

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

1. Brielle C. Stark, PhD Indiana University Bloomington
Department of Speech, Language, and Hearing Sciences
Program in Neuroscience
Cognitive Science Program
Bloomington, IN USA

2. Manaswita Dutta, PhD., Rush University
CCC-SLP Department of Communication Disorders and Sciences
Chicago, IL USA

3. Laura L. Murray, PhD Western University
School of Communication Sciences and Disorders
London, ON, CA

4. Davida Fromm, PhD Carnegie Mellon University
Department of Psychology
Pittsburgh, PA USA

5. Lucy Bryant, PhD University of Technology Sydney
Graduate School of Health
Sydney, NSW, AUS

6. Tyson G. Harmon, PhD, Brigham Young University
CCC-SLP Department of Communication Disorders
Provo, UT USA 84605

7. Amy E. Ramage, PhD University of New Hampshire
Department of Communication Sciences and Disorders
Durham, NH USA 03824

8. Angela C. Roberts, PhD Northwestern University
Pepper Department of Communication Sciences and Disorders
Evanston, IL USA 60208

Corresponding Author: Brielle Stark, bcstark@iu.edu, 770-548-7121, 200 S Jordan Avenue Bloomington IN 47405

Conflict of Interest: No authors have a conflict of interest.

Funding: American Speech Hearing (ASH) Foundation New Investigator Award (Stark), 1R21DC017255-01 (Roberts)

Abstract

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 non-mandatory 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 the current study will be used to guide development of process standardization in spoken discourse, and for the creation of a psychometric and normative property database.

Key words: Discourse, aphasia, “spontaneous speech,” “connected speech,” “neurogenic communication disorder”

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Introduction

Assessment of Spoken Discourse in Aphasia

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., 2020; Wilkinson, 2014) properties of language and communication. Conducting multi-level 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 semi-structured 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, 1994b; 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 et al., 2011; Fergadiotis & Wright, 2011; Stark, 2019; Ulatowska et al., 1981; Wright & Capilouto, 2009). To assess language constructed interactionally between two or more individuals, and also 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 (Elizabeth 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).

51 Methodological Issues Hampering Spoken Discourse Evidence

52 Despite its relevance for understanding the functional and pragmatic communication abilities of persons
53 with aphasia, and its potential to serve as a primary and important outcome measure, spoken discourse outcomes
54 have been excluded from core outcome sets in aphasia (an agreed, minimum set of outcomes for treatment work)
55 (Brady et al., 2016; Wallace et al., 2019). A key reason why outcome sets currently exclude spoken discourse
56 outcomes is due to a lack of standardization across data collection, analysis, and reporting. For the purposes of this
57 study, we defined reporting as the explicit documentation and communication of information regarding how
58 samples were recorded, inter-rater reliability, and other measures essential to replication, confidence, and
59 reproducibility of the data. Another limitation in using discourse measures in outcome sets is that studies often
60 report insufficient details around how language samples were collected and analyzed, which creates replication and
61 reproducibility challenges. With few exceptions (e.g., Leaman & Edmonds, 2019; Roberts & Post, 2018),
62 researchers have rarely (a) reported on how the raters responsible for transcribing, coding, and/or analyzing
63 language samples were trained, or (b) made rater training documents freely available (e.g., see Thompson, n.d.).
64 Furthermore, in addition to a large number of spoken discourse measures being reported in the aphasia literature,
65 only infrequently (e.g., Boyle, 2014, 2015; Brookshire & Nicholas, 1994; Pritchard et al., 2018) have researchers
66 intentionally studied the psychometric properties of specific discourse measures or developed robust normative data
67 (Dietz & Boyle, 2018; Pritchard et al., 2017; Wallace et al., 2018). These issues preclude comparison/meta-analyses
68 of discourse data and hinder the replication and reproducibility of findings, both in research and clinical settings.
69 As such, there needs to be a concerted effort to standardize assessment, analysis, and reporting procedures in the
70 field. Establishing and transparently reporting psychometric properties of discourse are necessary to instill
71 confidence in users (e.g., clinicians, researchers) that the task and outcomes are reliable and valid for meaningful
72 decision making.

73 Given the persistent heterogeneity in spoken discourse measurement and analysis procedures, their clinical
74 utility in aphasia rehabilitation remains limited. Recent surveys have shown that although speech-language
75 pathologists (SLPs) recognize the importance of targeting discourse in aphasia assessment and treatment, a large
76 majority of them experience a variety of barriers in efficiently implementing spoken discourse analysis in their
77 clinical practice. For instance, Bryant et al. (2017) surveyed 123 clinicians to better understand the contemporary
78 uses of discourse analysis in clinical settings. Data were collected from five English-speaking countries with the
79 majority of respondents being from Australia. The survey asked the SLPs about their discourse analysis practices
80 in terms of frequency of use, collection, transcription, and analysis methods; perceptions and attitudes of SLPs

81 regarding the use of discourse analysis in clinic were also queried. Over 85% of their study respondents indicated
82 that they used a variety of methods to examine discourse productions of persons with aphasia, but perceptual,
83 judgment-based approaches were most commonly used. The respondents noted that time to transcribe and analyze
84 discourse samples was the most significant barrier, followed by other factors such as lack of adequate training and
85 access to tools.

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

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

110 across different settings and form the basis towards establishing standards for discourse analysis to improve its
111 transparency, replicability, and clinical utility.

112 **Spoken Discourse Reporting Guidelines**

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

132 Like the Committee on Best Practice in Data Analysis and Sharing (COBIDAS), created by the
133 Organization for Human Brain Mapping (Nichols et al., 2017), the objective of creating and maintaining reporting
134 standards for spoken discourse in aphasia is to identify practices that maximize analytical stability and
135 generalizability of study findings. Given the growth in spoken discourse research in aphasia across the past few
136 decades (highlighted well in Bryant et al., 2016), we propose that the creation of reporting standards will: “(a)
137 encourage replication of studies; (b) ensure consistent reporting across studies; (c) recommend appropriate
138 statistical modeling, thereby ensuring the most appropriate statistical inferences; and, (d) overall, contribute to a
139 more homogeneous, rigorous and standardized process by which spoken discourse research is evaluated and

140 ultimately disseminated for clinical use” (p. 6; Stark et al., 2020). Reporting standards (and indeed, adherence to
141 set standards) will enable meta-analytic consolidation of evidence and more importantly, will have downstream,
142 direct clinical implications by improving practices for collecting, analyzing, and accurately interpreting initial
143 presenting status as well as changes in spoken discourse outcomes in aphasia.

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

153 **Goals of The Current Study**

154 The current survey was completed as part of the Best Practices Task Force within the FOQUSAphasia
155 working group (Stark et al., 2020). The broad goal of the study was to survey current researchers and clinicians as
156 a first step toward creating recommendations for field-wide standards in methods, analysis, and reporting of spoken
157 discourse outcomes, as has been done across other related disciplines (Nichols et al., 2017; Simmons-Mackie et al.,
158 2017; Wallace et al., 2019). We used a mixed-methods survey to examine the current practices in spoken discourse
159 collection, analysis, and interpretation undertaken by researchers and clinicians involved in aphasia assessment and
160 rehabilitation across the globe. Albeit a prior survey provided a foundation to understand discourse use in clinical
161 practice in aphasia (Bryant et al., 2017), the current survey focused on extending prior findings to a broader
162 (research and clinical) audience to work toward standardization of discourse reporting in both clinical and research
163 practice. That is, we created an extended survey to collect more detail regarding the current methods used for spoken
164 discourse data collection, analysis, and interpretation in both research and clinical contexts. Such information is
165 needed as a first step in creating guidelines because it can not only contribute further empirical rationale for the
166 need for such guidelines (e.g., documenting heterogeneity in the methods used), but also inform our understanding
167 of barriers that must be considered when developing such guidelines. Accordingly, the specific aims of our survey
168 study were to: (1) target concepts previously evaluated in prior related surveys (e.g., Bryant et al., 2017; Cruice et
169 al., 2020), thus expanding on previous surveys in a different sample of respondents (i.e., those working in research

170 and/or clinical settings), and (2) extend findings related specifically to data collection and analysis, with an emphasis
171 on the psychometric properties of spoken discourse outcomes. To do so, we focused on the following research
172 questions:

- 173 1) What are the current practices employed by clinicians and researchers using spoken discourse in their work
174 relative to discourse sample collection, analysis approaches, and consideration of data reliability and validity?
- 175 2) What are the barriers faced by clinicians and researchers in using spoken discourse in their work relative to
176 discourse sample collection, analysis, and data reliability and validity?

177 **Method**

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

181 **Participants**

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

189 **Survey Design and Procedure**

190 The survey design and data collection were completed using REDCap, secure web-based data management
191 application (Harris et al., 2009, 2019). In the first iteration, the first and second authors (BCS and MD) composed
192 a set of items, built the survey, and designed the order of the items within the survey. The third, fourth, and fifth
193 authors (LLM, DF, and LB) reviewed the initial draft and suggested revisions to the survey questions as well as the
194 order of items. Authors who contributed to designing the survey are certified speech-language pathologists who
195 have worked in clinical settings with the aphasia population (MD, LLM, DF, LB) and who have conducted prior
196 survey studies (LLM, LB; e.g., Bryant et al., 2017; Salis et al., 2018) and/or prior aphasia research focused on or
197 utilizing spoken discourse assessment (BCS, MD, LLM, DF, LB; e.g., (Fromm et al., 2017; Stark, 2019). Following
198 this, the survey was piloted by the remainder of the authors along with select researchers and clinicians at Indiana
199 University (N = 5 pilot participants, who were SLPs or communication sciences and disorders researchers reflecting

200 the survey's target audience). Based on feedback received during piloting, some questions were revised and
 201 rechecked by the first five authors prior to distribution of the survey (via distribution methods discussed above).
 202 The presentation of survey questions was kept the same for all respondents (i.e., questions were not presented in a
 203 random order). Note that safety measures were used in REDCap to prevent the same respondent taking the survey
 204 on multiple occasions. Mobile compatibility settings were also used to enable potential respondents to fill out the
 205 survey on a variety of devices (e.g., desktop, laptop, tablet, phone).

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

- 208 (1) Demographic information of participants (7 quantitative; 7 qualitative items), discussing
 209 respondent's geographic location, age, years of working with persons with aphasia, education, and
 210 role;
- 211 (2) Spoken discourse measurement in aphasia (8 quantitative; 8 qualitative items), determining the
 212 extent to which respondents measured spoken discourse in aphasia, their reasons for doing so, and
 213 their barriers to discourse collection, transcription, analysis, and interpretation;
- 214 (3) Data collection procedures (9 quantitative, 14 qualitative items), evaluating specific discourse data
 215 collection procedures, which expanded upon items asked in section 2 (Spoken discourse
 216 measurement in aphasia);
- 217 (4) Data analysis procedures (30 quantitative, 15 qualitative items), examining transcription, coding,
 218 and analysis of spoken discourse data, including information regarding reliability analyses; and,
- 219 (5) Psychometric properties and normative data (13 quantitative, 15 qualitative items), understanding
 220 common practices in and opinions about psychometric properties of discourse-derived outcomes
 221 (e.g., validity, stability, reliability) and the potential for normative data in discourse work.

222 Each section of the survey began with a short explanation of questions to be answered within that section.
 223 Question formats included multiple-choice, fill in the blank, yes/no, and open-ended response options. Note that
 224 some questions allowed for multiple answers (e.g., one could be both an SLP and a researcher). While most items
 225 followed a forced-response format, answering all questions was not mandatory. Consequently, whilst N=189
 226 completed the survey, some questions were answered by fewer than 189 respondents; we note these instances in
 227 our Results section. The survey employed branching logic, such that for several items, respondents were shown an
 228 additional question if they answered in a certain way (e.g., if "yes" a follow-up question appeared). The number of
 229 items per page varied, as might be expected given that we created five sections of questions (described above); the

230 maximum number of questions per page was 11. In total, the survey was 23 pages in length. A completeness check
231 was not instituted by REDCap or the survey authors. There were 1986 page views from respondents (and potential
232 respondents), including creating, updating, and viewing the survey on REDCap.

233 See Appendix for the full survey.

234 **Distribution of the Survey**

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

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

250 Potential respondents were told in the informed consent information section that the survey would take 30-
251 40 minutes to complete, and that it could be completed in more than one sitting. To continue completing the survey
252 at a future time, they entered their email address, and the survey generated a password for re-entry and access to
253 their saved survey responses at a later date. The email addresses of the respondents were not saved by REDCap
254 (i.e., the authors of this study could not see these email addresses). Respondents were allowed to change answers
255 to their questions at any time during the survey period. Respondents were requested to click 'Complete' when they
256 had fully completed the survey, or REDCap would automatically select this option if all quantitative and qualitative
257 sections had been completed. No identifying information was collected from participants during the completion of
258 the survey. The survey was distributed in English from September to November 2019, then closed for response
259 analysis. No incentives were given as a part of this survey.

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

264 **Data Analysis**

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

271 Qualitative analysis was completed on thirty-five open-ended survey questions that had a response rate of
272 20% or higher (median = 73%, range = 21-100%). This criterion excluded five open-ended survey questions from
273 the analysis (see Supplemental Material for response rates by item) that had extremely low response rates (i.e., high
274 non-response bias). Using a thematic analysis approach, informed by Braun and Clarke (2006), participant
275 responses were coded independently by the final three authors (TGH, AER, ACR) using a 5-step iterative process.
276 First, all three coders familiarized themselves with the dataset by reading through the open-ended participant
277 responses while reflecting on the data and taking notes. Second, the three coders independently assigned possible
278 codes to the responses using an inductive coding approach. Third, the three coders met to collaboratively discuss
279 their independently assigned codes and to draft and revise a codebook. During this process, coders iteratively
280 convened group discussion followed by independent review of a portion of the data until the codebook was deemed
281 appropriate for the dataset (i.e., three times). Fourth, the coders returned to the data and independently (i.e., masked
282 to the other raters' data) recoded responses based on the final codebook (see Tables 2 - 4 for themes and codes
283 used). Coding consensus was defined as responses where at least 2 of 3 raters independently assigned the same code
284 to a survey response. Discrepancies, responses for which coding consensus was not achieved, occurred on 40
285 (5.56%) survey responses. In the final step of the analysis, these discrepancies were discussed by the three raters
286 collectively until coding consensus agreement was achieved. The qualitative coders were blinded to the quantitative
287 survey responses and to the quantitative question prompts while coding and extracting themes from the open-ended
288 responses to minimize extracting themes that would have been biased by the summary quantitative question
289 statistics.

290

Results

291

Responses

292

293

294

295

296

297

298

299

300

301

302

303

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 non-completers. 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 the Supplemental Material. 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

304

305

306

307

308

309

310

311

312

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.

313

TABLE 1 HERE

314

Survey Section 2: Spoken Discourse Measurement in Aphasia

315

316

317

318

319

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

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

320 symptoms for clinical intervention purposes (72.1%); as an outcome measure for aphasia treatment in clinical
321 practice (53.9%); and/or for research (31.5%). There was no significant association between years working with
322 persons with aphasia and how often respondents collected spoken discourse (χ^2 [df = 140, N = 187] = 141.57, $p =$
323 .45), nor between age of respondents and how often respondents collected spoken discourse (χ^2 [df = 12, N = 189]
324 = 4.36, $p = .98$). Further, there was a non-significant association between primary work setting and how often
325 respondents collected spoken discourse (χ^2 [df = 32, N = 189] = 46.12, $p = .051$), though university and hospital
326 settings were the most common settings to ‘always’ collect discourse data.

327 The most commonly collected spoken discourse genre was a description of a single picture (e.g., Cookie
328 Theft picture from the *Boston Diagnostic Aphasia Examination* [BDAE], Goodglass et al., 2000) (89.1%), followed
329 by a conversation with a clinician and/or family member (70.9%) (N = 165). Other typically collected genres
330 included a personal recount (e.g., important life event, 67.9%), procedural narrative (e.g., how to make a peanut
331 butter and jelly sandwich, 57%) and an interview (biographical or otherwise, conducted by a clinician, 51.5%). To
332 collect spoken discourse data (N = 163), respondents reported relying most on standardized aphasia assessment
333 tools that included discourse generation tasks (e.g., *Western Aphasia Battery-Revised*, Kertesz, 2007) (74.8%), but
334 some also endorsed using protocols such as the Nicholas and Brookshire protocol (31.3%, Nicholas & Brookshire,
335 1993), the AphasiaBank protocol (15.3%, MacWhinney et al., 2011), or a self-developed protocol (20.2%).
336 Qualitative responses elaborated on the quantitative findings discussed above, with respondents expanding on
337 specific practices related to conversation elicitation (including both unscripted conversation interactions collected
338 through naturalistic tasks and formal scripted exchanges elicited using interview guides, questions, and barrier-style
339 tasks; these practices also included single-partner and group conversations), retellings or recounts (recounted
340 content from videos, wordless picture books, and current events), and narrative descriptions from visual information
341 (single, sequence, or wordless books) (see Table 2, theme one).

342 In terms of the number of samples (i.e., discourse tasks) typically collected and/or analyzed per individual
343 with aphasia, respondents (N = 147) most often collected 1-2 samples (41.5%) or 3-4 samples (38.1%), with
344 relatively few collecting 5-6 samples (11.6%) or more than six samples (8.8%). Additionally, respondents
345 mentioned that ideally, they would like to collect 3-4 samples (41.5%) or 5-6 samples (24.5%); few cited 1-2
346 samples (17.7%) or more than six samples (16.3%) as an ideal number.

347 To indicate the typical barriers to discourse sample collection, transcription, analysis, and interpretation,
348 respondents could choose more than one barrier per section (e.g., collection, transcription) (see Figure 1). The most
349 commonly selected barriers to discourse collection (N = 162) included: lack of access to tools and resources (e.g.,

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

350 computer hardware/software, recording equipment) (34.6%); inadequate training in discourse collection (25.9%);
351 and insufficient skills and/or knowledge in discourse collection (19.1%). Notably, 29.6% indicated no barriers to
352 discourse collection. Regarding discourse transcription (N = 162), a major barrier was lack of time/time constraints
353 (80.2%), with only 9.9% reporting no barriers. For discourse analysis (N = 161), respondents overwhelmingly
354 endorsed lack of time/time constraints as a major barrier (75.8%), with only 6.8% indicating no barriers. Finally,
355 the major barriers to discourse analysis interpretation (N = 161) included time constraints (50.9%), lack of training
356 (33.5%), and lack of skills and knowledge (26.7%). A small proportion of respondents (20.5%) cited no barriers to
357 discourse analysis interpretation.

358 FIGURE 1 HERE

359 TABLE 2 HERE

360 Open-ended question responses provided further and clarifying information regarding perceived barriers to
361 use of discourse assessment in aphasia (see Table 3). Responses were grouped by the following themes: resource
362 related barriers, clinician/researcher related barriers, patient/participant related barriers, and measurement related
363 barriers. Within resource related barriers, different aspects of time were cited as issues (e.g., lacking “time to devote
364 to self-training,” “it takes a long time to train [others],” general feeling of lack of time), as well as limits in personnel
365 and environment. Responses that highlighted clinician/researcher related barriers included perceived misalignment
366 with high priority outcomes (e.g., “[not] relevant to dissect a client’s verbal output—I prefer to look at the big
367 picture [and their] primary goals”), lack of training or knowledge, a belief that discourse analysis was not related to
368 their job position (e.g., “not my job”) and/or that discourse analysis was not pursued because of a lack of
369 professional interest, and historic or current practice patterns. Barriers were also endorsed by respondents in relation
370 to the patient/participant, including the severity of impairment (e.g., most severe language impairments, especially
371 those with concomitant motor speech disorders, may preclude usefulness or meaningfulness of discourse [“task will
372 be too challenging for the client if more severe,” “too little understandable speech to warrant an informative
373 analysis”]) and the perceived burden on the patient/participant of collecting discourse samples. Finally, a number
374 of measurement-related barriers were endorsed by respondents, including a general lack of psychometric data and
375 shared processes specific to discourse (e.g., lack of standard practice, lack of psychometric data, variability in
376 discourse collection and analysis methods [“[discourse measures ultimately] depend on the individuals doing the
377 assessments and how they are trained”]), lack of linguistic and culturally specific discourse methods/data (e.g., lack
378 of protocols, normative data, and other psychometric properties in languages other than English and in cultures
379 where monolingual speakers are not the majority), and lack of or difficulty finding empirical evidence.

380

TABLE 3 HERE

381

Survey Section 3: Data Collection Procedures

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

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%), 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%). Further, 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 on-line 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 on-line, only owing to the rater's skill or the low frequency of errors (see Table 2).

397

398

399

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

400

Survey Section 4: Data Analysis Procedures

401

402

403

404

405

406

407

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).

FIGURE 2 HERE

408 **Survey Sub-section 4.a. Transcription**

409 In terms of the frequency of transcriptions, respondents (N = 144) indicated *always* (33.3%), *usually*
 410 (31.3%), *sometimes* (18.8%), *rarely* (11.8%) or *never* (3.5%) undertaking sample transcriptions. When asked if
 411 samples were typically personally transcribed (i.e., by the person collecting the discourse data), respondents (N =
 412 144) answered as follows: *always* (41%), *usually* (11.1%), *sometimes* (18.1%), *rarely* (18.8%), *never* (6.9%).
 413 Respondents who indicated that they did not always personally transcribe samples (N = 79) reported that the
 414 following personnel were involved in transcriptions: a graduate student volunteer [unpaid] (31.6%), a
 415 clinician/speech-language pathologist (30.4%), a paid graduate-level research assistant (29.1%), a paid research
 416 assistant or lab manager [not a student] (26.6%), other (24.1%), a paid undergraduate research assistant (19%), a
 417 PhD student whose work involves the data collected (16.5%), an undergraduate student volunteer [unpaid] (24.1%)
 418 and/or a researcher with a PhD in a related field (8.9%). Open responses elaborated on these data (see Table 2,
 419 theme two). Respondents who did not transcribe data cited that they perceived themselves as having sufficient
 420 ability to detect features of interest on-line (i.e., in real time) or to use perceptual rating scales. Those who completed
 421 partial transcription described orthographically transcribing part of the sample verbatim or noting and transcribing
 422 errors only.

423 **Survey Sub-section 4.b. Coding**

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

427 Respondents (N = 143) indicated that they *always* (29.4%), *usually* (12.6%), *sometimes* (21%), *rarely*
 428 (9.8%) or *never* (14%) personally coded the samples. Further, the personnel reported (N = 81) being involved in
 429 coding included: a graduate student volunteer [unpaid] (32.1%), a clinician/speech-language pathologist (24.7%),
 430 a paid graduate-level research assistant (32.1%), a paid research assistant or lab manager [not a student] (22.2%),
 431 other (25.9%), a paid undergraduate research assistant (13.6%), a PhD student whose work involves the data
 432 collected (19.8%), an undergraduate student volunteer [unpaid] (19.8%) and/or a researcher with a PhD in a related
 433 field (11.1%). Notably, of those respondents providing explanations or open responses, two respondents indicated
 434 not knowing what the term ‘coding’ meant in reference to discourse analysis.

435 **Survey Sub-sections 4.c and 4.d. Analysis**

436 In terms of the frequency of data analysis, respondents (N = 139) indicated analyzing samples *always*
 437 (51.1%), *usually* (28.1%), *sometimes* (10.8%), *rarely* (6.5%), or *never* (3.6%). Furthermore, participants (N = 138)

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

438 reported that they *always* (55.8%), *usually* (18.8%), *sometimes* (10.9%), *rarely* (8.7%) or *never* (5.8%) personally
439 analyzed the samples. If the respondents did not themselves code the discourse samples (N = 56), the following
440 personnel were involved in the analysis: graduate student volunteer [unpaid] (35.7%), a clinician/speech-language
441 pathologist (21.4%), a paid graduate-level research assistant (32.1%), a paid research assistant or lab manager [not
442 a student] (25%), other (19.6%), a paid undergraduate research assistant (10.7%), a PhD student whose work
443 involves the data collected (23.2%), an undergraduate student volunteer [unpaid] (21.4%) and/or a researcher with
444 a PhD in a related field (21.4%). Of those who responded to open-ended questions regarding who was involved in
445 analysis, most indicated that they either worked alone or did not have access to trained personnel to support
446 discourse analysis and reliability procedures. For example, one respondent replied, “I work in the real world and do
447 it all myself.” In addition, some respondents mentioned receiving help from students, colleagues, required
448 coursework, research assistants, and lab managers (see Table 2, theme two).

449 To better understand the common practices in data analysis, we asked respondents *how* discourse samples
450 were typically analyzed (N = 133). A majority indicated that they used clinical judgment (69.9%), employed
451 standardized aphasia assessment ratings/scoring (62.4%), and/or used manual coding and analysis (e.g., main
452 concept analysis) (48.9%). Fewer respondents utilized computerized systems, such as Systematic Analysis of
453 Language Transcripts (SALT; Miller & Chapman, 1983) (14.3%), Computerized Language Analysis (CLAN;
454 MacWhinney, 2000) (11.3%), Praat (Boersma & Van Heuven, 2001) (3.8%), or Computer Analysis of Speech for
455 Psychological Research (CPIDR; Brown et al., 2007; Covington et al., 2007) (0.8%). Six percent indicated that they
456 did not use a specific protocol, and 11.3% indicated ‘other’ methods such as blinded listener ratings, pragmatic
457 analysis protocol, or spontaneous speech analysis. In general, when expanding on their responses, respondents
458 reported using both granular (e.g., specific coding of discourse features) and global (e.g., overall rating or singular
459 judgment score) analyses relating to language form and content, pragmatics, and functional as well as motor speech
460 measures (Table 2, theme five).

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

466 FIGURE 3 HERE

467 **Section 5: Psychometric Properties and Normative Data**

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

470 *Raters.* As noted earlier, different personnel were reported as being involved in the transcription, coding
471 and analyzing process. Accordingly, this section asked more detailed questions about raters and their training.
472 Respondents (N = 123) were split as to whether they personally trained raters: 39% did, 47.2% did not, and 13.8%
473 taught in conjunction with others (e.g., collaborative training of raters between multiple study personnel). Notably,
474 we want to acknowledge a limitation this result. This question made the assumption that other parties are involved
475 in rating, which may not be the case. Therefore, it may be the case that some respondents who do all of the discourse
476 work on their own were forced into an erroneous response option (e.g., “no”); we therefore urge readers to interpret
477 these results in this context.

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

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

492 FIGURE 4 HERE

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

496 FIGURE 5 HERE

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

497 *Psychometric properties: Rater reliability.* The majority of respondents indicated that they did not usually
498 collect data about rater agreement (53.4%), but some respondents (N = 116) indicated collecting rater agreement
499 during transcription (29.3%), coding (37.1%), and analysis (38.8%). If reliability was checked, the personnel
500 completing the reliability analysis were most commonly the respondents themselves (40.9%), a clinician/speech-
501 language pathologist (19.1%), or a paid graduate-level research assistant (18.3%) (N = 115).

502 We then asked what discourse measures were examined for rater reliability (multiple answer), with the
503 most common being all outcome measures of interest (29.6%), followed by total words (or tokens) (17.4%), only
504 some outcome measures of interest (16.5%), total utterances (14.8%), other (3.5%), or not applicable (51.3%) (N=
505 115). To quantify rater agreement, respondents (N = 115) frequently used percent agreement (29.6%) or intraclass
506 correlation coefficient (17.4%), with less-used metrics being correlation coefficient (13%) or other (5.2%) (51.3%
507 answered N/A and 9.6% answered ‘none’).

508 *Psychometric properties: Test-retest data.* Respondents (N = 110) typically never (36.4%) or sometimes
509 (40%) collected test-retest data for spoken discourse samples (often, 6.4%; usually, 10.9%; always, 10%).

510 *Psychometric properties of outcome measures.* A majority of respondents indicated that, in general, there
511 was inadequate availability of psychometric data (81.8%, N = 111) and normative data (51.8%, N = 110) for spoken
512 discourse outcome measures. Respondents (N = 110) further stated that they would find a database of psychometric
513 properties and/or normative data of discourse outcomes useful (93.6%).

514 We then asked if respondents looked for psychometric information of discourse outcome measures (e.g.,
515 reliability, validity, stability, acceptability): 39.6% said ‘no,’ 33.3% ‘yes,’ and 27% ‘sometimes’ (N = 116).
516 Respondents (N = 110) cited that major barriers to collecting psychometric data included time (82.7%), knowledge
517 and training (60.9%), funds (46.4%), personnel (42.7%) and other (10%). Respondents were also asked if they
518 looked for normative data for discourse outcome measures: 30% said ‘no,’ 32.7% ‘yes,’ and 37.3%, ‘sometimes’
519 (N = 110). Finally, participants had the opportunity to share their insights into ways to facilitate the collection,
520 analysis, and publication of spoken discourse data in aphasia (N = 29).

521 Qualitative data indicated that respondents saw the potential value and usefulness of standardized discourse
522 measures for comparison and interpretation, to ensure best practice, as a meaningful outcome measure, and for
523 reimbursement purposes (see Table 4). Specifically, psychometric properties of discourse data were thought to be
524 important for comparing and interpreting discourse measures across individuals and approaches (e.g., “without
525 adequate psychometric properties described, interpretation of results is problematic, and clinical application of
526 measures will be limited.”) and were thought to be useful for expressing “stable,” “reliable, valid and sensitive”

527 measures that are considered “best practice.” Additionally, respondents described valuing discourse measures
 528 related to clients’ goals, post-therapy change, and those that could be applied to “real life” and highlighted the
 529 importance of psychometrically sound measures for determining clinical outcomes and “gauging treatment effects.”
 530 Further, open ended responses spoke to important issues in psychometric data collection and use: that it is often not
 531 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
 532 collect discourse samples, so I’m not sure how I’d recruit someone to assess my reliability”) as well as that they
 533 acknowledge that psychometric data are important (e.g., “we re-test over time to test for treatment effects. We rely
 534 on the published reliability of the measure for the test-retest stability.”)

535 TABLE 4 HERE

536 **Comparison of Responses from Clinical and Research Settings**

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

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

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

557 and these were inadequate training and access to tools and resources), analyzing discourse data (specifically, clinical
558 group was more likely to select barriers, and these barriers were inadequate training and access to tools and
559 resources), and interpreting discourse data (specifically, clinical group was more likely to select barriers, and these
560 barriers were time, access to tools and resources, and in general, a lower frequency of checking the “no barrier”
561 box). Putting this together with the larger findings of the survey, it is interesting that “time” is only found to be a
562 significantly greater barrier in clinical settings during interpretation of data, whereas respondents from both research
563 and clinical settings are not significantly different in selecting “time” as a barrier during collection, transcription,
564 and analysis.

565 We then evaluated differences in clinical and research groups regarding the discourse data itself. The
566 clinical group tended to collect fewer samples than the research group (specifically, either 1-2 samples, or 3-4
567 samples) ($p = .001$). Notably, though, the groups did not demonstrate a significant difference in the number of ideal
568 discourse samples collected ($p = .18$), in that both groups preferred to collect more samples. There was a significant
569 difference between respondents who recorded (e.g., audiotaped or videoed) discourse data, in that respondents from
570 the research setting recorded discourse more often ($p = .00001$). Of those who did not record discourse data, there
571 was not a significant difference between groups for whether they transcribed live ($p = .048$) or analyzed live ($p =$
572 $.095$) (note that corrected p -value for significance for this comparison was $p < .025$, defined using Bonferroni
573 correction). In general, persons from research settings tended to transcribe ($p = .003$) and code ($p = .00003$)
574 discourse data more often as a part of their work, but the groups did not significantly differ on how often they
575 analyze discourse ($p = .12$). This likely reflects a difference in the choice to transcribe and code rather than to
576 perceptually analyze the discourse.

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

584 TABLE 5 HERE

585

586

Discussion

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

592 **Demographics of Participants**

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

600 **Spoken Discourse Data Collection Procedures**

601 We found that respondents working in university or hospital (i.e., acute care, rehabilitation, hospital-based
602 outpatient clinic) settings were most likely to report always collecting discourse information. However, this finding
603 may reflect a selection bias as respondents who worked in settings outside of universities or hospitals (e.g.,
604 community health, long-term care, private practice) were underrepresented in our survey sample. Confirming the
605 findings Bryant et al. (2017), our survey respondents reported frequently relying on single picture descriptions to
606 elicit spoken discourse samples from persons with aphasia. Stimuli from standardized aphasia assessments (e.g.,
607 BDAE and WAB) and well-established protocols (e.g., Nicholas & Brookshire (1993), AphasiaBank from
608 MacWhinney et al., (2011)) were most used during discourse collection. Despite long-standing recommendations
609 to collect and analyze discourse data from multiple genres to obtain a holistic and stable understanding of language
610 use across variable communicative contexts (Armstrong, 2000; Olness, 2006; Nicholas & Brookshire, 1994; Stark,
611 2020), our survey respondents (41.5%) most commonly collected 1-2 samples from a given persons with aphasia,
612 with relatively few respondents collecting more than 4 samples per person. These findings differ from Bryant et al.
613 (2016) who reported use of multiple genres and topics to elicit discourse samples in the aphasia research studies
614 they reviewed.

615 Interestingly, over 70% of our total sample reported collecting conversation samples with a clinician and/or
616 family member of persons with aphasia This is a promising finding as it indicates that by sampling discourse within

617 functional communicative interactions, SLPs and researchers are looking beyond impairment-level changes and are
618 also focused on capturing and improving participation, confidence, and quality of life for persons with aphasia
619 (Boyle, 2020; Kagan et al., 2008). However, only a limited number of respondents reported using formal
620 conversation analysis tools, which examine the interactional level. Because this survey was not optimized to query
621 the value and utility of conversational analysis, further investigation of conversational samples and analysis is
622 warranted.

623 Expanding further on common discourse collection methods, our survey found that respondents typically
624 collected discourse in a variety of environments, ranging from a quiet room to participants' homes. The endorsement
625 by respondents of a variety of environments may reflect differences in work settings, patient needs, and/or research
626 protocols. Additionally, the length of recorded discourse samples ranged between 1-5 minutes, with the time varying
627 based on the type of discourse task. In contrast to Bryant et al. (2016) and Cruice et al. (2020) who noted that only
628 38.7% (total N = 123) and 16-33% (N = 211) of their respondents recorded audio or video samples, respectively,
629 we found that approximately 78% of our total sample recorded discourse data, primarily using audio recordings. Of
630 those who did not record the samples, around 60% frequently transcribed the samples in real time or analyzed the
631 discourse output without transcribing. This latter finding reflects preferences reported by Cruice et al. (2020) who
632 found that 69% and 36% of their respondents favored transcribing in real time or analyzing in real time without
633 transcription, respectively. While real time analysis may be an efficient means of data analysis, there is little research
634 comparing the accuracy of on-line vs. off-line transcription approaches. Qualitatively some respondents indicated
635 that recording and transcribing spoken discourse samples was 'unnecessary' because they were able to evaluate
636 these data and make performance judgments concurrently while collecting samples. For example, one respondent
637 wrote, "I may make hash marks or take other notes of errors or successes vs. full transcription." Although we were
638 not able to ascertain the specific perceptual rating scales used by respondents here, it is important to note that there
639 is mounting evidence that perceptual rating can be a useful tool for discourse analysis in aphasia (e.g., Doyle et al.,
640 1996; Webster & Morris, 2019). However, there are relatively few validated perceptual rating scales specific to
641 spoken discourse in aphasia (e.g., Casilio et al., 2019; Kim & Wright, 2020), and use of non-validated perceptual
642 rating scales contributes to issues with reproducibility. In general, this speaks to a larger issue highlighted in our
643 survey section regarding the psychometric properties and normative data: discourse analysis is being employed in
644 a way that may not optimize its utility and may call into question its integrity and quality. We discuss this point in
645 more detail in the section regarding psychometric properties.

646 **Spoken Discourse Data Analysis Procedures**

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

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

676 protected, shared databases through TalkBank (<https://talkbank.org/>) to allow for maximal transparency as well as
677 maximal benefit from the time and effort invested by everyone involved in the data collection and management
678 process. Some open science frameworks have begun requiring investigators to create and make available a
679 videotape of all procedures utilized (to be hosted in storage and data management repositories such as Databrary).
680 These type of considerations and additions to methods sections of papers will be particularly useful in improving
681 the use of spoken discourse measures and indeed, all behavioral studies in the field.

682 **Common Barriers to Spoken Discourse Collection, Analysis, and Interpretation**

683 The common barriers to spoken discourse collection identified by respondents included insufficient
684 knowledge, training, and confidence in carrying out discourse collection, in addition to difficulty in implementing
685 and interpreting certain discourse collection protocols. The most common problem was lack of access to tools and
686 resources (e.g., computer hardware/software, recording equipment). The link between this lack of support by trained
687 individuals is clearly aligned with the resource barrier discussed earlier (i.e., a lack of resources for discourse
688 collection/analysis, especially in the clinical group). In general, more respondents from clinical settings were likely
689 to cite at least one barrier to discourse collection, transcription, analysis, and interpretation. The survey conducted
690 by Bryant et al. (2017) focused on discourse use in clinical practice. In contrast, the current survey was not explicitly
691 focused on clinical practice, and indeed, a portion of respondents self-identified as working in a research setting.
692 Although we cannot directly compare our findings with those of Bryant et al. (2017), both survey studies serve to
693 reflect a similar principle: Respondents felt that there were significant barriers to collection and analysis of discourse
694 across a variety of work settings. It is noteworthy that barriers were endorsed by respondents from both clinical and
695 research settings, together suggesting that barriers pose a critical hurdle to overcome in order to increase integration
696 of discourse into clinical and research settings, and to ensure that best practices are being used when spoken
697 discourse is being assessed. Given the uniformity of findings across studies relating to the barriers hindering more
698 widespread application of spoken discourse analysis, these barriers appear universal and persistent (Bryant et al.,
699 2017; Cruice et al., 2020).

700 A barrier raised by our study respondents that has received less attention in the literature is the need for
701 multicultural and multilingual spoken discourse elicitation materials and assessments. In an increasingly culturally
702 and linguistically multifarious patient or research participant pool, a focus on establishing and validating such
703 assessment tools and materials is a needed area of future research, as several respondents raised the lack of such
704 tools as a significant barrier to using discourse in their practice. While there are examples of culturally adapted

705 elicitation stimuli and assessment systems (e.g., Kong & Law, 2009; Pak-Hin & Law, 2004; Rousseaux et al., 2010),
706 their application to spoken discourse in individuals with aphasia has not been investigated.

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

717 **Psychometric Properties of and Normative Data for Spoken Discourse in Aphasia**

718 In the current study, respondents made clear the importance of psychometric properties of discourse data
719 in both open-ended and quantitative responses. Themes arising from the responses to open-ended questions included
720 that psychometric properties of discourse data were thought to be important for comparing and interpreting
721 discourse measures across individuals and approaches, and that spoken discourse was useful for expressing “stable,”
722 “reliable, valid and sensitive” measures that are considered “best practice.” Such themes were also reflected in the
723 quantitative results: Nearly 94% of respondents stated that they would find a database of psychometric properties
724 and/or normative data of discourse outcomes useful (93.6%) whilst also highlighting inadequate availability of
725 psychometric data and normative data. Interestingly, a surprising number of study participants reported that they
726 don’t look for psychometric properties (33%) or normative data (30%). This may reflect not only that it is well
727 known that this literature base is impoverished, but also a general thought bias that discourse does not need (or
728 needs less) psychometric validation. However, it is also important to note that respondents’ concerns over
729 ‘availability’ extended to difficulties locating information regarding discourse best practices and psychometric
730 properties in the extant literature, stating specifically that they would benefit from having this literature/information
731 consolidated in a way that was more accessible to the field. Our findings highlight the need not only to develop
732 more robust psychometric metrics for spoken discourse variables, but also the need to improve uniformity in
733 reporting (i.e., documenting and dissemination the procedures undertaken in discourse analysis and psychometric
734 information of the selected discourse measures) across studies, the development of a common nomenclature for use

735 in discourse studies, and the need to develop dissemination tools that are accessible to both clinicians and
736 researchers. This may also reflect the numerous and considerable barriers to psychometric data collection and
737 aggregation, which were endorsed by respondents: time, knowledge and training, funds, personnel, and other
738 aspects, like environment (e.g., not appropriate to do in their work environment) and belief (e.g., unfair to patient,
739 not in patient's best interest).

740 Psychometric properties are key for reproducibility and data aggregation across studies. Our survey results,
741 which highlight different approaches to rater reliability and collection and use of test-retest data, underscore that
742 consistency and transparency of collecting and reporting psychometric properties in spoken discourse in aphasia
743 remains an issue. Test-retest stability is one of the most important metrics for clinical research, and indeed, should
744 be established for research to be implemented in the clinical setting (Brookshire & Nicholas, 1994; Herbert et al.,
745 2008). For example, short interval sampling (testing and retesting within a short window of time [e.g., two weeks])
746 can determine the variability of a participant's baseline performance. Notably, a measure that varies widely within-
747 participants for a short interval *is not stable enough* to be used as a clinically meaningful outcome or assessment
748 measure (Boyle, 2014, 2015). Test-retest stability is paramount in treatment research, particularly given that data
749 acquired during short interval testing periods are prone to practice effects (i.e., participant behavior may improve
750 over testing sessions due to learning the discourse stimuli/procedure vs. the treatment). Stability is of particular
751 concern in persons with aphasia in whom language has long been characterized as highly variable from day to day
752 (Hula & McNeil, 2008; Murray, 1999). When no normative data exist for test-retest across discourse elicitation
753 methods in aphasia, it falls to the researcher to collect this information (but: see barriers), or to look to the literature
754 for standards. However, given that test-retest stability is reported uncommonly in the literature (Pritchard et al.,
755 2017), the direct result of these gaps is a lack of prioritization and dissemination of, and focus on, this psychometric
756 property.

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

765 for improved clinical justification of spoken discourse outcomes in aphasia (e.g., Boyle, 2020; Bryant et al., 2017;
766 Cruice et al., 2020).

767 **Study Limitations**

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

777 We also recognize that the length of the survey may have contributed to the 58% completion rate. The
778 completion rate for each question is provided in Supplemental Tables S1 and S2; indeed, question completion
779 declines over the course of the survey. The survey was lengthy because it included several questions with branching
780 logic, and we suggest that future surveys that extend on ours may reduce questions to encourage a higher completion
781 rate. Additionally, some survey questions and data considerations may have been more applicable to research rather
782 than clinical settings and vice versa, although we did not note any explicit trends in question answering between
783 the clinical and research groups. The driving factor seemed to be fewer questions answered with time, i.e., those at
784 the end of the survey were less likely to be answered, reflecting an issue in the length of the survey, or the interest
785 in filling out the section related to psychometric properties and normative data.

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

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

808 **Recommendations and Future Research Directions**

809 There have been many 'calls to arms' for addressing the spoken discourse evidence issues in the extant
810 aphasia literature, all of which highlighted the benefit of this kind of language sampling (Armstrong, 2000; Boyle,
811 2011; Dietz & Boyle, 2018; Kintz & Wright, 2017; Linnik et al., 2016; Prins & Bastiaanse, 2004; Pritchard et al.,
812 2018; Wallace et al., 2018). In consideration of the current study's findings, we propose some recommendations to
813 improve the state of the spoken discourse evidence in the aphasia literature. First, to address barriers related to
814 training, graduate education and clinician training in spoken discourse analysis must be emphasized. Second, to
815 improve the availability of resources, investigators should make available their study protocol, including all
816 documents used for transcription and coding training and, wherever possible, a video of their training procedures.
817 Additionally, (3) psychometric properties and normative data need to be established based on larger and
818 internationally diverse samples of spoken discourse outcomes, and be made freely available to clinicians and
819 researchers. Third, The aphasia field should focus on improving perceptual analysis and integrating training
820 regarding a variety of transcription and analysis methodologies (e.g., automatic transcription techniques) to combat
821 commonly endorsed barriers related to time in clinical and research settings. . Finally, adherence to 'best practice'
822 living documents should be advocated, wherein reviewers of papers and investigators assure that all necessary
823 components for procedure reproducibility are reported.

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

824 To begin addressing these recommendations, FOQUSAphasia (www.foqusaphasia.com) includes a Best
825 Practices task force, whose first initiative is to create a living, best practices document. Its second task force,
826 Methodological & Data Quality, is pursuing an initiative to collect a large database of test-retest data using the
827 AphasiaBank protocol. This database will be made available on AphasiaBank and will be critical for outlining the
828 psychometric properties of commonly used discourse metrics and for building a normative sample. Likewise,
829 findings from the current study can be used to guide development of process standardization in spoken discourse
830 and the creation of a psychometric and normative property database. Presently, members of the Best Practices task
831 force of FOQUSAphasia are conducting an e-Delphi study to gather expert consensus for best practices in this field.

832

Acknowledgments

833

We would like to acknowledge the members of the FOQUSAphasia working group

834

(www.foqusaphasia.com). We also acknowledge funding to authors Brielle C. Stark (ASHFoundation New

835

Investigator Award) and Angela Roberts (NIH/NIDCD 1R21DC017255-01).

836

References

- 837 Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2020). A unified model of post-stroke
 838 language deficits including discourse production and their neural correlates. *Brain*, *143*(5), 1541–1554.
 839 <https://doi.org/10.1093/brain/awaa074>
- 840 Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, *14*(9), 875–892.
 841 <https://doi.org/10.1080/02687030050127685>
- 842 Armstrong, Elizabeth. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, *14*(9), 875–892.
 843 <https://doi.org/10.1080/02687030050127685>
- 844 Beeke, S., Maxim, J., & Wilkinson, R. (2007). Using conversation analysis to assess and treat people with aphasia.
 845 In *Seminars in Speech and Language* (Vol. 28, Issue 2, pp. 136–147). Copyright © 2007 by Thieme Medical
 846 Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. <https://doi.org/10.1055/s-2007-970571>
- 847 Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT . *Glott International*, *5*(9/10), 341–347.
- 848 Boyle, M. (2011). Discourse treatment for word retrieval impairment in aphasia: The story so far. *Aphasiology*,
 849 *25*(11), 1308–1326. <https://doi.org/10.1080/02687038.2011.596185>
- 850 Boyle, M. (2014). Test-retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and*
 851 *Hearing Research*, *57*(June), 966–978. <https://doi.org/10.1044/2014>
- 852 Boyle, M. (2015). Stability of Word-Retrieval Errors With the AphasiaBank Stimuli. *American Journal of Speech-*
 853 *Language Pathology*, *24*(4). https://doi.org/10.1044/2015_AJSLP-14-0152
- 854 Boyle, M. (2020). Choosing Discourse Outcome Measures to Assess Clinical Change. *Seminars in Speech and*
 855 *Language*, *41*(1), 1–9. <https://doi.org/10.1055/s-0039-3401029>
- 856 Brady, M., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia
 857 following stroke. *Cochrane Database of Systematic Reviews*, *6*.
 858 <https://doi.org/10.1002/14651858.CD000425.pub4.www.cochranelibrary.com>
- 859 Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2),
 860 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- 861 Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and test-retest stability of connected speech
 862 measures for adults with aphasia. *Journal of Speech and Hearing Research*, *37*(2), 399–407.

- 863 <http://www.ncbi.nlm.nih.gov/pubmed/8028321>
- 864 Brown, C., Snodgrass, T., & Covington, M. (2007). *Computerized Propositional Idea Density Rater 3 (CPIDR 3)*.
865 CASPR Project, Artificial Intelligence Center.
- 866 Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the
867 literature. *Clinical Linguistics & Phonetics*, 30(7), 489–518. <https://doi.org/10.3109/02699206.2016.1145740>
- 868 Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of
869 language in aphasia. *Aphasiology*, 31(10), 1105–1126. <https://doi.org/10.1080/02687038.2016.1239013>
- 870 Cahana-Amitay, D., & Jenkins, T. (2018). Working memory and discourse production in people with aphasia.
871 *Journal of Neurolinguistics*, 48(April), 90–103. <https://doi.org/10.1016/j.jneuroling.2018.04.007>
- 872 Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-Perceptual Rating of
873 Connected Speech in Aphasia. *American Journal of Speech-Language Pathology*, 28(2), 550–568.
874 https://doi.org/10.1044/2018_AJSLP-18-0192
- 875 Cherney, L., Coelho, C. A., & Shadden, B. B. (1998). *Analyzing Discourse in Communicatively Impaired Adults -*
876 *Leora Reiff Cherney, Carl A. Coelho, Barbara Bennett Shadden - Google Books*. Aspen Publishers.
877 https://books.google.com/books/about/Analyzing_Discourse_in_Communicatively_I.html?id=DeMLAQAA
878 MAAJ
- 879 Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716–
880 aac4716. <https://doi.org/10.1126/science.aac4716>
- 881 Covington, M. A., Riedel, W. J., Brown, C., Congzhou He, Morris, E., Weinstein, S., Semple, J., & Brown, J.
882 (2007). Does ketamine mimic aspects of schizophrenic speech? *Journal of Psychopharmacology*, 21(3), 338–
883 346. <https://doi.org/10.1177/0269881107077729>
- 884 Cruice, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., & Dipper, L. (2020). UK speech and
885 language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. *International*
886 *Journal of Language and Communication Disorders*, 1–26. <https://doi.org/10.1111/1460-6984.12528>
- 887 Damico, J. S., Oelschlaeger, M., & Simmons-Mackie, N. (1999). Qualitative methods in aphasia research:
888 Conversation analysis. *Aphasiology*, 13(9–11), 667–679. <https://doi.org/10.1080/026870399401777>
- 889 Davidson, B., Worrall, L., & Hickson, L. (2003). Identifying the communication activities of older people with

- 890 aphasia: Evidence from naturalistic observation. *Aphasiology*, 17(3), 243–264.
 891 <https://doi.org/10.1080/729255457>
- 892 Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: have we reached the tipping point?
 893 *Aphasiology*, 32(4), 459–464. <https://doi.org/10.1080/02687038.2017.1398803>
- 894 Doedens, W. J., & Meteyard, L. (2020). Measures of functional, real-world communication for aphasia: a critical
 895 review. In *Aphasiology* (Vol. 34, Issue 4, pp. 492–514). Routledge.
 896 <https://doi.org/10.1080/02687038.2019.1702848>
- 897 Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The Communicative Informativeness and Efficiency of
 898 Connected Discourse by Adults With Aphasia Under Structured and Conversational Sampling Conditions.
 899 *American Journal of Speech-Language Pathology*, 4(4), 130–134. [https://doi.org/10.1044/1058-](https://doi.org/10.1044/1058-0360.0404.130)
 900 0360.0404.130
- 901 Doyle, P. J., Tsironas, D., Goda, A. J., & Kalinyak, M. (1996). The Relationship Between Objective Measures and
 902 Listeners' Judgments of the Communicative Informativeness of the Connected Discourse of Adults With
 903 Aphasia. *American Journal of Speech-Language Pathology*, 5(3), 53–60. [https://doi.org/10.1044/1058-](https://doi.org/10.1044/1058-0360.0503.53)
 904 0360.0503.53
- 905 Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-
 906 Surveys (CHERRIES). In *Journal of Medical Internet Research* (Vol. 6, Issue 3). Journal of Medical Internet
 907 Research. <https://doi.org/10.2196/jmir.6.3.e34>
- 908 Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse
 909 elicitation tasks. *Aphasiology*, 25(11), 1414–1430. <https://doi.org/10.1080/02687038.2011.603898>
- 910 Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types.
 911 *Aphasiology*, 25(10), 1261–1278. <https://doi.org/10.1007/s11103-011-9767-z>.Plastid
- 912 Fromm, D., Forbes, M., Holland, A., & Dalton, G. (2017). *Discourse Characteristics in Aphasia Beyond the*
 913 *Western Aphasia Battery Cutoff*. 26(August), 1–7. https://doi.org/10.1044/2016_AJSLP-16-0071
- 914 Galski, T., Tompkins, C., & Johnston, M. V. (1998). Competence in discourse as a measure of social integration
 915 and quality of life in persons with traumatic brain injury. *Brain Injury*, 12(9), 769–782.
 916 <https://www.researchgate.net/profile/Mark-Johnston->

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

- 917 [5/publication/13531365_Compentence_in_discourse_as_a_measure_of_social_integration_and_quality_of_li](https://doi.org/10.1016/j.jbi.2019.103208)
918 [fe_in_persons_with_traumatic_brain_injury/links/56bb9d2708ae3f9793153c2f/Competence-in-discourse-as-](https://doi.org/10.1016/j.jbi.2019.103208)
919 [a-measure-of-social-integration-and-quality-of-life-in-persons-with-traumatic-brain-injury.pdf](https://doi.org/10.1016/j.jbi.2019.103208)
- 920 Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination* (3rd ed.). Pearson.
- 921 Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O’Neal, L., McLeod, L., Delacqua, G., Delacqua,
922 F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of
923 software platform partners. In *Journal of Biomedical Informatics* (Vol. 95). Academic Press Inc.
924 <https://doi.org/10.1016/j.jbi.2019.103208>
- 925 Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data
926 capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research
927 informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381.
928 <https://doi.org/10.1016/j.jbi.2008.08.010>
- 929 Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid
930 assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, 22(2), 184–203.
931 <https://doi.org/10.1080/02687030701262613>
- 932 Hula, W. D., & McNeil, M. R. (2008). Models of attention and dual-task performance as explanatory constructs in
933 aphasia. *Seminars in Speech and Language*, 29(3), 169–187. <https://doi.org/10.1055/s-0028-1082882>
- 934 IBM Corp. (2019). *IBM SPSS Statistics for Windows Version 26*.
- 935 Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124.
936 <https://doi.org/10.1371/journal.pmed.0020124>
- 937 Jacks, A., Haley, K. L., Bishop, G., & Harmon, T. G. (2019). Automated Speech Recognition in Adult Stroke
938 Survivors: Comparing Human and Computer Transcriptions. *Folia Phoniatica et Logopaedica*, 71(5–6),
939 286–296. <https://doi.org/10.1159/000499156>
- 940 Kagan, A., Simmons-Mackie, N., Rowland, A., Huijbregts, M., Shumway, E., McEwen, S., Threats, T., & Sharp,
941 S. (2008). Counting what counts: A framework for capturing real-life outcomes of aphasia intervention.
942 *Aphasiology*, 22(3), 258–280. <https://doi.org/10.1080/02687030701282595>
- 943 Kertesz, A. (2007). *Western Aphasia Battery - Revised*. The Psychological Corporation.

- 944 Kim, H., & Wright, H. H. (2020). Concurrent Validity and Reliability of the Core Lexicon Measure as a Measure
 945 of Word Retrieval Ability in Aphasia Narratives. *American Journal of Speech-Language Pathology*, 29(1),
 946 101–110. https://doi.org/10.1044/2019_AJSLP-19-0063
- 947 Kintz, S., & Wright, H. H. (2017). Discourse measurement in aphasia research. *Aphasiology*, 00(00), 1–3.
 948 <https://doi.org/10.1080/02687038.2017.1398807>
- 949 Kong, A. P. H., & Law, S. P. (2009). A linguistic communication measure for monitoring changes in Chinese
 950 aphasic narrative production. *Clinical Linguistics and Phonetics*, 23(4), 255–269.
 951 <https://doi.org/10.1080/02699200802673234>
- 952 Le, D., Licata, K., & Mower Provost, E. (2018). Automatic quantitative analysis of spontaneous aphasic speech.
 953 *Speech Communication*, 100, 1–12. <https://doi.org/10.1016/j.specom.2018.04.001>
- 954 Le, D., & Provost, E. M. (2016). Improving automatic recognition of aphasic speech with AphasiaBank.
 955 *Proceedings of the Annual Conference of the International Speech Communication Association*,
 956 *INTERSPEECH, 08-12-Sept(March)*, 2681–2685. <https://doi.org/10.21437/Interspeech.2016-213>
- 957 Leaman, M. C., & Edmonds, L. A. (2019). Revisiting the Correct Information Unit: Measuring Informativeness in
 958 Unstructured Conversations in People With Aphasia. *American Journal of Speech-Language Pathology*,
 959 28(3), 1099–1114. https://doi.org/10.1044/2019_AJSLP-18-0268
- 960 Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: a current review of theoretical and
 961 methodological challenges. *Aphasiology*, 30(7), 765–800. <https://doi.org/10.1080/02687038.2015.1113489>
- 962 MacWhinney, B. (2000). *The CHILDES Project: Tools for Analysing Talk* (3rd ed.). Lawrence Erlbaum Associates
 963 Inc.
- 964 MacWhinney, B., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for Studying Discourse. *Aphasiology*,
 965 25(11), 1286–1307. <https://doi.org/10.1080/02687038.2011.589893>.AphasiaBank
- 966 Miller, J., & Chapman, R. (1983). *SALT: Systematic Analysis of Language Transcripts, User's Manual*. University
 967 of Wisconsin Press.
- 968 Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., &
 969 Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel
 970 group randomised trials. *BMJ (Clinical Research Ed.)*, 340. <https://doi.org/10.1136/bmj.c869>

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

- 971 Murray, L. L. (1999). Review Attention and aphasia: theory, research and clinical implications. *Aphasiology*,
972 13(June 2015), 91–111. <https://doi.org/10.1080/026870399402226>
- 973 Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the
974 connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36(2), 338–350.
975 <http://www.ncbi.nlm.nih.gov/pubmed/8487525>
- 976 Nichols, T., Das, S., Eickhoff, S., Evans, A., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M., Poldrack, R.,
977 Poline, J.-B., Proal, E., Thirion, B., Van Essen, D., White, T., & Yeo, B. T. T. (2017). Best practices in data
978 analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, 20(3), 299–303.
979 <https://doi.org/10.1038/nn.4500>
- 980 Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M. P.,
981 Poldrack, R. A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D. C., White, T., & Yeo, B. T. T. (2017). Best
982 practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, 20(3), 299–303.
983 <https://doi.org/10.1038/nn.4500>
- 984 Osborne, F., Hickin, J., Best, W., & Howard, D. (1998). Treating word-finding difficulties - Beyond picture naming.
985 *International Journal of Language and Communication Disorders*, 33(SUPPL. 1), 208–213.
986 <https://doi.org/10.3109/13682829809179424>
- 987 Pak-Hin, A. K., & Law, S.-P. (2004). A Cantonese linguistic communication measure for evaluating aphasic
988 narrative production: normative and preliminary aphasic data. *Journal of Multilingual Communication*
989 *Disorders*, 2(2), 124–146. <https://doi.org/10.1080/14769670310001659708>
- 990 Prins, R., & Bastiaanse, R. (2004). Review: Analysing the spontaneous speech of aphasic speakers. *Aphasiology*,
991 18(12), 1075–1091. <https://doi.org/10.1080/02687030444000534>
- 992 Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures
993 in aphasia. *International Journal of Language & Communication Disorders*, 1–44.
994 <https://doi.org/10.1111/1460-6984.12318>
- 995 Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2018). Psychometric properties of discourse measures in aphasia:
996 acceptability, reliability, and validity. *International Journal of Language & Communication Disorders*, 53(6),
997 1078–1093. <https://doi.org/10.1111/1460-6984.12420>

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

- 998 Prutting, C. A., & Kirchner, D. M. (1987). A clinical appraisal of the pragmatic aspects of language. *Journal of*
999 *Speech and Hearing Disorders*, 52(2), 105–119. <https://doi.org/10.1044/jshd.5202.105>
- 1000 Roberts, A., & Orange, J. (2013). Discourse in Lewy Body Spectrum Disorder. In B. Davis & J. Guendouzi (Eds.),
1001 *Pragmatics in Dementia Discourse*.
1002 https://www.researchgate.net/publication/311923808_DISCOURSE_IN_LEWY_BODY_SPECTRUM_DIS
1003 ORDER
- 1004 Roberts, A., & Post, D. (2018). Information Content and Efficiency in the Spoken Discourse of Individuals With
1005 Parkinson's Disease. *Journal of Speech, Language, and Hearing Research*, 61(9), 2259–2274.
1006 https://doi.org/10.1044/2018_JSLHR-L-17-0338
- 1007 Rousseaux, M., Daveluy, W., & Kozłowski, O. (2010). Communication in conversation in stroke patients. *Journal*
1008 *of Neurology*, 257(7), 1099–1107. <https://doi.org/10.1007/s00415-010-5469-8>
- 1009 Salis, C., Murray, L., & Bakas, K. (2018). An international survey of assessment practices for short-term and
1010 working memory deficits in aphasia. *American Journal of Speech-Language Pathology*, 27(2), 574–591.
1011 https://doi.org/10.1044/2017_AJSLP-17-0057
- 1012 Shadden, B. (1998). Obtaining the Discourse Sample. In L. Cherney, B. Shadden, & C. Coelho (Eds.), *Analyzing*
1013 *Discourse in Communicatively Impaired Adults* (pp. 9–34). Aspen Publishers.
- 1014 Shadden, B. B., Burnette, R. B., Eikenberry, B. R., & DiBrezzo, R. (1991). All Discourse Tasks Are Not Created
1015 Equal. *Clinical Aphasiology Conference*. <http://aphasiology.pitt.edu/153/>
- 1016 Sim, P., Power, E., & Togher, L. (2013). Describing conversations between individuals with traumatic brain injury
1017 (TBI) and communication partners following communication partner training: Using exchange structure
1018 analysis. *Brain Injury*, 27(6), 717–742. <https://doi.org/10.3109/02699052.2013.775485>
- 1019 Simmons-Mackie, N., Worrall, L., Murray, L. L., Enderby, P., Rose, M. L., Paek, E. J., Klippi, A., & on behalf of
1020 the Aphasia United Best Practices Working Group. (2017). The top ten: best practice recommendations for
1021 aphasia. *Aphasiology*, 31(2), 131–151. <https://doi.org/10.1080/02687038.2016.1180662>
- 1022 Stark, B. C. (2019a). A comparison of three discourse elicitation methods in aphasia and age-matched adults:
1023 Implications for language assessment and outcome. *American Journal of Speech-Language Pathology*, 28(3),
1024 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

- 1025 Stark, B. C. (2019b). A comparison of three discourse elicitation methods in aphasia and age-matched adults:
1026 Implications for language assessment and outcome. *American Journal of Speech-Language Pathology*, 28(3).
1027 https://doi.org/10.1044/2019_AJSLP-18-0265
- 1028 Stark, B. C., Dutta, M., Murray, L., Bryant, L., Fromm, D., MacWhinney, B., Ramage, A., Simmons-Mackie, N.,
1029 Roberts, A., Fergadiotis, G., Ouden, D.-B. den, Brock, K., Harmon, T., Paek, E. J., Themistokleous, C.,
1030 MicKinney-Bock, K., Yoo, H., Yoon, S. O., Sharma, S., ... Gutierrez, S. (2020). Standardizing assessment of
1031 spoken discourse in aphasia: A working group with deliverables. *American Journal of Speech-Language*
1032 *Pathology, Selected Papers from the Clinical Aphasiology Conference*.
1033 <https://doi.org/10.31234/OSF.IO/E6R3X>
- 1034 Tetnowski, J. T., Tetnowski, J. A., & Damico, J. S. (2020). Patterns of Conversation Trouble Source and Repair as
1035 Indices of Improved Conversation in Aphasia: A Multiple-Case Study Using Conversation Analysis.
1036 *American Journal of Speech-Language Pathology*, 1–18. https://doi.org/10.1044/2020_ajslp-19-00100
- 1037 Thompson, C. K. (n.d.). *NORTHWESTERN NARRATIVE LANGUAGE ANALYSIS (NNLA) THEORY AND*
1038 *METHODOLOGY*.
- 1039 Ulatowska, H. K., Allard, L., Reyes, B. A., Ford, J., & Chapman, S. (1992). Conversational discourse in aphasia.
1040 *Aphasiology*, 6(3), 325–330. <https://doi.org/10.1080/02687039208248602>
- 1041 Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse
1042 in aphasia. *Brain and Language*, 13(2), 345–371. [https://doi.org/10.1016/0093-934X\(81\)90100-0](https://doi.org/10.1016/0093-934X(81)90100-0)
- 1043 Wallace, S. J., Worrall, L. E., Rose, T., & Le Dorze, G. (2018). Discourse measurement in aphasia research: have
1044 we reached the tipping point? A core outcome set ... or greater standardisation of discourse measures?
1045 *Aphasiology*, 32(4), 479–482. <https://doi.org/10.1080/02687038.2017.1398811>
- 1046 Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M.,
1047 Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A.-C.,
1048 Marshall, J., ... Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus
1049 statement. *International Journal of Stroke*, 14(2), 180–185. <https://doi.org/10.1177/1747493018806200>
- 1050 Webster, J., & Morris, J. (2019). Communicative Informativeness in Aphasia: Investigating the Relationship
1051 Between Linguistic and Perceptual Measures. *American Journal of Speech-Language Pathology*, 28(3), 1115–

- 1052 1126. https://doi.org/10.1044/2019_AJSLP-18-0256
- 1053 Whitworth, A., Leitao, S., Cartwright, J., Webster, J., Hankey, G. J., Zach, J., Howard, D., & Wolz, V. (2015).
1054 NARNIA: a new twist to an old tale. A pilot RCT to evaluate a multilevel approach to improving discourse in
1055 aphasia. *Aphasiology*, 29(11), 1345–1382. <https://doi.org/10.1080/02687038.2015.1081143>
- 1056 Wilkinson, R. (2014). Intervening With Conversation Analysis in Speech and Language Therapy: Improving
1057 Aphasic Conversation. In *Research on Language and Social Interaction* (Vol. 47, Issue 3, pp. 219–238).
1058 Routledge. <https://doi.org/10.1080/08351813.2014.925659>
- 1059 Wright, H. H., & Capilouto, G. J. (2009). Manipulating task instructions to change narrative discourse performance.
1060 *Aphasiology*, 23(10), 1295–1308. <https://doi.org/10.1080/02687030902826844>
- 1061

Tables and Figures

1062

1063 Figure Legends

1064

1065 Figure 1: Barriers to data collection, transcription, analysis and interpretation. Respondents could select more than
 1066 one barrier.\

1067 * = No response option for 'data collection.'

1068 ^ = No response option for 'data interpretation.'

1069 Figure 2: Follow-up steps respondents reported taking after collecting a spoken discourse sample. Respondents
 1070 could select more than one option.

1071

1072

1073 Figure 3: Typically analyzed outcome measures. Respondents could select more than one option.

1074

1075 Figure 4: Utterances were delineated in a variety of ways by respondents. Respondents could select more than one
 1076 option.

1077

1078 Figure 5: Respondents indicated which discourse outcome measures they extracted based on a variety of factors.
 1079 Respondents could select more than one option.

1080

Table 1: Demographic information of respondents.

Demographic Information	Responses	Respondents (n)
Locations	USA (55%) United Kingdom (7.4%) Australia (19.6%) New Zealand (0.5%) Canada (6.9%) Other (10.6%)	189
Roles (could select more than one)	Researcher (43.4%) Academic/teacher (22.2%) Speech-language pathologist (81%) Student (9%) Other (3.7%)	189
Age years	< 25 years of age (4.2%) 26-40 years of age (47.6%) 41-55 years of age (30.7%) > 55 years of age (17.5%)	189
Gender	Female (93.1%)	189

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

	Male (5.8%) Other (0.5%)	
Terminal degree	Bachelor's (15.9%) Master's (51.3%) PhD (22.8%) Post-doctoral (5.3%) Clinical doctorate (1.1%) Other (3.7%)	189
Main area of data collection	Acute care (8%) 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%)	188
Years of working with people with aphasia	M = 14.17 (SD = 10.45), range 1-45	187

1081

1082

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

1083 Table 2. Number of Respondents Reporting/Endorsing Use of Specific Discourse Procedures in Their Clinical
 1084 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 (WAB); 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		

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND 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

1085

1086

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

1087 Table 3. Number of Respondents Reporting/Endorsing Specific Barriers to Using Discourse Analysis in Clinical
 1088 and/or Research Practice, Organized by Theme and Subtheme.

1089

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 HIPAA 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 is 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., AOS, 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

1090

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

1092 Table 4. Number of Respondents Reporting/Endorsing the Value of Standardized Spoken Discourse Measures in
 1093 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

1094

1095

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

1096 Table 5. A comparison of core survey responses (e.g., barriers) between clinical and research settings. Respondents
 1097 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 26-40 years 41-55 years More than 55 years	4 64 35 15	4 23 21 14	Kruskal-Wallis $\chi^2 = 5.98$, $df = 3$, $p = 0.11$
Years in aphasia	Continuous	17.58(11.35)	M=12.66(9.66)	W = 2660, $p = 0.004^*$
How often do you collect discourse	Always Usually Sometimes Rarely	28 34 27 11	23 22 11 3	$\chi^2 = 4.07$, $df = 3$, $p = 0.25$
Total barriers to collecting discourse data	Sum of categorical selections (yes/no)	1.54(1.15)	1.27(0.75)	Kruskal-Wallis $\chi^2 = 23.36$, $df = 5$, $p = 0.00028^*$
Insufficient skills	Yes/No	27/91	4/58	$\chi^2 = 6.59$, $df = 1$, $p = 0.01$
Inadequate training	Yes/No	37/81	5/57	$\chi^2 = 11.056$, $df = 1$, $p = 0.00088^*$
Tool & resource access	Yes/No	47/71	9/53	$\chi^2 = 11$, $df = 1$, $p = 0.00091^*$
Confidence	Yes/No	11/107	4/58	$\chi^2 = 0.14$, $df = 1$, $p = 0.71$
Protocol interpretation	Yes/No	28/90	17/45	$\chi^2 = 0.13$, $df = 1$, $p = 0.72$
No barriers	Yes/No	14/104	30/32	$\chi^2 = 27.41$, $df = 1$, $p = 1.645e-07^*$
Total barriers to transcribing discourse data	Sum of categorical selections (yes/no)	1.93(1.53)	1.52(0.94)	Kruskal-Wallis $\chi^2 = 23.54$, $df = 6$, $p = 0.00064^*$
Time	Yes/No	80/38	47/15	$\chi^2 = 0.899$, $df = 1$, $p = 0.34$
Insufficient skills	Yes/No	27/91	4/58	$\chi^2 = 6.59$, $df = 1$, $p = 0.01027$
Inadequate training	Yes/No	32/86	5/57	$\chi^2 = 7.91$, $df = 1$, $p = 0.004924^*$
Tool & resource access	Yes/No	41/77	7/55	$\chi^2 = 10.27$, $df = 1$, $p = 0.001355^*$
Confidence	Yes/No	15/103	4/58	$\chi^2 = 1.09$, $df = 1$, $p = 0.2966$
Protocol interpretation	Yes/No	24/94	15/47	$\chi^2 = 0.17$, $df = 1$, $p = 0.6847$
No barriers	Yes/No	4/114	9/53	Fisher's exact test, $p = 0.1201$
Total barriers to analyzing discourse data	Sum of categorical selections (yes/no)	2.08(1.61)	1.63(1.12)	Kruskal-Wallis $\chi^2 = 20.38$, $df = 6$, $p = 0.0024^*$
Time	Yes/No	77/41	43/19	$\chi^2 = 0.151$, $df = 1$, $p = 0.697$
Insufficient skills	Yes/No	35/83	10/52	$\chi^2 = 3.28$, $df = 1$, $p = 0.07$
Inadequate training	Yes/No	41/77	8/54	$\chi^2 = 8.72$, $df = 1$, $p = 0.0032^*$
Tool & resource access	Yes/No	42/76	7/55	$\chi^2 = 10.92$, $df = 1$, $p = 0.00095^*$
Confidence	Yes/No	15/103	7/55	$\chi^2 = 0.0014$, $df = 1$, $p = 0.97$
Protocol interpretation	Yes/No	26/92	16/46	$\chi^2 = 0.15$, $df = 1$, $p = 0.72$
No barriers	Yes/No	3/115	7/55	Fisher's exact test, $p = 0.034$
Total barriers to interpreting discourse analysis	Sum of categorical selections (yes/no)	1.68(1.36)	1.23(0.76)	Kruskal-Wallis $\chi^2 = 27.23$, $df = 5$, $p = 5.154e-05^*$
Time	Yes/No	64/54	16/46	$\chi^2 = 12.18$, $df = 1$, $p = 0.00048^*$
Insufficient skills	Yes/No	35/83	8/54	$\chi^2 = 5.39$, $df = 1$, $p = 0.02$
Inadequate training	Yes/No	43/75	10/52	$\chi^2 = 7.12$, $df = 1$, $p = 0.0076^*$
Tool & resource access	Yes/No	23/95	3/59	$\chi^2 = 5.93$, $df = 1$, $p = 0.015$
Confidence	Yes/No	20/98	10/52	$\chi^2 = 4.3672e-30$, $df = 1$, $p > .99$
No barriers	Yes/No	7/111	24/38	$\chi^2 = 28.37$, $df = 1$, $p = 9.999e-08^*$

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

Total barriers to psychometric data collection	Sum of categorical selections (yes/no)	1.29(1.43)	1.74(1.45)	Kruskal-Wallis $\chi^2 = 5.27$, $df = 4$, $p = 0.26$
Time		53/65	36/26	$\chi^2 = 2.31$, $df = 1$, $p = 0.13$
Funds		25/93	25/37	$\chi^2 = 6.495$, $df = 1$, $p = 0.01^*$
Personnel		25/93	21/41	$\chi^2 = 2.80$, $df = 1$, $p = 0.09$
Knowledge / training		41/77	24/38	$\chi^2 = 0.13$, $df = 1$, $p = 0.72$
Number of typical discourse samples collected	1-2 samples 3-4 samples 5-6 samples >6 samples	45 33 6 4	14 20 11 9	Kruskal-Wallis $\chi^2 = 15.52$, $df = 3$, $p = 0.0014^*$
Number of ideal discourse samples collected				Kruskal-Wallis $\chi^2 = 3.48$, $df = 2$, $p = 0.176$
Do you record the discourse data (e.g., audio, visual)?	Yes/No	56/32	54/0	$\chi^2 = 23.31$, $df = 1$, $p = 1.38e-06^*$
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 Usually Sometimes Rarely Never NA	20 26 17 17 3 2	26 17 9 2 0 0	Kruskal-Wallis $\chi^2 = 18.19$, $df = 5$, $p = 0.0027^*$
How often discourse samples are coded (e.g., phonetic coding, error marking)	Always Usually Sometimes Rarely Never NA	13 14 13 16 27 2	24 7 18 2 3 0	Kruskal-Wallis $\chi^2 = 33.37$, $df = 5$, $p = 3.173e-06^*$
How often collected discourse samples are analyzed	Always Usually Sometimes Rarely Never	40 23 6 7 5	29 13 9 2 0	Kruskal-Wallis $\chi^2 = 7.33$, $df = 4$, $p = 0.12$
How often respondents collect psychometric information	Always Usually Sometimes Rarely Never	1 9 7 8 42	6 17 11 6 5	Kruskal-Wallis $\chi^2 = 33.001$, $df = 4$, $p\text{-value} = 1.194e-06^*$
Do you seek out psychometric properties for discourse outcomes?	Yes Sometimes No	17 16 31	20 14 9	$\chi^2 = 8.69$, $df = 2$, $p\text{-value} = 0.01297^*$
Do you believe there is adequate psychometric data available for discourse?	Yes/No	12/52	6/36	$\chi^2 = 0.11$, $df = 1$, $p = 0.74$
Do you seek out normative data for discourse outcomes?	Yes Sometimes No	17 23 24	18 17 7	$\chi^2 = 5.94$, $df = 2$, $p = 0.05$
Do you believe there is adequate normative data available for discourse?	Yes Sometimes No I don't believe normative data are required	3 27 30 4	1 14 26 1	Fisher's exact test, $p = .51$

CURRENT PRACTICES IN SPOKEN DISCOURSE ASSESSMENT AND ANALYSIS

Would you use a normative and/or psychometric data database?	Yes/No	59/5	41/1	Fisher's exact test, p=.399
--	--------	------	------	-----------------------------

1098
 1099
 1100
 1101
 1102
 1103
 1104
 1105
 1106
 1107

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.

* = 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.

Appendices

We have attached the full survey as an Appendix.

Supplement information

Table S1. Response rates for quantitative questions (for all primary questions, e.g., not those that branched off)

1108 Table S2. Response rates for qualitative questions (for all primary questions, e.g., not those that branched off)