Research Article

A Clinical Training Program for Auditory-Perceptual Rating of Connected Speech in Aphasia

Marianne Casilio,[a] [iD] Manaswita Dutta,[b] [iD] Katherine Bryan,[a] [iD] Kelly Crouch,[a] [iD] Zachary DeWall,[a] Ilana Feld,[a] Jennifer Kile,[a] Keely McMahan,[a] Jacqueline Samora,[a] Kiiya Shibata,[a] [iD] Stacey Steel,[b] Anna V. Kasdan,[a] [iD] Lily Walljasper,[a] Michael de Riesthal,[a] [iD] and Stephen M. Wilson[a,c] [iD]

[a] Vanderbilt University Medical Center, Nashville, TN [b] Portland State University, OR [c] The University of Queensland, Brisbane, Australia

A R T I C L E   I N F O

A B S T R A C T

Purpose: Although discourse analysis has been receiving growing interest among aphasia researchers and speech-language pathologists (SLPs), several key challenges, such as limited time, insufficient training, and a lack of accessible tools, preclude its practical application in clinical settings. To address these barriers, we developed and evaluated a clinical training program for Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA; Casilio et al., 2019), a time-efficient and psychometrically robust transcriptionless system for quantifying a range of speech-language impairment features.

Method: A collaborative working group was established, comprising five researchers with expertise in discourse analysis in aphasia and eight SLPs who actively worked with individuals with aphasia in clinical or research roles. Through this partnership, SLP members provided feedback regarding their training needs, which informed the creation of a comprehensive training program that included a manual, webinar, and hybrid training sessions. Program effectiveness was measured by comparing SLPs' interrater agreement on six discourse samples before and after the training relative to expert consensus ratings, as well as posttraining perceptions on the program. Adoption of APROCSA into SLPs' work settings at approximately 1 year posttraining was also quantified.

Results: Interrater agreement improved as a function of training and newer clinicians appeared to benefit the most from the training. Posttraining perceptions were complementary to the interrater agreement findings, with nearly all SLPs indicating that the training was very helpful. At 1 year posttraining, most SLPs reported using APROCSA, primarily for assessment purposes, highlighting its effectiveness in helping them characterize spoken discourse in clinical contexts.

Conclusions: The APROCSA training program addresses key barriers to implementation, with preliminary evidence supporting its effectiveness and adoption into clinical work settings. Results overall support APROCSA's feasibility in clinical use and emphasize the benefits of targeted training for clinical discourse assessment.

Spoken discourse (or connected speech)—the use of naturalistic language beyond single words and sentences for specific meaningful purposes—is integral to everyday communication and ubiquitously impaired in individuals with aphasia (Armstrong, 2000; Halliday & Matthiessen,

2004). These discourse impairments can detrimentally affect social communication, life participation, and return to work (Hartman-Maeir et al., 2007; Haun et al., 2008), ultimately decreasing the quality of life of those living with aphasia (Dipper et al., 2021).

Recently, the evaluation of spoken discourse in aphasia has been gaining widespread empirical and clinical attention, as it provides a practical and comprehensive way to analyze language and quantify treatment outcomes (e.g., Brisebois et al., 2022; DeDe & Hoover, 2021; Dutta, Murray, Stark, & Bryant, 2024; Stark et al.,

Correspondence to Marianne Casilio: marianne.e.casilio@vanderbilt.edu and Manaswita Dutta: mdutta@pdx.edu. **Disclosure:** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

2023). Spoken discourse analysis reveals impairments across multiple levels of speech and language processing including discourse microstructure (e.g., lexical-semantic access, phonological encoding, morphosyntactic processing), macrostructure (e.g., cohesion, coherence, main concepts), and use (e.g., turn-taking, topic maintenance, repair; Armstrong, 2000; Dipper et al., 2021; Linnik et al., 2016; Prins & Bastiaanse, 2004). Traditional aphasia batteries or standalone measures (e.g., confrontation naming, word and sentence repetition) often fail to capture all the integrated complexities of speech and language processing, thus limiting their relevance to everyday functional communicative interactions (Bryant et al., 2016; Cavanaugh & Haley, 2020; Fromm et al., 2017; Holland et al., 2019). Accurate evaluation of spoken discourse in aphasia, therefore, is paramount to developing a comprehensive and person-centric profile of strengths and weaknesses, which is necessary to providing effective interventions and tailored education.

Implementation of spoken discourse analysis in aphasia in clinical settings, however, remains limited (Dutta, Murray, Stark, & Bryant, 2024; Kim & Wright, 2020; Prins & Bastiaanse, 2004; Stark & Dalton, 2024; Stark et al., 2020). Multiple recent studies have shown that clinicians consistently report time constraints as the most prominent barrier to clinical implementation, although other common barriers include (a) inadequate training, knowledge, and skills and (b) inconsistent access to tools and relevant resources across medical settings and clinical sites due to financial constraints (Bryant et al., 2017; Cruice et al., 2020; Maddy et al., 2016; Stark et al., 2021). When discourse analysis is implemented in clinical settings, practice patterns tend to diverge from those in research settings. Specifically, the majority of clinicians report using real-time transcription or direct analysis of discourse output without formal transcription (Stark et al., 2021). This divergence further recapitulates the fact that many existing discourse analysis systems are impractical for use in fast-paced clinical settings, given high caseloads and pressures to meet productivity expectations (de Riesthal & Diehl, 2018; Dietz & Boyle, 2018; Wallace et al., 2017). Thus, successful implementation of spoken discourse analysis in aphasia necessitates the use of analytic systems that are well positioned to be easily adopted into clinical work settings.

Two primary approaches are used to analyze spoken discourse in aphasia: transcription-based and transcriptionless analysis systems. *Transcription-based* systems involve orthographic and phonemic transcription of an elicited discourse sample, followed by extensive coding—either manual (Dalton & Richardson, 2019; Richardson & Dalton, 2016) or automated (MacWhinney, 2011)—that allows for the extraction of various speech-language features thought to be reflective of impairment across multiple discourse levels. Although transcription-based systems are widely used

in research (e.g., Dalton & Richardson, 2024; Vermeulen et al., 1989; Wilson et al., 2010), due to their relative objectivity and precision, transcription and coding is a lengthy process (Prins & Bastiaanse, 2004; Stark & Dalton, 2024) that is prohibitive to everyday clinical constraints, as noted above. In contrast, *transcriptionless* systems are inherently more time efficient (Stark et al., 2021), making use of rating scales or listener judgments to circumvent the need for laborious transcription and coding procedures. Although commonly used in clinical settings via their inclusion in common aphasia batteries (e.g., the fluency scale of the Western Aphasia Battery–Revised or the multifeature scales in the Boston Diagnostic Aphasia Examination or the Quick Aphasia Battery [QAB]; Goodglass & Kaplan, 1972; Kertesz, 2007; Wilson et al., 2018), these systems are not without drawbacks. Specifically, transcriptionless systems can introduce subjectivity, with scores sometimes varying considerably among raters (Gordon & Clough, 2022; Trupe, 1984), and they often only evaluate a limited range of features that do not account for all aspects of language processing or concomitant impairments in speech motor processing (e.g., apraxia of speech, dysarthria). Moreover, most transcriptionless systems were developed for stimulus-bound elicitation tasks, such as picture description (e.g., Kertesz, 2007; Richardson & Dalton, 2016) or narrative retell (Linnik et al., 2022), which are more removed from everyday speech-language use and tend to capture task-specific effects (Stark, 2019; Stark & Fukuyama, 2021).

We recently developed Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA), a novel approach for spoken discourse analysis (Casilio et al., 2019). Here, 27 commonly occurring microstructural impairment features (e.g., semantic paraphasias, paragrammatism) are each rated on a 5-point scale (see Figure 1A and Table 1). APROCSA is among the most psychometrically robust transcriptionless systems currently available (Stark & Dalton, 2024), possessing good-to-excellent interrater agreement and concurrent validity (Casilio et al., 2019, 2025). It is both clinically feasible and accessible, taking no more than 15 min to complete and requiring no specialized software or equipment. Additionally, APROCSA is among the most comprehensive transcriptionless systems, allowing for quantification of impairment across multiple speech and language domains, and is one of the few systems for spoken discourse analysis, more generally, that can be situated within established theory on spoken language production (e.g., Dell et al., 1997; Levelt et al., 1999). Specifically, its features can be distilled into four explanatory dimensions of performance as per an exploratory factor analysis (Casilio et al., 2019): paraphasia (misselection of words and sounds), logopenia (paucity of words), agrammatism (grammatical omissions), and motor

Figure 1. Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA). (A) Features and rating definitions of APROCSA, as described in detail in Casilio et al. (2019). (B) Neuroanatomical correlates of the four dimensions comprising APROCSA (top portion) and a schematic of a brain–behavior model of spoken discourse in aphasia, as measured using APROCSA (Casilio et al., 2025). Images were adapted from Figures 4 and 6 in Casilio et al. (2025).
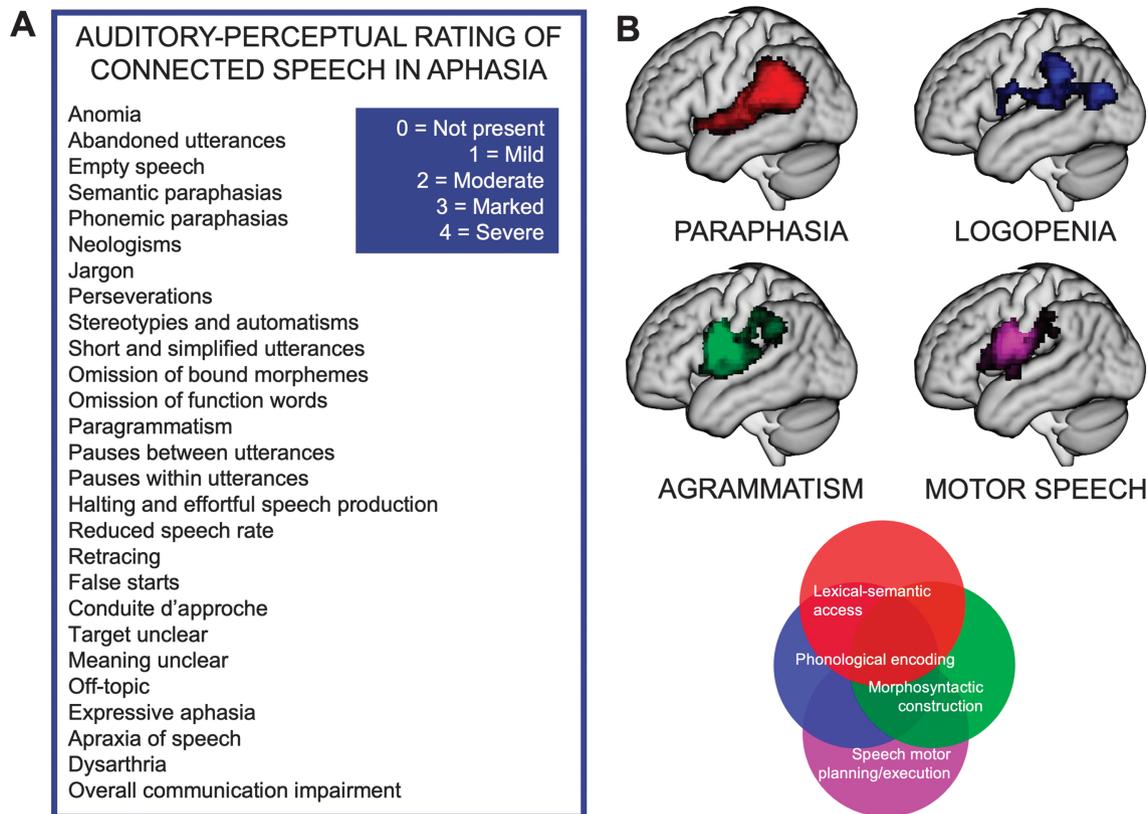
speech (impaired speech motor programming and execution). These four dimensions are associated with distinct brain regions across the language network, thus constituting the only brain–behavior model of spoken discourse in aphasia (Casilio et al., 2025; see Figure 1B).

Overall, the collective evidence related to APROCSA illustrate its strong potential as a future clinical tool (Casilio et al., 2019, 2025; Dutta et al., 2025; Ezzes et al., 2022; Stark et al., 2021) as also identified in a recent scoping review (Stark & Dalton, 2024). However, the creation of a training program is a necessary precursor to using APROCSA in clinical settings (Casilio et al., 2019).

The aim of the current project was to create and evaluate a training program for APROCSA to support its clinical implementation. Specifically, we aimed to (a) codesign materials and a protocol for the program with practicing clinicians; (b) evaluate the program's effectiveness by measuring clinicians' interrater agreement relative to expert consensus scores when using APROCSA before and after training, as well as their qualitative perceptions; and (c) quantify clinicians' adoption of APROCSA into

clinical practice. A secondary aim was to identify areas for future development in support of broader dissemination of the training program.

Method

The current project took a mixed-methods approach and was conducted within a well-established theoretical framework for implementation science, as described below.

Implementation Framework

Although not yet widely adopted in the field of speech-language pathology (Douglas et al., 2022), application of one or more theoretical frameworks is highly recommended for implementation-focused studies (e.g., Esmail et al., 2020; Tabak et al., 2012). We chose to situate our larger aim of facilitating the implementation of APROCSA in clinical practice within the RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) framework (Glasgow et al., 1999; Holtrop et al., 2021; https://www.

Table 1. Auditory-Perceptual Rating of Connected Speech in Aphasia features and abbreviated definitions.

Feature	Definition
Anomia	Overall impression of word-finding difficulties.
Abandoned utterances	Utterances are left incomplete.
Empty speech	Speech conveys little or no meaning.
Semantic paraphasias	Substitution of content words for other content words.
Phonemic paraphasias	Substitution of a clearly articulated phoneme or syllable.
Neologisms	Word forms that are not real English words.
Jargon	Fluent and prosodically correct but meaningless speech.
Perseverations	Repetition of words or utterances in inappropriate contexts.
Stereotypies and automatisms	Words, neologisms, or phrases produced with relative fluency.
Short and simplified utterances	Utterances are reduced in length or complexity.
Omission of bound morphemes	Bound morphemes are not used when they should be.
Omission of function words	Function words are not used when they should be.
Paragrammatism	Inappropriate juxtapositions of phrases and misuse of words.
Pauses between utterances	Unfilled or filled pauses that occur between utterances.
Pauses within utterances	Unfilled or filled pauses within utterances.
Halting and effortful	Speaking is labored and consequently uneven.
Reduced speech rate	The number of words per minute is reduced.
Retracing	Repetitions or revisions of one or more complete words.
False starts	Partial words are abandoned after one or two phonemes.
Conduite d'approche	Successive attempts at producing a target known to the speaker.
Target unclear	Features (place, manner, or voice) of phonemes are distorted.
Meaning unclear	It is not clear what the speaker is talking about.
Off-topic	It is not clear how the speaker's utterances relate to the context.
Expressive aphasia	Language production is disrupted.
Apraxia of speech	Speech is slow with atypical prosody or distorted substitutions.
Dysarthria	Speech is slurred, choppy, or mumbled.
Overall comm. impairment	The speaker is impaired in conveying their message.

Note. comm. = communication.

re-aim.org). Among the most widely applied frameworks for bridging the research–practice gap (Gaglio et al., 2013; Glasgow et al., 2019), RE-AIM provides a clear and well-specified roadmap for initiating, implementing, and maintaining health care initiatives (see Figure 2 for overview). In the context of the current project, our primary implementation aim (Initiative) was to train speech-language pathologists (SLPs) working with individuals with aphasia in APROCSA. Given the preliminary nature of our training program, we focused on the first three components of RE-AIM: (a) Reach, or the individuals targeted and participating in the initiative; (b) Effectiveness, or the impact of the initiative on clinical practice; and (c) Adoption, or the individuals and clinical settings implementing the initiative. These are outlined in Figure 2 and discussed in greater detail below.

Initiative

APROCSA Clinical Training Program

The APROCSA clinical training program was based directly on SLPs' perceptions of their training needs for an auditory-perceptual discourse system and was an expansion of initial materials created for training student raters in APROCSA (Casilio et al., 2019). All materials for the program are freely available on AphasiaBank (MacWhinney et al., 2011) at https://media.talkbank.org/open/APROCSA-training/index.html (see Data Availability Statement below for additional information).

Training program materials. A comprehensive training manual and a recorded webinar were created for the purposes of the present project using the information learned from responses to the survey outlined under the Reach section below. Consensus ratings on each of APROCSA's 27 features from a discourse sample of a single participant with aphasia were also derived.

Manual: We created a 21-page document that provided (a) comprehensive background information about both discourse and psychometrics of auditory-perceptual assessment; (b) detailed information about APROCSA, including its assessment approach, intended use, psychometric properties, and administration and scoring procedures; (c) updated definitions and examples for each of APROCSA's

Figure 2. Implementation framework. The top portion shows a graphical summarization of the five components of the implementation framework RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance; Glasgow et al., 1999). The components in color were those addressed in the current project; the components in gray were not addressed. The bottom portion provides an overview of the specific implementation initiative of the current project and how each component was addressed. APROCSA = Auditory-Perceptual Rating of Connected Speech in Aphasia.
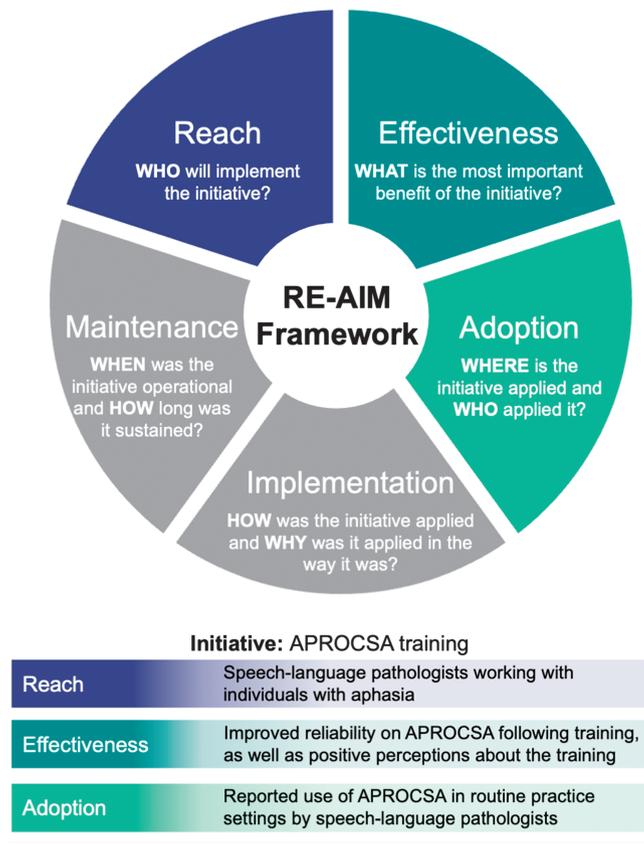
27 features; (d) a printable one-page copy of the scoring protocol; and (e) references to relevant research articles (see Figure 3A). The updated manual served as a comprehensive resource for both training in APROCSA and its application once training was completed.

Recorded webinar: We also created a 90-min recorded webinar (narrated by author M.C.), which consisted of a slide deck that first provided a brief overview of APROCSA and then focused on defining each of its 27 features (see Figure 3B). Features were grouped by dimension based on our prior work (Casilio et al., 2019, 2025) as a way of guiding SLPs in learning about behaviors that frequently co-occur with one another and have shared neural substrates.

Each feature definition was accompanied by a brief (~30–60 s) audiovisual example from discourse samples of one of eight individuals with chronic poststroke aphasia

who were recruited as part of a larger prior study (Wilson et al., 2018). The eight participants (of 16 total participants) were selected for their diverse yet scorable behavioral presentations, as quantified on the QAB (see Table 2 and supplementary materials of Wilson et al., 2018). Discourse samples from the eight participants were elicited by a trained SLP using the QAB's protocol (Wilson et al., 2018), which uses a bank of semistructured interview questions. This prior study was approved by the institutional review board at The University of Arizona (protocol # 1000000983) and all participants provided informed written consent.

Consensus ratings: We recruited one participant with chronic poststroke aphasia as part of a larger research project on recovery from aphasia after stroke (Wilson et al., 2023). This study was approved by the institutional review board at Vanderbilt University Medical Center (protocol # 160847) and the participant provided informed written consent, which included the free sharing of his audiovisual-recorded behavioral testing. We recruited this participant specifically for the development of this clinical training program as he presented with varying degrees of impairment across all language domains (see row pertaining to participant aprocsa2364a in Table 2 for his demographic and clinical background).

Spoken discourse was elicited by author M.C. following AphasiaBank's protocol (MacWhinney et al., 2011). The first 5 min of participant speech from the Free Speech portion of the elicitation protocol were clipped for analysis, following our prior work in APROCSA (Casilio et al., 2019). Consensus ratings were then derived collectively by the five researcher members of the working group. This involved the researchers watching the sample twice without pausing, independently rating all of the features, and then reviewing these independent ratings as a group and arriving at a consensus rating in the case of any disagreements (see Ezzes et al., 2022, for a complete description of this process). A brief rationale for each feature's consensus rating, including examples from the sample, was also created by the researcher members of the group to support clinicians' learning.

Training program structure. We created a structured procedure for the training program that all SLPs completed. First, SLPs spent approximately 2 hr independently reviewing the training materials (i.e., reading the manual and watching the recorded webinar). Then, SLPs participated in two 45-min hybrid synchronous training sessions facilitated by the researchers. During these sessions, SLPs completed a guided scoring practice using the discourse sample with APROCSA consensus ratings, as described above. This involved watching the sample twice, independently rating each of APROCSA's features, and then discussing SLPs' ratings in relation to the consensus ratings. The synchronous training was scheduled to accommodate clinic schedules

Figure 3. Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA) training materials. Examples of the manual (A) and recorded webinar (B) from the APROCSA clinical training program.

A

B
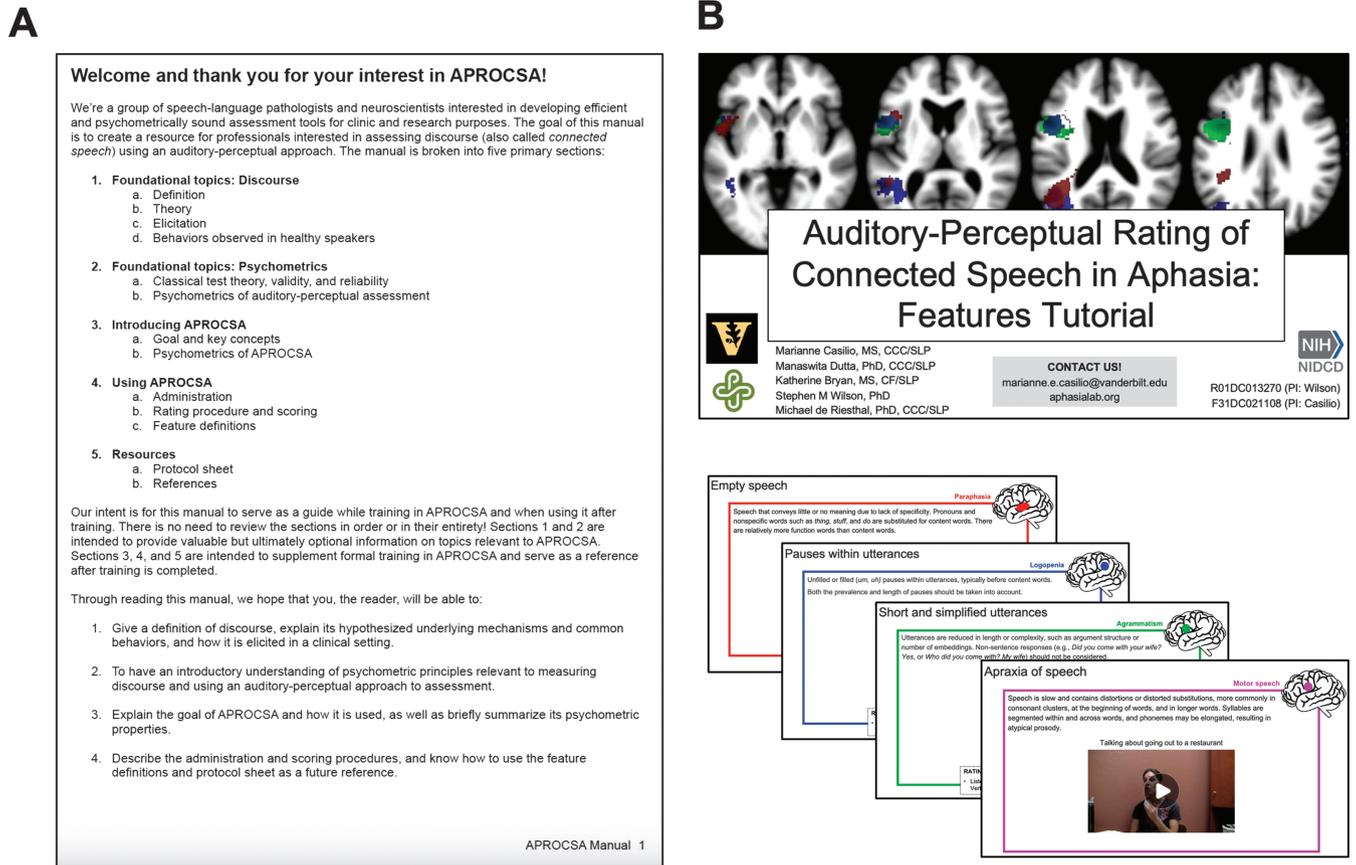
and all attendees were offered continuing education credits for their participation in these sessions.

Reach

Clinician–Researcher Working Group

As identified in our prior work (Casilio et al., 2019, 2025; Ezzes et al., 2022), we initially envisioned SLPs working with individuals with aphasia as the primary group to participate in APROCSA training. To maximize the reach of this training initiative and actively work toward bridging the divide between clinicians and researchers (Douglas et al., 2023), our APROCSA training program—as described above—was co-designed by a working group of SLPs and academic researchers.

Our working group was composed of five researchers (authors M.C., M.D., K.B., M.dR., and S.M.W.) with expertise in discourse analysis in aphasia and eight SLPs (authors K.C., Z.D., I.F., J.K., K.M., K.S., J.S., and S.S.) actively working with individuals with aphasia in either a clinical or research capacity. Researchers involved in clinical supervision (authors M.D. and M.d.R.) at their respective institutions (Portland State University and Vanderbilt University Medical Center) acted as liaisons for connecting researchers and clinicians. Relevant demographic information about members of the working group are displayed in Table 3.

Identification of Training Needs

Procedure. Prior to creating the clinical training program, researchers queried SLP members of the working group about their perceived needs for an APROCSA training program. This query, which was created by authors M.C. and M.dR., took the form of an online pretraining survey in Research Electronic Data Capture (REDCap; Harris et al., 2009). The survey solicited feedback from the respondents on the development of a spoken discourse training program using APROCSA, including information regarding training content, training activities, and training logistics. Question formats included multiple-choice, yes/no, rankings, and open-ended response options.

Table 2. Discourse sample details.

Participant	Age (years)	Sex	Race	Education (years)	Postonset (months)	QAB overall (/10)	Training purpose	Consensus scores
aprocsa1738a	72	Male	White	14	120	7.72	Pre-/postratings	Yes
aprocsa1944a	71	Female	Black	16	151	7.69	Pre-/postratings	Yes
aprocsa1713a	63	Female	White	14	23	8.84	Pre-/postratings	Yes
aprocsa1554a	46	Female	White	15	35	7.96	Pre-/postratings	Yes
aprocsa1833a	67	Male	White	14	18	7.52	Pre-/postratings	Yes
aprocsa1731a	48	Male	White	18	52	3.74	Pre-/postratings	Yes
aprocsa2364a	70	Male	White	20	247	5.44	Scoring practice	Yes
0131	72	Male	White	18	118	4.91	Training webinar	No
0182	71	Male	White	12	83	4.02	Training webinar	No
0190	55	Male	White	13	63	7.87	Training webinar	No
0199	79	Male	White	18	61	5.31	Training webinar	No
0308	54	Male	White	16	42	5.70	Training webinar	No
1295	32	Female	White	14	9	7.41	Training webinar	No
1351	52	Male	White	12	12	6.49	Training webinar	No
1356	32	Female	White	19	6	3.15	Training webinar	No

Note. All samples are or will be available on AphasiaBank. QAB = Quick Aphasia Battery.

Analysis. Survey responses were extracted from REDCap and evaluated qualitatively using descriptive statistics.

Effectiveness

Pre- and Posttraining Interrater Agreement

Procedure

To quantify the effectiveness of our clinical training program, clinicians used APROCSA both pre- and post-training to evaluate six discourse samples from participants with chronic poststroke aphasia (see Table 2). These participants were recruited as part of a prior study, which had the explicit intent of developing a bank of discourse samples to be used for training purposes for APROCSA (see Table 1 of Ezzes et al., 2022, for additional clinical and demographic information). In line with the procedures for the recruited participant described above, this study was approved by the institutional review board at Vanderbilt University Medical Center (Protocol # 160847) and provided informed written consent. More specifically, all participants from this study consented to having their audiovisual recordings of behavioral testing freely shareable;

Table 3. Demographic information of participants in the clinician–researcher working group.

Demographic information	Responses
Clinicians (*n* = 8)	
Years practicing as an SLP (*M ± SD*)	6.61 ± 4.53
Proportion of caseload involving individuals with aphasia	0%–20% (*n* = 2; 37.5%) 21%–40% (*n* = 3; 37.5%) 41%–60% (*n* = 2; 25%) 81%–100% (*n* = 1; 12.5%)
Approximate number of aphasia assessments performed per month	0–2 (*n* = 2; 37.5%) 3–5 (*n* = 5; 62.5%) 6–8 (*n* = 1; 12.5%)
Of the aphasia assessments performed, the amount of time discourse is assessed	61%–80% (*n* = 3; 37.5%) 81%–100% (*n* = 5; 62.5%)
Number of SLPs who incorporate discourse in aphasia treatment	Yes (*n* = 7; 87%) No (*n* = 1; 12.5%)
Researchers (*n* = 5)	
Years of aphasia research experience (*M ± SD*)	12.8 ± 8.81
Years of aphasia clinical experience (*M ± SD*)	8.4 ± 9.65
Years of aphasia instruction experience (*M ± SD*)	6.6 ± 6.94

Note. SLP = speech-language pathologist.

underwent the AphasiaBank discourse elicitation protocol; and had APROCSA consensus ratings from a group of expert researchers, as derived from clips containing the first 5 min of participant speech from the Free Speech portion of the AphasiaBank protocol (see Table 2 of Ezzes et al., 2022, for full details regarding sample time stamps and feature ratings).

Pre- and posttraining rating procedures for SLPs were identical to those previously described for APROCSA (Casilio et al., 2019). In brief, SLPs were instructed to listen to and watch each audiovisual clip twice and rate all language impairment features based on the criteria outlined in the APROCSA manual. They were permitted to take a brief break between the first and second viewing; however, they were advised not to pause the video during either viewing. Note taking, either during or between views, was encouraged. Ratings for a given sample were to be completed within a 15-min timeframe and all samples were to be rated within a 2-week time frame.

Analysis

For interrater agreement, we evaluated the extent to which SLPs, either individually or as a group, agreed with the consensus ratings to determine the extent to which clinical practitioners could approximate the performance of experts and whether this improved as a function of the training program. This was done using two-way random-effects intraclass correlation coefficients of absolute agreement (ICC(A,1); McGraw & Wong, 1996). Here, an ICC—across all 27 APROCSA features and all six discourse samples—was calculated for each clinician at both pretraining and posttraining. Bootstrapped 95% confidence intervals (CIs) for each ICC were also calculated with a bias-corrected accelerated method (DiCiccio & Efron, 1996; Efron, 1987) from 10,000 iterations of resampled data. ICCs were also averaged across the eight SLPs to obtain a group-level measure of agreement. To test for the effect of training, we took a nonparametric approach given that our sample size was small ($n = 8$) and the noninterval nature of ICCs (i.e., bounded from 0–1), and tested whether there was a median increase in ICCs between pre- and posttraining using a one-tailed Wilcoxon signed-ranks test (Higgins, 2004) at $p < .05$. Coefficients were interpreted qualitatively as poor (ICC < .40), fair (.40 ≤ ICC < .60), good (.60 ≤ ICC < .75), or excellent (ICC ≥ .75; Cicchetti, 1994). There were no missing data aside from ratings from one SLP (I.F.) for one of the six discourse samples at pretraining, which was accidentally skipped. All SLPs completed the pre- and posttraining ratings. Given the use of averaging across both samples and ICCs, coupled with ICCs' relative flexibility regarding the handling of missing data (Hallgren, 2012), no methods for imputing or deleting missing data were taken.

We additionally explored whether, in the case of SLP agreement with consensus ratings, certain aspects of their clinical background may be associated with their response to the training program. To do this, we correlated each SLP's change score in pre- and posttraining ICCs with (a) their number of years of clinical practice and (b) their pretraining rating of confidence in spoken discourse analysis (1–5, where higher scores indicated greater confidence). Again, accounting for our sample size and the nature of our ICC parameter estimates, we quantified this association using a nonparametric approach, specifically Kendall's rank correlations (Higgins, 2004) with bootstrapped 95% CIs, calculated in the same manner as described above.

Posttraining Perceptions

Procedure

Within 1 month of the synchronous training sessions, research members of the working group asked the SLP members to complete a posttraining survey, where they provided feedback on the training program and suggestions for future improvements. This survey was also disseminated via REDCap and contained similar question formats to the pretraining survey.

Analysis

As with the pretraining survey, responses were extracted from REDCap and then evaluated using descriptive statistics.

Adoption

Procedure

Approximately 1 year after completion of the training program, research members of the working group asked the SLP members to share information about their adoption of APROCSA into their work settings. As with the pre- and posttraining surveys, this information was obtained using a REDCap survey and included a variety of different question formats.

Analysis

Similar to the pre- and posttraining survey, responses were extracted from REDCap and then evaluated using descriptive statistics.

Results

Reach

Clinician–Researcher Working Group

Both researcher and SLP members of the working group were well situated to codesign a training program

for APROCSA (see Table 3). Regarding the research members, all were affiliated with universities with many years of research, clinical, and/or educational experience in aphasia. Regarding the SLP members, the majority endorsed having a substantial number of individuals with aphasia on their clinic caseload or research recruitment pipeline and reported frequent incorporation of spoken discourse into both their assessments and their interventions. Importantly, 75% (n = 6) of SLPs reported that targeting spoken discourse was "very important" for improving communication skills in individuals with aphasia, with the remaining 25% (n = 2) reporting it is "fairly important."

Identification of Training Needs

The pretraining survey was used to gather initial insights from participating SLPs, which informed the development of the APROCSA training materials. In terms of content to be included in the program, SLPs reported a mix of potential topics (see Table 4). As such, the training manual, as described above, included information on all possible topics. With regard to activities to be included in the training program, there was a clearer delineation of priorities (see Table 4). The majority of SLPs reported that a guided practice in rating with audiovisual examples would be the most useful, followed by didactic instruction on different discourse features; the remaining possible activities were generally considered to be of lower priority in terms of usefulness. Consequently, the training program consisted primarily of the (a) recorded webinar and (b) guided scoring practice described above. Lastly, pertaining to training logistics, 75% of SLPs (n = 6) reported that 3–4 hr of training was a reasonable time commitment for them given their schedules, although opinions

were mixed as to whether the training should be synchronous (n = 5) or asynchronous (n = 3). Consequently, we chose to create a training that was ~3 hr in duration with a mix of synchronous and asynchronous activities.

Effectiveness

#### Pre- and Posttraining Interrater Agreement

Before training, raters were in a *good* range. They had an average pretraining ICC of .704, reflecting the already high level of expertise and familiarity with discourse analysis among the eight SLP members (see Table 3). Training resulted in a statistically significant increase in clinicians' interrater agreement with the consensus ratings (V = 32, p = .027); the average posttraining ICC was .728, a .024 increase from pretraining (see Figure 4A). Posttraining interrater agreement continued to be in the *good* range.

Regarding interrater agreement for each SLP, the majority (six of the eight) showed an increasing trajectory in their agreement with the consensus ratings; the remaining two SLPs' performance was stable when considering their 95% CIs and had ICCs in the *good* range at both pre- and posttraining (see Figure 4A and Table 5). Of the SLPs with an increasing trajectory in ICCs, the majority (four of the six) were in the *good* range at both pre- and posttraining. These other two SLPs showed remarkable improvement as a function of training. One SLP (J.S.) had an ICC increase of .106, moving from the *fair* range pretraining to the *good* range posttraining; another SLP (S.S.) had an ICC increase of .099, with interrater agreement in the *excellent* range both pre- and posttraining. Notably, both of these SLPs' posttraining ICCs were above the upper bound of their 95% CIs at pretraining

Table 4. Speech-language pathologists' report of training program needs.

Training content	Most useful	Second most useful	Third most useful	Fourth most useful	Fifth most useful
Theories of discourse and spoken language production	1	0	6	1	0
Literature review on the importance of discourse in aphasia assessment/treatment	1	3	0	3	1
Methods and techniques for eliciting or collecting discourse samples	4	3	0	1	0
Auditory-perceptual diagnostic procedures for neurogenic communication disorders	2	2	0	2	2
Psychometric theory and determining whether a behavioral test is sufficiently normed or validated for clinical practice	0	0	2	1	5
Training activities					
Guided practice in rating using audiovisual examples	6	1	1	0	0
Q&A or "office hours" with trained experts in discourse assessment	0	0	1	3	4
Didactic education (e.g., lectures, readings) on different discourse behaviors	1	4	0	2	0
Small group scoring activities	1	1	4	1	1
Individualized feedback on scoring	0	2	2	1	3

Figure 4. Clinician agreement with consensus ratings. (A) Violin plots of the pre- and posttraining intraclass correlation coefficients (ICCs), where the dashed horizontal lines depict the lower and upper bounds of qualitatively *good* agreement per established guidelines by Cicchetti (1994). ICC(A,1) refers to the type of ICC, as described in the main text (McGraw & Wong, 1996). Results of the Wilcoxon signed-ranks test comparing pre- and posttraining agreement are shown in the upper-left corner. (B) Bivariate scatterplots showing the association between change in ICCs from pre- to posttraining and two clinical variables: (a) years of clinical practice experience (left) and (b) self-reported pre-training confidence in spoken discourse analysis, as rated on a 5-point scale (lower scores indicate greater confidence). The black line shows the association trend and results of the Kendall's rank correlations are shown in the upper-left corner. In both panels, color-coded dots reference each SLP (see legend for cross-referencing with rater initials).
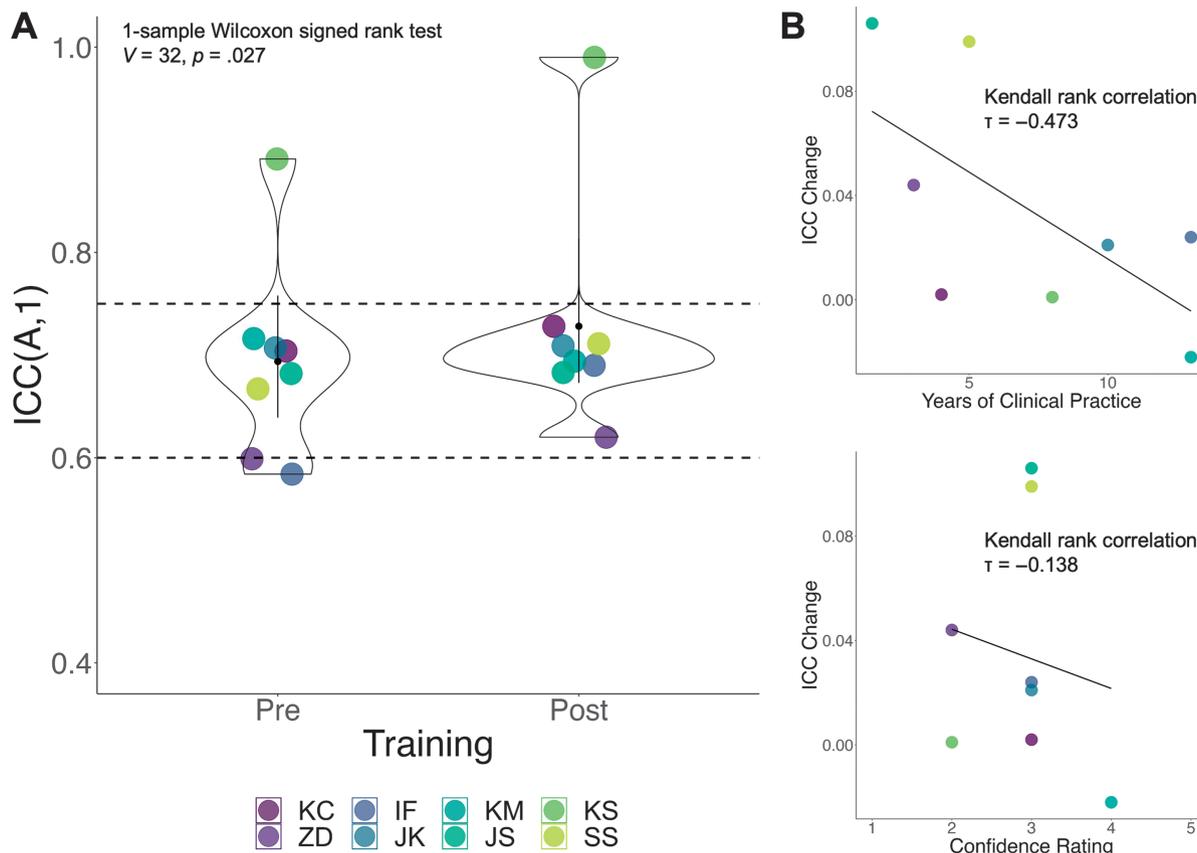
(see Table 5), indicating their improved performance is not the byproduct of measurement error.

Kendall's rank correlations (see Figure 4B) suggested that newer SLPs' agreement with the expert consensus may be more likely to improve with training, although this association was variable at the group level ($T = -0.473$, 95% CI [$-0.917, 0.333$]). At the individual level, the two SLPs with significantly improved agreement with consensus ratings were among the less experienced clinicians in the working group; both had been practicing for 5 years or less. Pre-training ratings of confidence in spoken discourse analysis, however, had negligible influence on improvement following training ($T = -0.138$, 95% CI [$-0.808, 0.722$]).

Posttraining Perceptions

Seven SLPs (88%) found the APROCSA training program to be "very helpful"; one respondent found it to be "fairly helpful." In terms of program feedback, SLPs reported that the feature descriptions, rating strategies, and video examples were particularly valuable in learning APROCSA. Four of the eight SLPs (50%) did report that some APROCSA features and their definitions—paragrammatism (two SLPs), retracing (one SLP), false starts (one SLP), conduite d'approche (one SLP), and target unclear (one SLP)—could benefit from further clarification or explanation. Some SLP group members also suggested providing more audiovisual examples, along with consensus scores, to help differentiate between points on the rating scale.

Adoption

At approximately 1 year postcompletion of the clinical training, six of the eight SLP members of the working group reported using APROCSA. The remaining two SLPs stated they were not using APROCSA because they

Table 5. Speech-language pathologist (SLP) interrater agreement with Auditory-Perceptual Rating of Connected Speech in Aphasia consensus scores before and after training.

Rater initials	Pretraining		Posttraining	
	ICC(A,1)	95% CI	ICC(A,1)	95% CI
I.F.	.704	[0.599, 0.790]	.728	[0.648, 0.794]
J.K.	.599	[0.473, 0.709]	.620	[0.503, 0.711]
J.S.	**.584**	**[0.469, 0.683]**	**.690**	**[0.582, 0.784]**
K.C.	.707	[0.608, 0.787]	.709	[0.615, 0.780]
K.M.	.716	[0.625, 0.792]	.694	[0.590, 0.781]
K.S.	.682	[0.564, 0.776]	.683	[0.574, 0.774]
S.S.	**.891**	**[0.799, 0.946]**	**.990**	**[0.971, 0.997]**
Z.D.	.667	[0.549, 0.760]	.711	[0.595, 0.794]

Note. Bold font face indicates cases where an SLP's posttraining ICC was beyond the upper bound of the 95% confidence interval of the pretraining ICC. ICC = intraclass correlation coefficient; CI = confidence interval.

either (a) worked in a clinical setting where the caseload had very few clients with aphasia (J.S.) or (b) worked in a research setting where APROCSA was not part of their data collection protocol (S.S.), although APROCSA will soon be adapted into their protocol in part due to the development of the training program described here. Of the six SLPs using APROCSA, all reported using it for assessment purposes. One SLP reported using it 80%–100% of the time. Two SLPs each reported using it 61%–80% of the time and 41%–60% of the time. The remaining SLPs reported using it 21%–40% of the time. Use of APROCSA for treatment monitoring purposes was relatively less: four SLPs reported using it 21%–40% of the time while the remaining two reported using it 0%–20% of the time.

SLPs also qualitatively shared their perspectives on how APROCSA has been useful to them in the year following training (see Table 6). Their appraisal of

Table 6. Examples of speech-language pathologist perspectives on the usefulness of Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA) since completing the clinical training program.

"It really helps me think through how this person's aphasia is presenting and our specific areas for improvement." — K.C.
"I love using APROCSA to describe features of my patient's discourse skills. It gives me new terms and language . . . Evaluating speech samples through APROCSA lenses helps me capture a more complete picture of my client's speech-language abilities." — J.K.
"It has helped me describe behaviors associated with aphasia to families and other professions." — K.S.
"The features and definitions on the APROCSA allow the graduate clinicians to listen for and document features that they would not otherwise recognize." — I.F.
"It is much easier to qualify and quantify expressive language deficits for milder aphasias." — K.M.

APROCSA was overwhelmingly positive. Most responses reflected the ways in which APROCSA provided them with a new and useful method by which to characterize spoken discourse features in a variety of contexts relevant to clinical practice. Among the many contexts identified were clinical assessment, family counseling, interprofessional communication, and student instruction. As part of these qualitative reports, SLPs also identified potential expansions that could maximize APROCSA's utility in their work settings. These included providing ready access to the training materials, creating a one-page handout or example reports using APROCSA features, and developing a short-form version of the tool.

Discussion

This project aimed to develop and evaluate a training program for APROCSA, a validated and time-efficient transcriptionless system for spoken discourse analysis in aphasia (Casilio et al., 2019). The resulting program included a detailed manual, a webinar, and hybrid training sessions and was designed with input from a clinician–researcher working group and feedback from SLPs on their training needs. Effectiveness of the training program was evaluated both quantitatively (i.e., change in interrater agreement as a function of training) and qualitatively (i.e., posttraining perceptions of the training program). Adoption of APROCSA into SLPs' work settings was also quantified. Overall, we found the training program—the first of its kind to be systematically evaluated for spoken discourse analysis in aphasia—resulted in improved interrater agreement relative to consensus ratings. This improvement appeared greatest for SLPs with fewer years of clinical experience, although all SLP members of the working group reported a qualitative benefit from undergoing the training program. Notably, the majority of SLPs endorsed posttraining adoption of APROCSA into their clinical work settings and nearly all reported using APROCSA for most of their aphasia assessments.

Applying the RE-AIM Framework

The RE-AIM framework was used in this project considering it is specifically designed to evaluate implementation outcomes and improve the adoption of best practices in clinical settings (Glasgow et al., 2019). This project provides the first initial evidence of systematically integrating RE-AIM (Glasgow et al., 1999), a well-established implementation science framework, into clinical research in speech-language pathology (Douglas et al., 2022). Our application of this framework suggests that it is applicable to the field and could prove to be a

productive means of carrying out implementation-focused research.

Initiative

To our knowledge, the training program we developed for APROCSA represents the first systematic characterization of a clinical implementation initiative for spoken discourse analysis in aphasia (Stark & Dalton, 2024). Although other discourse analysis training systems have been created, such as those available online for main concept analysis (Richardson & Dalton, 2016) or Computerized Language Analysis (MacWhinney et al., 2011), these have yet to be evaluated with regard to their effectiveness or feasibility of adoption into clinical practice.

Related research on training students and health care professionals in conversation partner training, an indirect intervention for both aphasia and acquired brain injury, has systematically evaluated different training program initiatives (Beeke et al., 2013; Kagan, 1998; Lock et al., 2001; Power et al., 2020, 2024; Power & Morrow, 2024; Rietdijk et al., 2019; Smith et al., 2011; Togher et al., 2011). These initiatives have many overlapping components with ours in that they often were relatively short in duration (e.g., Power et al., 2020, 2024) or involved analogous types of training materials, such as a manual or hands-on practice (e.g., Kagan, 1998; Lock et al., 2001). Findings from these studies were also similar, with significant posttraining improvements in perceptions of competence, as measured on a validated survey (Power et al., 2021), that were maintained for at least a few months (e.g., Power et al., 2020, 2024). It consequently stands to reason that training initiatives more generally in aphasia, as was the case in the current project and this prior body of work, yields positive outcomes.

Reach

The development of our APROCSA training program was driven by a strong clinician–researcher partnership. Recognizing the value of engaging stakeholders in aphasia rehabilitation (Cruice et al., 2022; Strong et al., 2023), we made clinician involvement a priority from the very beginning to ensure the training program was directly relevant to clinical practice (Hersh et al., 2021). Our goal was to ensure that the APROCSA training program was informed by both research evidence and practical clinical experience through the equal involvement of both groups (Wallace et al., 2017; Williamson et al., 2012). Researchers in the working group established the research agenda, whereas the SLP members played a critical role in providing essential feedback on their training needs and preferences that guided the program's development. By adopting this approach, our group afforded practicing SLPs the opportunity to provide input in research activities intended to have a direct impact on clinical practice (Douglas et al., 2023; Olswang & Goldstein, 2017). Our hope is that actively involving SLPs in training development and evaluation will increase the likelihood of APROCSA being adopted into clinical practice more broadly, as this codesign approach may increase buy-in, acceptability, approachability, and confidence.

Effectiveness

From a quantitative perspective, the training program we developed for APROCSA resulted in a statistically significant improvement in interrater agreement relative to expert consensus ratings with ~3 hr of training, a timeframe that was identified as feasible by the SLP members of the working group. This improved agreement with the consensus indicates that the training program resulted in increased mastery of certain theoretical perspectives (e.g., the view of discourse fluency as graded and multidimensional), along with improved operationalization of their underlying features (e.g., how the halting and effortful feature contributes to multiple manifestations of nonfluency; Casilio et al., 2025; Wilson et al., 2010).

It is worth noting that interrater agreement for the SLPs was already in the *fair*-to-*good* range prior to completing the training program. This finding suggests that APROCSA's features are well defined and capture characteristics of spoken discourse that are both readily observable and diagnostically meaningful. It is also consistent with interrater agreement reported for experienced clinicians without training in another auditory-perceptual rating scale for dysarthria (Bunton et al., 2007), thus suggesting that the expertise of the clinicians may play a role. Given that all SLP members of the working group were experienced in working with individuals with aphasia, we consequently would anticipate greater variability in pretraining agreement among a larger group of SLPs who are generalists or who specialize in other areas of adult medical clinical practice (e.g., voice disorders).

With regard to who benefited most from the clinical training program, exploratory correlations suggested that, at least as measured relative to consensus ratings, SLPs with fewer years of experience may experience the greatest improvement in agreement as a function of the training program. Although this finding ultimately was not significant at the group level, individual-level evidence still suggested that a subset of newer SLPs showed substantial improvements posttraining. Thus, it may be that the training program may serve as a "boost" to expertise in aphasic discourse that otherwise would have been acquired through the accumulation of everyday clinical interactions.

Interestingly, pretraining perceptions of confidence in spoken discourse analysis was not associated with

improved agreement relative to consensus ratings despite the importance of confidence to implementing discourse analysis in aphasia (Stark et al., 2021) and clinical training in speech-language pathology more generally (Edwards et al., 2023; McBride, 2022). This lack of association may be related to statistical power, given our relatively small sample size ($n = 8$), or relative lack of variability in SLPs' self-report of confidence (range: 2–4 on a 1–5 scale).

From a qualitative perspective, the program yielded positive perceptions from all SLPs regarding the information gained from completing the training activities. Thus, even though interrater agreement for some SLPs did not change substantially as a function of the training, all endorsed that undergoing the training was nonetheless valuable. Moreover, the posttraining survey also yielded valuable information about how to improve the training program in the future, a critical part of the clinical implementation process (Glasgow et al., 1999). Notably, there was a clear emphasis on the desire for more discourse examples, suggesting that clinical education—like clinical intervention—requires several exposures across contexts (e.g., identifying and quantifying paragrammatism in individuals who are fluent vs. nonfluent) to acquire new knowledge and skills.

Adoption

At approximately 1 year posttraining, the majority of SLPs in the working group endorsed using APROCSA in their work settings and those who were not were not actively working with individuals with aphasia. Although not reported directly in the survey, those SLPs also made efforts to optimize and maintain their adoption of APROCSA, including the creation of Epic SmartPhrases (a feature that allows for quick population of clinical information in a common electronic health record system). Remarkably, those using APROCSA were implementing it for the majority of their aphasia assessments, suggesting that APROCSA has the potential to not simply be a useful adjuvant but a dominant approach to clinical diagnostic practice patterns. Qualitative comments on the utility of APROCSA in clinical practice further support this interpretation, with SLPs stating that APROCSA provides them with a useful system by which to quantify behaviors relevant to everyday language use (see Table 6). This utility is supported by the empirical evidence on APROCSA outlined in the introduction, where it has been shown to be not only reliable but also useful in understanding the linguistic and neural underpinnings of spoken discourse in aphasia (see Figure 1B).

Notably, one SLP in the working group (I.F.) commented specifically on the usefulness of implementing the training program to teach graduate SLP students not only about APROCSA but about discourse in aphasia more generally. Since its development, multiple members of the working group have reported using elements of the training program to educate students in both clinical and research contexts across six academic institutions in both the United States and Australia. Thus, although the training program was designed with practicing clinicians in mind, these anecdotes suggest there is potential for expanding the program to serve clinical trainees and research assistants, filling a critical gap given the scarcity of research on clinical education more generally (Edwards et al., 2021, 2023; Tolsgaard et al., 2015) and as done in related work on conversation partner training mentioned previously (e.g., Power et al., 2020, 2024).

Limitations and Future Directions

It is important to consider some limitations of the current project. First, our working group consisted of a relatively small number of SLPs ($n = 8$) with clinical expertise in aphasia and strong buy-in for spoken discourse analysis prior to the training. Future work on the training program, as well as clinical implementation of APROCSA more generally, should ideally involve a larger group of SLPs with more varied clinical backgrounds and perspectives. Nonetheless, we view the skill and experience of all members of our working group as a strength, given our focus on codesigning a useful training program.

Second, we did not randomize the rating order of the pre- and posttraining discourse samples, thus, there is the potential that learning effects may have been a contributor to our findings on interrater agreement. Future research should consider the potential impact of learning effects in training, as well as other sample-specific factors (e.g., the overall impairment severity and profile of the individual with aphasia).

Third, although APROCSA may be a useful framework for capturing salient features of spoken discourse in aphasia across etiologies, the training program described here uses examples exclusively from poststroke aphasia. Although this is the most common etiology of aphasia, future iterations of the training program ideally should include examples from other etiologies (e.g., tumor, neurodegeneration), as well as other potentially relevant scoring considerations (e.g., behavioral or cognitive changes often observed concomitant to language decline in primary progressive aphasia).

Fourth, the training program as it is currently developed relies on multiple doctorate-level researchers experienced in aphasia to guide SLPs through the synchronous portion of the training. Although this has the advantage of allowing trainers to draw on their wealth of expertise, an important aspect of clinical education (DeRuiter & Ginsberg, 2020), having clinician trainers will be essential to broader dissemination of the training program. Thus, future work will need to evaluate whether and how a

clinician who has undergone the training could act as an APROCSA trainer (similar to the training model for Supported Conversation for Adults with Aphasia; see https://www.aphasia.ca/health-care-providers/education-training/training-programs-workshops/). In the interim, SLPs are encouraged to independently make use of the available training materials (e.g., manual, recorded webinar), which are freely accessible (see Data Availability Statement below), and contact the author team for further support.

Fifth and relatedly, SLP members of the working group aptly identified multiple areas for both improving and expanding on the training program, as noted in the Results section above. Importantly, there was a collective emphasis on having a more varied pool of discourse samples—with consensus ratings—to showcase all of the potential features on APROCSA and to calibrate to its rating scale. We concurred that a larger data set of freely shareable audiovisual discourse samples with consensus scores would be a beneficial adjuvant to the program. As such, we expanded on our open-source data set ($n = 6$; Ezzes et al., 2022) by recruiting 14 additional participants with aphasia, six of whom were tested longitudinally (see Wilson et al., 2023, for a complete description of the protocol), and obtained expert consensus scores using the same method described here and in Ezzes et al. (2022). All were recruited under a larger study approved by the institutional review board at Vanderbilt University Medical Center (protocol # 160847) and provided informed written consent that included free sharing of their audiovisual-recorded testing. It is our hope that this expanded data set will support future learners of APROCSA in calibrating to its feature set and rating scale.

The long-term vision of this work is broad dissemination of the APROCSA training program among the community of professionals serving individuals with aphasia, encompassing research personnel, clinicians, and clinical trainees. This is a significant undertaking, and we are cognizant of the potential financial and operational factors that will undoubtedly impact dissemination. Accordingly, the future development of our dissemination plan will require ongoing stakeholder engagement. Moreover, we are committed to accessibility and open science practices, and, as a result, we are committed to ensuring that all key resources supporting APROCSA will always be available to end users without financial or other barriers (see Data Availability Statement below for more information about accessing program materials).

Conclusions

Limited time and insufficient training are significant barriers to implementing spoken discourse analysis in clinical practice. Our clinical training program for APROCSA, a time-efficient and validated system, addresses both of these barriers, with initial results suggesting the program is both effective and leads to the adoption of APROCSA in relevant work settings. It is our view that the training program in its current form could be immediately effective and useful in supporting SLPs working with individuals with aphasia in both research and clinical contexts.

Data Availability Statement

APROCSA feature ratings and survey responses from SLP members of the working group, as well as analysis code, are freely available on https://github.com/mcasilio/aprocsatraining. Training program materials (manual, recorded webinar, audiovisual recording, and consensus ratings of the discourse sample used in the guided scoring practice) are available on AphasiaBank at https://media.talkbank.org/open/APROCSA-training/index.html. The expanded open-source data set and corresponding consensus ratings will soon be available as part of the APROCSA corpus on AphasiaBank at https://talkbank.org/aphasia/access/English/Protocol/APROCSA.html.

Acknowledgments

References

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

Beeke, S., Sirman, N., Beckley, F., Maxim, J., Edwards, S., Swinburn, K., & Best, W. (2013). *Better conversations with aphasia: An e-learning resource.* University College London.

Brisebois, A., Brambati, S. M., Boucher, J., Rochon, E., Leonard, C., Désilets-Barnabé, M., Desautels, A., & Marcotte, K. (2022). A longitudinal study of narrative discourse in post-stroke aphasia. *Aphasiology, 36*(7), 805–830. https://doi.org/10.1080/02687038.2021.1907295

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics, 30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology, 31*(10), 1105–1126. https://doi.org/10.1080/02687038.2016.1239013

Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. F. (2007). Listener agreement for auditory-perceptual ratings of dysarthria. *Journal of Speech, Language, and Hearing Research, 50*(6), 1481–1495. https://doi.org/10.1044/1092-4388(2007/102)

Casilio, M., Kasdan, A. V., Bryan, K., Shibata, K., Schneck, S. M., Levy, D. F., Entrup, J. L., Onuscheck, C., de Riesthal, M., & Wilson, S. M. (2025). Four dimensions of naturalistic language production in aphasia after stroke. *Brain, 148*(1), 291–312. https://doi.org/10.1093/brain/awae195

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

Cavanaugh, R., & Haley, K. L. (2020). Subjective communication difficulties in very mild aphasia. *American Journal of Speech-Language Pathology, 29*(1S), 437–448. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0222

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cruice, M., Aujla, S., Bannister, J., Botting, N., Boyle, M., Charles, N., Dhaliwal, V., Grobler, S., Hersh, D., Marshall, J., Morris, S., Pritchard, M., Scarth, L., Talbot, R., & Dipper, L. (2022). Creating a novel approach to discourse treatment through coproduction with people with aphasia and speech and language therapists. *Aphasiology, 36*(10), 1159–1181. https://doi.org/10.1080/02687038.2021.1942775

Cruice, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., & Dipper, L. (2020). UK speech and language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. *International Journal of Language & Communication Disorders, 55*(3), 417–442. https://doi.org/10.1111/1460-6984.12528

Dalton, S. G. H., & Richardson, J. D. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology, 28*(1S), 293–320. https://doi.org/10.1044/2018_AJSLP-17-0166

Dalton, S. G. H., & Richardson, J. D. (2024). Clinically feasible analysis of discourse: Current state and looking forward. In A. P.-H. Kong (Ed.), *Spoken discourse impairments in the neurogenic populations: A state-of-the-art, contemporary approach* (pp. 305–316). Springer.

DeDe, G., & Hoover, E. (2021). Measuring change at the discourse-level following conversation treatment: Examples from mild and severe aphasia. *Topics in Language Disorders, 41*(1), 5–26. https://doi.org/10.1097/TLD.0000000000000243

de Riesthal, M., & Diehl, S. K. (2018). Conceptual, methodological, and clinical considerations for a core outcome set for discourse. *Aphasiology, 32*(4), 469–471. https://doi.org/10.1080/02687038.2017.1398805

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

DeRuiter, M., & Ginsberg, S. M. (2020). Conscious clinical education: The evidence-based education—clinical education model. *Seminars in Speech and Language, 41*(4), 279–288. https://doi.org/10.1055/s-0040-1713779

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189–228. https://doi.org/10.1214/ss/1032280214

Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia: Consensus and caveats. *Aphasiology, 32*(4), 487–492. https://doi.org/10.1080/02687038.2017.1398814

Dipper, L., Marshall, J., Boyle, M., Botting, N., Hersh, D., Pritchard, M., & Cruice, M. (2021). Treatment for improving discourse in aphasia: A systematic review and synthesis of the evidence base. *Aphasiology, 35*(9), 1125–1167. https://doi.org/10.1080/02687038.2020.1765305

Douglas, N. F., Feuerstein, J. L., Oshita, J. Y., Schliep, M. E., & Danowski, M. L. (2022). Implementation science research in communication sciences and disorders: A scoping review. *American Journal of Speech-Language Pathology, 31*(3), 1054–1083. https://doi.org/10.1044/2021_AJSLP-21-00126

Douglas, N., Hinckley, J., Grandbois, K., Schliep, M., Wonkka, A., Oshita, J., & Feuerstein, J. (2023). How a power differential between clinicians and researchers contributes to the research-to-practice gap. *American Journal of Speech-Language Pathology, 32*(2), 803–810. https://doi.org/10.1044/2022_AJSLP-22-00207

Dutta, M., Murray, L. L., Park, H., Burklow, E., Bose, A., Kim, H., Greenslade, K., Ramage, A. E., Combs, C., Balasubramanian, A., & Casilio, M. (2025). Let's chat about spoken discourse: A tutorial to support use of spoken discourse analysis when providing aphasia clinical services. *American Journal of Speech-Language Pathology, 34*(4), 1931–1966. https://doi.org/10.1044/2025_AJSLP-24-005

Dutta, M., Murray, L. L., Stark, B. C., & Bryant, L. (2024). FOQUSAphasia: An initiative to facilitate research of spoken discourse in aphasia and its translation into improved evidence-based practice for discourse treatment. In A. P.-H. Kong (Ed.), *Spoken discourse impairments in the neurogenic populations: A state-of-the-art, contemporary approach* (pp. 305–316). Springer.

Edwards, A., Froude, E., Dark, L., & Carding, P. (2023). Factors that influence success when training videofluoroscopic swallowing study analysts. *Speech, Language and Hearing, 26*(3), 223–234. https://doi.org/10.1080/2050571X.2023.2171952

Edwards, A., Froude, E., Sharpe, G., & Carding, P. (2021). Training for videofluoroscopic swallowing analysis: A systematic review. *International Journal of Speech-Language Pathology, 23*(5), 529–539. https://doi.org/10.1080/17549507.2020.1861327

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*(397), 171–185. https://doi.org/10.1080/01621459.1987.10478410

Esmail, R., Hanson, H. M., Holroyd-Leduc, J., Brown, S., Strifler, L., Straus, S. E., Niven, D. J., & Clement, F. M. (2020). A scoping review of full-spectrum knowledge translation theories, models, and frameworks. *Implementation Science, 15*(1), Article 11. https://doi.org/10.1186/s13012-020-0964-5

Ezzes, Z., Schneck, S. M., Casilio, M., Fromm, D., Mefferd, A. S., de Riesthal, M., & Wilson, S. M. (2022). An open dataset of connected speech in aphasia with consensus ratings of

auditory-perceptual features. *Data, 7*(11), Article 148. https://doi.org/10.3390/data7110148

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

Gaglio, B., Shoup, J. A., & Glasgow, R. E. (2013). The RE-AIM framework: A systematic review of use over time. *American Journal of Public Health, 103*(6), e38–e46. https://doi.org/10.2105/AJPH.2013.301299

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health, 89*(9), 1322–1327. https://doi.org/10.2105/AJPH.89.9.1322

Glasgow, R. E., Harden, S. M., Gaglio, B., Rabin, B., Smith, M. L., Porter, G. C., Ory, M. G., & Estabrooks, P. A. (2019). RE-AIM planning and evaluation framework: Adapting to new science and practice with a 20-year review. *Frontiers in Public Health, 7,* Article 64. https://doi.org/10.3389/fpubh.2019.00064

Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders.* Lea & Febiger.

Gordon, J. K., & Clough, S. (2022). How do clinicians judge fluency in aphasia? *Journal of Speech, Language, and Hearing Research, 65*(4), 1521–1542. https://doi.org/10.1044/2021_JSLHR-21-00484

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *The Quantitative Methods for Psychology, 8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed.). Hodder Arnold.

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

Hartman-Maeir, A., Soroker, N., Ring, H., Avni, N., & Katz, N. (2007). Activities, participation and satisfaction one-year post stroke. *Disability and Rehabilitation, 29*(7), 559–566. https://doi.org/10.1080/09638280600924996

Haun, J., Rittman, M. R., & Sberna, M. (2008). The continuum of connectedness and social isolation during post stroke recovery. *Journal of Aging Studies, 22*(1), 54–64. https://doi.org/10.1016/j.jaging.2007.03.001

Hersh, D., Israel, M., & Shiggins, C. (2021). The ethics of patient and public involvement across the research process: Towards partnership with people with aphasia. *Aphasiology.* Advance online publication. https://doi.org/10.1080/02687038.2021.1896870

Higgins, J. J. (2004). *An introduction to modern nonparametric statistics.* Brooks/Cole.

Holland, A., Forbes, M., Fromm, D., & MacWhinney, B. (2019). Communicative strengths in severe aphasia: The Famous People Protocol and its value in planning treatment. *American Journal of Speech-Language Pathology, 28*(3), 1010–1018. https://doi.org/10.1044/2019_AJSLP-18-0283

Holtrop, J. S., Estabrooks, P. A., Gaglio, B., Harden, S. M., Kessler, R. S., King, D. K., Kwan, B. M., Ory, M. G., Rabin, B. A., Shelton, R. C., & Glasgow, R. E. (2021). Understanding and applying the RE-AIM framework: Clarifications and resources. *Journal of Clinical and Translational Science, 5*(1), Article e126. https://doi.org/10.1017/cts.2021.789

Kagan, A. (1998). Supported conversation for adults with aphasia: Methods and resources for training conversation partners. *Aphasiology, 12*(9), 816–830. https://doi.org/10.1080/02687039808249575

Kertesz, A. (2007). *Western Aphasia Battery–Revised.* PsychCorp.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

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

Linnik, A., Bastiaanse, R., Stede, M., & Khudyakova, M. (2022). Linguistic mechanisms of coherence in aphasic and non-aphasic discourse. *Aphasiology, 36*(2), 123–146. https://doi.org/10.1080/02687038.2020.1852527

Lock, S., Wilkinson, R., & Bryan, K. (2001). *SPPARC: Supporting Partners of People with Aphasia in Relationships and Conversations.* Winslow Press.

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

Maddy, K. M., Howell, D. M., & Capilouto, G. J. (2016), Current practices regarding discourse analysis and treatment following non-aphasic brain injury. *Journal of Interactional Research in Communication Disorders, 6*(2), 211–236. https://doi.org/10.1558/jircd.v7i1.25519

McBride, E. (2022). The development of clinical self-efficacy in speech-language pathology graduate training: A longitudinal study. *American Journal of Speech-Language Pathology, 31*(2), 770–789. https://doi.org/10.1044/2021_AJSLP-21-00134

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Olswang, L. B., & Goldstein, H. (2017). Collaborating on the development and implementation of evidence-based practices: Advancing science and practice. *Evidence-Based Communication Assessment and Intervention, 11*(3–4), 61–71. https://doi.org/10.1080/17489539.2017.1386404

Power, E., Attard, M. C., Lanyon, L. E., & Togher, L. (2024). Efficacy of online communication partner training package for student healthcare professionals. *International Journal of Language & Communication Disorders, 59*(1), 304–326. https://doi.org/10.1111/1460-6984.12947

Power, E., Falkenberg, K., Barnes, S., Elbourn, E., Attard, M., & Togher, L. (2020). A pilot randomized controlled trial comparing online versus face-to-face delivery of an aphasia communication partner training program for student healthcare professionals. *International Journal of Language & Communication Disorders, 55*(6), 852–866. https://doi.org/10.1111/1460-6984.12556

Power, E., Falkenberg, K., Elbourn, E., Attard, M., & Togher, L. (2021). The test–retest reliability of the Aphasia Attitudes, Strategies and Knowledge (AASK) survey with student health professionals. *Aphasiology, 35*(9), 1190–1206. https://doi.org/10.1080/02687038.2020.1787943

Power, E., & Morrow, R. (2024). Digital, co-created implementation of communication partner training programs for stroke, brain injury, and dementia: Past, present, and future. *International Journal of Speech-Language Pathology, 26*(3), 317–333. https://doi.org/10.1080/17549507.2024.2362856

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

Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology, 30*(1), 45–73. https://doi.org/10.1080/02687038.2015.1057891

Rietdijk, R., Power, E., Brunner, M., Attard, M., McDonald, S., Tate, R., & Togher, L. (2019). *TBIconneCT clinician manual: Connecting people living with traumatic brain injury to conversation training.* ASSBI.

Smith, E. R., Broughton, M., Baker, R., Pachana, N. A., Angwin, A. J., Humphreys, M. S., Mitchell, L., Byrne, G. J., Copland, D. A., Gallois, C., Hegney, D., & Chenery, H. J. (2011). Memory and communication support in dementia: Research-based strategies for caregivers. *International Psychogeriatrics, 23*(2), 256–263. https://doi.org/10.1017/S1041610210001845

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

Stark, B. C., Bryant, L., Themistocleous, C., den Ouden, D.-B., & Roberts, A. C. (2023). Best practice guidelines for reporting spoken discourse in aphasia and neurogenic communication disorders. *Aphasiology, 37*(5), 761–784. https://doi.org/10.1080/02687038.2022.2039372

Stark, B. C., & Dalton, S. G. (2024). A scoping review of transcription-less practices for analysis of aphasic discourse and implications for future research. *International Journal of Language & Communication Disorders, 59*(5), 1734–1762. https://doi.org/10.1111/1460-6984.13028

Stark, B. C., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., Ramage, A. E., Roberts, A., den Ouden, D. B., Brock, K., McKinney-Bock, K., Paek, E. J., Harmon, T. G., Yoon, S. O., Themistocleous, C., Yoo, H., Aveni, K., Gutierrez, S., & Sharma, S. (2020). Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. *American Journal of Speech-Language Pathology, 30*(1S), 491–502. https://doi.org/10.1044/2020_AJSLP-19-00093

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., Ramage, A. E., & Roberts, A. C. (2021). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *Journal of Speech, Language, and Hearing Research, 64*(11), 4366–4389. https://doi.org/10.1044/2021_JSLHR-20-00708

Stark, B. C., & Fukuyama, J. (2021). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Language, Cognition and Neuroscience, 36*(5), 562–585. https://doi.org/10.1080/23273798.2020.1862258

Strong, K. A., Douglas, N. F., Johnson, R., Silverman, M., Azios, J. H., & Archer, B. (2023). Stakeholder-engaged research: What our friendship in aphasia team learned about processes and pitfalls. *Topics in Language Disorders, 43*(1), 43–56. https://doi.org/10.1097/TLD.0000000000000302

Tabak, R. G., Khoong, E. C., Chambers, D. A., & Brownson, R. C. (2012). Bridging research and practice: Models for dissemination and implementation research. *American Journal of Preventive Medicine, 43*(3), 337–350. https://doi.org/10.1016/j.amepre.2012.05.024

Togher, L., McDonald, S., Tate, R., Power, E., Ylvisaker, M., & Rietdijk, R. (2011). *TBI express: A communication training program for everyday communication partners of people with traumatic brain injury (TBI) (Manual)*. ASSBI.

Tolsgaard, M. G., Tabor, A., Madsen, M. E., Wulff, C. B., Dyre, L., Ringsted, C., & Nørgaard, L. N. (2015). Linking quality of care and training costs: Cost-effectiveness in health professions education. *Medical Education, 49*(12), 1263–1271. https://doi.org/10.1111/medu.12882

Trupe, E. H. (1984). *Reliability of rating spontaneous speech in the Western Aphasia Battery: Implications for classification* [Paper presentation]. Clinical Aphasiology Conference, Seabrook Island, SC, United States.

Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language, 36*(2), 252–274. https://doi.org/10.1016/0093-934X(89)90064-3

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., Kong, A. P. H., Simmons-Mackie, N., Scarinci, N., & Alary Gauvreau, C. (2017). Which outcomes are most important to people with aphasia and their families? An international nominal group technique study framed within the ICF. *Disability and Rehabilitation, 39*(14), 1364–1379. https://doi.org/10.1080/09638288.2016.1194899

Williamson, P. R., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., & Tugwell, P. (2012). Developing core outcome sets for clinical trials: Issues to consider. *Trials, 13,* Article 132. https://doi.org/10.1186/1745-6215-13-132

Wilson, S. M., Entrup, J. L., Schneck, S. M., Onuscheck, C. F., Levy, D. F., Rahman, M., Wiley, E., Casilio, M., Yen, M., Brito, A. C., Kam, W., Taylor Davis, L., de Riesthal, M., & Kirshner, H. S. (2023). Recovery from aphasia in the first year after stroke. *Brain, 146*(3), 1021–1039. https://doi.org/10.1093/brain/awac129

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLOS ONE, 13*(2), Article e0192773. https://doi.org/10.1371/journal.pone.0192773

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., Miller, B. L., & Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain, 133*(7), 2069–2088. https://doi.org/10.1093/brain/awq129