Discourse-Level Communication Success in Aphasia: Unveiling Its Significance through Observer's Ratings

Amy E. Ramage, Ph.D.,[1,2] **Abigail L. Rowe, B.S.,**[1] and
Kathryn J. Greenslade, Ph.D.[1]

ABSTRACT

Audrey Holland (1982) compared test scores to observers' ratings of conversational communicative success in people with aphasia (PWA). This springboarded a body of evidence employing observers to rate discourse. We review the utility of those ratings for assessing PWA's communication success. A traditional literature review identified 16 articles involving naive or trained raters assessing PWAs' communicative success across discourse genres. Another 10 articles reported ratings over time. Collectively, these studies evaluated 349 PWAs. Four studies utilized observers to rate the success of PWA's conversations. Eight studies that reported observers' ratings on other discourse genres found that multimodal communication and facilitative contexts improved success, and ratings of informativeness and comfort related to objective discourse analysis measures. Nine of 10 studies examining treatment effects found that communicative success ratings captured improvements. Observers' ratings provide social validity by reliably assessing the discourse-level communicative success of PWA. Ratings correlated with standardized diagnostic and objective discourse metrics but provided a window into factors that affect communicative success, including the degree to which communication is interactive, multimodal, and contextual. Integrating observers' ratings of discourse success at pretreatment may help identify supports or barriers to successful communication, facilitate individualization of treatments, and offer social validity of change.

KEYWORDS: functional communication, discourse, aphasia, raters, observers

[1]Department of Communication Sciences and Disorders, University of New Hampshire, Durham, New Hampshire; [2]Interdisciplinary Program in Neuroscience and Behavior, University of New Hampshire, Durham, New Hampshire.

Address for correspondence: Amy E. Ramage, Ph.D., CCC-SLP, Department of Communication Sciences and Disorders, University of New Hampshire, 4 Library Way, Durham, NH 03824 (e-mail: Amy.ramage@unh.edu).

Learning Outcomes: As a result of this activity, the reader will be able to:

- Recognize discourse genres that may be used in evaluating communicative success in PWA.
- Identify methods for using raters to assess discourse success and describe the benefits/drawbacks of different methods.
- Consider how ratings of discourse-level communication in PWA may be used to facilitate intervention development.

The long history of aphasiology spans many fields from linguistics to psychology, to neurology and speech pathology, in part because human brain mapping has focused on language processing in the brain. In the field of aphasiology, like communication sciences and disorders (CSD), both basic and clinical research endeavors drive the evolution of the field. Audrey Holland was at home in both domains but certainly was a driving force for over 60 years in clinical aphasiology.

In a historical overview of aphasia therapy, Code (2008, p. 20) highlighted the transitions in approaches to aphasia over time, including Gutzman as the "father of aphasia therapy" in the late 1800s, Wepman and Schuell as the revolutionizers for a focus on the *use* of language as essential for intervention in the United States, and Helm-Estabrooks as the systematizer of treatments to align with aphasia classification schemes. It was Holland and Sarno who were instrumental in introducing functional approaches, with the inclusion of social and psychosocial perspectives to processes of recovery in aphasia. While Audrey has been referred to as an aphasia Goddess (Anbar, 2022), a trailblazer of aphasia treatment (https://aphasiacenter.net/remembering-dr-audrey-holland/), and a true legend in the field (https://slhs.arizona.edu/news/remembering-audrey-holland), we propose that her mark is as the superhero for people with aphasia (PWA). Like many great scholars, Audrey found her work with PWA so satisfying that it was hardly a job, and she continued her work until the moment she passed.

APHASIOLOGY AND COMMUNICATION

Historically the focus of aphasiology has been on describing the language characteristics of people who have suffered damage to the head/brain (c.f., Edwin Smith Papyrus in van Middendorp et al., 2010). Hippocrates foreshadowed the future of aphasia syndromes with the dichotomy of expressive (*aphonos*—"without voice") and receptive (*anaudos*—"without hearing") aphasias (Ardila, 2014). Whether traveling through time via France (Broca) and Germany (Wernicke, Lichtheim, Freud), through Russia (Luria), or the United States (Geschwind, Goodglass), aphasiology has attended to the deficit by categorizing it. From this line of inquiry has grown so much knowledge about how components of language are processed by the brain (c.f., Dell et al., 1999; Henderson, 1992); how one maps incoming linguistic information to the lexicon to comprehend a message (Hickok & Poeppel, 2007, 2015; Poeppel et al., 2004); and how messages are formulated and encoded for sentence production (Bock & Levelt, 1994; Levelt, 2001). It is from this wealth of knowledge about language processing that many foundational approaches to intervention for aphasia have developed (c.f., Boyle & Coelho, 1995; Edmonds et al., 2009; Linebaugh et al., 2005; Martin & Dell, 2019; Schuell et al., 1969; Schwartz, 2013; Schwartz et al., 2011; Thompson & Shapiro, 2005).

The fruitful reports and theoretical accounts that have grown from the work cited earlier have contributed scientific roots to the field of aphasiology. However, most of this work focused on the impairment of aphasia—restoring the function of broken language processes. The approaches have emphasized the linguistic and neurological aspects of aphasiology and underemphasized its psychological aspects. Audrey Holland once told the first author (A.E.R.) that her clinical plan for a person with aphasia that centered on improving word retrieval using a theoretically driven approach highlighting lemma selection was "arcane" (or, maybe it was "archaic"). From that moment on, the first author, who admits a fascination with the linguistic and neurological aspects of aphasiology, *never* approached diagnostics or treatment without recognizing the

need to *couch it in the context of the PWA participating in life*. This exchange epitomizes what Audrey Holland was instrumental in parading into the field.

Nonetheless, approaches to understanding recovery following brain damage have tended to focus on the characteristics of the damage to the brain (Hartwigsen et al., 2013) and the presenting speech-language or cognitive deficits (Gilmore et al., 2019). These factors, particularly the role of the lesion, are paramount for predicting at least 50% of the variance seen for the persistence of aphasia into the chronic phase (Busby et al., 2023; Plowman et al., 2012; Wilson et al., 2023). This leaves 50% of the variance to prognostic factors that include age, sex, comorbid medical history, and other underspecified factors (Fig. 1). Holland et al. (2017) noted that some potential contributors to outcomes are outside of the direct effects of the injury to the brain, including factors like changing post-stroke lifestyles, accessibility of benefits to support aphasia intervention, fatigue, and lack of motivation to engage in the therapeutic process. In most cases, the success of intervention and enhancement of generalization is likely predicted by these, or other, underspecified factors. Thus, they need to be specified and incorporated into the focus of therapies.

Over 20 years ago, Holland and Ramage (2000) proposed the need to think outside the (black) box, acknowledging that a paradigm shift is needed to recognize the interplay of brain and language as a dynamic set of processes. They noted that "we are long overdue for innovative, synthesizing, iconoclastic, and possibly even outrageous thinking about the activities in language processing" (p. 93). There are many ways that this thinking has been shifted, but one way that has seen a burst of activity grew from the work of Holland (1982) and the effort to observe the success of PWAs communicating in their daily lives. That is, there have been major advances in using observers to rate communicative success within situated language (Doedens & Meteyard, 2022).

Foremost, there have been extensions from the typical standardized assessment to include evaluation of situated language (modified from Clark, 2005), recognizing communication as interactive, multimodal, and contextual. Situated language is the interaction between at least two people who are utilizing words, prosody, gestures, and/or other nonverbal modes of conveying a message. In a real-time interaction, common ground, based on the situational and communicative contexts as well as shared personal and cultural knowledge, facilitates the co-construction of context throughout the interaction (Clark, 2005). To analyze situated language, Hengst (2020) argues that we must consider (1) the activities in which participants are engaging and the goals/objectives of those activities, (2) the participants' patterns of participation (e.g., being a "ratified participant" vs. someone eavesdropping), and (3) the multimodality (visual, verbal, aural, tactile, olfactory, and proprioceptive) resources being used to communicate. Furthermore, Armstrong and Ferguson (2010) suggested that speech-language pathologists should consider key aspects of the context, including partner characteristics (e.g., familiarity, authority relationship), discourse topic (e.g., familiarity, technical vs. lay topic), and channel of communication (e.g., face-to-face, phone call, text message; oral, written). These factors each contribute to the success of a specific communicative interaction.

Importantly, situated language contrasts with the decontextualized "ex vacuo" language that is commonly assessed on standardized diagnostic measures of aphasia (e.g., the *Western Aphasia Battery – Revised* [WAB-R]; Kertesz, 2007) which use tasks such as single word naming, verbal fluency, or matching words/ sentences to corresponding pictures. Such decontextualized measures do not correspond with how well a PWA communicates (Doyle et al., 1996), likely due to the removal of all three aspects of situated language: (1) the minimal requirement for co-construction/interaction, (2) the reliance on a single mode of communication, and (3) the relatively simple and static context. Furthermore, this focus on form rather than function limits the assessment's ability to capture the nuanced meanings of content words and applications of grammatical structures when language is used within interactions, all of which impact communicative success (Armstrong & Ferguson, 2010).

Holland (2010) noted that it was situated language that fascinated her about PWA, as there was much to be learned through
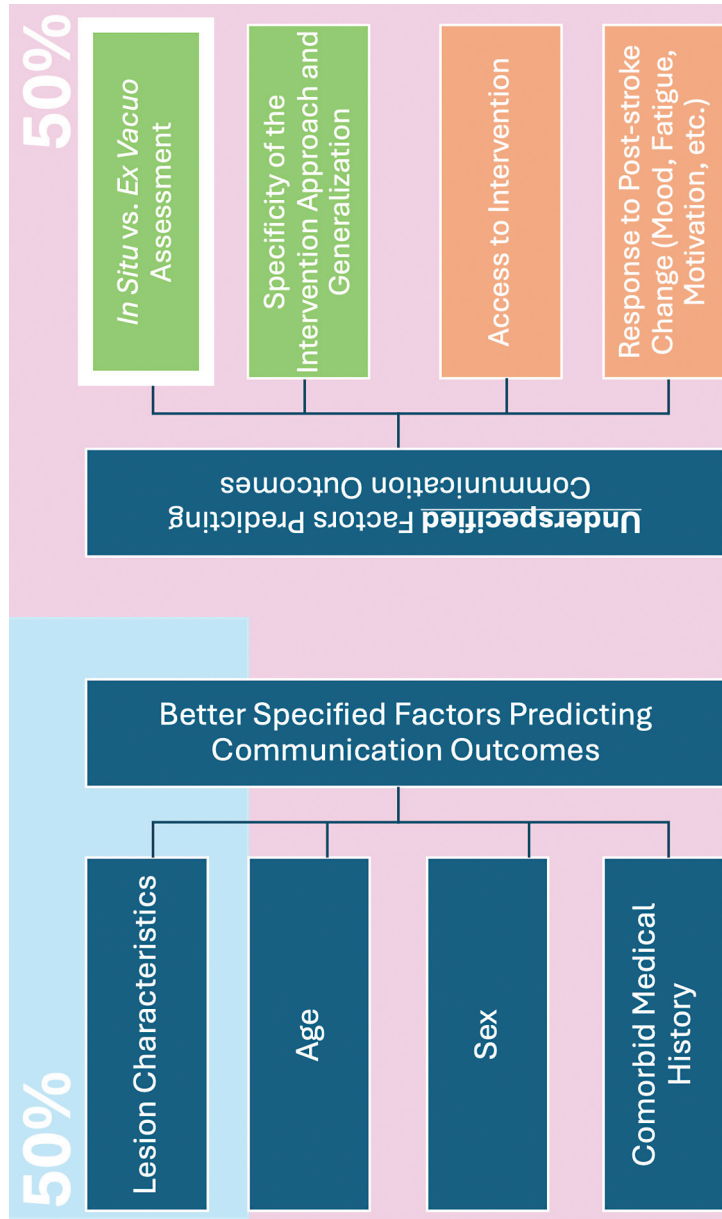
Figure 1   While there are strong data suggesting that characteristics of the inciting lesion contribute up to 50% of the variance in outcomes in aphasia (top left in blue), the other 50% (pink) of the variance is accounted for by biological/neurological and health-related factors (left) that may be considered biological and not "treatable" as part of standard SLP practice. Other *underspecied* factors (right) are within the control of the SLP in practice (green) or the transdisciplinary approaches in the purview of rehabilitation (orange). This is a nonexhaustive list of some potential factors that are underspecified in the literature regarding communication outcomes in aphasia. Addressing many of these potential factors has been the focus of studies over the past several years (Busby et al., 2023; Holland et al., 2017; Mayer et al., 2024; Plowman et al., 2012; Wilson et al., 2023), but the focus of this article is on the use of naive or trained raters in evaluations of communication success in discourse.

observation of interactions with PWA—including how "… interactions with strangers – store clerks, bank employees, people occupying the adjoining seats on buses – are managed" (p. 13). Although Holland did not use the term "situated language," she talked about characterizing aphasia based on functional communication (Holland, 1982; Holland et al., 2018), defined by Doedens and Meteyard (2022) as the skills necessary to convey one's ideas, needs, wants, etc., in the varying situations encountered daily. For example, functional communication may be defined by successful interactions through reliance on multimodal communication of intentions, rather than with sole reliance on spoken words, and within goal-directed activities that require active participation. Functional communication is an inherently joint activity, requiring more than one person to create meaning and maintain common ground to share knowledge or beliefs with another. That is, functional communication is situated language.

Another area of growth in the study of functional communication, in part due to the work of Audrey Holland, is the new and innovative approaches to observing discourse-level functional communication (c.f., virtual reality; Bryant et al., 2020; Garcia et al., 2007), utilizing Language Environment Analysis system (LENA; Language Environment Analysis Foundation, Boulder, CO, as piloted in Suting & Mozeiko, 2021) or analyzing conversational samples for specific features (e.g., topic initiations [Leaman & Edmonds, 2020], conversational repairs [Azios et al., 2022]). However, few have evaluated how successful communicative exchanges are for PWA (Azios et al., 2022), particularly from the perspective of an observer or a communication partner without aphasia. This gap limits the field's understanding of functional communication success in aphasia and how well that success correlates with decontextualized, objective measures of aphasia. Here, we focus on what Holland (1982) used as an additional element for evaluating the success of functional communication—using observers' ratings. While ideally communicative success would be evaluated within a co-constructed conversation, it may also be valuable to examine observers' ratings of communicative success within other discourse genres. Although other discourse genres offer a weak proxy to the interactional aspect of situated language, observers' ratings of communicative success within such tasks can provide a window into (1) the success of conveying information with and without interaction/co-construction of messages, (2) the value of multimodal communication attempts through the use of audio or video recordings, and (3) the effect of contextual factors, such as a rater's knowledge of aphasia or the nature of common ground/shared context. As such, this review focuses on published studies of PWAs' discourse (conversational or narrative) that utilized ratings from observers on aspects of communicative success, accounting for the degree to which language samples were situated.

METHODS

Literature Search, Filtering, and Selection

To identify published evidence that addresses the communicative success of PWAs, this traditional literature review focused on studies that included an evaluation of discourse-level productions (e.g., conversation, story retell, picture description) by observers (naive or otherwise) for communicative success, sometimes referred to as informativeness or effectiveness. Self-ratings or analyses that did not focus on broad perceptual judgments of success (e.g., ratings of specific communicative features like word finding; counts of content information units, main concepts, etc.; measures resulting from in-depth conversational analysis, such as ratings of conversational repair or topic initiation) were excluded. Studies that examined whether observers' ratings of communicative success changed over time and/or were associated with changes in standardized metrics of communicative ability over time were also collected.

The literature search spanned from 1982, when Audrey Holland's publication on observations of PWAs in discourse was published, and extended to March 2024. As a traditional literature review, no systematic guidelines were followed for conducting the search. The searches started in PubMed/NCBI, PsycInfo, and other search engines to find articles. Using a few

key articles to identify terms, the following terms were searched: aphasia, discourse *or* narrative *or* conversation *or* picture description, content *or* informativeness *or* communicat* success, perceptual *or* judge* *or* listener *or* rater *or* observer. Articles were included if they were in English, sampled the discourse of PWA, *and* included observers/raters who heard or watched discourse samples of PWAs and rated the communicative success, including informativeness, effectiveness, favorability, comfort, ease, and other variables that did not capture specific language features (e.g., syntactic accuracy). The original searches in PubMed and PsycInfo yielded 180 studies. Only peer-reviewed experimental articles were included in the review. Seven of the 180 articles were reviews and were excluded. The authors screened the abstracts which resulted in the exclusion of articles that focused on individuals with primary progressive aphasia (e.g., Gallée et al., 2024; Haley et al., 2021; Quaranta et al., 2022) or discourse in other nonaphasia or neurodegenerative populations (Carlomagno et al., 2005; Saldert et al., 2010), or utilized augmentative/alternative tools to facilitate discourse production (i.e., Bartlett et al., 2007; Dietz et al., 2018; Fink et al., 2008).

These database searches identified nine studies utilizing raters to evaluate aspects of communicative success or informativeness as judged by a rater. Hand searching through these and other relevant works and their references identified an additional seven studies that utilized raters to evaluate the discourse of PWAs. In total, 16 articles reported listener (audio) or watcher (video) ratings of PWAs' discourse productions across a broad range of narrative types. Ten additional studies used observer's ratings at two or more time points to capture change in communicative success—and, in some cases, to examine the relationship between ratings and standardized language measures.

Of note, the traditional literature search yielded many studies that sought to establish the reliability of communicative success/informativeness ratings of discourse-level communication to allow accessible analysis options for clinicians and raters (Casilio et al., 2019; Copeland, 1989; Hula et al., 2003; Kim & Wright, 2020). For example, information units (Hula et al., 2003) or CoreLex scores (Kim & Wright, 2020) now have established reliability for a range of picture description and story-retelling tasks that may be used almost as checklists for a rater to reliably judge content (c.f., Stark & Dalton, 2024). Thus, these studies are providing standardized elicitation methods to extract valuable information from raters, which may increase the ability to use raters. However, these studies were excluded here as their intent was not to evaluate the communication success, informativeness, or effectiveness of the PWAs directly.

RESULTS

Table 1 summarizes 16 studies in which observers rated the communicative success of the discourse of PWAs. The studies are organized from highly situated (interactive, multimodal, contextual) to minimally situated. Details about the PWAs, raters, discourse genre/task, and the nature of ratings are also summarized. Table 2 summarizes similar information for an additional 10 studies that included ratings acquired over at least two time points and the relationship between ratings and other standardized measures of language and discourse structure.

This traditional literature review of studies that involved elicitation of discourse from PWAs and included raters (ranging from naive to expert) indicates that some have utilized a highly situated approach using conversation, while most have constrained their observations to standardized elicitation approaches—semistructured interviews, personal narratives, picture descriptions or sequences, story retells and procedural discourse. Four of the studies utilized data that followed the AphasiaBank protocol, which standardizes the instructions to the participant, the materials used, and the transcription approach for analysis (Macwhinney et al., 2011). These collective studies report data from 349 PWAs, most of whom were in the chronic phase of recovery and presented with varying aphasia classifications (nonfluent and fluent, ranging in severity). Raters evaluated audio- or video-recorded samples for variables including success, completeness, informativeness, comprehensibility, favorability, comfort, efficiency, and ease/skill of conveying a message. Most ratings used Likert scales, visual analog scales, or direct magnitude estimation,

Table 1 Studies employing raters who judged some aspect of discourse elicited from people with aphasia

Study	n	Sex	Age range	Months post onset	Raters	Ratings	Data capture	Task type	Rating scale
Conversation ratings based on exchanges with a high degree of interaction, multimodal communication, and context									
Holland (1982)	40	30 M 10 F	32–74	3–86	NR	Communicative success	Live	Conversation	Categorical (success/failure)
Ramsberger and Rende (2002)	14	6 M 8 F	26–78	6–153	2 experts	Comprehensiveness	Video	Conversational story retell	5-point scale
Leaman and Edmonds (2019)	8	3 M 5 F	42–75	8–163	3 CSD graduate students	Communicative success/clarity	Video	Conversation	4-point scale
Leaman and Edmonds (2021)	20	14 M 6 F	38–77	6–232	3 CSD graduate students	Communicative success/ease	Video	Conversation, structured narrative	4-point scale
Discourse ratings based on manipulation of multimodal communication or context									
de Beer et al. (2017)[a]	10	7 M 3 F	38–76	12–118	60 naive	Communicative success by modality	Audio and video	Personal narrative	Open-ended and multiple-choice questions about content
Rose et al. (2017)[a]	13	9 M 4 F	47–81	9–360	67 naive	Communicative success by modality	Audio and video	Personal narrative	Open-ended and multiple-choice questions about content
Tompkins et al., 2006	10	NR	NR	NR	6 naive	Informativeness	Video	Word description	Categorical (+/−)
Hogrefe et al., 2012	24	16 M 8 F	33–68	3–53	18 naive	Informativeness	Video	story retell	Raters matched sample to correct story
Discourse ratings based on audio recordings, limiting multimodal communication									
Groenewold et al., 2014	10	6 M 4 F	24–82	3–77	37 listeners	Liveliness and comprehensibility	Audio	Semistructured interview	Likert scales
Behrns et al., 2009	8	6 M 2 F	28–63	6–100	60 naive	Favorability	Audio	Personal narrative	Visual analog scale

(Continued)

Table 1 (*Continued*)

Study	n	Sex	Age range	Months post onset	Raters	Ratings	Data capture	Task type	Rating scale
Harmon et al., 2016[a]	9	4 M 5 F	41–76	NR	18 naive, 18 students	Favorability	Audio	Personal narrative	Likert scale
Ulatowska et al., 1983	15	10 M 5 F	24–71	3–220	3 SLPs	Informativeness	Audio	Multiple narratives	2- to 3-point scales
Doyle et al. 1996	25	23 M 2 F	48–77	1–203	11 naive	Informativeness	Audio	Picture description	Direct magnitude estimation
Webster and Morris, 2019	20	17 M 3 F	36–82	4–120	11 naive	Informativeness	Audio	Picture description	Direct magnitude estimation
Kong and Wong, 2018[a]	68	49 M 19 F	30–70+	NR	25 naive	Completeness, listener comfort	Audio	Story retell	5-point Likert scale
Kong et al., 2018[a]	15	13 M 2 F	43–72	≥ 6	20 naive	Coherence	Audio	Picture sequence, procedural narrative	Likert scale

Abbreviations: CSD, communication sciences and disorders; F, female; M, male; MPO, months post onset of stroke; n, number of participants in the study with aphasia; NR, not reported; SLP, speech-language pathologist.

Notes: People with aphasia (PWA) participants in these studies ranged in severity and aphasia types. Studies are arranged from top to bottom as those that are highly situated (conversation) to those that are less situated. Range: aphasia classifications in the study varied across a range of types; Other: in Hickey and Rondeau (2005), raters judged seven dimensions of conversation: the comfort of both members of the dyad, PWA expressed a lot of information, volunteer communicated well with PWA, turn-taking, topic continuation, quality of conversation; and in Kong and Wong (2018), raters rated for five questions—two about speech output, two about speaker's attributes, and one about listener's feelings. All students were in communication sciences and disorders programs.

[a]Utilized PWA discourse data from the AphasiaBank repository.

but several were categorical (e.g., present/absent). The raters varied in their knowledge of aphasia/communication disorders, their experience interacting with PWA, and their familiarity with/exposure to the PWA whom they were rating. Specifically, raters included naive listeners, undergraduate and graduate students in CSD programs, speech-language pathologists, or researchers working on the topic.

OBSERVATION OF COMMUNICATIVE SUCCESS IN THE CONVERSATIONS OF PWA

Four studies (Holland, 1982; Leaman & Edmonds, 2019, 2021; Ramsberger & Rende, 2002) rated the communicative success of PWA in conversation, the most highly situated discourse genre examined. To study the conversational success of PWAs, Holland (1982) applied field observation techniques to study interactions between PWAs and conversational partners in their homes, doing errands, and other daily events. Observers categorically judged communicative exchanges as successes (any response that communicated the PWA's intended message) or failures (responses that did not). This approach was unique given its focus on not only what is said in a conversation (form/content) but also on what constitutes successful communication (function). Observed exchanges were interactive; multimodal (utterances, gestures, writing, nonverbal behaviors); and included a variety of partners, settings, topics and goals, and modes of communication. Pairs of observers, who rated communicative success *in vivo*, demonstrated high interrater reliability ($r = 0.85$). However, field observation has limitations. Because the activities/environmental contexts, topics and communicative goals, numbers of exchanges, and the modes of communication used naturally varied across PWAs, those elements of the communication could not be experimentally controlled while capturing the real-life aspect of such a study.

Closely following this approach, Leaman and Edmonds (2019, 2021) developed the Social Conversation Collection Protocol for sampling casual, social conversations. Samples were analyzed for communicative success using a 4-point scale, where 4 represents "good success" and 1 represents "no success." Protocol descriptions are available for either a "home partner," who shares substantial personal and cultural knowledge with the PWA, or a "speech-language pathologist (SLP) partner" who would have knowledge about communication disorders. Multimodal communication during conversation samples is encouraged and captured in the observer's ratings. This approach has pioneered the use of unstructured conversation to establish reliable and valid measures of communicative success, along with measures of microstructural elements (e.g., percent content information units in conversation, subject–verb–[object] integrity, the relevance of words in and outside of the subject–verb–(object) frame, referential and global coherence). Interrater reliability, $r = 0.95$ (Leaman & Edmonds, 2021) and test–retest stability, $r > 0.90$, (Leaman & Edmonds, 2019), and $r = 0.798$ (Leaman & Edmonds, 2021) for these metrics have been good to excellent, boding well for the future of evaluating PWA's communicative success.

Finally, Ramsberger and Rende (2002) developed ratings for communicative success based on semistructured conversations that involved co-constructing a story retell of an episode of *I Love Lucy*. In this study, PWAs first watched the episode of *I Love Lucy*, and then had a conversation with an unfamiliar, naive (did not watch the episode), non-PWA partner about the episode. Multimodal communication was encouraged within task instructions. The goal was for the non-PWA partner to retell the episode's events to an experimenter following the conversation. Although the PWA's communication was not directly rated in this case, the non-PWA partner's retell was rated on a 5-point scale, documenting the comprehensiveness of their retell (0 = no main ideas communicated, 4 = all main ideas communicated). Importantly, the conversation was designed to be interactive, with both partners using multimodal resources as they worked together toward their joint goal.

OBSERVATION OF COMMUNICATIVE SUCCESS IN OTHER DISCOURSE GENRES IN PWA

While conversation is the discourse genre that best represents situated language, ratings of

communicative success based on other discourse genres may help us understand some *components* of situated language, for example, the benefit of using multimodal resources to convey a message. These other genres, including personal narratives, picture descriptions or sequences, story retell procedural discourse, or word descriptions, offer the advantage of standardizing the context, controlling for many of the factors that could not be controlled in Holland's (1982) field observation approach. It must be acknowledged, though, that this benefit comes at the cost of interaction (i.e., most genres generate monologues rather than dyadic exchanges) and social validity (i.e., conversations occur far more frequently in daily interactions than fictional narratives). Direct comparison between unstandardized conversation and standardized narrative monologues in PWA are reported in Leaman and Edmonds (2021.) Furthermore, the type of recording (audio vs. video) can impact raters' ability to account for the use of multimodal communication; specifically, audio recordings do not allow raters to observe the PWA's use of gestures, facial expressions, writing, drawing, etc. Such factors should be considered when determining how situated a language sample is and how this may impact the validity of measurement.

Discourse Ratings Based on Manipulation of Multimodal Communication or Context

As summarized in Table 1, four studies examined ratings of discourse-level communicative success when either multimodal communication or context was manipulated. Three of these studies (de Beer et al., 2017; Hogrefe et al., 2012; Rose et al., 2017) manipulated raters' access to multimodal communicative attempts to determine the relative contribution of audible attempts and nonverbal attempts to communicative success in PWA. Both de Beer et al. (2017) and Rose et al. (2017) compared ratings for audio recordings of personal narratives, where raters only had access to audible communicative attempts (e.g., words, prosody, sound effects), to video recordings where raters had access to multimodal communicative attempts. In both cases, naive raters rated

communicative success based on open-ended questions and multiple-choice questions related to the content conveyed (de Beer et al., 2017; Rose et al., 2017). In contrast, Hogrefe et al. (2012) had raters view videos of story retells with the sound muted, eliminating access to audible communicative attempts and forcing raters to focus on nonverbal communicative attempts. The goal of ratings was to match each PWA's sample to the correct story retell after having viewed videos of each story themselves. Findings indicated that multimodal communicative attempts were rated as more successful than attempts where one or more modalities were removed, and the success of gestural communicative attempts varied based on the PWA's apraxia severity. In the final study, Tompkins et al. (2006) investigated the effect of manipulating context/shared knowledge on communicative success. Specifically, naive observers were shown video clips of PWA engaging in multimodal self-cuing during a word description task, where the final response (regardless of accuracy) was removed. Observers attempted to predict the word that the PWA was attempting to say, with half of the raters being provided no contextual information to support their prediction, and the other half provided with four options (the target and three semantically/phonetically related or unrelated foils). As predicted, having more context/shared knowledge improved the observers' success in predicting the target word.

Discourse Ratings Based on Audio Recordings

Eight studies had raters judge the discourse-level communicative success of PWA based on audio recordings, providing access only to audible modes of communication (words, prosody, sound effects). In all cases, samples involved little to no interaction, as most tasks produced monologues, and topics were prescribed by the discourse task's elicitation. Five of these studies (Doyle et al., 1996; Kong et al., 2018; Kong & Wong, 2018; Ulatowska et al., 1983; Webster & Morris, 2019) had raters indicate a sample's overall informativeness or accuracy/completeness within various discourse genres. Specifically, Ulatowska and colleagues (1983) had

listeners rate PWAs' personal narratives, picture sequences, story retelling, and procedural discourse for content (2- to 4-point scale, depending on the task) and clarity (3-point scale). SLP listeners reported poorer content and clarity for the PWA relative to healthy controls, but the main points and message were conveyed, nonetheless (Ulatowska et al., 1983). Doyle et al. (1996) and Webster and Morris (2019) had naive listeners rate the informativeness of picture descriptions/sequences of PWA using direct magnitude estimation methods. Both studies found that raters' judgments of informativeness were highly correlated with objective measures (e.g., correct information units, accurate/complete main concepts, number of propositions) and, to a lesser extent, with overall aphasia severity (in Doyle et al., 1996). Finally, Kong and colleagues had listeners rate the accuracy and completeness of a story retell (Kong & Wong, 2018) or the completeness, clarity, and understandability of a picture sequence and procedural discourse (Kong et al., 2018). In both studies, the discourse of PWA was rated as less understandable and clear than that of controls. Also, ratings showed strong correlations with objective discourse analysis measures (i.e., ratings of completeness with informative words and main concept score; ratings of accuracy with informative words and communication efficiency; ratings of clarity/understandability with word-level errors and coherence variables; Kong et al., 2018; Kong & Wong, 2018). These findings support the use of raters as valid sources in the evaluation of the overall effectiveness of PWAs' communication akin to the metrics derived from careful analysis of macro- and microstructural features of discourse.

Furthermore, four studies (Behrns et al., 2009; Groenewold et al., 2014; Harmon et al., 2016; Kong & Wong, 2018) examined naive listeners' ratings of discourse favorability, including ratings of liveliness, ease, comfort, or preference for listening to PWAs' personal narratives, picture description/sequence, story retells and/or procedural discourse. Behrns et al. (2009) investigated listeners' judgments about PWAs' spoken discourse, reporting that they preferred written to spoken productions from PWAs, thought that the PWAs did not like

telling stories, and generally perceived PWA's narratives to be less interesting than those of controls. Kong and Wong (2018) found that raters' comfort in listening to the story retells of PWA was associated with an objective measure of communicative efficiency. Groenewold and colleagues (2014) found that when PWA used direct speech (i.e., reporting on what another speaker says), raters judged their personal narratives as more lively, but this style did not influence listeners' understanding of what was being said. Finally, Harmon et al. (2016) reported that listeners preferred samples that simulated fluency in nonfluent speakers (i.e., by removing disfluencies like pauses longer than 0.4 seconds, fillers, filled pauses, repetitions, and revisions), suggesting that the lack of fluency may be a source of discomfort or general negative perceptions in listener ratings. These findings highlight the potential biases that naive raters may bring to communication with PWAs.

CHANGES IN RATINGS SPECIFIC TO CONVERSATIONAL SUCCESS OF PWA OVER TIME

Table 2 presents findings for studies that included ratings or judgments by observers that are used as outcomes for documenting change in language measures over time. Nine studies report on ratings of discourse productions prior to and following communication interventions to provide social validity of the treatment approaches; the remaining study (Ross & Wertz, 1999) documented changes in communicative ability over time in the absence of a specific treatment. The nine intervention studies had varied treatment targets, from focusing on syntax (Ballard & Thompson, 1999; Goral & Kempler, 2009; Jacobs, 2001) or lexical retrieval (Cupit et al., 2010) to communication strategy training (Lustig & Tompkins, 2002) or conversational coaching/conversational partner training (Hickey et al., 2004; Hickey & Rondeau, 2005; Hopper et al., 2002; Kagan et al., 2001). In all but one study (Kagan et al., 2001, who used expert raters), naive raters or students with some exposure to communication disorders (i.e., enrolled in CSD programs) were asked to indicate the communicative success

Table 2 Studies employing raters who judged some aspect of discourse at two time points, most with intervening treatment

Study	Assessed	Treatment	No. of PWA	Sex	Age	MPO	Raters	Rating	Data capture	Task type	Rating scale	Summary results
Lustig and Tompkins, 2002	Pre/Post	Written communication strategy training	1	1 F	52	10	9 GS	Informativeness	Video	Conversation	DME	Post > Pre
Hickey and Rondeau, 2005	Pre/Post	Conversational partner training	1	1 M	72	37	10 Naive 10 GS 10 SLPs	Informativeness, comfort (partner/PWA), overall quality	Video	Conversation	VAS	Post > Pre
Hickey et al., 2004	Pre/Post	Conversational partner training	2	1 M 1 F	72, 77	37, 72	10 UG, 5 GS	Effectiveness, comfort (partner/PWA)	Video	Conversation	VAS	Post > Pre
Kagan et al., 2001	Pre/Post	SCA	40	25 M 15 F	M = 70, SD = 11	12–178	1 expert +1 expert for reliability	Interaction, transaction	Video	Conversational semistructured interview	9-point rating scale	Post > Pre
Hopper et al., 2002	Pre/Post	Conversational coaching	2	2 M	76, 41	36, 24	16 UG	Content, pre vs. post video	Video	Conversational story retell	% main concepts, categorical	Post > Pre
Jacobs, 2001	Pre/Post	Linguistic specific treatment	5	3 M 2 F	44–79	19–198	10 GS	Effectiveness, informativeness, listener comfort	Audio	Story retell	DME	Post > Pre (not significant) ↑ CIU: ↑ ratings
Ballard and Thompson, 1999	Pre/Post and 4 wk	Linguistic specific treatment	5	3 M 2 F	38–69	10–168	10 GS	Informativeness and efficiency	Audio	Story retell	Composite score of ratings	Post > Pre for 3 of 5
Cupit et al., 2010	Pre/Post	Phonological components analysis	11	8 M 3 F	50–84	18–204	10 SLPs, 20 naive	Informativeness, skill and ease	Audio	Story retell	Likert à composite score	Post > Pre ↑ propositions ↑ ratings
Goral and Kempler, 2009	Pre/Post	CIAT targeting verbs	1	1 M	60	144	20 UG + GS	Favorability	Audio	Personal narrative	Likert	Post > Pre
Ross and Wertz, 1999	M = 4.8 mo between sessions	No Treatment	22	18 M 4 F	32–78	3–210	10 GS	Communicative ability of 2nd compared to 1st sample	Video	Picture description	Categorical (1 = worse, 2 = same, 3 = better)	Changes in aphasia severity correlated with differences in ratings

Abbreviations and notes: All students who served as raters were in communication sciences and disorders programs. SLP, speech-language pathologist; SCA, supported conversation for adults with aphasia (Kagan, 1999), phonological components analysis (Leonard et al., 2008); DME, direct magnitude estimation scale; F, female; M, male; PWA, people with aphasia; VAS, visual analog scale; CIAT, constraint-induced aphasia treatment; GS, graduate students; UG, undergraduate students. In Goral and Kempler (2009), favorability reflected ease of interacting, stress of interacting, awkwardness, success, listener comfort, etc.

of a PWA's discourse. Five studies explicitly asked raters to indicate the effectiveness of the PWAs in conveying information (Cupit et al., 2010; Goral & Kempler, 2009; Hickey & Rondeau, 2005; Jacobs, 2001; Lustig & Tompkins, 2002); three had raters directly compare samples from two time points (Ballard & Thompson, 1999; Hopper et al., 2002; Ross & Wertz, 1999); one had raters compare conversations between PWA and a conversational partners who had or had not been trained in supportive communication strategies (Hickey et al., 2004). In all but one of the treatment studies, ratings from naive raters at posttreatment correlated with improvements in objective outcome measures associated with the interventions. In Jacobs (2001), there was not a significant increase in the ratings of informativeness of the PWAs' story retell but the increases in content information units following linguistic specific treatment correlated with ratings of overall informativeness by graduate students.

Five studies (Hickey et al., 2004; Hickey & Rondeau, 2005; Hopper et al., 2002; Kagan et al., 2001; Lustig & Tompkins, 2002) rated communicative success in conversations collected before and after treatment. Hickey et al. (2004) and Hickey and Rondeau (2005) had raters judge multiple dimensions of conversations (e.g., comfort of partner, comfort of PWA, effectiveness, informativeness, overall quality) between PWAs and volunteers before and after training the volunteers in supportive communication skills. Topics of conversation varied based on the communication partners' preferences, and multimodal communication was encouraged. Raters observed the conversations through video recordings, considering all modes of communication. Hickey et al. (2004) found higher ratings across all dimensions in posttreatment conversations compared to baseline conversations. Hickey and Rondeau (2005) found that naive raters were more likely to rate pretraining exchanges lower for comfort and quality of the interaction than raters who had experience with aphasia, highlighting that raters' knowledge about communication and aphasia may influence their perceptions of these dimensions. Nonetheless, ratings of both the experienced and naive raters documented improved overall conversation quality after communication partner training (Hickey & Rondeau, 2005).

Kagan et al. (2001) took a different approach to rating a conversation-like task. Specifically, to allow exchanges to be more comparable across participants, Kagan et al. (2001) used a conversational semistructured interview with preselected topics, echoing the types of exchanges a PWA might have with a health care provider. Video-recorded interviews between PWA and unfamiliar volunteers were collected before and after the volunteers received communication partner training (supported conversation for adults with aphasia; Kagan, 1999). One experimental rater used a 9-point Likert scale (0.5-point increments) to rate the PWA's interaction and transaction (see Kagan et al.'s Appendix B for definitions) from poor (0) to high (4). A second expert rater rated 25% of interviews, documenting a high intraclass correlation for transaction ($r = 0.84$) and acceptable agreement for interaction ($r = 0.65$). Similar to Hickey et al. (2004) and Hickey and Rondeau (2005), interaction and transaction ratings for PWA documented significant improvements after partners were trained; furthermore, higher ratings were found for interviews with trained as compared to untrained volunteers.

Hopper et al. (2002) trained two PWAs and their communication partners (spouses) using conversational coaching (Hopper et al., 2002; Simmons-Mackie et al., 2014). Using a task similar to Ramsberger and Rende (2002), Hopper et al. (2002) video-recorded co-constructed story retells between the PWA and their partner prior to and after coaching. Undergraduate CSD raters unfamiliar with aphasia or with conversational partners identified the number of main concepts that were successfully conveyed and indicated whether they thought each video was collected before or after treatment. The raters identified more accurate main concepts in the posttreatment conversations for both dyads and correctly identified the pre- and posttreatment clips (100% for both of dyad 1's retells, 100 and 68% for dyad 2's two retells).

Lustig and Tompkins (2002) collected conversation samples before and after one PWA received written communication strategy training, which targeted the use of writing to

converse with unfamiliar partners. In the conversations, the PWA and unfamiliar communication partners actively and meaningfully participated in exchanges about six topics, which involved multiple conversational goals. Using a direct magnitude estimation scale, nine CSD graduate students rated the conversations for understandability of the PWA as well as for difficulty on the part of the PWA, effort on the part of the partner, their estimation of the PWA's satisfaction with communication, and how pleasant the conversation was to watch. Overall, the raters indicated that after treatment, they better understood the PWA and felt more comfortable watching the conversation. Ratings also indicated that posttreatment, the PWA delivered their message more easily, was better understood by the partner, and seemed more satisfied with her communication.

An additional four studies had raters judge the communicative success of PWA based on audio recordings of personal narratives (Goral & Kempler, 2009) or story retells (Ballard & Thompson, 1999; Cupit et al., 2010; Jacobs, 2001). In these studies, the discourse samples often yielded monologues, with minimal to no requirement for interaction, and multimodal communication was not captured, as audio recordings only provided access to audible modes of communication (e.g., words, prosody, sound effects). The context in all cases involved a prescribed topic based on the corresponding discourse elicitation. Raters in Ballard and Thompson (1999), Goral and Kempler (2009), and Jacobs (2001) were CSD students with associated knowledge of communication disorders. In contrast, Cupit et al. (2010) compared the ratings of SLPs and naive listeners, who lacked knowledge of communication disorders. Cupit et al. (2010) and Goral and Kempler (2009) had raters use 7-point Likert scales to rate various features of communicative success (e.g., informativeness, ease of retelling the story/ interacting, skill in transmitting the message, comfort as a listener). Ballard and Thompson (1999) created a composite score of ratings across several domains, including informativeness of the content and efficiency, while raters in Jacobs (2001) used direct magnitude estimation to rate the effectiveness and informativeness of the story retell as well as the comfort of the listener. Based on comparisons of pre/post observer's ratings, Cupit et al. (2010) and Goral and Kempler (2009) documented improved discourse-level communicative success following phonological components analysis and constraint-induced aphasia treatment, respectively. Jacobs (2001) and Ballard and Thompson (1999) showed some evidence of higher ratings following linguistic specific treatment, though improvements were not statistically significant in Jacobs (2001) and were found only in three of five participants in Ballard and Thompson (1999). Also of note, both Jacobs (2001) and Cupit et al. (2010) found that changes in ratings of communicative success were associated with changes in objective discourse analysis measures (e.g., effectiveness, informativeness, and comfort ratings increased in line with increases in CIUs; informativeness ratings improved in line with the use of more propositions per minute).

Finally, one study (Ross & Wertz, 1999) evaluated narratives over two time points, but without an intervention between them. In this study, raters were shown two videos of a PWA describing the WAB-R Picnic Scene acquired with an intervening ~5 months between recordings. Raters were asked to directly contrast the two recordings by indicating whether the second sample was worse than, the same as, or better than the first. These ratings were compared to PWA's change in scores on the *Porch Index of Communicative Ability* (Porch, 1967), WAB-R Aphasia Quotient, and *Communicative Abilities in Daily Living* (Holland, 1980). Findings indicated that change in communication ability on the diverse skills assessed across these standardized tests correlated with the observer's ratings of overall communicative change.

DISCUSSION
Aphasiology has expanded over time to encompass not only the study of language in individuals who have suffered a stroke in the left hemisphere but also to include their communication ability more broadly. Audrey Holland was instrumental in the shift to observing communication as a whole—situated in the context of daily life. With this shift in learning about aphasia through observing it as situated language (Armstrong & Ferguson, 2010; Clark,

2005; Doedens & Meteyard, 2022; Hengst, 2020) has come a plethora of approaches to observing and evaluating PWA in their homes, during errands, and in natural conversations with their partners. However, observations of fully situated language bring their own challenges, including achieving high levels of reliability among coders and accounting for natural variations in activities/environmental contexts, topics and communicative goals, numbers of exchanges, and modes of communication used across PWAs. Such factors have limited the application of situated observations to date.

Situated language recognizes the joint interplay of exchanges with a communication partner, the multiple (verbal and nonverbal) modes through which a message can be conveyed, and the contextual variables that contribute to, or distract from, discourse-level communication success. Ultimately, considering these aspects of situated language in assessment and in documenting intervention outcomes will improve the ecological and social validity of our methods and keep the "PWA participating in life" at the forefront of our efforts.

The current literature review documents considerable gains that have been made in addressing the inherent complexities of measuring real-time interaction through ratings of communicative success (e.g., Leaman & Edmonds, 2021). Yet, the extant literature on communicative success ratings continues to rely predominantly on standardized protocols for eliciting and evaluating discourse genres other than conversation. Here, a review of the use of naive and informed raters in evaluating the discourse-level communication success of PWAs identifies the state of what has been, and can be, learned about communication success by appreciating the role of the observer and their ratings for how they may fit in a situated language framework.

The present review unveiled considerable diversity in the methods used to rate discourse-level communication success, ranging from highly situated conversational tasks to considerably less situated discourse tasks. Yet, each approach contributes to our understanding of language use in aphasia. Holland (1982) involved the most highly situated language use, as the PWAs interacted with communication partners using multimodal communication in real-life situations with abundant opportunities to build common ground. This study recognized that PWAs can be successful communicators despite the severity of their aphasia profiles, particularly when they use total communication strategies and when the function of the informational exchange is prioritized. This approach's benefits, unfortunately, lead directly to its drawbacks: the conversations happened in real time and were not recorded; the metrics were not standardized; and the concept of *success* was not objectively defined. Since then, attempts to systematize highly situated conversational observations have yielded protocols that provide experimental control (Leaman & Edmonds, 2019, 2021) and rating scales for communication success (e.g., Ramsberger & Rende, 2002) that demonstrate good to excellent psychometric properties. Using live and video-recorded conversations and co-constructed story retells observers in these four studies have documented the critical, positive impact that each PWA's repertoire of multimodal communication strategies can have on communicative success. Also, these investigations have highlighted that conversational partners (and observers) who are more familiar/knowledgeable with the PWA and who foster co-constructed exchanges also foster more successful communication.

Thus, also building upon the foundation of Holland's (1982) conversational work has been the development of treatment approaches designed to train PWA and/or their communication partners to implement strategies that facilitate supportive, co-constructed interactions. Hickey et al. (2004), Hickey and Rondeau (2005), Hopper et al. (2002), and Kagan et al. (2001) found overwhelming evidence documenting the success of communication partner training based on the ratings of observers who were unfamiliar with aphasia, intervention, or the transactional nature of situated language.

While the field has made considerable progress in rating the communicative success of PWA in conversations, such ratings predominantly continue to be used within more structured, less situated discourse tasks. The discourse tasks included in this review varied widely in terms of genre or type (personal narratives, picture descriptions/sequences, story

retells procedural recitations, word descriptions), speaker (monologue vs. joint construction), spontaneity (spontaneous vs. scripted), and content familiarity (unfamiliar, familiar, overlearned). Furthermore, the variables rated in less situated tasks were not referred to as "communicative success"; instead, common metrics used across studies included informativeness, effectiveness, favorability, and comfort of the PWA, the partner, and/or the listener/rater. While these studies had low to no requirement for interaction (i.e., were monologues), many manipulated the other aspects of situated language—by either allowing/disallowing multimodal communication or by varying the amount of shared context. For example, de Beer et al. (2017) and Rose et al. (2017) directly contrasted how well a rater understood the information shared in personal narratives when exposed to audio only (speech), visual only (gesture), or full video (speech and gesture), which corroborated findings from conversation that unfamiliar observers perceive greater communicative success/access more information from the PWA when multimodal strategies are used. This finding was particularly true for PWAs who struggle with fluent verbal expression, highlighting the importance of using more situated assessments to fully capture the communicative abilities of such individuals.

There is also value in considering discourse-level communication success when tasks remove multimodal communication, as it provides information about how much an observer can glean from the least situated communication environment. Such studies have illustrated that when raters focus either on narrow aspects of discourse (e.g., content and clarity in Ulatowska et al., 1983; coherence in Kong et al., 2018) or on broad perceptions/feelings about the speaker with aphasia (e.g., liveliness [Groenewold et al., 2014] or favorability [Behrns et al., 2009; Harmon et al., 2016]), raters can detect differences between PWA and controls. Furthermore, these ratings may be valid metrics that align with results from objective macro/microstructural discourse analyses. Also, the degree to which observers share contextual knowledge with the speaker with aphasia may also impact the success of communicative attempts (e.g., knowing possible target words improves prediction of those words

when a PWA is having difficulty with word finding; Tompkins et al., 2006).

One limitation to the use of raters in evaluating discourse-level communication success is implicit bias. In some studies, observers indicated their preferences or perceptions about speakers with aphasia, revealing their biases and how those biases may influence the meaningfulness of their ratings. For example, raters prefer more fluent verbal expression (Harmon et al., 2016); they also assume that PWAs do not like to speak, or their discourse is uninteresting relative to controls (Behrns et al., 2009). While these biases may point to possible treatment targets, such as improving fluency, they also highlight a potential disincentive for partners to engage with PWAs. Thus, the preconceived assumption of communication partners entering situated exchanges with PWA may pose a barrier to successful communication. Similar to results from ratings of conversations, ratings of less situated discourse likewise suggest that training communication partners is essential to easing and improving communication quality and success for PWA. While few communication partners may receive such training, limiting PWAs' communicative participation across daily activities, enriching these interactions through training may have far-reaching benefits to their social well-being. This premise is similar to many successful intervention approaches for individuals with cognitive-communication disorders (c.f., Togher et al., 2023).

Limitations of Observer's Ratings of Communication Success

Doyle et al. (1987) noted that "social validation measures… are crucial to evaluating the effectiveness of treatment programs" and "they have value as pretreatment measures for selecting responses to be targeted for intervention" (p. 153). The social validation measures used in Doyle et al's study were ratings of sentence production adequacy (i.e., stimuli designed to elicit trained structures) in PWAs who participated in a syntax treatment. Given that ratings were not based on discourse, this study was excluded from the present literature review. However, the study's results highlight a key limitation of observer's ratings that was not

apparent in included studies: ratings of syntactic adequacy did not reflect improvements following treatment, whereas studies with ratings of informativeness, effectiveness, comfort, ease, and skill did find associations between improved ratings and improvement on objective discourse analysis metrics (see Table 2). While Doyle and colleagues hypothesized that raters' lack of sensitivity to treatment improvements related to the relatively poor outcome of one PWA, another explanation is that naive raters are not as sensitive to syntactic adequacy as they are to the overall conveyance of meaning and content. That is, because adequacy was defined for raters as the communication of an unambiguous description of the elicitation stimulus, naive listeners may not have felt that the PWA's use of a particular syntactic structure was essential to clearly describing the picture. This explanation is speculative, but it illustrates naive listeners' utility in identifying key aspects of a person's message that are essential to successful communication. In the present review, raters were sensitive not only to the informativeness of content included in PWAs' discourse but also to the skill, ease, and comfort with which they speak. Thus, the use of naive raters as socially valid metrics may be best suited to certain aspects of language (e.g., communicative success, informativeness) as compared to others (e.g., syntax). By nature, this may reflect the degree to which a language sample is situated, as certain elements (e.g., form) may not be essential when raters focus on the function of interactive, multimodal communication that is supported by a context of considerable shared knowledge/common ground.

CONCLUSION

Audrey Holland's legacy in aphasiology is vast but centered on the need for clinicians and researchers to focus on the person with aphasia. This traditional literature review built upon one of her many works, which demonstrated the value of all aspects of situated language from the perspective of the PWA, their communication partner, and the observer. Audrey's emphasis in 1982 was not on the observer, per se, but on successful communication. Ultimately, it is the PWA and their frequent communication partners that are central to that success, and

Audrey Holland wholeheartedly supported the focus on those partners, as seen in the research of Leaman and Edmonds (2024), Hickey et al. (2004), Hickey and Rondeau (2005), Kagan and colleagues (2001), and many others. The current study highlighted the key role observer's ratings play in exploring factors that contribute to PWAs' communicative success when engaging with people they encounter in daily life. We look forward to observing how future superheroes for PWA will continue to use observers' ratings to help PWA participate more fully in their lives.

CONFLICT OF INTEREST
A.E.R., A.L.R., and K.J.G. declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

Anbar, S. (2022). *Goddess Aphasia: A Stroke Survivor and His Dual Muse*. Stillwater River Publications

Ardila, A. (2014). *Aphasia Handbook*. Florida International University. Accessed August 15, 2024 at: https://aalfredoardila.files.wordpress.com/2013/07/ardila-a-2014-aphasia-handbook-miami-fl-florida-international-university2.pdf

Armstrong, E., & Ferguson, A. (2010). Language, meaning, context, and functional communication. *Aphasiology*, *24*(4), 480–496

Azios, J. H., Archer, B., Simmons-Mackie, N., Raymer, A., Carragher, M., Shashikanth, S., & Gulick, E. (2022). Conversation as an outcome of aphasia treatment: a systematic scoping review. *American Journal of Speech–Language Pathology*, *31*(6), 2920–2942

Ballard, K. J., & Thompson, C. K. (1999). Treatment and generalization of complex sentence production in agrammatism. *Journal of Speech, Language, and Hearing Research: JSLHR*, *42*(3), 690–707

Bartlett, M. R., Fink, R. B., Schwartz, M. F., & Linebarger, M. (2007). Informativeness ratings of messages created on an AAC processing prosthesis. *Aphasiology*, *21*(5), 475–498

Behrns, I., Wengelin, A., Broberg, M., & Hartelius, L. (2009). A comparison between written and spoken narratives in aphasia. *Clinical Linguistics & Phonetics*, *23*(7), 507–528

Bock, K., & Levelt, W. (1994). Language production: grammatical encoding. In: M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). Academic Press

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic

dysnomia. *American Journal of Speech-Language Pathology*, *4*(4), 94–98

Bryant, L., Brunner, M., & Hemsley, B. (2020). A review of virtual reality technologies in the field of communication disability: implications for practice and research. *Disability and Rehabilitation. Assistive Technology*, *15*(4), 365–372

Busby, N., Hillis, A. E., Bunker, L., Rorden, C., Newman-Norlund, R., Bonilha, L., Meier, E., Goldberg, E., Hickok, G., Yourganov, G., & Fridriksson, J. (2023). Comparing the brain-behaviour relationship in acute and chronic stroke aphasia. *Brain Communications*, *5*(2), fcad014

Carlomagno, S., Santoro, A., Menditti, A., Pandolfi, M., & Marini, A. (2005). Referential communication in Alzheimer's type dementia. *Cortex*, *41*(4), 520–534

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

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Studies*, *7*(4–5), 507–525

Code, C. (2008). A short history of the past and future of aphasia therapy. University of Exeter. Doi: 10.13140/2.1.3036.3840

Copeland, M. (1989). An assessment of natural conversation with Broca's aphasics. *Aphasiology*, *3*(4), 301–306

Cupit, J., Rochon, E., Leonard, C., & Laird, L. (2010). Social validation as a measure of improvement after aphasia treatment: its usefulness and influencing factors. *Aphasiology*, *24*(11), 1486–1500

de Beer, C., Carragher, M., van Nispen, K., Hogrefe, K., de Ruiter, J. P., & Rose, M. L. (2017). How much information do people with aphasia convey via gesture? *American Journal of Speech-Language Pathology*, *26*(2), 483–497

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: lexical access and grammatical encoding. *Cognitive Science*, *23*(4), 517–542

Dietz, A., Vannest, J., Maloney, T., Altaye, M., Holland, S., & Szaflarski, J. P. (2018). The feasibility of improving discourse in people with aphasia through AAC: clinical and functional MRI correlates. *Aphasiology*, *32*(6), 693–719

Doedens, W. J., & Meteyard, L. (2022). What is functional communication? A theoretical framework for real-world communication applied to aphasia rehabilitation. *Neuropsychology Review*, *32*(4), 937–973

Doyle, P. J., Goldstein, H., & Bourgeois, M. S. (1987). Experimental analysis of syntax training in Broca's aphasia: a generalization and social validation study. *The Journal of Speech and Hearing Disorders*, *52*(2), 143–155

Doyle, P. J., Tsironas, D., Goda, A. J., & Kalinyak, M. (1996). The relationship between objective measures and listeners' judgments of the communicative infor-

mativeness of the connected discourse of adults with aphasia. *American Journal of Speech-Language Pathology*, *5*(3), 53–60

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of verb network strengthening (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, *23*(3), 402–424

Fink, R. B., Bartlett, M. R., Lowery, J. S., Linebarger, M. C., & Schwartz, M. F. (2008). Aphasic speech with and without SentenceShaper: two methods for assessing informativeness. *Aphasiology*, *22*(7-8), 679–690

Gallée, J., Volkmer, A., Whitworth, A., Hersh, D., & Cartwright, J. (2024). Applications of the R.A.I.S.E. Assessment framework to support the process of assessment in primary progressive aphasia. *American Journal of Speech-Language Pathology*, 1–11. Doi: 10.1044/2024_AJSLP-24-00085

Garcia, L. J., Rebolledo, M., Metthé, L., & Lefebvre, R. (2007). The potential of virtual reality to assess communication in aphasia. *Topics in Language Disorders*, *27*(3), 272–288

Gilmore, N., Dwyer, M., & Kiran, S. (2019). Benchmarks of significant change after aphasia rehabilitation. *Archives of Physical Medicine and Rehabilitation*, *100*(6), 1131–1139.e87

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston Diagnostic Aphasia Examination (BDAE-3)* (3rd ed.). Pro-Ed

Goral, M., & Kempler, D. (2009). Training verb production in communicative context: evidence from a person with chronic non-fluent aphasia. *Aphasiology*, *23*(12), 1383–1397

Groenewold, R., Bastiaanse, R., Nickels, L., & Huiskes, M. (2014). Perceived liveliness and speech comprehensibility in aphasia: the effects of direct speech in auditory narratives. *International Journal of Language & Communication Disorders*, *49*(4), 486–497

Haley, K. L., Jacks, A., Jarrett, J., Ray, T., Cunningham, K. T., Gorno-Tempini, M. L., & Henry, M. L. (2021). Speech metrics and samples that differentiate between nonfluent/agrammatic and logopenic variants of primary progressive aphasia. *Journal of Speech, Language, and Hearing Research: JSLHR*, *64*(3), 754–775

Harmon, T. G., Jacks, A., Haley, K. L., & Faldowski, R. A. (2016). Listener perceptions of simulated fluent speech in nonfluent aphasia. *Aphasiology*, *30*(8), 922–942

Hartwigsen, G., Saur, D., Price, C. J., Baumgaertner, A., Ulmer, S., & Siebner, H. R. (2013). Increased facilitatory connectivity from the pre-SMA to the left dorsal premotor cortex during pseudoword repetition. *Journal of Cognitive Neuroscience*, *25*(4), 580–594

Henderson, V. W. (1992). Sigmund Freud and the diagram-maker school of aphasiology. *Brain and Language*, *43*(1), 19–41

Hengst, J. A. (2020). *Understanding Everyday Communicative Interactions: Introduction to Situated Discourse Analysis for Communication Sciences and Disorders.* Taylor & Francis Group

Hickey, E. M., Bourgeois, M. S., & Olswang, L. B. (2004). Effects of training volunteers to converse with nursing home residents with aphasia. *Aphasiology*, *18* (5–7), 625–637

Hickey, E. M., & Rondeau, G. (2005). Social validation in aphasiology: Does judges' knowledge of aphasiology matter? *Aphasiology*, *19*(3–5), 389–398

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393–402

Hickok, G., & Poeppel, D. (2015). Neural basis of speech perception. *Handbook of Clinical Neurology*, *129*, 149–160

Hogrefe, K., Ziegler, W., Weidinger, N., & Goldenberg, G. (2012). Non-verbal communication in severe aphasia: Influence of aphasia, apraxia, or semantic processing? *Cortex*, *48*(8), 952–962. https://doi.org/10.1016/j.cortex.2011.02.022

Holland, A., Fromm, D., Forbes, M., & MacWhinney, B. (2017). Long-term recovery in stroke accompanied by aphasia: a reconsideration. *Aphasiology*, *31*(2), 152–165

Holland, A. L. (1980). *Communicative Abilities in Daily Living (CADL): A Test of Functional Communication.* University Park Press

Holland, A. L. (1982). Observing functional communication of aphasic adults. *The Journal of Speech and Hearing Disorders*, *47*(1), 50–56

Holland, A. L. (2010). Lessons from a clinical life in aphasia. *Topics in Stroke Rehabilitation*, *17*(1), 13–19

Holland, A. L., Fromm, D., & Wozniak, L. (2018). *CADL3 Communication Activities of Daily Living* (3rd ed.). Pro-ed. Accessed August 15, 2024 at: https://www.proedinc.com/Products/14650/cadl3-communication-activities-of-daily-livingthird-edition.aspx

Holland, A. L., & Ramage, A. E. (2000). Thinking outside the (black) box. *Brain and Language*, *71*(1), 93–95

Hopper, T., Holland, A., & Rewega, M. (2002). Conversational coaching: treatment outcomes and future directions. *Aphasiology*, *16*(7), 745–761

Hula, W., McNeil, M., Doyle, P., Rubinsky, H., & Fossett, T. (2003). The inter-rater reliability of the story retell procedure. *Aphasiology*, *17*(5), 523–528

Jacobs, B. J. (2001). Social validity of changes in informativeness and efficiency of aphasic discourse following linguistic specific treatment (LST). *Brain and Language*, *78*(1), 115–127

Kagan, A. (1999). *Supported Conversation for Adults with Aphasia: Methods and Evaluation.* University of Toronto

Kagan, A., Black, S. E., Duchan, F. J., Simmons-Mackie, N., & Square, P. (2001). Training volunteers as conversation partners using "supported conversation for adults with aphasia" (SCA): a controlled trial. *Journal of Speech, Language, and Hearing Research: JSLHR*, *44*(3), 624–638

Kertesz, A. (2007). *Western Aphasia BatteryTM–Revised (WAB–RTM).* Pearson. Accessed August 15, 2024 at: https://www.pearsonclinical.ca/en/products/product-master/item-176.html

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

Kong, A. P.H., Linnik, A., Law, S.P., & Shum, W. W.-M. (2018). Measuring discourse coherence in anomic aphasia using Rhetorical Structure Theory. *International Journal of Speech-Language Pathology*, *20*(4), 406–421

Kong, A. P.H., & Wong, C. W.Y. (2018). An integrative analysis of spontaneous storytelling discourse in aphasia: relationship with listeners' rating and prediction of severity and fluency status of aphasia. *American Journal of Speech-Language Pathology*, *27*(4), 1491–1505

Leaman, M. C., & Edmonds, L. A. (2019). Conversation in aphasia across communication partners: exploring stability of microlinguistic measures and communicative success. *American Journal of Speech-Language Pathology*, *28*(1S), 359–372

Leaman, M. C., & Edmonds, L. A. (2020). "By the way"… how people with aphasia and their communication partners initiate new topics of conversation. *American Journal of Speech-Language Pathology*, *29* (1S), 375–392

Leaman, M. C., & Edmonds, L. A. (2021). Assessing language in unstructured conversation in people with aphasia: methods, psychometric integrity, normative data, and comparison to a structured narrative task. *Journal of Speech, Language, and Hearing Research: JSLHR*, *64*(11), 4344–4365

Leaman, M. C., & Edmonds, L. A. (2024). Pilot results for ECoLoGiC-Tx: a new conversation-level intervention improving language in people with moderate to severe aphasia. *American Journal of Speech-Language Pathology*, *33*(1), 153–172

Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology*, *22*(9), 923–947. https://doi.org/10.1080/02687030701831474

Levelt, W. J. M. (2001). Spoken word production: a theory of lexical access. *Proc. Natl. Acad. Sci. USA*, *98* (23), 13464–13471

Linebaugh, C. W., Shisler, R. J., & Lehner, L. (2005). CAC classics: cueing hierarchies and word retrieval: a therapy program. *Aphasiology*, *19*(1), 77–92

Lustig, A. P., & Tompkins, C. A. (2002). A written communication strategy for a speaker with aphasia and apraxia of speech: treatment outcomes and social validity. *Aphasiology*, *16*(4–6), 507–521

Macwhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: methods for studying discourse. *Aphasiology*, *25*(11), 1286–1307

Martin, N., & Dell, G. S. (2019). Maintenance versus transmission deficits: the effect of delay on naming performance in aphasia. *Frontiers in Human Neuroscience*, *13*, 406

Mayer, J. F., Madden, E. B., Mozeiko, J., Murray, L. L., Patterson, J. P., Purdy, M., Sandberg, C. W., & Wallace, S. E. (2024). Generalization in aphasia treatment: a tutorial for speech-language pathologists. *American Journal of Speech–Language Pathology*, *33*(1), 57–73

McNeil, M. R., & Prescott, T. E. (1978). *Revised Token Test*. PRO-ED

Plowman, E., Hentz, B., & Ellis, C., Jr. (2012). Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. *Journal of Evaluation in Clinical Practice*, *18*(3), 689–694

Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia*, *42*(2), 183–200

Porch, B. E. (1967). *Porch Index of Communicative Ability*. Consulting Psychologists Press

Quaranta, D., Di Tella, S., Marra, C., Gaudino, S., L'Abbate, F., & Silveri, M. C. (2022). Neuroanatomical correlates of semantic features of narrative speech in semantic and logopenic variants of primary progressive aphasia. *Brain Sciences*, *12*(7), 910

Ramsberger, G., & Rende, B. (2002). Measuring transactional success in the conversation of people with aphasia. *Aphasiology*, *16*(3), 337–353

Rose, M. L., Mok, Z., & Sekine, K. (2017). Communicative effectiveness of pantomime gesture in people with aphasia. *International Journal of Language & Communication Disorders*, *52*(2), 227–237

Ross, K. B., & Wertz, R. T. (1999). Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. *Aphasiology*, *13*(2), 113–124

Saldert, C., Fors, A., Ströberg, S., & Hartelius, L. (2010). Comprehension of complex discourse in different stages of Huntington's disease. *International Journal of Language & Communication Disorders*, *45*(6), 656–669

Schuell, H., Shaw, R., & Brewer, W. (1969). A psycholinguistic approach to study of the language deficit in aphasia. *Journal of Speech and Hearing Research*, *12*(4), 794–806

Schwartz, M. F. (2013). Theoretical analysis of word production deficits in adult aphasia. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1634), 20120390

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., Mirman, D., & Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proc. Natl. Acad. Sci. USA*, *108* (20), 8520–8524

Simmons-Mackie, N., Savage, M. C., & Worrall, L. (2014). Conversation therapy for aphasia: a qualitative review of the literature. *International Journal of Language & Communication Disorders*, *49*(5), 511–526

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*; Doi : 10.1111/1460-6984.13028

Suting, L. B., & Mozeiko, J. (2021). Analysis of real-world language use in a person with Wernicke's aphasia. *Perspectives of the ASHA Special Interest Groups*, *6*(3), 553–565

Thompson, C. K., & Shapiro, L. P. (2005). Treating agrammatic aphasia within a linguistic framework: treatment of underlying forms. *Aphasiology*, *19*(10-11), 1021–1036

Togher, L., Douglas, J., Turkstra, L. S., Welch-West, P., Janzen, S., Harnett, A., Kennedy, M., Kua, A., Patsakos, E., Ponsford, J., Teasell, R., Bayley, M. T., & Wiseman-Hakes, C. (2023). INCOG 2.0 Guidelines for cognitive rehabilitation following traumatic brain injury, Part IV: cognitive-communication and social cognition disorders. *The Journal of Head Trauma Rehabilitation*, *38*(1), 65–82

Tompkins, C. A., Scharp, V. L., & Marshall, R. C. (2006). Communicative value of self cues in aphasia: a re-evaluation. *Aphasiology*, *20*(7), 684–704

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

van Middendorp, J. J., Sanchez, G. M., & Burridge, A. L. (2010). The Edwin Smith papyrus: a clinical reappraisal of the oldest known document on spinal injuries. *European Spine Journal*, *19*(11), 1815–1823

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

Wilson, S. M., Entrup, J. L., Schneck, S. M., Onuscheck, C. F., Levy, D. F., Rahman, M., Willey, E., Casilio, M., Yen, M., Brito, A. C., Kam, W., Davis, L. T., de Riesthal, M., & Kirshner, H. S. (2023). Recovery from aphasia in the first year after stroke. *Brain*, *146*(3), 1021–1039