able to ascertain the specific perceptual rating scales used by respondents here, it is important to note that there is mounting evidence that perceptual rating can be a useful tool for discourse analysis in aphasia (e.g., Doyle et al., 1996; Webster & Morris, 2019). However, there are relatively few validated perceptual rating scales specific to spoken discourse in aphasia (e.g., Casilio et al., 2019; Kim & Wright, 2020), and use of nonvalidated perceptual rating scales contributes to issues with reproducibility. In general, this speaks to a larger issue highlighted in our survey section regarding the psychometric properties and normative data: Discourse analysis is being employed in a way that may not optimize its utility and may call into question its integrity and quality. We discuss this point in more detail in the section regarding psychometric properties.

Spoken Discourse Data Analysis Procedures

In line with findings from Bryant et al. (2017), our survey results indicated that over 60% of our respondents "usually" or "always" transcribed recorded discourse samples. In contrast, Cruice et al. (2020) found that only 5% of the clinicians in their study reported frequently transcribing language samples. These differences in study findings could relate to the fact that Cruice and colleagues primarily surveyed SLPs, whereas the current survey included both clinical SLPs and researchers. Approximately, 15% of our respondents "rarely" or "never" transcribed samples. There has been a considerable push toward more automated methods of transcription to alleviate the time burden of transcription (e.g., Jacks et al., 2019; Le et al., 2018; Le & Provost, 2016). Indeed, time burden was the most cited barrier to transcription in our study and in Bryant et al. (2017), and time was cited by both clinical and research groups. Notably, the only significant difference for time found between the clinical and research groups was that the clinical group cited a lack of time for interpretation of the findings more often than the research group. Thanks to the insight of a helpful anonymous reviewer, the fact that time barriers were not found to be significantly different between the clinical and the research group may have been because clinicians were found to collect fewer samples, record samples less often, and transcribe/code data less frequently. For this reason, clinicians may be contributing less time overall, which might make them less likely to cite time as a barrier in the categories of data collection, transcription, and analysis. It therefore follows that, if clinicians are expected to use the most robust, evidence-based practices (which are typically identified in the research setting prior to being implemented clinically), time may become a larger barrier for clinicians.

Among those respondents who reported collecting language samples in this study, 41% indicated that they did the transcriptions themselves. Similar to the findings of Cruice et al. (2020), in cases when our respondents did not personally complete transcriptions, trained SLPs or graduate-level research personnel (e.g., students, paid research assistants) were most commonly involved in the transcription process. The discrepancies across respondents highlighted in this survey (e.g., status [undergraduate, graduate, PhD, other]; training [linguistics, speech-language pathology]; availability of protocol) lead to questions regarding experiment fidelity and reproducibility/replicability of studies. It is paramount that authors detail rater demographics and also detail how raters were trained, providing freely available protocols wherever possible. Indeed, transparent, consistent reporting of this type of information is a cornerstone driving the creation of best practices documents and checklists (e.g., EQUATOR network). Without transparent, consistent reporting of this type of information, it is difficult for other studies to replicate or reproduce results because of possible errors at an upstream step (e.g., transcription, coding, analysis). Ideally, transcripts and media files could be contributed and archived in one of the password-protected, shared databases through TalkBank (https://talkbank.org/)

to allow for maximal transparency as well as maximal benefit from the time and effort invested by everyone involved in the data collection and management process. Some open science frameworks have begun requiring investigators to create and make available a video of all procedures utilized (to be hosted in storage and data management repositories such as Databrary). These type of considerations and additions to methods sections of papers will be particularly useful in improving the use of spoken discourse measures and, indeed, all behavioral studies in the field.

Common Barriers to Spoken Discourse Collection, Analysis, and Interpretation

The common barriers to spoken discourse collection identified by respondents included insufficient knowledge, training, and confidence in carrying out discourse collection, in addition to difficulty in implementing and interpreting certain discourse collection protocols. The most common problem was lack of access to tools and resources (e.g., computer hardware/software, recording equipment). The link between this lack of support by trained individuals is clearly aligned with the resource barrier discussed earlier (i.e., a lack of resources for discourse collection/analysis, especially in the clinical group). In general, more respondents from clinical settings were likely to cite at least one barrier to discourse collection, transcription, analysis, and interpretation. The survey conducted by Bryant et al. (2017) focused on discourse use in clinical practice. In contrast, the current survey was not explicitly focused on clinical practice, and indeed, a portion of respondents self-identified as working in a research setting. Although we cannot directly compare our findings with those of Bryant et al. (2017), both survey studies serve to reflect a similar principle: Respondents felt that there were significant barriers to collection and analysis of discourse across a variety of work settings. It is noteworthy that barriers were endorsed by respondents from both clinical and research settings, together suggesting that barriers pose a critical hurdle to overcome in order to increase integration of discourse into clinical and research settings and to ensure that best practices are being used when spoken discourse is being assessed. Given the uniformity of findings across studies relating to the barriers hindering more widespread application of spoken discourse analysis, these barriers appear universal and persistent (Bryant et al., 2017; Cruice et al., 2020).

A barrier raised by our study respondents that has received less attention in the literature is the need for multicultural and multilingual spoken discourse elicitation materials and assessments. In an increasingly culturally and linguistically multifarious patient or research participant pool, a focus on establishing and validating such assessment tools and materials is a needed area of future research, as several respondents raised the lack of such tools as a significant barrier to using discourse in their practice. While there are examples of culturally adapted elicitation stimuli and assessment systems (e.g., Kong & Law, 2009; Pak-Hin &

Law, 2004; Rousseaux et al., 2010), their application to spoken discourse in individuals with aphasia has not been investigated.

A somewhat surprising finding that emerged from the qualitative responses was that some perceived that spoken discourse was not well aligned with function-focused communication goals or outcome measurement needs. This is counter to studies reporting the use of linguistic and interactional discourse for the purpose of measuring functional outcomes (for a review, see Doedens & Meteyard, 2020). These findings highlight the need for more education around spoken discourse in aphasia and align with the identification of "knowledge" as a barrier to implementing spoken discourse in clinical and research practice. The development of best practice guidelines and validation in multicultural or multilingual persons with aphasia may help overcome these barriers, as they can make more explicit the knowledge, training, and resource needs required to implement discourse procedures. Our findings also indicate that there is an appetite for instructions regarding how to adapt discourse stimuli and analysis procedures for a broad spectrum of cultures and languages.

Psychometric Properties of and Normative Data for Spoken Discourse in Aphasia

In this study, respondents made clear the importance of psychometric properties of discourse data in both open-ended and quantitative responses. Themes arising from the responses to open-ended questions included that psychometric properties of discourse data were thought to be important for comparing and interpreting discourse measures across individuals and approaches and that spoken discourse was useful for expressing "stable," "reliable, valid, and sensitive" measures that are considered "best practice." Such themes were also reflected in the quantitative results: Nearly 94% of respondents stated that they would find a database of psychometric properties and/or normative data of discourse outcomes useful (93.6%) while also highlighting inadequate availability of psychometric data and normative data. Interestingly, a surprising number of study participants reported that they don't look for psychometric properties (33%) or normative data (30%). This may reflect not only that it is well known that this literature base is impoverished but also a general thought bias that discourse does not need (or needs less) psychometric validation. However, it is also important to note that respondents' concerns over "availability" extended to difficulties locating information regarding discourse best practices and psychometric properties in the extant literature, stating specifically that they would benefit from having this literature/information consolidated in a way that was more accessible to the field. Our findings highlight the need not only to develop more robust psychometric metrics for spoken discourse variables but also to improve uniformity in reporting (i.e., documenting and dissemination the procedures undertaken in discourse analysis and psychometric information of the selected discourse measures) across studies, the development of a common nomenclature for use in discourse studies, and the need to develop dissemination tools that are accessible to both clinicians and researchers. This may also reflect the numerous and considerable barriers to psychometric data collection and aggregation, which were endorsed by respondents: time, knowledge and training, funds, personnel, and other aspects, such as environment (e.g., not appropriate to do in their work environment) and belief (e.g., unfair to patient, not in patient's best interest).

Psychometric properties are key for reproducibility and data aggregation across studies. Our survey results, which highlight different approaches to rater reliability and collection and use of test–retest data, underscore that consistency and transparency of collecting and reporting psychometric properties in spoken discourse in aphasia remains an issue. Test–retest stability is one of the most important metrics for clinical research and, indeed, should be established for research to be implemented in the clinical setting (Brookshire & Nicholas, 1994; Herbert et al., 2008). For example, short interval sampling (testing and retesting within a short window of time [e.g., 2 weeks]) can determine the variability of a participant's baseline performance. Notably, a measure that varies widely within participants for a short interval *is not stable enough* to be used as a clinically meaningful outcome or assessment measure (Boyle, 2014, 2015). Test–retest stability is paramount in treatment research, particularly given that data acquired during short interval testing periods are prone to practice effects (i.e., participant behavior may improve over testing sessions due to learning the discourse stimuli/procedure vs. the treatment). Stability is of particular concern in persons with aphasia in whom language has long been characterized as highly variable from day to day (Hula & McNeil, 2008; Murray, 1999). When no normative data exist for test–retest across discourse elicitation methods in aphasia, it falls to the researcher to collect this information (but see "Barriers" section) or to look to the literature for standards. However, given that test–retest stability is reported uncommonly in the literature (Pritchard et al., 2017), the direct result of these gaps is a lack of prioritization and dissemination of, and focus on, this psychometric property.

In conclusion, our survey identifies clear gaps and important future directions related to the psychometric properties and normative data of spoken discourse outcome measures. An important step is the aggregation of preexisting psychometric data into a single access port, to overcome issues related to the disparate nature of reporting critical aspects of data collection and analysis that are essential for replication, confidence in the findings, and reproducibility. A second critical step is the creation of, and adherence to, a set of best practice standards, which we highlight in more detail below (see "Future Directions" section). A focus on psychometric properties, and indeed on best practices in general, will overcome some of the challenges inherent to implementation science (moving from research to clinical practice). Surveys such as the one we report here have already been instrumental in pushing for improved clinical justification

of spoken discourse outcomes in aphasia (e.g., Boyle, 2020; Bryant et al., 2017; Cruice et al., 2020).

Study Limitations

There are limitations to this study. We acknowledge selection biases. First, answers to this survey came from those with an interest in discourse and aphasia and those who regularly use discourse sampling and analyses, thus creating a convenience sample. Therefore, the results may not necessarily give a clear representation of the use of spoken discourse or perceived barriers to its use by the whole population of professionals who provide clinical services to persons with aphasia or research aphasia. Second, selection bias is evident from the demographic information collected from participants. Although our survey was distributed to a large number of countries, the majority of respondents were from the United States of America and also based in hospital and university settings. At present, this means that our survey may not be wholly capturing current practice (if many people are not using at all). An extension of this survey will enhance its sample representativeness of those working in discourse and aphasia.

We also recognize that the length of the survey may have contributed to the 58% completion rate. The completion rate for each question is provided in Supplemental Materials S2 and S3; indeed, question completion declines over the course of the survey. The survey was lengthy because it included several questions with branching logic, and we suggest that future surveys that extend on ours may reduce questions to encourage a higher completion rate. Additionally, some survey questions and data considerations may have been more applicable to research rather than clinical settings and vice versa, although we did not note any explicit trends in question answering between the clinical and research groups. The driving factor seemed to be fewer questions answered with time; that is, those at the end of the survey were less likely to be answered, reflecting an issue in the length of the survey or the interest in filling out the section related to psychometric properties and normative data.

This survey made assumptions about respondents' knowledge of certain terminology or used language that was not shared across respondent groups (e.g., those in a purely research setting vs. those in a clinical setting). For example, all terms were not explicitly defined, such as those used when asking respondents how they delineated utterances (e.g., C-units). An example from the qualitative responses that illustrates this limitation is, "Not sure if collecting CIUs (correct information units) and number of complete phrases/sentences counts as 'coding'?" Although the decision not to define all terms was based on the demographic being sampled (i.e., individuals already working on spoken discourse in aphasia), we acknowledge that this choice may have contributed to additional noise in the data.

In the demographics section of the survey, we asked about the primary setting in which respondents collected discourse data, but we did not ask respondents in which

primary role they collected discourse data. For example, a respondent could have been working at a university setting in the roles of a researcher, academic/teacher, and SLP but may only be collecting spoken discourse data for research purposes. This is a limitation that may prevent us from wholly appreciating differences in data collection in clinical versus research roles. We have attempted to address this limitation by conducting post hoc analyses to evaluate results stratified by primary data collection setting, which enables us to speculate on differences in clinical (i.e., acute care, rehabilitation, community health, long-term care facility, private practice, hospital-based outpatient clinic) and research (university research lab or clinic) settings. Next, the survey included respondents who were either previously or currently (at the time of participating in the study) involved in discourse collection and/or analysis. However, we did not specify a time frame for "previously" or "in the past." As a result, it remains unclear whether our findings reflect current rather than old practices or a mix of both. This limitation could have been avoided by specifying a time frame for being involved in discourse analysis "previously" or "in the past" (e.g., within the last 5 years) in the informed consent and demographic information sections of the survey.

Recommendations and Future Research Directions

There have been many "calls to arms" for addressing the spoken discourse evidence issues in the extant aphasia literature, all of which highlighted the benefit of this kind of language sampling (Armstrong, 2000; Boyle, 2011; Dietz & Boyle, 2018; Kintz & Wright, 2017; Linnik et al., 2016; Prins & Bastiaanse, 2004; Pritchard et al., 2018; Wallace et al., 2018). In consideration of this study's findings, we propose some recommendations to improve the state of the spoken discourse evidence in the aphasia literature. First, to address barriers related to training, graduate education and clinician training in spoken discourse analysis must be emphasized. Second, to improve the availability of resources, investigators should make available their study protocol, including all documents used for transcription and coding training and, wherever possible, a video of their training procedures. Additionally, three psychometric properties and normative data need to be established based on larger and internationally diverse samples of spoken discourse outcomes and be made freely available to clinicians and researchers. Third, the aphasia field should focus on improving perceptual analysis and integrating training regarding a variety of transcription and analysis methodologies (e.g., automatic transcription techniques) to combat commonly endorsed barriers related to time in clinical and research settings. Finally, an adherence to "best practice" living documents should be advocated, wherein reviewers of papers and investigators assure that all necessary components for procedure reproducibility are reported.

To begin addressing these recommendations, FOQUSAphasia (http://www.foqusaphasia.com) includes a Best Practices task force, whose first initiative is to create a living, best practices document. Its second task force,

Methodological & Data Quality, is pursuing an initiative to collect a large database of test–retest data using the AphasiaBank protocol. This database will be made available on AphasiaBank and will be critical for outlining the psychometric properties of commonly used discourse metrics and for building a normative sample. Likewise, findings from this study can be used to guide development of process standardization in spoken discourse and the creation of a psychometric and normative property database. Presently, members of the Best Practices task force of FOQUSAphasia are conducting an e-Delphi study to gather expert consensus for best practices in this field.

Acknowledgments

References

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2020). A unified model of post-stroke language deficits including discourse production and their neural correlates. *Brain, 143*(5), 1541–1554. https://doi.org/10.1093/brain/awaa074

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology, 14*(9), 875–892. https://doi.org/10.1080/02687030050127685

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

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International, 5*(9/10), 341–347.

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

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

Boyle, M. (2015). Stability of word-retrieval errors with the aphasia bank stimuli. *American Journal of Speech-Language Pathology, 24*(4), S953–S960. https://doi.org/10.1044/2015_AJSLP-14-0152

Boyle, M. (2020). Choosing discourse outcome measures to assess clinical change. *Seminars in Speech and Language, 41*(1), 1–9. https://doi.org/10.1055/s-0039-3401029

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 2016*(6), CD000425. https://doi.org/10.1002/14651858.CD000425.pub4

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and test–retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research, 37*(2), 399–407. https://doi.org/10.1044/jshr.3702.399

Brown, C., Snodgrass, T., & Covington, M. (2007). *Computerized Propositional Idea Density Rater 3 (CPIDR 3)*. CASPR Project, Artificial Intelligence Center.

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

Cahana-Amitay, D., & Jenkins, T. (2018). Working memory and discourse production in people with aphasia. *Journal of Neurolinguistics, 48,* 90–103. https://doi.org/10.1016/j.jneuroling.2018.04.007

Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology, 28*(2), 550–568. https://doi.org/10.1044/2018_AJSLP-18-0192

Cherney, L. R., Coelho, C. A., & Shadden, B. B. (1998). *Analyzing discourse in communicatively impaired adults.* Aspen Publishers.

Covington, M. A., Riedel, W. J., Brown, C., He, C., Morris, E., Weinstein, S., Semple, J., & Brown, J. (2007). Does ketamine mimic aspects of schizophrenic speech? *Journal of Psychopharmacology, 21*(3), 338–346. https://doi.org/10.1177/0269881107077729

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

Damico, J. S., Oelschlaeger, M., & Simmons-Mackie, N. (1999). Qualitative methods in aphasia research: Conversation analysis. *Aphasiology, 13*(9–11), 667–679. https://doi.org/10.1080/026870399401777

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

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

Doyle, P. J., Tsironas, D., Goda, A. J., & Kalinyak, M. (1996). The relationship between objective measures and listeners' judgments of the communicative informativeness of the connected discourse of adults with aphasia. *American Journal of Speech-Language Pathology, 5*(3), 53–60. https://doi.org/10.1044/1058-0360.0503.53

Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research, 6*(3), e34. https://doi.org/10.2196/jmir.6.3.e34

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25*(11), 1414–1430. https://doi.org/10.1080/02687038.2011.603898

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology, 25*(10), 1261–1278. https://doi.org/10.1007/s11103-011-9767-z

Fromm, D., Forbes, M., Holland, A., Dalton, S. G., Jessica, R., & MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech-Language Pathology, 26*(3), 762–768. https://doi.org/10.1044/2016_AJSLP-16-0071

Galski, T., Tompkins, C., & Johnston, M. V. (1998). Competence in discourse as a measure of social integration and quality of life in persons with traumatic brain injury. *Brain Injury, 12*(9), 769–782.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination* (3rd ed.). Pearson.

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95,* 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

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

Hula, W. D., & McNeil, M. R. (2008). Models of attention and dual-task performance as explanatory constructs in aphasia. *Seminars in Speech and Language, 29*(3), 169–187. https://doi.org/10.1055/s-0028-1082882

IBM Corporation. (2019). *IBM SPSS Statistics for Windows Version 26*.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jacks, A., Haley, K. L., Bishop, G., & Harmon, T. G. (2019). Automated speech recognition in adult stroke survivors: Comparing human and computer transcriptions. *Folia Phoniatrica et Logopaedica, 71*(5–6), 286–296. https://doi.org/10.1159/000499156

Kagan, A., Simmons-Mackie, N., Rowland, A., Huijbregts, M., Shumway, E., McEwen, S., Threats, T., & 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

Kertesz, A. (2007). *Western Aphasia Battery–Revised*. The Psychological Corporation.

Kim, H., & Wright, H. H. (2020). Concurrent validity and reliability of the core lexicon measure as a measure of word retrieval ability in aphasia narratives. *American Journal of Speech-Language Pathology, 29*(1), 101–110. https://doi.org/10.1044/2019_AJSLP-19-0063

Kintz, S., & Wright, H. H. (2017). Discourse measurement in aphasia research. *Aphasiology, 32*(4), 472–474. https://doi.org/10.1080/02687038.2017.1398807

Kong, A. P. H., & Law, S. P. (2009). A linguistic communication measure for monitoring changes in Chinese aphasic narrative production. *Clinical Linguistics & Phonetics, 23*(4), 255–269. https://doi.org/10.1080/02699200802673234

Le, D., Licata, K., & Mower Provost, E. (2018). Automatic quantitative analysis of spontaneous aphasic speech. *Speech Communication, 100,* 1–12. https://doi.org/10.1016/j.specom.2018.04.001

Le, D., & Provost, E. M. (2016). Improving automatic recognition of aphasic speech with AphasiaBank. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH,* 2681–2685. https://doi.org/10.21437/Interspeech.2016-213

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

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

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

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

Miller, J., & Chapman, R. (1983). *SALT: Systematic Analysis of Language Transcripts, User's Manual*. University of Wisconsin Press.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340,* c869. https://doi.org/10.1136/bmj.c869

Murray, L. L. (1999). Review attention and aphasia: Theory, research and clinical implications. *Aphasiology, 13*(2), 91–111. https://doi.org/10.1080/026870399402226

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

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M. P., Poldrack, R. A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D. C., White, T., & Yeo, B. T. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience, 20*(3), 299–303. https://doi.org/10.1038/nn.4500

Olness, G. S. (2006). Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology, 20*(2–4), 175–187. https://doi.org/10.1080/02687030500472710

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Osborne, F., Hickin, J., Best, W., & Howard, D. (1998). Treating word-finding difficulties - Beyond picture naming. *International Journal of Language & Communication Disorders, 33*(Suppl. 1), 208–213. https://doi.org/10.3109/13682829809179424

Pak-Hin, A. K., & Law, S.-P. (2004). A Cantonese linguistic communication measure for evaluating aphasic narrative production: Normative and preliminary aphasic data. *Journal of Multilingual Communication Disorders, 2*(2), 124–146. https://doi.org/10.1080/14769670310001659708

Prins, R., & Bastiaanse, R. (2004). Review. *Aphasiology, 18*(12), 1075–1091. https://doi.org/10.1080/02687030444000534

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

Prutting, C. A., & Kirchner, D. M. (1987). A clinical appraisal of the pragmatic aspects of language. *Journal of Speech and Hearing Disorders, 52*(2), 105–119. https://doi.org/10.1044/jshd.5202.105

Roberts, A., & Orange, J. (2013). Discourse in Lewy body spectrum disorder. In B. Davis & J. Guendouzi (Eds.), *Pragmatics in dementia discourse* (pp. 153–211). Cambridge Scholars Publishing.

Roberts, A., & Post, D. (2018). Information content and efficiency in the spoken discourse of individuals with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 61*(9), 2259–2274. https://doi.org/10.1044/2018_JSLHR-L-17-0338

Rousseaux, M., Daveluy, W., & Kozlowski, O. (2010). Communication in conversation in stroke patients. *Journal of Neurology, 257*(7), 1099–1107. https://doi.org/10.1007/s00415-010-5469-8

Salis, C., Murray, L., & Bakas, K. (2018). An international survey of assessment practices for short-term and working memory deficits in aphasia. *American Journal of Speech-Language Pathology, 27*(2), 574–591. https://doi.org/10.1044/2017_AJSLP-17-0057

Shadden, B. (1998). Obtaining the discourse sample. In L. Cherney, B. Shadden, & C. Coelho (Eds.), *Analyzing Discourse in Communicatively Impaired Adults* (pp. 9–34). Aspen Publishers.

Shadden, B. B., Burnette, R. B., Eikenberry, B. R., & DiBrezzo, R. (1991). All discourse tasks are not created equal. *Clinical Aphasiology Conference*. Pro-Ed. http://aphasiology.pitt.edu/153/

Sim, P., Power, E., & Togher, L. (2013). Describing conversations between individuals with traumatic brain injury (TBI) and communication partners following communication partner training: Using exchange structure analysis. *Brain Injury, 27*(6), 717–742. https://doi.org/10.3109/02699052.2013.775485

Simmons-Mackie, N., Worrall, L., Murray, L. L., Enderby, P., Rose, M. L., Paek, E. J., & Klippi, A. (2017). The top ten: Best practice recommendations for aphasia. *Aphasiology, 31*(2), 131–151. https://doi.org/10.1080/02687038.2016.1180662

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

Stark, B. C., Dutta, M., Murray, L., Bryant, L., Fromm, D., MacWhinney, B., Ramage, A., Simmons-Mackie, N., Roberts, A., Fergadiotis, G., den Ouden, D.-B., Brock, K., Harmon, T., Paek, E. J., Themistocleous, C., MicKinney-Bock, K., Yoo, H., Yoon, S. O., Sharma, S., . . . Gutierrez, S. (2020). Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. *American Journal of Speech-Language*

Pathology, 30(1S), 491–502. https://doi.org/10.1044/2020_AJSLP-19-00093

Stark, B. C., & Fukuyama, J. (2020). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Cognition and Neuroscience, 36*(5), 562–585. https://doi.org/10.1080/23273798.2020.1862258

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

Thompson, C. K. (n.d.). *Northwestern Narrative Language Analysis (NNLA) theory and methodology*. https://anr.northwestern.edu/wp-content/uploads/2016/09/NNLA_REVISED.Fall2015.pdf

Ulatowska, H. K., Allard, L., Reyes, B. A., Ford, J., & Chapman, S. (1992). Conversational discourse in aphasia. *Aphasiology, 6*(3), 325–330. https://doi.org/10.1080/02687039208248602

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

Wallace, S. J., Worrall, L. E., Rose, T., & Le Dorze, G. (2018). Discourse measurement in aphasia research: Have we reached the tipping point? A core outcome set … or greater standardisation of discourse measures? *Aphasiology, 32*(4), 479–482. https://doi.org/10.1080/02687038.2017.1398811

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

Webster, J., & Morris, J. (2019). Communicative informativeness in aphasia: Investigating the relationship between linguistic and perceptual measures. *American Journal of Speech-Language Pathology, 28*(3), 1115–1126. https://doi.org/10.1044/2019_AJSLP-18-0256

Whitworth, A., Leitao, S., Cartwright, J., Webster, J., Hankey, G. J., Zach, J., Howard, D., & Wolz, V. (2015). NARNIA: A new twist to an old tale. A pilot RCT to evaluate a multilevel approach to improving discourse in aphasia. *Aphasiology, 29*(11), 1345–1382. https://doi.org/10.1080/02687038.2015.1081143

Wilkinson, R. (2014). Intervening with conversation analysis in speech and language therapy: Improving aphasic conversation. *Research on Language and Social Interaction, 47*(3), 219–238. https://doi.org/10.1080/08351813.2014.925659

Wright, H. H., & Capilouto, G. J. (2009). Manipulating task instructions to change narrative discourse performance. *Aphasiology, 23*(10), 1295–1308. https://doi.org/10.1080/02687030902826844