REVIEW

International Journal of Language & Communication Disorders

Towards efficient, ecological assessment of interaction: A scoping review of co-constructed communication

Marcella Carragher[1,2] 🄾 | **Zaneta Mok**[3,4] | **Gillian Steel**[1,2] | **Paul Conroy**[5] | **Kathryn Pettigrove**[1,2] | **Miranda L. Rose**[1,2] | **Leanne Togher**[1,6]

[1]Centre for Research Excellence in Aphasia Recovery and Rehabilitation, Melbourne, Victoria, Australia

[2]School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, VIC, Australia

[3]School of Allied Health, Faculty of Health Sciences, Australian Catholic University, Melbourne, VIC, Australia

[4]Speech Pathology, Alfred Health, Melbourne, VIC, Australia

[5]School of Health Sciences, The University of Manchester, Manchester, UK

[6]School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

Correspondence
Marcella Carragher, Centre for Research Excellence in Aphasia Recovery and Rehabilitation, Australia
Email: m.carragher@latrobe.edu.au

Funding information
La Trobe University; School of Allied Health, Human Services and Sport

Abstract

Background: The complexity of communication presents challenges for clinical assessment, outcome measurement and intervention for people with acquired brain injury. For the purposes of assessment or treatment, this complexity is usually managed by isolating specific linguistic functions or speech acts from the interactional context. Separating linguistic functions from their interactional context can lead to discourse being viewed as a static entity comprised of discrete features, rather than as a dynamic process of co-constructing meaning. The ecological validity of discourse assessments which rely on the deconstruction of linguistic functions is unclear. Previous studies have reported assessment tasks that preserve some of the dialogic features of communication, but as yet, these tasks have not been identified as a distinct genre of assessment. We suggest the term 'co-constructed communication' to describe tasks which are specifically designed to capture the dynamic, jointly produced nature of communication within a replicable assessment task.

Aims: To identify and summarize how co-constructed communication has been assessed with individuals with non-progressive acquired communication disability regarding task design, measures and psychometric robustness.

Methods: A scoping review methodology was used to identity relevant studies. Systematic database searches were conducted on studies published before July 2021. Studies in the yield were assessed against eligibility criteria, with 37 studies identified as eligible for inclusion.

Main contribution: This is the first time that co-constructed communication has been defined as a genre of discourse assessment for stroke and traumatic brain injury populations. Co-constructed communication has been assessed for 144 individuals with aphasia and 111 with cognitive–communication disability. Five categories of co-constructed communication tasks were identified, ranging in complexity. Variability exists in how these assessment tasks are labelled

and measured. Assessment measures require further psychometric profiling, specifically regarding test–retest reliability and validity.

Conclusions: Co-constructed communication is a discourse genre which offers researchers and clinicians a replicable method to assess language and communication in an experimentally rigorous way, within an ecologically valid context, bridging the gap between experimental and ecological assessment approaches.

What this paper adds

What is already known on this subject

- Standardized assessments of language skills and monologue offer reliable, replicable ways to measure language. However, isolating language from an interactional context fundamentally changes the behaviour under study. This raises questions about the ecological validity of the measures we routinely use to determine diagnoses, guide treatment planning and measure the success of treatment.

What this study adds to the existing knowledge

- This review highlights studies that conceptualize, and often quantify, interaction by combining experimental rigour and aspects of everyday dialogue. This is the first time this genre of discourse assessment has been identified. We propose the term 'co-constructed communication' to describe this genre and provide an operational definition for the term.

What are the practical and clinical implications of this study?

- Co-constructed communication assessment tasks require refinement, particularly regarding aspects of psychometric robustness. In the future, these tasks offer pragmatic, meaningful ways to capture the effect and impact of aphasia and cognitive–communication disability within interaction.

BACKGROUND

In its broadest definition, communication includes all means by which 'one mind can affect another' (Weaver, 2007: 27). Communication is a dynamic, hybrid process drawing on language, cognitive resources and social knowledge to connect, share and receive information, manage disruptions, display identity, shift between social roles, and negotiate relationships. In clinical and research settings, the complexity of communication presents challenges for assessment, outcome measurement and intervention (e.g., Doedens & Meteyard, 2019). In aphasiology in particular, traditional approaches to assessment and

intervention usually seek to limit the inherent complexity of communication by isolating specific linguistic functions (such as naming) or speech acts (such as requesting). There is a logic to deconstructing communication into its component parts: decontextualized language samples can reduce any potential confounding influence from sources such as variable skill of communication partners, familiarity between the dyad, and the individual's knowledge and interest of the topic. However, neurologically intact individuals adapt language and interaction to suit their social context and communication partner (Gumperz, 1992; Simmons-Mackie, 2018). Therefore, it is not surprising that people with aphasia also show adaptation

by performing differently across structured assessment and everyday contexts (Beeke et al., 2003, 2007; Leaman & Edmonds, 2021) and across different discourse genres (Alyahya et al., 2020; Armstrong et al., 2011). It is essential that the design of the elicitation task is carefully considered when sampling connected speech.

Assessment of interaction falls within the study of 'discourse', which is itself a tricky term (Hengst, 2020). Armstrong (2000) outlines the different philosophical perspectives used to conceptualize discourse, broadly classified as structuralist or functionalist (or a mixture of both, as in the cognitivist perspective). A structuralist perspective conceptualizes discourse as a unit of language above sentence level (e.g., Harris, 1988; Schwartz et al., 1994). Within this approach, analysis focuses on the sentences, phrases and words that make up the discourse, that is, the microstructure of discourse. In contrast, a functionalist approach is interested in how meaning is constructed and organized. A functionalist-oriented analysis takes a more semantic, macro-structure interest and acknowledges the integral role of context within the discourse, for example, measures of communicative effectiveness, topic maintenance and turn-taking. These various philosophical perspectives influence whether discourse is conceptualized either as a *product* of the constituent words, phrases and sentences, or as a *process* that emerges from interaction (Hengst, 2020). Hengst and Sherrill (2021) point out that in most everyday situations, discourse is a complex *process* that emerges from interaction between individuals and groups. This distinction is illustrated by Clark's (2021) framework of '*in vacuo*' and 'in situ ' language, which facilitates us to think about how we assess and treat communication in people with acquired communication disabilities. *In vacuo* refers to language that has been stripped from the interactional context and is useful for considering components such as syntax or morphology. This describes much of the impairment-level assessment of aphasia involving picture naming, picture description, spoken word to picture-matching or producing a decontextualized monologue. However, to understand the interpersonal process of speaking and understanding, Clark (2021) argues it is necessary to study language in situ. Clark's (1996) framework sets out 'situated language use' as a face-to-face interaction between at least two participants using multiple communicative modalities, interacting in real time, where the interaction is embedded within the context of the immediate environment and the dyad's relationship. Doedens and Meteyard (2022) further refine the differences between *in vacuo* and in situ language in terms of task complexity, goal, context and channel of communication. The authors argue that situated language tasks are necessarily more multifaceted than *in vacuo* tasks (complexity); that *in*

vacuo tasks usually have a narrow goal of verification while in situ tasks expand to include both interaction and transaction (the goal of the task); that context is usually absent in *in vacuo* tasks but accumulates in situated language tasks (context); and that *in vacuo* tasks usually a single communication modality while situated language is multi-modal (the channel of communication).

Focusing assessment and intervention solely on *in vacuo* components may mean we miss the opportunity to understand and ultimately alleviate the social consequences of aphasia for the individual and their close others (Togher, 2003). This has implications for therapy planning, evaluating outcomes and measuring generalization to everyday communication. Assessing language in isolation creates problems further downstream when we try to reintegrate component behaviours into a rich everyday interactional context. Generalization from impairment-level interventions to everyday language use is inconsistent (Best et al., 2011; Boyle et al., 2022; Conroy et al., 2009; Hickin et al., 2022). Springer and colleagues highlight how therapy sessions are designed so that the individual can focus on one problem at a time (as is the case in many impairment-level interventions), whereas in conversation they are confronted with linguistic, cognitive and emotional processing demands simultaneously (Springer et al., 2000). While people with aphasia can learn targeted skills in a structured environment, they struggle to shift these skills into more typical everyday communication (e.g., Ballard & Thompson, 1999; Purdy et al., 1994), and require additional prompting (Purdy et al., 1994) or specific generalization-focused training (e.g., Coelho, 1990; Garrett et al., 1989; Robson et al., 1998). Similarly, individuals with communication disabilities following traumatic brain injury (TBI) show little evidence of generalization from decontextualized drills (Politis & Norman, 2016) and may require coaching that is context specific (Kennedy, 2017) and opportunities to target communication skills in natural settings (Ylvisaker et al., 2005). in essence, lack of generalization to everyday communication may relate to the nature of the intervention itself; if therapeutic targets are stripped of their typical interactional context, generalization from grammatical therapies to everyday communication may be unlikely (Best et al., 2016). Therapy protocols do not usually require that the component parts are sequenced together and practiced as a whole (one exception is whole-part learning described by Milman et al., 2014). In the context of cognitive–communication disability, Meulenbroek and colleagues argue that generalization should be planned for (Meulenbroek et al., 2019). Another challenge is the absence of valid, reliable, clinically feasible tools that directly quantify outcomes in conversation (Kurland et al., 2021). While there is an expansive range of discourse measures available (Bryant et al., 2016;

Pritchard et al., 2017), these typically focus on monologue tasks using counts of linguistic behaviours, which do not examine 'the social devices and strategies that help us craft social interaction' (Simmons-Mackie, 2000: 166).

Another challenge relates to how we design tasks to elicit interactive data within clinical settings. Togher (2003) points out the artificiality of conversation within clinical settings, where interactions more closely resemble interviews due to clinicians' use of imperatives, prompts and initiation–response–evaluation sequences (Leahy, 2004; Simmons-Mackie & Damico, 1999). How we interact as clinicians with our clients is something that can change, given the right training and a philosophy to interact as partners rather than as experts (Hengst & Duff, 2007). The status quo, however, means that 'asymmetrical power relations' position the clinician to maintain the role of expert in the conversation, while the client assumes a more passive, compliant role (Hengst & Duff, 2007: 47). This relationship differential has potential implications for the representativeness of sample obtained as well as the types of linguistic functions or speech acts (e.g., Beeke et al., 2003).

It is possible to design assessment and intervention tasks that preserve features of everyday communication, such as conversational discourse. For example, Ramsberger and Menn (2003) describe a 'conversation under glass' task using an 'observable, controllable, analysable contrived task [that] is realistic enough so that it affords a valid index of the unobservable, uncontrollable, unanalysable behaviour that we really want to know about' (288). Other studies have drawn on a barrier task originally designed by Clark and colleagues (Clark & Wilkes-Gibbs, 1986b), later adapted and developed for people with communication disability (Doedens et al., 2021; Hengst, 2003). These tasks share common features of preserving aspects of everyday communication (such as situating the interaction within the context of the environment) within observable, controllable and analysable tasks (Ramsberger & Menn, 2003). We suggest the term 'co-constructed communication' to broadly capture these tasks. Although co-constructed communication tasks are not new in the literature, they are dispersed over evidence bases. In the absence of a consensus definition, we have suggested an operational definition of co-constructed communication (Table 1) that builds upon Clark's description of situated language. Co-constructed communication is semi-structured conversation that is based on a task, where the shared goal is to reach mutual understanding and the target is known and is verifiable. The interaction is a dialogue between two (or more) speakers and occurs at the level of connected discourse. Both communication partners play a shared role in the interaction. A separate systematic review has investigated *intervention* targeting co-constructed communication (Hall et al., 2023).

Thinking about discourse more broadly may help to contextualize co-constructed communication as a genre. Inspired by Webster et al. (2015) and Doedens (2022), Figure 1 outlines a ladder of discourse complexity from picture description to unstructured dialogue. Each step represents an incremental (although not necessarily equal) increase in linguistic, cognitive and social demands. Co-constructed communication is a type of semi-structured dialogue (step 4 in Figure 1). In comparison to more structured discourse genres (steps 1–3), co-constructed communication involves increased availability of context, interaction with a communication partner and options for multi-modal communication.

Aims

The purpose of this scoping review is to systematically identify and summarize how co-constructed communication has been assessed for individuals with communication disability. As a preliminary exploration, scoping was restricted to two broad clinical populations: post-stroke aphasia and cognitive–communication disability following TBI. Given the dispersed nature of these assessments and the gaps in our knowledge of discourse measures in general, a scoping review was considered the most appropriate methodology. This review will address the following research questions:

1. What **tasks** have been used to elicit co-constructed communication involving individuals with aphasia or cognitive–communication disability?
2. What **measures** have been used to assess co-constructed communication?
3. What information is available on the **reliability and validity** of co-constructed communication assessment?

Methodology

The methodology for this scoping review was informed and guided by best practice (Arksey & O'Malley, 2005; Levac et al., 2010; Peters et al., 2017). Reporting in this paper follows the guidelines set out in the PRISMA scoping review extension (Tricco et al., 2018).

Database search to identify relevant studies

Five electronic databases were systematically searched up to July 2021: MEDLINE, CINAHL (EBSCO), Cochrane

TABLE 1 Operational definition of co-constructed communication tasks

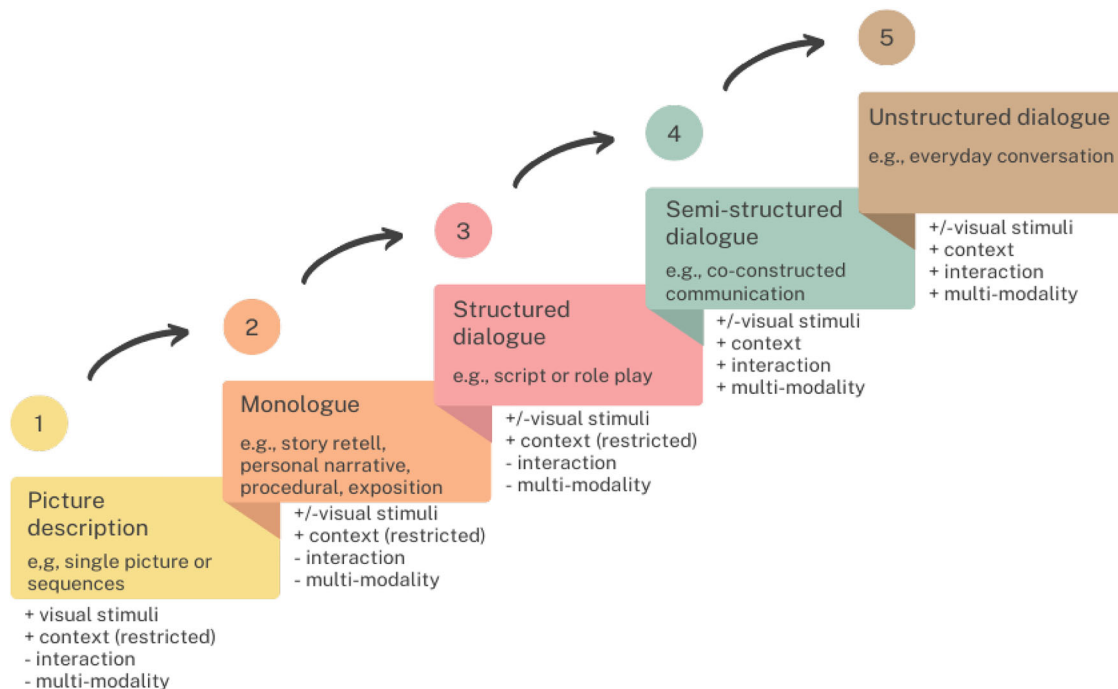| Characteristic	Description
The interaction occurs within a task	The interaction occurs within the context of a task that is set up by the researcher/clinician. The goal of the task is made clear to the participants. A stimulus is provided and forms the trigger for the interaction
Participants have a shared, collaborative goal:	Participants work towards a shared, collaborative goal of reaching a mutual understanding or solving a problem. To do this, participants may need to negotiate, clarify and check each other's understanding and repair misunderstandings
The target of the task is known and is verifiable	The stimulus generates discussion between the dyad regarding target content that can be verified. As such, the interaction can be scored, and performance quantified for the purposes of description or outcome measurement
The interaction is dyadic (or multi-party)	The task stimulates the participants to engage in a dialogue. While the genre of the task imposes certain boundaries on the interaction, participants are free to choose how they interact and respond to one another (contributions are not predetermined in any way) and context is an available resource to the dyad. Contributions from either/both individuals may be the focus of analysis
The level of interest is connected discourse	The interaction occurs at the level of connected speech and goes beyond an exchange of adjacency pairs, that is, two-part exchanges where the first part requires a specific second part (Schegloff & Sacks, 1973), e.g., 'How much does it cost?' 'Twenty dollars'
The communication partner has a shared role	The contributions of the partner are crucial in ensuring an ecologically valid interaction between the dyad. As such, the communication partner is not confined to the role of 'interested listener', they are not the expert in the interaction nor are they primarily engaged in a cueing activity (e.g., test questions). Instead, the partner is a free to use an unrestricted range of communication behaviours (back-channelling, questions, clarification, suggestions) to collaborate in the construction of the message. Both participants have an equal opportunity to contribute to the interaction
The burden of communication may vary (task spectrum)	Either participant may hold the information that needs to be communicated or neither member holds the information. Therefore, one participant may be blinded to the target content or both participant are blinded

FIGURE 1 Stepped discourse complexity.
Sources: Inspired by Webster et al. (2015) and Doedens (2022).
[Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Database search terms.

Population		Communication profile		Task		Focus
stroke **OR** traumatic brain injur* **OR** TBI **OR** acquired brain injur* **OR** ABI **OR** brain injur* **OR** head injur*	**AND**	aphasia **OR** dysphasia **OR** language impair* **OR** acquired language dis* **OR** cognitive language **OR** cognitive linguistic **OR** cognitive pragmatic **OR** high* level language **OR** higher order language **OR** complex language **OR** cognitive communicat* **OR** social communication	**AND**	co-constructed communication **OR** coconstructed communication **OR** co-constructed narrative **OR** coconstructed narrative **OR** co-constructed interaction **OR** coconstructed interaction **OR** joint* produc* **OR** situated language **OR** socially constructed narrative **OR** everyday communication **OR** socially oriented assess* **OR** interactive elicited narrative **OR** transaction* success **OR** problem solving interaction **OR** structured conversation **OR** co-narrat* **OR** dialogue **OR** narrative **OR** discourse **OR** referential communication	**AND**	assess* **OR** measur* **OR** analys* **OR** quantif*

library, EMBASE, PsycINFO. No restrictions were imposed regarding year of publication. Search terms were designed to reflect the clinical populations of interest as well as the likely variability in terms used to describe the assessment of interest. Key search terms were developed on the basis of studies that were recognized as targeting co-constructed communication and which formed the impetus for the scoping review. These included the assessments described by Ramsberger and Rende (2002) and Kilov et al. (2009). The search was piloted with the guidance of a university librarian; the terms was refined based on the search yields. The final search terms are outlined in Table 2.

Eligibility criteria and study selection

Studies were considered eligible if they met the following criteria:

- Original data published in a peer-reviewed journal with the full text available in English
- Adult participants (more than 18 years) with aphasia following stroke (any severity) or cognitive–communication disability (any severity) following TBI
- Any communication partner (including family or professional)
- Participants engaged in a co-constructed dialogue task, that is, (1) a task-oriented interaction, (2) consists of dialogue, (3) where the target is known, (4) a stimulus prompts the interaction, and (5) the interaction occurs at the level of connected speech.

The search yield was imported into Endnote and Covidence; manual and automatic checks identified duplicates which were subsequently removed. Figure 2 outlines the PRISMA flow diagram. Two authors (MC and KP) piloted the eligibility criteria during the title and abstract screening, independently screening 20 papers. They achieved 85% agreement and comparison of decisions led to small refinements to the eligibility criteria (i.e., specifying the exclusion of studies that recruited people with a progressive language/communication disability; and adding detail to the definition of 'co-constructed communication'). For title and abstract screening, 20% of citations were double screened (MC and KP), achieving 88% agreement. For full text screening, 20% of citations were double screened (MC and KP), achieving 98% agreement. Disagreements were resolved through discussion and a consensus reached.

Data charting

The first author developed a draft data extraction template based on the research questions of the scoping review. The draft template was piloted and subsequently refined by two authors (MC and ZM) (Table 3). For 50% of the studies, data were extracted by two authors (MC and GS) and checked by a third author (ZM); for the remaining studies, data were extracted by one author (MC). As outlined in Table 3, for each paper, contextual information was extracted to enable adequate description of the yield in terms of the clinical sample (columns 1 and 2). The third column contains prompts on information to extract in order to answer
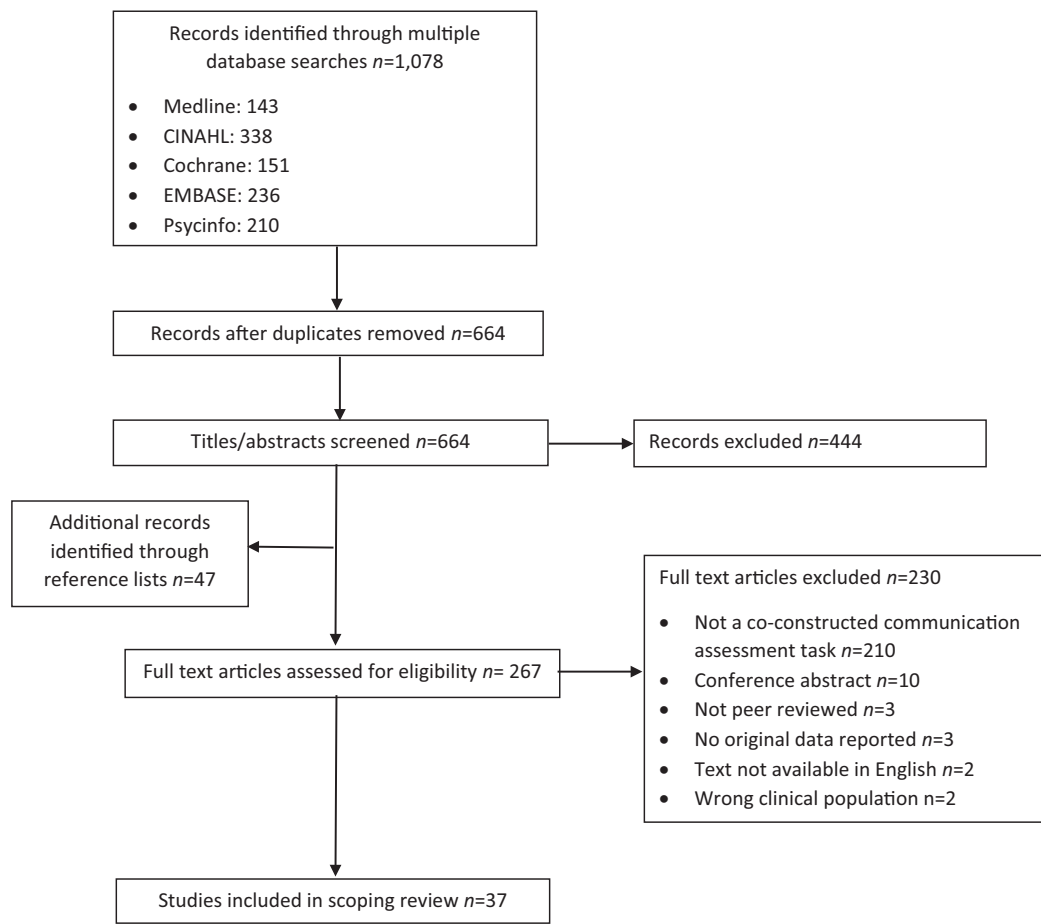
```
┌─────────────────────────────────────┐
│ Records identified through multiple  │
│    database searches n=1,078         │
│                                      │
│  • Medline: 143                      │
│  • CINAHL: 338                       │
│  • Cochrane: 151                     │
│  • EMBASE: 236                       │
│  • Psycinfo: 210                     │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│ Records after duplicates removed     │
│              n=664                   │
└─────────────────────────────────────┘
                  │
                  ▼
┌──────────────────────────────┐      ┌──────────────────────────┐
│ Titles/abstracts screened     │─────▶│ Records excluded n=444   │
│           n=664               │      └──────────────────────────┘
└──────────────────────────────┘
      │
┌──────────────────────┐
│ Additional records    │
│ identified through    │──────┐
│ reference lists n=47  │      │
└──────────────────────┘      │
                  │           │
                  ▼
┌──────────────────────────────────┐    ┌─────────────────────────────────────┐
│ Full text articles assessed for   │───▶│ Full text articles excluded n=230   │
│      eligibility n= 267           │    │                                     │
└──────────────────────────────────┘    │ • Not a co-constructed communication│
                  │                      │   assessment task n=210             │
                  ▼                      │ • Conference abstract n=10          │
┌──────────────────────────────────┐    │ • Not peer reviewed n=3             │
│ Studies included in scoping       │    │ • No original data reported n=3     │
│      review n=37                  │    │ • Text not available in English n=2 │
└──────────────────────────────────┘    │ • Wrong clinical population n=2     │
                                         └─────────────────────────────────────┘
```

FIGURE 2 PRISMA flow diagram of the selection of sources of evidence.

TABLE 3 Data extraction template.

Study	Participants	Task
• Author(s) • Year of publication • Country where the study was published/conducted • Study design • Aim of the study • Population • Sample size	• Age • Gender • Education • Time post-onset • Severity and subtype • Cognitive profile • Communication partner (relationship to the participant, age, education)	• Description of task • Aim of the task • Stimuli used • Location • Number of trials • Role of the participant versus partner • Blinding of either participant • Data collection (video or audio recording, online, checklist) • Transcription • Description of the therapy study (aim of therapy, was co-constructed communication used to evaluate treatment effect or generalization) • Control data (absent/present, matched) • Analysis (macro/micro level of analysis; blinding of assessors) • Measures of co-constructed communication and their psychometric properties • Assessment of the relationship between measures of co-constructed communication and a standardized assessment • Findings (for the clinical and control samples)

the research questions, that is, task description, measures, as well as reliability and validity data reported.

RESULTS

Summary

A total of 37 studies were included from database and reference list searches, published between 1980 and July 2021, consisting of 27 assessment studies and 10 treatment studies (Carlomagno et al., 2005a, 2005b, 2013; Carragher et al., 2015; Correll et al., 2010; Devanga et al., 2021; Doedens et al., 2021; Feyereisen et al., 1988; Flowers & Peizer, 1984; Gordon & Duff, 2016; Guo & Togher, 2008; Gupta et al., 2011, 2012; Hengst, 2003, 2006b; Hengst et al., 2008, 2010; Hopper et al., 2002; Hux et al., 2010; Jorgensen & Togher, 2009; Kilov et al., 2009; Linebaugh et al., 1985; Lyon et al., 1997; Marsh & Knight, 1991; Marshall et al., 1997; Nykänen et al., 2013; Ramsberger & Rende, 2002; Rousseaux et al., 2010; Simmons-Mackie et al., 2005; Togher & Hand, 1998; Togher et al., 1996, 1997a, 1997b, 2004; Tu et al., 2011; Wambaugh et al., 1990; Yorkston et al., 1980). Data were reported for a total of 255 participants: 144 with aphasia and 111 with cognitive–communication disability. The studies were mostly from the United States ($n = 18$) and Australia ($n = 10$), with nine studies published elsewhere (Italy, England, Belgium, France, Finland and New Zealand).

What tasks have been used to elicit co-constructed communication?

Before reviewing the tasks reported in the literature, it is worth nothing the variety of labels used to describe tasks of co-constructed communication. A total of 26 unique labels were used to describe the assessment task, with some studies using multiple labels (Table 4). The most common label was 'collaborative referencing/referential task'. The research team categorized tasks into subcategories based on the nature of the task and stimuli used; five subcategories were identified. The most common task in the literature was a referential communication task ($n = 17$). Other tasks included a message exchange task ($n = 8$), seeking specific information during a telephone call ($n = 6$), a joint problem-solving task ($n = 5$) and a naming task ($n = 1$). See Appendix A for a description of the assessment tasks by individual study.

Referential communication tasks were often based on or adapted from tasks developed for neurologically healthy participants (Clark & Wilkes-Gibbs, 1986a; Glucksberg et al., 1975). These tasks were used in both the aphasia and cognitive–communication disability litera-

ture. In the aphasia literature, referential communication tasks were sometimes described as PACE-like (promoting aphasics' communicative effectiveness; Davis, 2005; Davis & Wilcox, 1985), a treatment designed to incorporate elements of everyday communication such as equal participation within the dyad, exchange of new information, availability of multimodality communication and feedback between the sender and receiver. The general procedure of these referential communication tasks consisted of the dyad sitting facing each other and communicating to distinguish between stimuli and/or to manipulate the order so that each other's workspaces match. One participant takes the role of 'director', instructing their partner ('matcher') to select or arrange the stimuli in a particular order. The dyad may swap the roles of director/matcher but not always (Feyereisen et al., 1988; Flowers & Peizer, 1984; Gordon & Duff, 2016; Linebaugh et al., 1985; Yorkston et al., 1980). Common stimuli in these tasks were black and white abstract tangram cards and the stimuli remained available to each participant throughout the interaction. A low barrier was often (but not always) present to allow for non-verbal communication, whilst concealing the stimuli and maintaining the communicative imperative. There were usually no constraints placed on the interaction, with two exceptions: one group of studies set a time limit of 3 min for the interaction (Flowers & Peizer, 1984; Yorkston et al., 1980); in another group of studies, the communication partner provided feedback in an incremental fashion, that is, providing progressively more specific request/feedback with each failed attempt, culminating in making a guess (Carlomagno et al., 2005a, 2005b, 2013). Typically, the communication partners were family members of the clinical participants. Studies using referential communication tasks cited a range of influencing theories such as Clark's theory of common ground (Clark, 1996), Ylvisaker and colleagues' collaborative and contextualized model of communication (Ylvisaker & Feeney, 1998; Ylvisaker et al., 2003), and evidence of people with aphasia maintaining communication competence despite linguistic disability (e.g., Holland, 1977). Studies using referential communication tasks tended to collect control data from neurologically intact participants, either matched to clinical participants on factors such as gender, age and education (Doedens et al., 2021; Gordon & Duff, 2016; Gupta et al., 2011, 2012; Rousseaux et al., 2010) or unmatched (Carlomagno et al., 2005a, 2005b, 2013; Feyereisen et al., 1988; Linebaugh et al., 1985; Wambaugh et al., 1990).

Message exchange tasks were used exclusively in the aphasia literature and consisted of the clinical participant being presented with novel information to communicate to a blinded communication partner. In these tasks, the clinical participant was always the 'holder' of the new

TABLE 4 Names and types of assessment tasks.

Task	Study	Name of assessment task
Referential communication tasks	Carlomagno et al. (2005a, 2005b, 2013); Linebaugh et al. (1985); Wambaugh et al. (1990); Feyereisen et al. (1988)	Referential communication task
	Doedens et al. (2021)	Collaborative, referential communication task
		Goal-directed communication task
	Flowers and Peizer (1984); Yorkston et al. (1980)	Interaction task
		Information exchange task
	Gordon and Duff (2016); Gupta et al. (2011, 2012)	Collaborative referencing task
		Dynamic, collaborative interactions
	Hengst (2003, 2006); Hengst et al. (2008, 2010)	Game-like barrier task
		Referencing/referential communication task
		Collaborative barrier task
	Rousseaux et al. (2010)	Subtest of the Lille Communication Test (Rousseaux et al., 2001) described as a promoting aphasic's communication effectiveness (PACE) situation
Message exchange tasks (picture, video or verbal stimuli)	Lyon et al. (1997)	Stimulus scenarios
		Communication probe scenarios
	Marshall et al. (1997)	Message exchange task
	Nykänen et al. (2013)	Couple Communication Scale (CCS) (Nykänen et al., 2013)
	Carragher et al. (2015)	Interactive storytelling
	Correll et al. (2010)	Video retelling task
	Hopper et al. (2002)	Videotaped stories as conversational topics
	Ramsberger and Rende (2002)	Semi-structured conversation
		Transactional conversation
	Hux et al. (2010)	Visual scene display (VSD)
		Conversational interactions involving information transfer
Telephone enquiry tasks	Guo and Togher (2008); Togher et al. (Togher et al. (1996, 1997a, 1997b, 2004); Togher and Hand (1998)	Telephone service enquiries or encounters
		Specific telephone enquiry
		Telephone calls, exchanges or interactions
		Information exchange
Joint problem-solving tasks	Kilov et al. (2009); Tu et al. (2011)	Shared problem-solving task/discussion
		Information exchange task
	Marsh and Knight (1991)	Problem-solving tasks
		Naturalistic assessment procedure
	Jorgensen and Togher (2009)	Jointly produced narrative task
	Simmons-Mackie et al. (2005)	Discussion of television programs
Collaborative naming task	Devanga et al. (2021)	Collaborative confrontation naming (CCN)

information and so the dyad had to collaborate to establish common ground; there was no alteration of roles. Stimuli included video clips from Mr Bean, Real TV and I Love Lucy (Carragher et al., 2015; Correll et al., 2010; Hopper et al., 2002; Ramsberger & Rende, 2002), scenarios of increasing complexity (Lyon et al., 1997), a graded hierarchy of stimuli from picture/object stimuli to discussion of current affairs (Nykänen et al., 2013) and a personally relevant conversation on the topic of vintage car restoration (Hux et al., 2010). For 7/8 studies, visual stimuli was not available to the participants during the interaction; the remaining study experimentally manipulated the presence/absence of visual stimuli (Hux et al., 2010). Communication partners were mostly familiar to the clinical participant, with the exception of two studies that recruited unfamiliar participants (Hux et al., 2010;

Ramsberger & Rende, 2002) and one study that did not report the relationship between the dyad (Marshall et al., 1997). Generally, there were no constraints on the dyad with regards to how they communicated; the only exceptions to this included a time limit on the interaction of 4.5 min (Hux et al., 2010) and disallowing the clinical participant to draw when the target stimuli consisted of a picture (Nykänen et al., 2013). Studies in this category cited influencing theories/philosophies including conversational coaching (Hopper et al., 2002), thinking for speaking (Marshall, 2009), story grammar (Rumelhart, 1975), social consequences of aphasia (Kagan, 1995; Sarno, 1993; Schiffrin, 1988) and co-construction of communication (Goodwin, 1995). Control data were reported by two studies, unmatched by the clinical participants (Carragher et al., 2015; Lyon et al., 1997).

Telephone enquiry tasks have been used exclusively in the cognitive–communication disability literature. The communication partner in these interactions was always a genuine person in a professional role, such as a bus service telephone operator or police personnel. In these tasks, the clinical participant was always in the role of requesting information, for example, making an enquiry to the police station about obtaining/regaining a driver's licence or regarding an employer criminal check to return to work. The stimuli consisted of the prompt for the telephone call, which was based on participants' real-life needs; these were discussed between the participant and researcher prior to the telephone call. It is unclear if participants had access to the prompts during the interaction. No constraints were placed on the interaction but the nature of the telephone call naturally restricted non-verbal communication. Theories drawn on included rules of cooperation and politeness (Grice, 1975), establishing and maintaining 'face' (Brown & Levinson, 1987), a systemic functional grammar which views language as a resource within a social context (e.g., Halliday, 1970), the power imbalance in a therapist/client dynamic (Silvast, 1991), and the structures of service encounters between neurologically healthy participants (Ventola, 1987). The same participants were reported across a number of papers (Togher & Hand, 1998; Togher et al., 1996, 1997a, 1997b); control data were collected from neurologically intact participants matched for age, gender and education (Togher & Hand, 1998; Togher et al., 1996, 1997a, 1997b, 2004).

Joint problem-solving tasks were used in both the aphasia and cognitive–communication disability literature. These tasks were characterized by both members of the dyad being privy to the novel information, with the aim of the task being for the dyad to work together to solve a problem or to discuss the information together. In one type of problem-solving task, the dyad was presented with an unfamiliar, low-frequency object (such as

an aid prescribed by occupational therapists to someone with a hemiplegia) and asked to work out the name and function of the object (Kilov et al., 2009; Tu et al., 2011). In another task, the dyad watched a video clip together and recounted it to an apparently blinded researcher, who wanted to know if the video would be useful for other clients (Jorgensen & Togher, 2009). Stimuli consisted of physical objects (Kilov et al., 2009; Tu et al., 2011), video clips (Jorgensen & Togher, 2009; Simmons-Mackie et al., 2005) and picture stimuli (Marsh & Knight, 1991). Presence of the visual stimuli during joint problem-solving tasks was mixed: some tasks were designed to provide participants with access to the stimuli during the interaction (Kilov et al., 2009; Marsh & Knight, 1991; Tu et al., 2011), while in other tasks, the stimuli were withdrawn during the interaction (Jorgensen & Togher, 2009; Simmons-Mackie et al., 2005). Joint problem-solving tasks were characterized by shared roles between the clinical participant and their communication partner, as they both had equal access to the stimulus content, and the task did not impose roles on the interaction (e.g., director versus matcher). The only constraints were related to maximum time for discussion (Kilov et al., 2009; Simmons-Mackie et al., 2005; Tu et al., 2011). Communication partners in these tasks were always familiar to the clinical participant, which is consistent with the premise of aiming to assess everyday problem-solving. Studies in this category drew on theories including systemic functional grammar (e.g., Halliday, 1970), the role of participants (co-tellers and active listeners) on story-telling (Norrick, 2000), skills involved in jointly producing narrative (Hartley, 1995), and the impact of scaffolding supports from familiar communication partners in narratives and conversations of children with TBI (Ylvisaker et al., 1998). The same participants were reported across two studies (Jorgensen & Togher, 2009; Kilov et al., 2009) and control data reported for neurologically intact participants match for age, gender, education (Jorgensen & Togher, 2009; Kilov et al., 2009) and premorbid IQ (Marsh & Knight, 1991).

The sole **collaborative naming task** (a probe task within an aphasia treatment study; Devanga et al., 2021) was similar to a traditional confrontation naming task, but with a twist: the clinical participant was asked to name picture stimuli using any form of expression and they could ask their communication partner for help to name the target. No constraints were placed on the interaction. The 'answer' consisted of single words that could be scored as correct/incorrect, but the process of arriving at the answer could involve a conversation between the participants in the dyad. The stimuli consisted of personally relevant referencing targets (people, places, locations, activities), with four variations of each target (half of which were used in treatment, half as probe items). Visual stimuli were

available to participants throughout the task. The communication partner was the researcher and the dyad did not alternate roles. The study drew on theories of communication and referencing as a collaborative process (Clark, 1992; Hanks, 1990) as well as a model of rich communicative environments (Hengst et al., 2019).

A total of 23 studies (62%) transcribed the assessment data for analysis. Nine studies did not use transcription, instead conducting analysis from video recordings of the data, live scoring during the session or coding video data using specialized software. Two studies did not report on transcription.

What measures have been used to assess co-constructed communication?

Measures were identified and mapped against existing frameworks. Many of the identified measures fell within Armstrong's (2000) definition of functional measures, that is, concerned with the meaning and how meanings are organized within the discourse. In order to tease apart these measures, Armstrong's classification was expanded by including Clark's (1996) framework of situated language use. Therefore, measures were identified as predominantly assessing aspects of interaction, multimodal expression, contextual/common ground, or cognitive–linguistic function (the latter drawing on Armstrong's structural versus structural–functional classification).

A total of 95 unique measures were identified, detailed in Appendix B. The majority of measures ($n = 72$) were categorized as assessing an aspect of interaction. The remaining measures consisted of cognitive–linguistic measures ($n = 13$), measures of common ground ($n = 6$), and multimodal expression ($n = 8$). Most of the measures are objective ($n = 76$) including counting frequency and proportions of behaviours of interest. Only four measures were published assessments: the Lille Communication Test (Rousseaux et al., 2001), the Measure of Skill in Supported Conversation and the Measure of Participation in Supported Conversation (Kagan et al., 2004), and the Global Social Impression Ratings Scales (Bond & Godfrey, 1997). The majority of measures were bespoke and developed for a specific study or group of studies.

Measures of interaction ($n = 72$) demonstrated the most variety and were further categorized as measures related to type of verbal contributions ($n = 21$), accuracy ($n = 17$), communicative ease/effectiveness ($n = 16$), repair ($n = 10$), time ($n = 6$), and verbal play ($n = 2$) (Table 5). Measures of interaction focused on behaviours related to the clinical participant, the partner, the dyad or a combination of these.

- The 'type of verbal contributions' category captures a range of measures relating to how participants contributed within the dialogue, for example, the overall number of turns/moves taken by the participant and the type of that contribution (e.g., number of conversational turns, frequency of interruptions, frequency of politeness markers as well as measures drawing on exchange structure analysis such as K1 moves, K2 moves and dynamic moves) (see Appendix B for definitions). This category of measures simply quantifies the type of verbal contribution; subsequent interpretation of the measure requires consideration of the context. For example, a low score on the 'number of conversational turns' might be appropriate for an individual with TBI in dialogue with a police officer, but out of character for a person with aphasia in dialogue with their spouse.

- A second category focused on accuracy of contributions within the interaction. Accuracy was conceptualized in various ways related to the clinical participant (e.g., production of wrong or irrelevant information, the number of key ideas communicated), the partner (e.g., the partner's understanding was operationalized as ratings of their understanding, the content they produced or the efficiency of their response) or the dyad (e.g., collaborative naming negotiated by the dyad, co-construction of main concepts, and number of misunderstandings).

- A third category of interactive measures, communicative ease/effectiveness, included measures that captured qualitative aspects of the interaction (such as communicative burden, listener ratings of ease and/or effectiveness) as well as others that focused on support provided by the partner (such as offering interpretations, initiating repair), and composite rating scales that included scores for accuracy, completeness, speed and relevance of the interaction (e.g., Lyon et al., 1997; Marshall et al., 1997). One measure (card placement sequences) specifically measured collaborative effort: this was operationalized as how efficiently (i.e., number of turns, the need for revisions or extra turns) a dyad completed a sequence of identifying a target card and placing it in the correct order so that the participants' workspaces matched.

- Measures of repair included number of breakdowns (at the level of the dyad), self- and other-initiated repair (for the participant and partner), and frequency of repair (at the level of the dyad). This category also included frequency of negative teaching instances, defined as instances when the partner corrected a successful communication attempt (Simmons-Mackie et al., 2005).

- The various measures of time all related to the dyad and quantified the efficiency of the dyad's interaction.

- Less common measures included those related to qualitative aspects of the interaction, that is, instances of

TABLE 5 Measures of co-constructed communication focused on the clinical participant, communication partner and/or dyad.

			Participant	Partner	Dyad
Interaction	Type of verbal contribution	Miniturns or Turns at Talk score (Carlomagno et al. 2005a, 2005b, 2013)			×
		Number of conversational turns (Hux et al., 2010)		×	
		Total number of moves (Guo & Togher, 2008; Togher et al. (1996, 2004; Kilov et al., 2009)	×	×	
		Dynamic moves/min (Guo and Togher, 2008; Togher et al. (1997a)	×	×	
		Percentage of moves which composed each GSP[a] element (Togher et al., 1997b, 2004; Kilov et al., 2009)	×	×	
		Types of moves (synopic or dynamic) by each participant (Togher et al., 1996)	×	×	
		Type of exchange lead (Togher et al., 1996)	×	×	
		Number and % of exchanges initiated by the participant (Togher et al., 1996)	×		
		% K1 moves (Jorgensen & Togher, 2009)	×		
		Interactional turns (Gupta et al., 2011, 2012; Gordon & Duff, 2016; Hengst, 2003)	×	×	
		Total number of exchanges (Guo & Togher, 2008)			×
		K1 moves/min (Guo & Togher, 2008; Togher et al., 1997a; Tu et al., 2011)	×		
		K2 moves/min (Guo and Togher, 2008; Togher et al., 1997a; Tu et al., 2011)	×		
		Frequency of politeness markers per total number of clauses (Togher & Hand, 1998)	×	×	
		Number of speaker initiations and responses (Hux et al., 2010)	×	×	
		Number of interruptions (Simmons-Mackie et al., 2005)		×	
		Number of convergent questions (Simmons-Mackie et al., 2005)		×	
		Form of the message (Flowers & Peizer, 1984)		×	
		Apparent purpose of the message (Flowers & Peizer, 1984)		×	
		Number and % of exchanges initiated by the partner (Togher et al., 1996)		×	
		Communicative functions including regulation, statement, exchange, personal, conversation and miscellaneous categories (Wambaugh et al., 1990)	×	×	
	Accuracy	Collaborative confrontation naming scored for accuracy, responsiveness, specificity, promptness and efficiency (Devanga et al., 2021)			×
		Task accuracy (Doedens et al., 2021; Flowers & Peizer, 1984)			×
		Accuracy of card placement (Gordon & Duff, 2016; Gupta et al., 2011, 2012; Devanga et al., 2021; Hengst et al., 2010)			×
		Errors in referent selection (Linebaugh et al., 1985)			×
		Production of irrelevant information (Carlomagno et al., 2005a, 2005b)	×		
		Production of wrong information (Carlomagno et al., 2005a, 2005b)	×		
		Number and % of main concepts successfully co-constructed (Hopper et al., 2002)			×
		Number of salient ideas communicated by the PWAs (Carragher et al., 2015)	×		
		Number of salient ideas communicated by the partner (Carragher et al., 2015)		×	×
		Semantic content of initial descriptions (Gupta et al., 2011)	×		
		Content units (Hux et al., 2010)		×	
		Crucial information score (Carlomagno et al. 2005a, 2005b, 2013)	×		
		Number of main ideas in the partner's retelling (Ramsberger & Rende, 2002)		×	
		Accuracy of the partner's understanding (Yorkston et al., 1980)		×	×
		Number of misunderstandings (Carlomagno et al. 2005a, 2005b, 2013)			×
		% of aberrant moves in Service Request and Service Compliance elements (Togher et al., 2004)	×	×	
		% essential units of information (Jorgensen & Togher, 2009)	×		

(Continues)

TABLE 5 (Continued)

			Participant	Partner	Dyad
	Communicative ease/ effectiveness	Card placement sequences (Devanga et al., 2021; Hengst, 2003; Hengst et al., 2010)			×
		Perception of communicative ease and effectiveness (Hux et al., 2010)	×		
		Communicative efficiency of the partner's response (Feyereisen et al., 1988)		×	×
		Communicative burden (Marshall et al., 1997)		×	
		Listener comfort ratings (Guo & Togher, 2008)		×	
		Lille Communication Test—PACE subtest (Rousseaux et al., 2010)	–	–	–
		Communication interaction (Lyon et al., 1997)			×
		Communicative effectiveness composite score (Carlomagno et al., 2005a)			×
		Verbal social skills (Marsh & Knight, 1991)	×	×	×
		Couple Communication Scale (CCS) (Nykänen et al., 2013)			×
		Communicative efficiency: completeness, clarity, speed (Marshall et al., 1997)	×		
		Measure of Participation in Supported Conversation (MPC) (Correll et al., 2010; Tu et al., 2011)	×		
		The Adapted Global Social Impression Rating Scales (Tu et al., 2011)			×
		Social validation measure (Hopper et al., 2002)			×
		Measure of Skill in Supported Conversation (MSC) (Correll et al., 2010; Tu et al., 2011)		×	
		The amount and type of conversational support provided by the listener (Carlomagno et al., 2013)		×	
	Repair	Refashioning (Gupta et al., 2011)			×
		Self-initiated repair (Doedens et al., 2021)	×		
		Clarification requests or other initiated repair (Doedens et al., 2021)	×	×	
		Number of communication breakdowns (Linebaugh et al., 1985)			×
		% breakdowns which were repaired (Linebaugh et al., 1985)			×
		Number of contingent query-revision sequences per breakdown (Linebaugh et al., 1985)	×	×	
		Frequency distributions of contingent queries and revisions (Linebaugh et al., 1985)	×	×	
		Query-revision distributions and communicative success (Linebaugh et al., 1985)			×
		Number of negative teaching instances (Simmons-Mackie et al., 2005)		×	
		Challenging moves per minute (Tu et al., 2011)	×	×	
	Time	Trial time (Doedens et al., 2021)			×
		Task duration (Flowers & Peizer, 1984; Yorkston et al., 1980)			×
		Total time of the interaction (Feyereisen et al., 1988)			×
		Time to completion (Gordon & Duff, 2016; Gupta et al., 2011, 2012)			×
		Total time (Guo & Togher, 2008)			×
		Duration of each transcript (Togher et al., 2004)			×
	Verbal play	Playful or mischievous episodes and the resources used (Hengst, 2006)	×	×	
		Number and types of verbal play episodes (Gupta et al., 2012)			×
Common ground	Communicative context	Types of initiating referential expressions for each card (Hengst, 2003)	×	×	
		Repetition of card labels (Hengst et al., 2010)			×
		Use of definite reference (Gupta et al., 2012)	×	×	
		Specific referencing expressions for each target (Devanga et al., 2021)			×
		Initial description word count (Gordon & Duff, 2016; Gupta et al., 2011, 2012)	×		
	Situational context (plus multimodality expression)	Card displays (Hengst, 2003)			×

(Continues)

TABLE 5 (Continued)

			Participant	Partner	Dyad
Cognitive linguistic	Structural (word finding)	Number of paraphasias (Carlomagno et al., 2005a, 2005b)	×		
	Functional (cohesion)	% complete cohesive ties (Jorgensen & Togher, 2009)	×		
	Structural (productivity)	Number of words (Carlomagno et al., 2005a, 2005b; Gupta et al., 2011, 2012; Hengst, 2003)	×		
	Structural (complexity)	Conceptual complexity of speaker-initiated turns (Hux et al., 2010)	×	×	
	Structural/functional (content)	Total no. of C-units (Jorgensen & Togher, 2009)	×		
		Words per C-unit (Jorgensen & Togher, 2009)	×		
	Functional (story grammar)	% story grammar elements (Jorgensen & Togher, 2009)	×		
		Narrative sequence (Carragher et al., 2015)	×		
	Memory	Retention of card labels after the trial (Gordon & Duff, 2016)	×		
Multimodal expression		Partner monitoring through eye gaze (Gupta et al., 2011)		×	
		Channel use (Feyereisen et al., 1988)	×		
		Manner of delivery (Flowers & Peizer, 1984)	×	×	
		Number of judges (n = 10) who identified the target referent from participants' gestures (Feyereisen et al., 1988)	×		
		Number of (informative) gestures (Carlomagno et al., 2005a, 2005b; Hengst, 2003)	×		
		Classification of gestures (Carlomagno et al., 2013)	×		
		Number of coverbal gestures (Carlomagno et al., 2013)	×		
		Type of voice (Hengst et al., 2008)	×	×	

Note: [a]Generic Structure Potential (GSP) is an analytic framework applicable to various discourse genres. GSP analyses moves to describe and measure participation within the discourse activity, that is, for a service encounter these include Greeting, Address, Service Initiation, Service Request, etc. For more information, see Appendix B.

play within the interaction, for example, making jokes, teasing, sound effects.

Measures of common ground (n = 6; Table 5) are used in studies employing referential communication tasks, that is, where the dyad communicate about often abstract, ambiguous stimuli, usually with a barrier to create a communication gap. Measures of common ground focused on what labels the dyad used to refer to an abstract stimulus and how their referencing labels changed through repeated trials; that is, how the dyad established common ground (Clark & Wilkes-Gibbs, 1986b). Some measures focused on the joint action of the dyad. For example, one measure focused on repetition of card label within the dyad; that is, one participant offered a name for a picture stimuli (card label), this name was picked up and repeated by the other participant, and the continual repetition of the name led to it being refined to suit the individual stimulus picture. In this sense, repetition between the dyad was seen as a collaborative act of developing and refining specific labels, and therefore supporting the dyad to communicate about specific labels/stimuli. Development of common ground was also conceptualized as moving from indefinite to definite references, for example, 'looks like a person standing' to 'the line' (Gupta et al., 2012). In some tasks, participants were permitted to show their partner the target card over the barrier (card display measure), thereby

making use of items within the environment or perceptual co-presence (Kronmüller & Barr, 2015). With one exception, all measures of common ground included the joint actions of the dyad, rather than the clinical participant in isolation; the exception was the measure of 'initial description word count' (Gordon & Duff, 2016; Gupta et al., 2011, 2012) which quantified only the director's first attempt to describe the target stimulus.

Measures of cognitive–linguistic skill (n = 13; Table 5) included functional–structural measures (number of words, number of paraphasias, number of C-units, words per C-unit), functional measures (percentage of complete cohesive ties, percentage of story grammar elements), content measures (content units, number of salient ideas communicated by the speaker, percentage of essential units of information, semantic content of initial descriptions, crucial information score), using language to present complex ideas (conceptual complexity of speaker-initiated turns) and memory (retention of card labels). On the whole, these measures focused on the clinical participant's output within the interaction.

Measures of multimodality (n = 8; Table 5) predominantly focused on gestures (number of informative gestures, classification of gestures, number of coverbal gestures, number of judges who identified the target from participants' gestures). Other measures included classifying how participants delivered a message (through

verbal, gesture, written/graphic or non-verbal means) and the type of dialogic voice used (e.g., social voice, representing others' words/actions, personalized). Measures of multimodality typically focused on the clinical participant.

What information is available on the reliability and validity of co-constructed communication assessment?

A total of 28 studies (76%) investigated an aspect of reliability or validity. Out of 99 measures, interrater reliability was reported for 53 measures; of these, intra-rater reliability was also reported for 20 measures. Inter- and intra-rater reliability was mostly calculated using percentage agreement; only two studies reported the intraclass correlation coefficient (Doedens et al., 2021; Guo & Togher, 2008). Of those studies that investigated inter- or intra-rater reliability, 43 measures scored > 80% agreement.

Interrater reliability > 80%: For those measures that achieved > 80% rater agreement, most measures were investigated and reported by one study; only four measures achieved satisfactory inter- and intra-rater reliability across multiple studies. These measures include interactional turns (Gordon & Duff, 2016; Gupta et al., 2011, 2012; Hengst, 2003), number of words (Gordon & Duff, 2016; Gupta et al., 2011, 2012; Hengst, 2003), initial description word count (Gordon & Duff, 2016; Gupta et al., 2011, 2012) and total number of moves (Guo & Togher, 2008; Togher et al., 1997a).

Interrater reliability < 80%: Six measures either reported reliability data below 80% agreement or there was a large range around the mean. These included the complexity level of speaker-initiated turns (77% interrater reliability; Hux et al., 2010); number of paraphasias (78% interrater reliability; Carlomagno et al., 2005b); ratings of the comprehensiveness of conversations (71% interrater reliability; Ramsberger & Rende, 2002); ratings of verbal social skills including language, speech delivery, conversation structure and style, (rater reliability ranged from 0.53 to 0.92 with a median of 0.85, Marsh & Knight, 1991); self-initiated repair (ICC = 0.74, CI = 0.51–0.87, Doedens et al., 2021); communicative efficiency defined as the completeness, clarity and speed of information exchange (interrater reliability was relatively poor while intra-rater reliability ranged from 82% to 86%; Marshall et al., 1997).

Rater reliability across studies: Three measures were reported to have mixed reliability across studies. These included counting the number of gestures: Carlomagno et al. (2005b) reported 78% interrater agreement compared with 97.9% reported by Hengst (2003); intra-rater reliability was also poor for this measure (77.4%; Hengst, 2003). Assigning GSP elements to moves also demon-

strated mixed interrater reliability, with reports of 77%, 90.4% and 95%, respectively (Kilov et al., 2009; Togher et al., 1997b; Togher et al., 2004, respectively). Variability was also seen in listener comfort ratings, where intra-rater agreement ranged from 33.3% to 100% with a mean of 81.1% (Guo & Togher, 2008). For interrater reliability 77.0% of the ratings were in agreement by ±1 (range = 57.3−89.1%). The ICC for the group was 0.96 (range = 0.91−0.99) while the individual rater reliability coefficient was 0.66 (range = 0.45−0.87).

Test–retest reliability: This was reported for two measures: Wambaugh et al. (1990) repeated a referential communication task after a 6-week interval with control participants; t-statistics for paired samples were calculated for each communicative function. The authors found no statistically significant differences in the proportional use of communicative functions between the two samples. Ramsberger and Rende (2002) repeated a video retelling task after a 6-month interval; the total number of main ideas from each administration was used to calculate a coefficient of stability. The correlation was moderately high but not significant based on the performance of four participants (rs = 0.80; ps > 0.01).

Internal consistency/evidence of equivalence: Two studies investigated consistency within a test or equivalence across parallel forms of a test. Nykänen et al. (2013) assessed stability within the Couple Communication Scale (CCS): consistency ranged from unacceptable to acceptable across the three subtests (Cronbach's alpha = 0.44–0.77 for object/action picture stimuli; 0.30–0.52 for complex pictures; and 0.61–0.77 for discussion of current affairs). Ramsberger and Rende (2002) investigated (1) the relationship between the number of ideas communicated in conversation for two complex video episodes and two simple episodes, to determine the coefficient of equivalence, and (2) the relationship between comprehensiveness ratings of the simple and complex conversations. The coefficients of equivalence (rs = 0.79; ps < 0.01 for main-idea analysis; rs = 0.77; ps < 0.01 for comprehensiveness ratings) suggest that the different methods of measuring transactional success were moderately high, providing evidence of the equivalence of these alternate forms of testing transactional success.

Validity: Three studies assessed validity: Ramsberger and Rende (2002) assessed construct validity, content validity and concurrent criterion-related validity; Guo and Togher (2008) and Devanga et al. (2021) assessed social validity. Ramsberger and Rende measured transactional success by firstly having the dyad discuss the stimulus (in this case, an episode of I Love Lucy), then the participant without aphasia was asked to retell their understanding of the story. The authors argue that assessing story retellings that are based on conversation provides

acceptable content validity given the evidence of this type of interaction in research and clinical practice. To measure construct validity Ramsberger and Rende drew on Messick's (1989) definition of 'the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores' (13); the authors created a multi-trait–multimethod matrix (Campbell & Fiske, 1959) to measure the construct of interest via multiple methods and to evaluate convergent and discriminant validity. The measure of interest related to transactional success, operationalized as the amount of correct information (number of main ideas) communicated in a retelling of a video episode. For convergent evidence, a different measure of transactional success was used: two judges rated the comprehensiveness/completeness of the partner's retelling of the video on a 5-point scale (a score of 0 indicated that no main ideas were conveyed and a score of 4 indicated all the main ideas had been communicated). For discriminant evidence, a different construct was measured: participants with aphasia completed a picture description monologue task, which was rated using the aforementioned the 5-point scale. Coefficients of equivalence (Spearman rank order) were obtained for the two measures of transactional success (main ideas and scores on the rating scale) in the story retellings conversations and monologue descriptions. The measures of transactional success in story retellings showed a strong positive correlation ($rs = 0.96$; $ps < 0.01$), suggesting this is a valid construct to capture conversational success. The poor correlation between measures of the picture description monologue and story retelling conversations (ratings of picture description and main ideas in conversation: $rs = 0.36$; $ps > 0.01$; ratings of transactional success in conversation with main ideas of picture description: $rs = 0.41$; $ps > 0.01$) suggests that success in conversation differs to that in a monologue task. Ramsberger and Rende (2002) also evaluated concurrent criterion-related validity using a measure of functional communication, the Communication Activities of Daily Living (CADL-2, Holland et al., 1999); they found a moderate but not statistically significant relationship between the number of main ideas transferred in conversation and the CADL−2 stanine scores ($rs = 0.60$; $ps = 0.012$).

Guo and Togher (2008) evaluated social validity by calculating the correlation between listener comfort ratings (in response to listening to a telephone enquiry task where one speaker presented with cognitive–communication disability and dysarthria) and standardized intelligibility test scores (the Assessment of Intelligibility of Dysarthric Speech, Yorkston & Beukelman, 1981). The authors found a significant correlation between listener comfort ratings and the standardized assessment; listener comfort could be predicted with 61% accuracy. Devanga et al. (2021) used a

qualitative approach to investigate social validity of their intervention by conducting interviews with participants immediately post-intervention. Participants with aphasia, their close other and clinician partners were asked to share their views on the social significance of the outcome measures, their perception of any changes on the measures, their perceptions of the intervention and impact of the intervention. Content and thematic analyses were completed on the interview data, with predominantly positive comments made by participants (although the authors note potential investigator influence as clinician partners and moderators participated in the interviews). Separately, the authors argue that validity was demonstrated by way of positive treatment effects for three participants (internal validity) and replication of data across participants (external validity).

DISCUSSION

This review represents the first time that co-constructed communication assessments have been defined and systematically collated within aphasia and cognitive–communication disability literature bases (for a review of co-constructed communication as a therapy task, see Hall et al., 2023). Unlike traditional assessments, co-constructed communication assessments are designed to focus on communication as a joint activity (e.g., Barnes & Bloch, 2019; Clark, 1996): meaning is constructed, negotiated and shaped by the active participation of both participants within the dyad. Co-constructed communication assessments are dispersed across literature bases and use many different labels and descriptions, which dilutes the visibility of the genre. The 37 studies in the current paper employ a range of tasks (mostly commonly, a referential communication task), stimuli, participant roles and measures. Assessment stimuli range in type, complexity, personalization and level of abstraction. The availability of stimuli to both participants in the dyad differs between tasks: at times, the communication partner is blind to the target to create a communicative imperative while at other times, the partner is privy to the target stimulus. The 'rules' of the task define and influence the role of each participant (e.g., which participant is the director versus the matcher) and the actions necessary to complete the task (e.g., directing, negotiating, compromising); beyond this, further constraints on the interaction are not common.

The communication partner within the dyad is often a family member or friend, thus avoiding any potential for institutional power dynamics to influence the interaction (Hengst & Duff, 2007). Even when the communication partner is a researcher, the 'rules' of the task mean they are

an active participant in the interaction as opposed to taking the traditional role of interviewer, expert or interested-but-tacit listener. The skill of the communication partner is not usually a variable of interest in assessment of co-constructed communication; yet, partners do vary in their competence as facilitators (Simmons-Mackie & Kagan, 1999) and the communicative success of the dyad can be altered by training the partner (e.g., Simmons-Mackie et al., 2016). Emerging evidence suggests that some aspects of performance of people with aphasia may be sensitive to the role they play in the interaction, although to date no consensus has been reached regarding the effect of interacting with a familiar versus unfamiliar communication partner (Doedens et al., 2021; Leaman & Edmonds, 2019, 2021). Future studies to refine assessment measures may consider baseline characteristics of the communication partner while further research is needed to clarify the effect of familiarity on the interaction.

The visual stimuli may be present during the interaction or withdrawn (e.g., typically present in referential communication tasks, not typically present in the message exchange tasks). While Doyle et al. (1998) found no effect of visual stimuli on monologue production, there is some evidence that the presence of visual stimuli facilitates interaction. Hux et al. (2010) compared dyads discussing a topic across three experimental conditions: shared visual stimuli, non-shared visual stimuli versus no visual stimuli. They found the shared stimuli condition resulted in more conversational turns, more complex utterances by the communication partners, more content produced by the speaker with aphasia, and higher self-reports of ease, transfer and understandability. The presence/absence of visual stimuli likely supports or taxes cognitive skills such as short-term memory; the scaffolding provided by visual stimuli may free up resources for additional linguistic processing (Boo & Rose, 2011). The availability and type of visual stimuli may need to be carefully considered in future experimental studies. For clinical practice, the different task designs and varying complexity of visual stimuli may present a natural spectrum of difficulty.

A key finding of the review relates to the measurement of co-constructed communication. A relatively small number of studies ($n = 37$) has generated a large number of measures ($n = 95$). This echoes findings elsewhere of a proliferation of measures of discourse (Bryant et al., 2016; Pritchard et al., 2017) and conversation (Azios et al., 2022). In the current review, measures included those from a formal analytic approach (such as Generic Structure Potential) as well as informal measures developed for the specific study. In a recent review of conversation outcome measures, Azios et al. (2022) note not only a lack of consensus regarding what to measure, but also disagreement as to *how* to measure the behaviour of interest.

The authors highlighted that some measures of conversation are based on raw data while others are analysed and reported as proportional data. For the latter, decisions regarding which denominator to use will potentially influence findings (Herbert et al., 2008). Another key finding of the current review relates to a relative absence of psychometric profiling: most measures of co-constructed communication require further investigation of rater reliability, test–retest reliability and validity; again, this echoes findings from the broader discourse literature (Azios et al., 2022; Pritchard et al., 2017). Interrater reliability data are available for 53 measures; however, these were usually based on percentage agreement, which is no longer the accepted standard (Feng, 2014). Intra-rater reliability and test–retest reliability were investigated less often in the studies. Based on current data, the most promising measure identified was a measure of transactional success (Ramsberger & Rende, 2002) which demonstrated interrater reliability (88% agreement) with good construct validity, but fell short of satisfying statistical criteria for test–retest reliability and concurrent criterion-related validity. For reliability across raters, the most promising measures (albeit based on percentage agreement) include number of interactional turns (Gordon & Duff, 2016; Gupta et al., 2011, 2012; Hengst, 2003), number of words (Gordon & Duff, 2016; Gupta et al., 2011, 2012; Hengst, 2003), initial description word count (Gordon & Duff, 2016; Gupta et al., 2011, 2012) and total number of moves (Guo & Togher, 2008; Togher et al., 1997a). Test–retest reliability was demonstrated for only one measure, that is, communicative functions within a referential communication task (Wambaugh et al., 1990). Ongoing psychometric profiling of measures of co-constructed communication is essential to identify a suite of valid measures that offer reliability across raters and time.

A relative strength of the studies included in the review relates to the collection of control data: 19 studies used normative data to compare the performances of dyads (Carlomagno et al., 2005a, 2005b, 2013; Carragher et al., 2015; Doedens et al., 2021; Feyereisen et al., 1988; Gordon & Duff, 2016; Gupta et al., 2011, 2012; Jorgensen & Togher, 2009; Linebaugh et al., 1985; Lyon et al., 1997; Marsh & Knight, 1991; Rousseaux et al., 2010; Togher & Hand, 1998; Togher et al., 1996, 1997a, 1997b; Wambaugh et al., 1990). Normative data may provide a useful quantitative framework for interpreting data from dyads where one speaker has aphasia or cognitive–communication disability. One way this could be operationalized is by building a bank of normative reference data against which we can compare clinical performances. At least some dyads will adapt to communication disability in creative ways not seen (or needed) in neurotypical interactions; these adaptations may relate to factors such as executive function,

emotional intelligence shared within the dyad, or familiarity. Therefore, normative data could guide interpretation and scoring of co-constructed communication without limiting options for exploring findings unique to interactions which include neurodiverse populations.

One important gap in the knowledge base relates to including the experiences of those living with aphasia or cognitive–communication disability in the development of assessment measures. Further work is needed to refine psychometrically robust measures that have been validated by people with lived experience, and to define the margins for what constitutes clinically meaningful change (Breitenstein et al., 2022). Given clinician and consumer priorities for treatment to improve everyday communication (e.g., Wallace et al., 2017a, 2017b), it is likely that assessment of co-constructed communication will be of interest to clinicians and those with lived experience of communication disability. The demands on clinicians' skill and time related to the transcription and analysis of discourse generally have been identified as barriers to clinical implementation (Cruice et al., 2020; Stark et al., 2021). Therefore, future research should consider the barriers and facilitators to implementing this genre of assessment in standard clinical practice. Given the complexity of communication and the many ways to conceptualize communication success (as demonstrated by 95 unique measures identified in this review), it may be likely that a suite of assessment tasks of varying complexity and measures are needed to capture the multimodal and interactive nature of these data. Data in the current study relates to individuals who have experienced stroke aphasia or cognitive–communication disability following TBI; future studies would benefit from including a wider range of acquired communication disabilities, including right hemisphere damage and dementias.

CONCLUSIONS

Co-constructed communication assessment builds upon Clark's framework, with a view to guiding the design of tasks and measures for use within research and clinical settings. This genre of discourse aims to preserve aspects of everyday communication within observable, controllable and analysable tasks (Ramsberger & Menn, 2003). With further development and refinement, co-constructed communication assessments could increase the sophistication of how we manage the complexity of communication for the purposes of assessment, outcome measurement and treatment. The result could be a suite of tasks and measures that validly and reliably capture the effect of intervention as well as some of the interactional and social consequences of communication disability. The findings

of this review and the application of co-constructed communication may be relevant to other clinical populations such as those with motor speech disorders, dementias or right hemisphere damage. It is hoped that this review will serve to highlight the presence of co-constructed communication as a genre of discourse, to guide future work to refine task design and measures, and to influence how the discourse process is conceptualized and operationalized in future intervention studies.

LIMITATIONS

Given the diversity of labelling in the literature, it is possible that our search strategy did not identify some texts that could have met the operational definition of 'co-constructed communication'. Therefore, despite a robust search strategy and piloting of the search, some relevant studies may have been unintentionally omitted from the review.

CONFLICT OF INTEREST DISCLOSURE
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that supports the findings of this study are available in the supplementary material of this article.

PATIENT CONSENT STATEMENT
Not applicable.

PERMISSION TO REPRODUCE MATERIAL FROM OTHER SOURCES
Not applicable.

ORCID
Marcella Carragher   https://orcid.org/0000-0002-7200-6968

REFERENCES
Alyahya, R.S.W., Halai, A.D., Conroy, P. & Lambon Ralph, M.A. (2020) A unified model of post-stroke language deficits including discourse production and their neural correlates. *Brain*, 143, 1541–1554.

Arksey, H. & O'Malley, L. (2005) Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32.

Armstrong, E. (2000) Aphasic discourse analysis: the story so far. *Aphasiology*, 14, 875–892.

Armstrong, E., Ciccone, N., Godecke, E. & Kok, B. (2011) Monologues and dialogues in aphasia: some initial comparisons. *Aphasiology*, 25(11), 1347–1371.

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 review. *American Journal of Speech–Language Pathology*, 16(6), 2920–2942.

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

Barnes, S. & Bloch, S. (2019) Why is measuring communication difficult? A critical review of current speech pathology concepts and measures. *Clinical Linguistics & Phonetics*, 33(3), 219–236.

Beeke, S., Wilkinson, R. & Maxim, J. (2003) Exploring aphasic grammar 2: do language testing and conversation tell a similar story? *Clinical Linguistics and Phonetics*, 17, 109–134.

Beeke, S., Wilkinson, R. & Maxim, J. (2007) Individual variation in agrammatism: a single case study of the influence of interaction. *International Journal of Language & Communication Disorders*, 42, 629–647.

Best, W., Grassly, J., Greenwood, A., Herbert, R., Hickin, J. & Howard, D. (2011) A controlled study of changes in conversation following aphasia therapy for anomia. *Disability & Rehabilitation*, 33(3), 229–242.

Best, W., Maxim, J., Heilemann, C., Beckley, F., Johnson, F., Edwards, S.I., et al. (2016) Conversation therapy with people with aphasia and conversation partners using video feedback: a group and case series investigation of changes in interaction. *Frontiers in Human Neuroscience*, 10, 562.

Bond, F. & Godfrey, H.P. (1997) Conversation with traumatically brain injured individuals: a controlled study of behavioural changes and their impact. *Brain Injury*, 11, 319–329.

Boo, M. & Rose, M.L. (2011) The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology*, 25, 154–175.

Boyle, M., Akersb, C.M., Cavanaugh, R., Hula, W.D., Swiderskic, A.M. & Elman, R.J. (2022) Changes in discourse informativeness and efficiency following communication-based group treatment for chronic aphasia. *Aphasiology*, 563–597. https://doi.org/10.1080/02687038.2022.2032586

Breitenstein, C., Hilari, K., Menahemi-Falkov, M., Rose, M.L., Wallace, S.J., Brady, M.C., et al. (2022) Operationalising treatment success in aphasia rehabilitation. *Aphasiology*.

Brown, P. & Levinson, S.C. (1987) *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.

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.

Campbell, D.T. & Fiske, D.W. (1959) Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 56, 81–105.

Carlomagno, S., Pandolfi, M., Marini, A., Di Iasi, G. & Cristilli, C. (2005a) Coverbal gestures in Alzheimer's type dementia. *Cortex*, 41(4), 535–546. doi: http://doi.org/10.1016/S0010-9452(08)70193-X

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

Carlomagno, S., Zulian, N., Razzano, C., De Mercurio, I. & Marini, A. (2013) Coverbal gestures in the recovery from severe fluent aphasia: a pilot study. *Journal of Communication Disorders*, 46, 84–99.

Carragher, M., Sage, K. & Conroy, P. (2015) Preliminary analysis from a novel treatment targeting the exchange of new information within storytelling for people with nonfluent aphasia and their partners. *Aphasiology*, 29(11), 1383–1408. https://doi.org/10.1080/02687038.2014.988110

Clark, H. (1996) *Using Language*. Cambridge, UK: University Press.

Clark, H.H. (1992) *Arenas of language use*. University of Chicago Press.

Clark, H.H. (1996) *Using language*. Cambridge: Cambridge University Press.

Clark, H.H. (2021) Anchoring utterances. *Topics in Cognitive Science*, 13, 329–350.

Clark, H.H. & Wilkes-Gibbs, D. (1986a) Referring as a collaborative process. *Cognition*, 22, 1–39.

Clark, H.H. & Wilkes-Gibbs, D. (1986b) Referring as a collaborative process. *Cognition*, 22(1), 1–39.

Coelho, C.A. (1990) Acquisition and generalisation of simple manual sign grammars by aphasic subjects. *Journal of communication disorders*, 23(6), 383–400.

Conroy, P., Sage, K. & Ralph, M.L. (2009) Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalise to connected speech? *International Journal of Language & Communication Disorders*, 44(6), 1036–1062.

Correll, A., van Steenbrugge, W. & Scholten, I. (2010) Judging conversation: how much is enough? *Aphasiology*, 24(5), 612–622.

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), 416–442.

Davis, G.A. (2005) PACE revisited. *Aphasiology*, 19(1), 21–38.

Davis, G.A. & Wilcox, M.J. (1985) *Adult aphasia rehabilitation: applied pragmatics*. San Diego, CA: College-Hill Press.

Devanga, S.R., Sherrill, M. & Hengst, J.A. (2021) The efficacy of collaborative referencing intervention in chronic aphasia: A mixed methods study. *American Journal of Speech–Language Pathology*, 30, 407–424.

Doedens, W. (2022) Everyday communication in aphasia rehabilitation. Paper presented at the Online seminar series.

Doedens, W. & Meteyard, L. (2019) Measures of functional, real-world communication for aphasia: a critical review. *Aphasiology*, 34(4). https://doi.org/10.1080/02687038.2019.1702848

Doedens, W. & Meteyard, L. (2022) What is functional communication? A theoretical framework for real-world communication applied to aphasia rehabilitation. *Neuropsychology Review*, 32(4), 937–973. https://doi.org/10.1007/s11065-021-09531-2

Doedens, W.J., Bose, A., Lambert, L. & Meteyard, L. (2021) Face-to-face communication in aphasia: the influence of conversation partner familiarity on a collaborative communication task. *Frontiers in Communication*, 6, 574051.

Doyle, P.J., McNeil, M.R., Spencer, K.A., Jackson Goda, A., Cottrell, K. & Lustig, A.P. (1998) The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. *Aphasiology*, 12(7/8), 561–574.

Feng, G.C. (2014) Intercoder reliability indices: disuse, misuse, and abuse. *Quality & Quantity*, 48(3), 1803–1815.

Feyereisen, P., Barter, D., Goossens, M. & Clerebaut, N. (1988) Gestures and speech in referential communication by aphasic subjects: channel use and efficiency. *Aphasiology*, 2(1), 21–31.

Flowers, C.R. & Peizer, E.R. (1984) *Strategies for obtaining information form aphasic persons*. Paper presented at the Clinical Aphasiology Conference Seabrook Island, SC.

Garrett, K., Beukelman, D.R. & Low-Morrow, D. (1989) A comprehensive augmentative communication system for an adult with Broca's aphasia. *Augmentative and Alternative Communication*, 5, 55–61.

Glucksberg, S., Krauss, R.M. & Higgins, E.T. (1975) The development of referential communication skill. In F. E. Horowitz (Ed.), *Review of child developmental research* (Vol. 4). Chicago: University of Chicago Press.

Goodwin, C. (1995) Co-constructing meaning in conversations with an aphasic man. *Research on Language and Social Interaction*, 28, 233–260.

Gordon, R.G. & Duff, M.C. (2016) Incorporating principles of the collaborative contextualised intervention approach with the empirical study of learning and communication in traumatic brain injury. *Aphasiology*, 30(12), 1461–1482.

Grice, P. (1975) Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: speech acts* (pp. 41–58). New York: Academic Press.

Gumperz, J.J. (1992) Contextualization and understanding. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: language as an interactive phenomenon* (pp. 229–252): Cambridge University Press.

Guo, Y.E. & Togher, L. (2008) The impact of dysarthria on everyday communication after traumatic brain injury: a pilot study.

Gupta, R., Duff, M.C. & Tranel, D. (2011) Bilateral amygdala damage impairs the acquisition and use of common ground in social interaction.

Gupta, R., Tranel, D. & Duff, M.C. (2012) Ventromedial prefrontal cortex damage does not impair the development and use of common ground in social interaction: Implications for cognitive theory of mind. *Neuropsychologia*, 50, 145–152.

Hall, Z., Elbourne, E., Togher, L. & Carragher, M. (2023) Co-constructed communication therapy for individuals with acquired brain injury: a systematic review. *International Journal of Language & Communication Disorders*. doi: https://doi.org/10.1111/1460-6984.12841

Halliday, M.A.K. (1970) Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics* (pp. 322–361). Harmondsworth Penguin.

Hanks, W.F. (1990) *Referential practice*. The University of Chicago Press.

Harris, Z. (1988) *Language and information*. NY: Columbia University Press.

Hartley, L. (1995) *Cognitive–communicative abilities following brain injury: a functional approach*. San Diego, CA: Singular Publishing Group.

Hengst, J.A. (2003) Collaborative referencing between individuals with aphasia and routine communication partners. *Journal of Speech, Language & Hearing Research*, 46(4), 831–848. Retrieved from https://ez.library.latrobe.edu.au/login?url = https://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=106702349&login.asp&site=ehost-live&scope=site

Hengst, J.A. (2006) 'That mea::n dog': linguistic mischief and verbal play as a communicative resource in aphasia. *Aphasiology*, 20(2–4), 312–326.

Hengst, J.A. (2020) *Understanding everyday communicative interactions: introduction to situated discourse analysis for communication sciences and disorders*. New York and London: Routledge.

Hengst, J.A. & Duff, M.C. (2007) Clinicians as communication partners: developing a mediated discourse elicitation protocol. *Topics in Language Disorders*, 27(1), 37–49.

Hengst, J.A., Duff, M.C. & Dettmer, A. (2010) Rethinking repetition in therapy: repeated engagement as the social ground of learning. *Aphasiology*, 24(6–8), 887–901.

Hengst, J.A., Duff, M.C. & Jones, T.A. (2019) Enriching communicative environments: Leveraging advances in neuroplasticity for improving outcomes in neurogenic communication disorders. *American Journal of Speech–Language Pathology*, 28, 216–229.

Hengst, J.A., Duff, M.C. & Prior, P.A. (2008) Multiple voices in clinical discourse and as clinical intervention. *International Journal of Language and Communication Disorders*, 43, 58–68.

Hengst, J.A. & Sherrill, M.J. (2021) Augmenting communicative environments for people with acquired neurogenic disorders: Exploring situated discourse analysis. *Topics in Language Disorders*, 41(1), 27–46.

Herbert, R., Hickin, J., Howard, D., Osborne, F. & Best, W. (2008) Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, 22(2), 184–203.

Hickin, J., Cruice, M. & Dipper, L. (2022) A feasibility study of a novel computer-based treatment for sentence production deficits in aphasia, delivered by a combination of clinician-led and self-managed treatment sessions. *Aphasiology*, 37(10). https://doi.org/10.1080/02687038.2022.2116928

Holland, A. (1977) Some practical consideration in aphasia rehabilitation. In M. Sullivan & M. S. Kommers (Eds.), *Rationale for adult aphasia therapy*. Nebraska: University of Nebraska Medical Center.

Holland, A., Frattali, C. & Fromm, D. (1999) *Communication activities of daily living—2 (CADL-2)*. Austin, TX: Pro-Ed.

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

Hux, K., Buechter, M., Wallace, S. & Weissling, K. (2010) Using visual scene displays to create a shared communication space for a person with aphasia. *Aphasiology*, 24(5), 643–660.

Jorgensen, M. & Togher, L. (2009) Narrative after traumatic brain injury: a comparison of monologic and jointly-produced discourse.

Kagan, A. (1995) Revealing the competence of aphasic adults through conversation: a challenge to health professionals. *Topics in Stroke Rehabilitation*, 2, 15–28.

Kagan, A., Winckel, J., Black, S., Duchan, J., Simmons-Mackie, N. & Square, P. (2004) A set of observational measures for rating support and participation in conversation between adults with aphasia and their conversation partners. *Topics in Stroke Rehabilitation*, 11(1), 67–83.

Kennedy, M.R.T. (2017) *Coaching college students with executive function problems*, The Guilford Press.

Kilov, A.M., Togher, L. & Grant, S. (2009) Problem solving with friends: discourse participation and performance of individuals with and without traumatic brain injury.

Kronmüller, E. & Barr, D.J. (2015) Referential precedents in spoken language comprehension: a review and meta-analysis. *Journal of Memory and Language*, 83, 1–19.

Kurland, J., Liu, A. & Stokes, P. (2021) Phase I test development for a brief assessment of transactional success in aphasia: methods and preliminary findings of main concepts in non-aphasic participants. *Aphasiology*, 37(1). doi: https://doi.org/10.1080/02687038.2021.1988046

Leahy, M.M. (2004) Therapy talk: analyzing therapeutic discourse. *Language, Speech and Hearing Services in Schools*, 35(1), 70–81.

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

Leaman, M.C. & Edmonds, L. (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*, 64(11), 4344–4365.

Levac, D., Colquhoun, H. & O'Brien, K.K. (2010) Scoping studies: advancing the methodology. *Implementation Science*, 20, 5, 69.

Linebaugh, C.W., Margulies, C.P. & Mackisack, E.L. (1985) *Contingent queries and revisions used by aphasic individuals and their most frequent communication partners*. Paper presented at the Clinical aphasiology: Conference proceedings, Minneapolis.

Lyon, J.G., Cariski, D., Keisler, L., Rosenbek, J., Levine, R., Kumpula, J., et al. (1997) Communication partners: enhancing participation in life and communication for adults with aphasia in natural settings. *Aphasiology*, 11(7), 693–708.

Marsh, N.V. & Knight, R.G. (1991) Behavioral assessment of social competence following severe brain injury. *Journal of Clinical Experimental Neuropsychology*, 13, 729–740.

Marshall, J. (2009) Framing ideas in aphasia: the need for thinking therapy. *International Journal of Language & Communication Disorders*, 44, 1–14.

Marshall, R.C., Freed, D.B. & Phillips, D.S. (1997) Communicative efficiency in severe aphasia. *Aphasiology*, 11(4–5), 373–384.

Messick, S.L. (1989) Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education/Macmillan Publishing Company.

Meulenbroek, P., Ness, B., Lemoncello, R., Byom, L., MacDonald, S., O'Neil-Pirozzi, T.M. & Moore Sohlberg, M. (2019) Social communication following traumatic brain injury part 2: identifying effective treatment ingredients. *International Journal of Speech–Language Pathology*, 21(2), 128–142.

Milman, L., Vega-Mendoza, M. & Clendenen, D. (2014) Integrated training for aphasia: an application of part–whole learning to treat lexical retrieval, sentence production, and discourse-level communications in three cases of nonfluent aphasia. *American Journal of Speech–Language Pathology*, 23, 105–119.

Norrick, N.R. (2000) *Conversational narrative: storytelling in everyday talk*. Amsterdam, Philadelphia: John Benjamins Pub.

Nykänen, A., Nyrkkö, H., Nykänen, M., Brunou, R. & Rautakoski, P. (2013) Communication therapy for people with aphasia and their partners (APPUTE).

Peters, M.D.J., Godfrey, C., McInerney, P., Baldini Soares, C., Khalil, H. & Parker, D. (2017) Chapter 11: scoping reviews. In E. Aromataris & Z. Munn (Eds.), *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute.

Politis, A.M. & Norman, R.S. (2016) Computer-based cognitive rehabilitation for individuals with traumatic brain injury: a systematic review. *Perspect ASHA Special Interest Groups*, 1(2), 18–46. https://doi.org/10.1044/persp1.SIG2.18

Pritchard, M., Hilari, K., Cocks, N. & Dipper, L. (2017) Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders*, 52(6), 689–732.

Purdy, M., Duffy, R.J. & Coelho, C.A. (1994) An investigation of the communicative use of trained symbols following multimodality training. *Clinical Aphasiology*, 22, 345–356.

Ramsberger, G. & Menn, M. (2003) Co-constructing lucy: adding a social perspective to the assessment of communicative success in aphasia. In G. Goodwin (Ed.), *Conversation and brain damage*. New York: Oxford University Press.

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

Robson, J., Pring, T., Marshall, J., Morrison, S. & Chiat, S. (1998) Written communication in undifferentiated jargon aphasia: a therapy study. *International Journal of Language & Communication Disorders*, 33(3), 305–328.

Rousseaux, M., Delacourt, A., Wyrzykowski, N. & Lefeuvre, M. (2001) *TLC: test lillois de communication*. Isbergue: Ortho Edition.

Rousseaux, M., Verigneaux, C. & Kozlowski, O. (2010) An analysis of communication in conversation after severe traumatic brain injury.

Rumelhart, D. (1975) Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: studies in cognitive science*. New York, NY: Academic Press.

Sarno, M.T. (1993) Aphasia rehabilitation: psychosocial and ethical considerations. *Aphasiology*, 7, 321–334.

Schegloff, E.A. & Sacks, H. (1973) Opening up closings. *Semiotica*, 7(4), 289–327.

Schiffrin, D. (1988) Conversation analysis. In F. J. Newmeyer (Ed.), *Language: the socio-cultural context* (Vol. IV, pp. 251–276). Cambridge, UK: Cambridge University Press.

Schwartz, M.F., Saffran, E.M., Bloch, D.E. & Dell, G.S. (1994) Disordered speech production in aphasic and normal speakers. *Brain and Language*, 47, 52–88.

Silvast, M. (1991) Aphasia therapy dialogues. *Aphasiology*, 383–390.

Simmons-Mackie, N. (2000) Social approaches to the management of aphasia. In L. Worrall & C. Frattali (Eds.), *Neurogenic communication disorders: a functional approach*. New York: Thieme.

Simmons-Mackie, N. (2018) Communication partner training in aphasia: reflections on communication accommodation theory. *Aphasiology*, 32(10), 1215–1224.

Simmons-Mackie, N. & Damico, J.S. (1999) Social role negotiation in aphasia therapy: competence, incompetence and conflict. In D. Kovarsky, F. J. Duchan & M. Maxwell (Eds.), *Constructing (in)competence: disabling evaluations in clinical and social interactions* (pp. 313–342). Mahwah, NJ: Erlbaum.

Simmons-Mackie, N. & Kagan, A. (1999) Communication strategies used by 'good' versus 'poor' speaking partners of individuals with aphasia. *Aphasiology*, 13(9–11), 807–820.

Simmons-Mackie, N., Kearns, K.P. & Potechin, G. (2005) Treatment of aphasia through family member training. *Aphasiology*, 19(6), 583–593.

Simmons-Mackie, N., Raymer, A. & Cherney, L.R. (2016) Communication partner training in aphasia: an updated systematic review. *Archives of Physical Medicine and Rehabilitation*, 97(12), 2202–2221.

Springer, L., Huber, W., Schlenck, K.J. & Schlenck, C. (2000) Agrammatism: deficit or compensation? Consequences for aphasia therapy. *Neuropsychological Rehabilitation*, 10(3), 279–309.

Stark, B.C., Dutta, M., Murray, L., Fromm, D., Bryant, L., Harmon, T.G., … 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.

The Language Archive. (2019) *(Version 5.7).* T. L. ELAN: Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://archive.mpi.nl/tla/elan

Togher, L. (2003) Do I have green hair? 'Conversations' in aphasia therapy. In S. Parr, J. Duchan & C. Pound (Eds.), *Aphasia inside out* (pp. 65–79). Berkshire, England: Open University Press.

Togher, L. & Hand, L. (1998) Use of politeness markers with different communication partners: an investigation of five subjects with traumatic brain injury. *Aphasiology*, 12(7–8), 755–770.

Togher, L., Hand, L. & Code, C. (1996) A new perspective on the relationship between communication impairment and disempowerment following head injury in information exchanges. *Disability and rehabilitation*, 18(11), 559–566.

Togher, L., Hand, L. & Code, C. (1997a) Analysing discourse in the traumatic brain injury population: telephone interactions with different communication partners. *Brain injury*, 11(3), 169–190.

Togher, L., Hand, L. & Code, C. (1997b) Measuring service encounters with the traumatic brain injury population. *Aphasiology*, 11(4–5), 491–504.

Togher, L., McDonald, S., Code, C. & Grant, S. (2004) Training communication partners of people with traumatic brain injury: a randomised controlled trial. *Aphasiology*, 18(4), 313–335. https://doi.org/10.1080/02687030344000535

Togher, L., Power, E., McDonald, S., Tate, R.L. & Rietdijk, R. (2010) Measuring the social interactions of people with traumatic brain injury and their communication partners: The adapted Kagan scales. *Aphasiology* 24, 914–927.

Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., et al. (2018) PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of Internal Medicine*, 169, 467–473.

Tu, L.V., Togher, L. & Power, E. (2011) The impact of communication partner and discourse task on a person with traumatic brain injury: the use of multiple perspectives. *Brain injury*, 25(6), 560–580. https://doi.org/10.3109/02699052.2011.571655

Ventola, E. (1987) *The structure of social interaction: a systemic approach to the semiotics of service encounters*. London: Pinter.

Wallace, S.J., Worrall, L., Rose, T. & Le Dorze, G. (2017a) Which treatment outcomes are most important to aphasia clinicians and managers? An international e-Delphi consensus study. *Aphasiology*, 31(6), 643–673.

Wallace, S.J., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., et al. (2017b) 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.

Wambaugh, J.L., Thompson, C.K., Doyle, P.J. & Camarata, S. (1990) Conversational discourse of aphasic and normal adults: an analysis of communicative functions. *Clinical Aphasiology*, 20, 343–353.

Weaver, W. (2007) The mathematics of communication. In C. D. Mortensen (Ed.), *Communication theory* (2nd ed., pp. 27–38). Taylor & Francis Group.

Webster, J., Whitworth, A. & Morris, J. (2015) Is it time to stop 'fishing'? A review of generalisation following aphasia intervention. *Aphasiology*, 29(11), 1240–1264.

Ylvisaker, M. & Feeney, T.J. (1998) *Collaborative brain injury intervention: positive everyday routines*. San Diego, CA: Singular Publishing Group, Inc.

Ylvisaker, M., Hanks, R. & Johnson-Greene, D. (2003) Rehabilitation of children and adults with cognitive–communication disorders after brain injury. *ASHA Supplement*, 23, 59–72.

Ylvisaker, M., Sellars, C. & Edelman, L. (1998) Rehabilitation after traumatic brain injury in preschoolers. In M. Ylvisaker (Ed.), *Traumatic brain injury rehabilitation: children and adolescents* (pp. 303–329). Newton, MA: Butterworth-Heinemann.

Ylvisaker, M., Turkstra, L.S. & Coelho, C. (2005) Behavioral and social interventions for individuals with traumatic brain injury: a summary of the research with clinical implications. *Seminars in Speech and Language*, 26(4), 256–267.

Yorkston, K.M. & Beukelman, D.R. (1981) *Assessment of intelligibility of dysarthric speech*. Austin: TX: Pro-Ed.

Yorkston, K.M., Beukelman, D.R. & Flowers, C.R. (1980) Efficiency of information exchange between aphasic speakers and communication partners. Paper presented at the Clinical Aphasiology Conference.

APPENDIX A

TABLE A1 Description of the assessment tasks

Referential communication tasks

Study	Description of the assessment task, instructions and the conversation communication partner (CP)	Restrictions: time, modality, role	Use of barrier or blinding	Stimuli	Data transcribed
Carlomagno et al. (2005a, 2005b, 2013)	**Description/instructions:** describe picture stimuli to a listener, so that they can identify the target picture from a choice of four **CP:** researcher **Roles:** alternate speaker–listener roles (25 trials each)	No restrictions of modalityThe CP provides structured feedback: 1. After failed first attempt, a general request for more information 2. After failed second attempt, a request for specific information plus a guess 3. After failed third attempt, a random guess	A low barrier prevents the dyad from seeing each other's picture stimuli	Black and white line drawings (n = 50)Each trial consists of four stimuli (one target + three distractors)Stimuli are controlled for the amount of information needed to distinguish the target from foils: • Simple stimuli (10× objects, 10× actions) distinguishable by one piece of information • Complex stimuli (persons, objects or animals interacting with or positioned relative to an animal, object or person) distinguishable from foils by one (n = 10), two (n = 10) or three pieces of information (n = 10)	Yes: Carlomagno et al. (2005a)No: Carlomagno et al. (2005b, 2013)
Doedens et al. (2021)	**Description/instructions:** dyads are asked to communicate as they normally would to replicate the set-up of the instructor's room. The researcher leaves the room, returning when the dyad indicate they have completed the task **CP:** familiar CP such as family member or friend **Roles:** alternate instructor–listener roles	None	A low barrier prevents participants from seeing each other's room set-upParticipants are asked not to show items to their partner or to look over the barrier	Playmobile rooms with six pieces of furniture. The location of some items/figures remain constant across trials, while other move. In each trial, one piece is a distractor and is not required in the set-up of the room	No, coded in ELAN (The Language Archive, 2019)
Feyereisen et al. (1988)	**Description/instructions:** the dyad face each other, both with an identical set of cards in front of them, in differing arrangements. The task of the sender is to select a picture and to refer to it by any means so that the receiver can identify the corresponding picture **CP:** researcher **Roles:** the dyad do not alternate roles: the PWA takes the role of sender	None	A low barrier prevents participants seeing each other's cards	Four sets of nine unrelated pictures of common objects (e.g., ring, glass, broom) from Snodgrass and Vanderwart (1980)The last picture is not described since there is no uncertainty at this time	Not reported
Flowers and Peizer (1984)Yorkston et al. (1980)	**Description/instructions:** the dyad sit facing each other with a picture visible only to the PWA. The CP is given a prompt question to ask, i.e., 'What do you see?' or 'What are the people doing?'. The dyad discuss until the CP feels they can write down the answer **CP:** • Flower and Peizer: familiar (n = 5) or unfamiliar (n = 4) • Yorkston et al.: speech pathologists and students, ranging in familiarity and experience **Roles:** the dyad do not alternate roles	3 min limit for each trial/picture	The stimulus is positioned so that it is visible only to the PWA	Flower and Peizer: picture stimuli of objects and actions (n = 20)Yorkston et al.: picture stimuli: • nouns (n = 5) • to elicit a description of why something had happened (n = 5) • not described (n = 5)	Yes: Flowers and Peizer (1984)Yorkston et al. (1980)

(Continues)

TABLE A1 (Continued)

Study	Description of the assessment task, instructions and the conversation partner (CP)	Restrictions: time, modality, role	Use of barrier or blinding	Stimuli	Data transcribed
Gordon and Duff (2016) Gupta et al. (2011, 2012)	**Description**: participants each have a playing board numbered 1–12 and identical sets of cards. The 'director' communicates to the 'matcher' which card to place in each numbered spot so that at the end of the trial, the boards are identical. The participants end each trial when they have placed all cards **Instructions**: The dyad are encouraged to have fun while completing the task and told the accuracy of card placements is more important than time **CP**: familiar partners such as family members or friends **Roles**: • Gordon & Duff (2016), Gupta et al. (2011, 2012); the dyad do not alternate roles. The clinical participant (or the matched control participant) is the director, the CP is the matcher • Hengst (2006, 2003), Hengst et al. (2008, 2010): alternated director–matcher roles	None	A low barrier prevents participants seeing each other's cards	Abstract tangrams (*n* = 12)	Yes: Gordon and Duff (2016), Gupta et al. (2011), Gupta et al. (2012)
Hengst (2003, 2006) Hengst et al. (2008, 2010)				• Hengst (2006): tangrams or familiar photo cards or tangrams—not clearly described • Hengst (2003): black and white tangram shapes, and personalized photo cards of familiar locations • Hengst et al. (2010): colour photographs of familiar people ×10, local street intersections ×10, area buildings ×10) • Hengst et al. (2008): photographs of people and places familiar to the participant	Yes: Hengst (2003, 2006) Hengst et al. (2008, 2010)
Linebaugh et al. (1985)	**Description/instructions**: the CP describes picture stimuli to the PWA, who selects the target from a choice of four pictures (the target plus three foils). The PWA can request additional information **CP**: family member **Roles**: the dyad do not alternate roles	None	A low barrier obscures the target picture from the PWA The PWA is blinded to the target stimulus	Picture stimuli (*n* = 10) Each foil differs from the target by one or two distinguishing features	Yes
Wambaugh et al. (1990)	**Description/instructions**: participants each have a 4 × 4 matrix and 16 small objects. They are asked to arrange the objects so that their matrices match. **CP**: spouse **Roles**: no restrictions on what could be said, who should give instructions or on time	None	Barrier present	Small objects (*n* = 16), e.g., comb, stick of gum, paper clip)	Yes
Rousseaux et al. (2010)	**Description/instructions**: the task forms part of the Lille communication test. The dyad sit facing each other, each with a similar set of images. One participant has the responsibility of communicating a specific image to the partner; the listener must identify which card the speaker is describing **CP**: researcher **Roles**: alternate speaker–listener roles	None	In the listener role, the CP is blinded to the target picture	Picture stimuli from the Lille Communication Test	No, assessments scored live and corrected from audio-video recording

(Continues)

TABLE A1 (Continued)

Message exchange: picture, video or verbal stimuli

Study	Description of the assessment task, instructions and the conversation communication partner (CP)	Restrictions: time, modality, role	Use of barrier or blinding	Stimuli	Data transcribed
Carragher et al. (2015)	**Description**: the PWA recounts a video clip to the CP **Instructions**: the CP is instructed that the PWA has viewed a video clip: they are asked to find out what happened and to later explain their interpretation to the researcher **CP**: spouse **Roles**: the dyad do not alternate roles	None	CP is blinded to the video clip	Video clips of a humorous fictional character (Mr Bean) Stimuli controlled for complexity (simple ×1, complex ×1)	Yes
Correll et al. (2010)	**Description**: the PWA recounts a video clip to the CP **Instructions**: participants are instructed to convey as much detail as possible, that the CP can assist the participant and for them to stop when they do not think they can progress further **CP**: spouse **Roles**: the dyad do not alternate roles	None	CP is blinded to the video clip	Video clip of two men undertaking a sequence of everyday activities including a shopping excursion (4 min)	No, judges viewed the video recordings
Hopper et al. (2002)	**Description/instructions**: the PWA watches a video clip and verifies their understanding by answering 3–5 yes/no questions. The CP enters the room and the dyad discuss the video, communicating as they normally would **CP**: spouse **Roles**: the dyad do not alternate roles	None	CP is blinded to the video clip	2–3 min video segments of humorous, adventurous or dangerous real-life events from the television programme Real TV Stimuli controlled for difficulty (all stimuli rated as moderately difficult by a group speech pathologists and students)	No, judges watched the video recordings and wrote a summary of the information communicated within the dyad
Ramsberger and Rende (2002)	**Description**: the PWA watches a randomly selected video clip, then recounts the story to the CP **Instructions**: CPs are instructed to work together to help the PWA tell the story. The dyads are instructed to continue their conversation until they feel they have achieved maximal exchange of information and to use any means of communication. They are told the metric of success is the number of correct ideas reported by the CP at the end of the task **CP**: each PWA paired with four unfamiliar communication partners (total of 56 dyads) **Roles**: the dyad do not alternate roles	None	CP is blinded to the video clip but knows it is an episode of I Love Lucy	Four episodes of I Love Lucy, varying in complexity (i.e., number of goals and resolutions)	Yes

(Continues)

TABLE A1 (Continued)

Study	Description of the assessment task, instructions and the conversation communication partner (CP)	Restrictions: time, modality, role	Use of barrier or blinding	Stimuli	Data transcribed
Lyon et al. (1997)	**Description/instructions**: the SP presents a randomly selected scenario to the PWA, using verbal output, writing, gesture and drawing. Once the SP has verified the PWA's understanding, the CP enters the room and the dyad discuss the scenario. **CP**: familiar CP (e.g., family member or friend) or a community volunteer. **Roles**: the dyad do not alternate roles	None	The CP is blinded to the target scenario	Scenarios are hypothetical situations varying in length and complexity from levels 1 to 5, e.g., level 1 'You would like to go to a hardware store'; level 5 'There is a worry that we have too much space debris; there is a concern that spaceships orbiting the earth might collide with it'	Not reported
Marshall et al. (1997)	**Description/instructions**: the PWA is provided with a message to communicate to the CP. The CP attempts to work out the message by paraphrasing, verifying and questioning. **CP**: not reported. **Roles**: the dyad do not alternate roles	None	The CP is blinded to the target message	Ten messages, each containing 4–6 key elements. For example, 'The black dog is drinking milk'. Unclear how the stimuli were presented (e.g., verbally or using multimodality channels)	No, ratings completed from video recordings
Nykänen et al. (2013)	**Description/instructions**: the PWA is shown a picture or given verbal information. Once understanding has been verified, the CP enters the room and the PWA attempts to convey the target information. **Instructions**: the PWA may use any means of communication to relay the message to the CP. The dyad is encouraged to take the initiative. **CP**: spouse. **Roles**: the dyad do not alternate roles	For tasks involving picture stimuli, the PWA is not allowed to draw but can draw in the air. The CP can use verbal and non-verbal means (including drawing) to find out the target information	The CP is blinded to the target message	Two versions (A and B) developed, each consisting of tasks of varying difficulty (maximum score 100): • six pictures of individual objects or actions, e.g., brush, key, getting dressing, yawning (maximum of 24 points) • two situational pictures, e.g., building a house, shoplifting (maximum of 32 points) • two verbal assignments, e.g., news item on a housing evacuation due to an electrical fire (maximum of 44 points)	No
Hux et al. (2010)	**Description/instructions**: the dyad discuss a topic of interest to the PWA, with interactions capped at 4.5 min. The presence/absence of visual stimuli is experimentally manipulated. Following the conversation, the CP provides a detailed description of what they have found out. **Instructions**: CPs instructed to find out as much as possible about the car's history and restoration. **CP**: unfamiliar participants recruited to the study. **Roles**: the dyad do not alternate roles	Interactions concluded after 4.5 min		Topic personalized to the PWA: his acquisition and restoration of an antique car. The visual scene display consisted of two photographs (one of the PWA and his antique car and one of a general scene related to antique cars) and 18 written words/phrases related to antique cars	Yes

(Continues)

TABLE A1 (Continued)

Telephone enquiry tasks

Study	Description of the assessment task, instructions and the conversation communication partner (CP)	Restrictions: time, modality, role	Use of barrier or blinding	Stimuli	Data transcribed
Guo and Togher (2008) Togher et al. (1996, 1997a, 1997b, 2004) Togher and Hand (1998)s	**Description:** the participant initiates a telephone enquiry to or an unfamiliar CP in a specific role: • Telephone enquiry to the police to obtain information on regaining a driver's licence (Togher et al., 1996, 1997a, 1997b, 2004; Togher and Hand, 1998) and completing employer criminal checks to return to work (Togher et al., 2004) • Telephone enquiry to a bus timetable service to obtain information about route, times and cost (Togher et al., 1996, 1997a, 1997b; Togher and Hand, 1998; Guo and Togher, 2008) • Telephone enquiry to the participant's mother (Togher et al., 1996, 1997b; Togher and Hand, 1998) • Telephone enquiry to the participant's therapist (Togher et al., 1996, 1997b; Togher and Hand, 1998) **CP:** telephone operator**Roles:** the participant is the requestor of information, the operator the provider of information	None but the nature of a phone call restricts non-verbal communication	None	No visual stimuli; the prompts are discussed between the participant and researcher (Togher et al., 1996, 1997b; Togher and Hand, 1998): the participant phones four CPs (varying in familiarity): • bus timetable service regarding a group outing • police regarding regaining a driver's licence • their mother regarding details of their weekly schedule • therapist regarding details about activities and programs for an insurance report Guo and Togher (2008): the participant phones a bus timetable service to find out: • which bus goes to a shopping centre • departure time to arrive at a particular time • Togher et al. (1997a): the participant phones: • the bus timetable information service regarding a specific bus route, times and cost • the police regarding regaining a driver's licence • Togher et al. (2004): the participant phones the police regarding: regaining a driver's licence (target information: need to get medical clearance; undergo a driver's assessment through a specialist driving school; and approach the Roads and Traffic Authority to reapply for a licence) • completing employer criminal checks to return to work (target information: giving consent to a criminal record search; a specific operational section of the police service can disclose such information; and information is strictly confidential and only made available to the personnel office of a potential employer)	Yes: Guo and Togher (2008), Togher (1997a, 1997b), Togher and Hand (1998), Togher et al. (1996, 2004)

(Continues)

TABLE A1 (Continued)

Joint problem-solving tasks

Study	Description of the assessment task, instructions and the conversation partner (CP)	Restrictions: time, modality, role	Use of barrier or blinding	Stimuli	Data transcribed
Kilov et al. (2009) Tu et al. (2011)	**Description:** dyads problem-solve the name and function of unfamiliar objects, with clues from the researcher **Instructions:** • Kilov et al.: 'I have been given these aids for people who have had a brain injury. We would like to know how difficult it is to work out what they are used for without instructions. Could you have a look at them for me, work out what they are used for, and when you have worked it out, come and get me and explain to me how you would use them. You may need to show me while you are explaining. If you are still stuck after 5 min, come out and I will give you a clue which might help you, but we would like you to try together to work it out for at least five minutes' • Tu et al.: 'I want you to both have a chat' **CP:** • Kilov et al.: friend • Tu et al.: mother and a paid carer **Roles:** equal roles, no blinding	Kilov et al.: at least 5-min discussion before the second clue is provided. For each subsequent clue, dyads are given 2 min to discuss Tu al.: 3-min discussion before the second clue is provided. For each subsequent clue, the dyad is given 3 min	Both members of the dyad are blinded as to the name and function of the stimuli	Unfamiliar real objects: • Kilov et al.: tap turner and belliclamp • Tu et al.: aids prescribed to assist in activities of daily living, e.g., a dressing stick to assist a person with hemiplegia to dress	Yes: Kilov et al. (2009), Tu et al. (2011)
Marsh and Knight (1991)	**Description:** the dyad is presented with problem-solving tasks and asked to reach a consensus decision **Instructions:** • Picture task: 'I want you to choose the picture which you think would look best hanging in this room. You can only choose one picture, so you have to agree together which one it is going to be' • Raffle task: 'I want you to imagine that you have both gone halves in a raffle ticket which has won. I have here a list of prizes and you must choose together which one item you would take as your prize' **CP:** familiar CP **Roles:** equal roles, no blinding	None	None: the dyad has equal access to the stimuli	Picture task: reproductions of eight different paintings Raffle task: a list of 10 raffle prizes of equal monetary value	No, ratings completed from video recordings

(Continues)

TABLE A1 (Continued)

Study	Description of the assessment task, instructions and the conversation partner (CP)	Restrictions: time, modality, role	Use of barrier or blinding	Stimuli	Data transcribed
Jorgensen and Togher (2009)	**Description/instructions**: dyads watch a video clip with the aim of retelling the story to the researcher. The participants are told the researcher has not seen the video and wants to know what it was about, in order to decide whether it will be useful for other clients **CP**: friend **Roles**: equal roles, no blinding	None	None: the dyad watch the video clip together	Video clips from shows about holidays and home improvements	Yes
Simmons-Mackie et al. (2005)	**Description**: the dyad view 3-min video clips together and discuss **CP**: spouse **Roles**: equal roles, no blinding	Discussion for 5 min	None: the dyad watch the video clip together	3-min video clips of television news programs, sports and a popular talk show	No, ratings completed from video recordings
Collaborative naming task					
Devanga et al. (2021)	**Description** (probe task): similar to a confrontation naming task. The dyad are presented with probe cards (*n* = 12) and the participant is asked to name the card using any means of output. The CP can assist if the participant asks for help or is unable to name the target **Instructions**: 'I will give 12 pictures to you, one after the other, and I would like you to give me a name for each card. You can say the name out loud or write it down or use your device. Your clinical partner can jump in, but only if you need any help' **CP**: researcher **Roles**: the dyad do not alternate roles in the probe task	None	None	Personally relevant referencing targets (*n* = 30) chosen by each participant (people, places, location, activities). There are four variations ('views') of each referencing target (total = 120 photos), aimed to reduce automaticity of responses. Views A and B are used as probe cards (views C and D used in treatment)	Yes

TABLE B1 Description of measures relating to domain, operational definition and reliability assessment.

Domain	Name	Operational definition	Reliability
Common ground (communicative context)	**Types of initiating referential expressions for each card placement** Hengst (2003)	The type of initiating referential expressions for each card placement were coded as descriptive, elementary, episodic, provisional, instalment, placeholder, proxy, or other (referencing expressions that were primarily non-verbal)	Point-by-point comparisons: • interrater reliability: 96% • intra-rater reliability: 100%
Interaction (accuracy)	**% of aberrant moves in Service Request and Service Compliance elements** Togher et al. (2004)	Aberrant moves defined as a repetition of information (due to misunderstanding, not following the information provided or forgetting) or use of inappropriate/incomplete elements (due to a delayed response or absence of response) by either participant	Interrater reliability for coding Service Request and Service Compliance: percentage agreement 94% and 92%, respectively
Common ground (communicative context)	**Repetition of card labels** Hengst et al. (2010)	Conceptualized as collaborative repetition as part of learning within the task (therefore, interpreted as an index of learning on the task). Coding of labels during the task: • agreed-upon target label (ATL) • non-agreed-upon target label (NATL) • repetition of agreed-upon target label (R-ATL) • repetition of non-agreed-upon target (R-NATL). • The first use of an ATL by either participant within a CPS was coded as ATL and any subsequent uses within that CPS coded as R-ATL. The first use of any other reference for the target card was coded as NATL and any subsequent uses coded as R-NATL. Repetition types were tallied across sessions	Two teams of coders used a consensus procedure (at least three checks through the data and resolution of all disagreements) to code all labels. If coders did not agree on a code, the referencing expression was coded NATL
Common ground (communicative context)	**Use of definite reference** Gupta et al. (2012)	Shifting from predominately indefinite references (e.g., looks like a person standing) to definite references (e.g., the line) was interpreted as a reference becoming part of shared knowledge or common ground. Each initial description was coded as an indefinite or definite referential expression	Point-to-point agreement on 12% of the data: • interrater reliability: 95% • intra-rater reliability: 96%
Common ground (communicative context)	**Specific referencing expressions for each target** Devanga et al. (2021)	During the intervention, dyads were instructed to create labels for each target card and that a maximum of 3 points could be earned: for using a specific verbal label to refer to the card (1 point), repeating the label to verify the card (1 point), and correct placement of the card on the matcher's board (1 point)	Not assessed
Common ground (situational context), multimodality expression	**Card displays** Hengst (2003)	Card displays were identified as a communicative resource used by dyads. These were counted only when they were clearly visible above the barrier	Interrater reliability: 100% Intra-rater reliability: 100%
Common ground (communicative context)	**Initial description word count** Gordon and Duff (2016) Gupta et al. (2011, 2012)	Defined as a director's entire first attempt at describing each target card (including verbal and non-verbal communication, before the matcher responds. Excluded from this word count were utterances that did not directly relate to referencing the target card, i.e, task management, mazing and discourse markers. The rate of reduction in the word count was interpreted as a critical index of learning	Point-to-point agreement on 12% of the data: • interrater reliability: 98% (Gordon & Duff, 2016), 97% (Gupta et al., 2012), 97% (Gupta et al., 2011) • intra-rater reliability: 99% (Gordon & Duff, 2016), 98% (Gupta et al., 2012), 98% (Gupta et al., 2011)

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (verbal play)	**Playful or mischievous episodes and the resources used** Hengst (2006)	Instances of verbal play were identified using set criteria: • a playful episode may be one utterance or several utterances • several utterances within one episode must be thematically related • laughter alone was not counted as a playful episode but could be part of an episode when it was a response to a playful utterance/action Descriptive codes were applied to characterise the resources participants used (verbal, prosodic, gestural, other) and the communicative functions these playful episodes served (narrative, teasing, referencing, other)	Reliability assessment not appropriate to the study design. Coding was completed by a team of research assistants and the author (the latter was the consensus coder for all analyses)
Interaction (collaborative effort, feedback, co-construction, backchannels)	**Card placement sequences (CPS)** Devanga et al. (2021) during the intervention task Hengst (2003) Hengst et al. (2010)	Card placement sequence (CPS) is interpreted as a measure of collaborative effort. A CPS is the sequence of identifying the target card and placement number, verifying the information and placing each card on the board. Types of CPS include: • basic exchange: dyads achieve the CPS in less than four turns, without needing to revise or refashion the referencing expression • modified basic exchange: referencing of the card and place number is accomplished without any refashioning but the CPS exceeds four turns • verbal only: those CPSs whereby when direct referencing of the target card and placement did not include non-verbal resources	Interrater reliability: • basic exchange or modified basic exchange: 85% (Hengst, 2003) • verbal-only CPS: 94% (Hengst, 2003) Intra-rater reliability: • basic exchange or modified basic exchange: 97% (Hengst, 2003) • verbal-only CPS: 97% (Hengst, 2003)
Interaction (feedback, co-construction, backchannels)	**Task accuracy** Doedens et al. (2021) Flowers and Peizer (1984)	Doedens et al. (2021): accuracy measured by the placement of items in the listener's room, scored as correct/incorrect based on the item's location and orientation Flowers & Peizer (1984): task accuracy was used as a measure of proficiency. When a participant gave an accurate, complete response to the partner's question, a score of 2 was given. For related or incomplete responses, a score of 1 was given (maximum score: 40)	Not assessed
Interaction (accuracy, feedback, co-construction, backchannels)	**Accuracy of card placement** Gordon and Duff (2016) Gupta et al. (2011, 2012) Devanga et al. (2021) Hengst et al. (2010)	At the end of each trial in the intervention sessions, the moderator checked the matcher's board and counted the number of cards placed correctly (maximum: 12 per trial, 72 per session) (Devanga et al., 2021) Hengst et al. (2010) used a scoring system, with a maximum of 36 points per trial: • 1 point for each card correctly placed on the matcher's board • 1 point for each time the dyad used their agreed-upon target label (ATL) to reference the card • 1 point for the matcher's repetition of the director's label	Research assistants verified the counts from video recordings of the sessions (Devanga et al., 2021) Reliability of accuracy generally not assessed, potentially as participants tend to be highly accurate, reaching ceiling quickly
Interaction (self-rating of communicative ease/effectiveness)	**Perception of communicative ease and effectiveness** Hux et al. (2010)	A debriefing interview with the PWA was conducted to find out his perceptions about the ease, amount, and success of information transfer during experimental sessions. The PWA indicated his responses on a 5-point Likert scale; responses were averaged for each of the three statements across conditions	Not assessed

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (communicative ease/effectiveness)	**Lille Communication Test: promoting aphasia communicative effectiveness (PACE) situation** Rousseaux et al. (2010)	Assessments were scored according to the instructions of the Lille communication test. No detail on how the PACE activity was scored. Unclear if there was any analysis of the target was achieved or not (transactional success)	Not assessed
Interaction (communicative ease/effectiveness)	**Communication interaction** Lyon et al. (1997)	Communication probe scenarios were used to stimulate an interaction between the dyad. The interaction was scored on an 8-point ordinal scale to assess accuracy, completeness, speed of identification and relevance of the interaction. No further information reported	Not assessed
Interaction (turns/moves)	**Miniturns or Turns at Talk score** Carlomagno et al. (2005a, 2005b, 2013)	The number of turn-taking events taken by the dyad to complete the task. The (mini)turn includes single or multiple utterances produced by the PWA, as well as repair initiated by the listener For the Turns at Talk score, the minimum number possible was 25 turns (i.e., no request for repair) while the maximum was 75 turns (i.e., 50 examiner prompts were needed)	Interrater reliability: percentage agreement = 96% (Carlomagno et al., 2005b)
Interaction (communicative ease/effectiveness)	**Communicative effectiveness composite score** Carlomagno et al. (2005a)	Quantified participants' conceptual organization of information. Operationalized as a combination of the Minitum score and the Misunderstandings scores	Not assessed
Interaction (accuracy)	**Errors in referent selection** Linebaugh et al. (1985)	Not defined	Not assessed
Interaction (communicative ease/effectiveness)	**Communicative functions** Wambaugh et al. (1990)	Each utterance or gesture was coded in terms of its communicative function: • Regulation (attention, request, vocative) • Statement (naming, description, information) • Exchange (giving, receiving) • Personal (doing, determination, refusal, protest, covert disagreement) • Conversation (imitation, answer, follow-on, question, request for repetition, request for confirmation) • Miscellaneous (other) Operationalized as the proportionate use of each of the communicative function categories per speaker	Test–retest reliability: assessment was administered a second time (6-week interval) to 25% of control participants and *t* statistics for paired samples calculated for each communicative function. No statistically significant differences in the proportional use of communicative functions between the two samplesClassification of the communicative function of each utterance: • Interrater reliability mean point-to-point agreement = 92% (range = 84–100%) • Intra-rater reliability point-to-point agreement = 95% (range = 90–100%) Transcription reliability in relation to transcription errors that caused utterances to be recoded as a different communicative function: • Interrater reliability: mean percentage of changed communicative functions: 0% • Intra-rater reliability: mean percentage of changed communicative functions: 1%

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (communicative ease/effectiveness)	**Couple Communication Scale (CCS)** Nykänen et al., 2013	The Couple Communications Scale (CCS) measures the amount of information communicated within the dyad. Two versions (A and B) were developed. Tasks include six pictures of objects or actions (maximum score 24 points), two situational pictures (maximum score 32 points), and two verbal assignments (maximum score 44 points)	Internal consistency: • tasks with pictures of individual objects and actions had a reliability (Cronbach's alpha) of 0.44–0.77 • situational pictures had a reliability of 0.30–0.52 • verbal assignments had a reliability of 0.61–0.77
Interaction (accuracy)	**Number and % of main concepts successfully co-constructed by the dyad** Hopper et al. (2002)	Main concepts were defined as the gist of the story that were successfully co-constructed by the dyad (i.e., not just successfully verbalized by one partner). Co-construction was defined as meaning negotiated by the dyad	Interrater reliability: point-to-point agreement on 34% of all data conditions: 100% for dyad Y and 97% for dyad G
Interaction (accuracy)	**Production of irrelevant information** Carlomagno et al. (2005a, 2005b)	Information that is not crucial to distinguishing the target stimuli from distractor stimuli	Not assessed
Interaction (accuracy)	**Production of wrong information** Carlomagno et al. (2005a, 2005b)	Information that leads to the listener selecting an incorrect stimulus. These errors were interpreted as impaired pragmatic/conceptual elaboration of information, thereby violating Grice's (1975) maxims of quality, quantity and relevance	Not assessed
Interaction (communicative ease/effectiveness)	**Measure of Participation in Supported Conversation (MPC)** Kagan et al. (2004) Correll et al. (2010) **Adapted Measure of Participation in Conversation** (Togher et al., 2010) Tu et al. (2011)	Correll et al. (2010): includes *Interaction* and transactional labelled *Transaction* scales. Each scale has a 9-point Likert scale, with five anchor points and half points between them. The scale anchors were adapted for use with conversation partners who are family members Tu et al. (2011): two subscales used to measure the skills of the person with TBI's participation in the content and social components of the conversation	Interrater reliability: 22 of the 24 ratings were point-to-point or within 0.5 point difference on a 9-point scale. All ratings were within 1.0 point difference (Tu et al., 2011)
Interaction (turns/moves)	**Number of conversational turns** Hux et al. (2010)	A turn was defined as continuous speech by one person followed either by the other person talking or a silence of >3 sec	Not assessed
Interaction (turn/moves)	**Total number of moves** Guo and Togher (2008) Togher et al. (1996, 2004) Kilov et al. (2009)	Moves are defined as units of information, which include a verb or process. The number of moves in each transcript was used to measure discourse participation	Interrater reliability: percentage agreement for the division of the transcripts into moves was 94% (Togher et al., 2004) Percentage agreement for the type of move and the division of exchange averaged • Intra-rater reliability: 94% (Guo & Togher, 2008) • Interrater reliability: 89% (Guo & Togher, 2008)

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (turns/moves)	**Dynamic moves/min** Guo and Togher (2008) Togher et al. (1997a)	Defined as the rate of negotiating and tracking of information required to facilitate successful information exchange (e.g., clarifying, checking, providing feedback)	Reliability of dynamic moves/min not directly assessed. Intra-rater reliability of both the type of move and the division of exchange: • average 94% (Guo & Togher, 2008) • range 82–96% for type of move, and 84–95% for division of exchanges (Togher et al., 1997a) Interrater reliability of both the type of move and the division of exchange: • average 89% (Guo & Togher, 2008) • average 86% for the type of move and 95% for the division of exchanges (Togher et al., 1997a)
Interaction (turns/moves)	**Percentage of moves which composed each GSP element** Togher et al. (2004) Kilov et al. (2009) **Mean % of moves which composed each GSP element** Togher et al. (1997b)	Generic Structure Potential (GSP) is an analytic framework that can be applied to various discourse genres. GSP analyses moves with regards to core structural elements to describe and measure discourse participation, i.e., for a service encounter these include Greeting, Address, Service Initiation, Service Request, Service Enquiry, Service Compliance, Closing remarks, Goodbye, Calls for attention and Action, and Unrelated moves. This measure calculated the proportion of each transcript that consisted of specific structural elements	Interrater reliability on 25% of transcripts: percentage agreement of assigning GSP elements to moves: • 95% (Togher et al., 2004) • 77% (Kilov et al., 2009) • 90.4 %, ranging from 80 % to 100% (Togher et al., 1997b) Intra-rater reliability of assigning GSP elements to moves: 92% (Kilov et al., 2009)
Interaction (turns/moves)	**Types of moves (synopic or dynamic) used by each participant** Togher et al. (1996)	Synopic moves defined as those needed for an exchange of information to be completed. These include asking for or providing information or actions. Dynamic moves defined as instances where confirmation, checking or channelling are used to progress or repair an interaction, so that the exchange of information or action can be completed	Not assessed
Interaction (turns/moves)	**Type of exchange lead** Togher et al. (1996)	Exchange lead is defined as the purpose of a move. For example, a K1 lead indicates the interaction begins with giving information, while a K2 lead indicates the interaction begins with asking for information	Not assessed
Interaction (turns/moves)	**Number and % of exchanges initiated by the participant** Togher et al. (1996)	Defined as the extent to which the participant controlled the exchange. This was based on the number of exchanges initiated by the participant	Not assessed
Interaction (turns/moves)	**% K1 moves** Jorgensen and Togher (2009)	K1 indicates the person who holds the information; a K1 move serves to provide information. The measure, % K1 moves, was operationalized by dividing the number of K1 moves produced by a participant by the total number of moves × 100	Interrater reliability of 20% of discourse transcripts from all participants: percentage agreement was 87%
Interaction (turns/moves)	**Challenging moves per minute** Tu et al. (2011)	Challenging moves are a subtype of dynamic moves. Speakers use challenges to query the content or relevance of an utterance or the authority of the other speaker. This measure quantified the rate of challenges per speaker	Interrater reliability for coding moves was not assessed for the problem-solving task

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (turns/moves)	**Interactional turns** Gupta et al. (2011, 2012) Gordon and Duff (2016) Hengst (2003)	Defined as utterances and non-verbal communication produced by one participant before the other participant spoke. Boundaries of interactional turns were identified using characteristics such as syntactic and semantic features, and intonational contours	Point-to-point agreement on 12% of the data: • interrater reliability: 99% (Gupta et al., 2011), 97% (Gupta et al., 2012), 99% (Gordon & Duff, 2016) • intra-rater reliability: 99% (Gupta et al., 2011), 99% (Gupta et al., 2012), 98% (Gordon & Duff, 2016) Interrater reliability on 4% of the data for number of turns: 99.6% (Hengst, 2003) Intra-rater reliability on 4% of the data for number of turns: 97.1% (Hengst, 2003)
Cognitive–linguistic skill: functional (story grammar)	**Narrative sequence** Carragher et al. (2015)	Defined as the order in which participants delivered the target narrative, guided by control participants' narratives	Not assessed
Interaction (turns/moves)	**Total number of exchanges** Guo and Togher (2008)	Interactions were divided into exchanges guided by intonational cues. An exchange consists of a series of moves, which are units of information	Intra-rater agreement = percentage agreement for the type of move and the division of exchange averaged 94% (Guo & Togher, 2008) Interrater reliability: percentage agreement for the type of move and the division of exchange averaged 89% (Guo & Togher, 2008)
Interaction (turns/moves)	**K1 moves/min** Guo and Togher (2008) Togher et al. (1997a) Tu et al. (2011)	K1 indicates the person who holds the information; a K1 move serves to provide information. K1 moves/minute measures the rate of information provider by a speaker	Reliability of K1 moves/min not directly assessed. Intra-rater agreement of both the type of move and the division of exchange: • average 94% (Guo & Togher, 2008) • range 82–96% for type of move, and 84–95% for division of exchanges (Togher et al., 1997a) Interrater agreement of both the type of move and the division of exchange: • average 89% (Guo & Togher, 2008) • average 86% for the type of move and 95% for the division of exchanges (Togher et al., 1997a) Tu et al. (2011): interrater reliability for coding moves was not assessed for the problem-solving task

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (turns/moves)	**K2 moves/min** Guo and Togher (2008) Togher et al. (1997a) Tu et al. (2011)	K2 indicates the person who is seeking the information. K2 moves are types of synoptic moves. K2 moves/minute measures the rate at which information is requested or received by a speaker	Reliability of K2 moves/min not directly assessed. Intra-rater agreement of both the type of move and the division of exchange: • average 94% (Guo & Togher, 2008) • range 82–96% for type of move, and 84–95% for division of exchanges (Togher et al., 1997a) Interrater agreement of both the type of move and the division of exchange: • average 89% (Guo & Togher, 2008) • average 86% for the type of move and 95% for the division of exchanges (Togher et al., 1997a) Tu et al. (2011): interrater reliability for coding moves was not assessed for the problem-solving task
Interaction (communicative ease/effectiveness)	**Communicative efficiency** Marshall et al. (1997)	Defined as the completeness, clarity and speed with which a participant conveys information. Communicative efficiency was rated by three groups of raters ($n = 24$) on a 10 mm visual analogue scale with two anchors indicating 'no attempt to provide any information' and 'complete information'.	Interrater reliability: visual analogue scale ratings were converted to numerical values (0–100) using a 10 mm ruler. Point-to-point agreement was 99%. Interrater reliability of communicative efficiency: relatively poor. Variability in ratings across messages and participants. Intra-rater reliability: this was operationalized as repeated ratings being within 10. For the three groups of raters, intra-rater reliability was 86%, 84% and 82%
Interaction (communicative ease/effectiveness)	**Communicative burden** Marshall et al. (1997)	Defined as the degree to which the partner is required to infer, question and guess to verify the accuracy of the information. Communicative burden was rated by three groups of raters ($n = 24$) on a 10 mm visual analogue scale with two anchors indicating 'partner assumes no communicative burden and 'partner assumes all communicative burden'.	Interrater reliability: visual analogue scale ratings were converted to numerical values (0–100) using a 10 mm ruler. Point-to-point agreement was 99% Interrater reliability of communicative burden: relatively poor. Variability in ratings across messages and participants Intra-rater reliability: this was operationalized as repeated ratings being within 10. For the three groups of raters, intra-rater reliability was 85%, 82% and 84%

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (communicative ease/effectiveness)	**Listener comfort ratings** Guo and Togher (2008)	Judges (×12) listened to excerpts of participant during the task and answered the question 'How comfortable would you feel interacting with this person?', from 1 (very uncomfortable) to 7 (very comfortable)	Intra-rater reliability: samples were repeated for 30% of speakers ($n = 3$) and the judges' values were compared. The percentage agreement was calculated by counting the number of comparisons in which the assigned scale value did not differ by more than ±1, dividing this number by three for each speaker and multiplying by 100. Mean intra-rater agreement = 81.1% (range = 33.3% to 100%) Interrater reliability: the value assigned by each judge was compared with the value assigned by every other judge; 77.0% of the ratings were in agreement by ±1 (range = 57.3–89.1%). The intraclass correlation: the group reliability coefficient was 0.96 (range = 0.91–0.99); the individual rater reliability coefficient was 0.66 (range = 0.45–0.87)
Interaction (communicative ease/effectiveness)	**Verbal social skills** Marsh and Knight (1991)	Raters assessed verbal social skills (7-point Likert type scales based on the BRISS): language, speech delivery, conversation structure, conversation content, personal conversational style and partner-directed behaviour	Generalizability analysis of raters, occasions and raters by occasions. Rater reliability ranged from 0.53 to 0.92 (median = 0.85). Ratings ranges from 0.60 to 0.91 (median = 0.78)
Interaction (communicative ease/effectiveness)	**The Adapted Impression Rating Scales** Bond & Godfrey (1997) Tu et al. (2011)	A 9-point Likert scale (scores ranging from 0 to 4) was used to quantify interactions in terms of how appropriate, effortful, interesting, and rewarding they were, and whether the task was completed	Interrater reliability: 26 of the 28 ratings were point-to-point or within 0.5 point difference on a 9-point scale
Interaction (communicative ease/effectiveness)	**Social validation measure** Hopper et al. (2002)	Video recordings of each dyad's assessment interactions were viewed by 16 blinded raters. Each rater (1) wrote a summary of their understanding of the story, and (2) made a judgement as to whether the conversation occurred before or after treatment. The raters' written transcripts were scored to identify correct and incorrect main concepts, i.e., as a measure of how much the raters understood of the conversations	Interrater reliability of rater summaries: percentage agreement was 0.84 (calculated by agreements divided by agreements plus disagreements)
Interaction (turns/moves)	**Frequency of politeness markers** Togher and Hand (1998)	Politeness markers include finite modal verbs; modal adjuncts; comment adjuncts; yes/no tags; and incongruent realization of the interrogative form. The frequency of politeness markers was operationalized by dividing the total number of politeness markers by the total number of clauses	Interrater reliability: point-to-point agreement on 25% of TBI and control data averaged 97% (range = 91–100%)

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (turns/moves)	**Number speaker-initiations and responses** Hux et al. (2010)	Utterances were classified as an initiation and/or a response. Initiations were subdivided into obliges (utterances that require a response, e.g., a question) and comments (utterances that do not require a response). Speaker responses must occur after obliges, but are optional after comments	Interrater reliability on one third of the data: percentage agreement was calculated for utterance-by-utterance comparison and yielded 80% agreement for identification of initiation and responses
Interaction (accuracy)	**Collaborative confrontation naming (CCN)** Devanga et al. (2021)	The moderator presented probe cards ($n = 12$) to the dyad and asked the participant to name each card (verbally or non-verbally). The partner was instructed to offer help only if the participant asked for help or if they were unable to name the card. There were 15 CCN probes in total, during the baseline ($n = 5$), treatment ($n = 5$), and maintenance phases ($n = 5$)The dyad's referencing expressions were recorded and scored on a 15-point multidimensional rating scale adapted from the Naming subtest of the Porch Index of Communicative Ability (PICA; Porch, 1971). Responses were scored as accurate or inaccurate. Accurate responses were assessed for responsiveness, specificity, promptness, and efficiency; inaccurate responses were assessed for intelligibility and type of errors. Scores on the scale were interpreted as reflecting participants' accuracy, independence and efficiency of naming: • Accuracy, i.e., a correct referencing expression of the target (via verbal or non-verbal communication) • Responsiveness, i.e., responding without needing additional help from self (e.g., circumlocution) or the partner (e.g., cues or repetition) • Specificity, i.e., providing a label-like referencing expression specific to the target and by effectively using oral resources only (e.g., words, sentences, vocalizations) • Promptness, i.e., producing responses without requiring additional time (e.g., pauses in the middle of a sentence or prior to producing the label) • Efficiency, i.e., the production of responses without any distortions	Interrater reliability: point-by-point agreement calculated by dividing the number of score agreements by the total number of score agreements and disagreements. Score agreement was defined by whether two raters obtained the same score (agreement) or a different score (disagreement) on the 15-point scale for each probe card per session. The mean interrater reliability ($n = 4$ participants) was 93.75%
Interaction (repair)	**Refashioning (follow-up analysis)** Gupta et al. (2011)	Refashioning is needed if the listener rejects the speaker's description; the director/speaker must subsequently refashion the description, i.e., expand, repair, abandon or replace it with a new description. Operationalized by counting expansions and repairs of the initial description and a director's attempt at a new description	Not assessed
Interaction (repair)	**Self-initiated repair** Doedens et al. (2021)	Occasions when the participant attempted to repair or modify their output. Three types of repair were coded: revised repair (speaker repeats the main clause with modifications), addition repair (speaker expands upon the main clause), and word finding repair (speaker experiences word-finding difficulties). The success of the repair was not coded	Interrater reliability: moderate intraclass correlation coefficient (ICC = 0.74, CI = 0.51–0.87, $p < 0.001$

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (repair)	**Clarification requests or other-initiated repair** Doedens et al. (2021)	Occasions when a speaker communicates that they have not understood what the partner said. Five types of repair were coded: request for elaboration or clarification; statement of not understanding; partial or complete repetition; insertion; indirect request for clarification	Not assessed
Interaction (repair)	**Number of communication breakdowns** Linebaugh et al. (1985)	Communication breakdowns or repair not defined	Interrater reliability: mean point-by-point agreement for number of breakdowns and repairs: 97.8% (range = 86.7–100% for individual dyads)
Interaction (repair)	**% breakdowns which were repaired** Linebaugh et al. (1985)		Not directly assessed
Interaction (repair)	**Number of contingent query-revision sequences per breakdown** Linebaugh et al. (1985)	Contingent queries include nonspecific request for repetition; specific request for affirmation; specific request for specification; and potential request for additional information Revisions include affirmations; non-affirmations; complete or	Interrater reliability: mean point-by-point agreement for categorizing contingent queries and revisions: 93.4% (range = 89.2–100% for individual dyads)
Interaction (repair)	**Frequency distributions of contingent queries and revisions** Linebaugh et al. (1985)	partial repetition; addition or deletion of information; revisions (syntactic, phonetic, inappropriate, inaccurate or ambiguous); recapitulation; off-query response; no response)	Not assessed
Interaction (repair)	**Query-revision distributions and communicative success** Linebaugh et al. (1985)		Not assessed
Interaction (verbal play)	**Number and types of verbal play episodes** Gupta et al. (2012)	Verbal play was defined as telling humorous stories or jokes, making puns, teasing, self-deprecating humour, use of playful voices, song-like intonations, and use of sound effects and gestures. Verbal play may occur as an episode (i.e., single or multiple utterances sharing a common theme) or within a series across trials or sessions. Each verbal play episode was characterized in terms of its communicative resources (verbal, prosodic, or gestural), functions (narrative, teasing, referencing, or other), and interactional forms (simple or extended)	Point-to-point agreement on 12% of the data: • interrater reliability: 88% • intra-rater reliability: 88%

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (accuracy)	**Number of main ideas in the partners' retelling** Ramsberger and Rende (2002)	Main ideas were defined as those corresponding with the research team's retelling of the target stories	Test–retest reliability: four participants with aphasia completed the I Love Lucy retells at two intervals 6 months apart. The total number of main ideas from each administration was used to calculate a coefficient of stability. A moderately high (but not significant) correlation was seen between the test–retest performance of four participants ($rs = 0.80$; $ps > 0.01$)Interrater reliability point-to-point agreement: • measure of main ideas in conversation: 88% • comprehensiveness ratings of conversations: 71% • measure of main ideas in the picture description tasks: 86% (ADP) and 100% (WAB) • comprehensiveness ratings of the picture descriptions: 79% (ADP) and 85% (WAB) • Consistency of scores over alternate forms of a test: • the coefficients of equivalence ($rs = 0.79$; $ps < 0.01$ for main-idea analysis; $rs = 0.77$, $ps < 0.01$ for comprehensiveness ratings) suggest that the parallel methods of measuring transactional success in conversations were moderately high • for the two picture description tasks, the coefficients of equivalence for main-idea analysis ($rs = 0.87$; $ps < 0.01$) and comprehensiveness ratings ($rs = 0.66$; $ps < 0.01$) are significant
Interaction (accuracy)	**Number of salient ideas communicated by the partner** Carragher et al. (2015)	Salient ideas were defined as content words that at least 50% of control participants produced in their retelling of the assessment stimuli. The threshold of 50% was used to distil the target story into the essential components or ideas. Transcripts of the participant and partner were analysed to identify instances when the partner produced a word that corresponded to the core salient idea	Not assessed
Interaction (accuracy)	**Accuracy of the partner's understanding** Yorkston et al. (1980)	Answers written by the partner were scored on a 0–4 scale for relevancy and accuracy. The accuracy score was the mean accuracy rating across 15 exchanges	Interrater reliability: percentage agreement was > 95%

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (accuracy)	**Communicative efficiency** Feyereisen et al. (1988)	The partner's response was coded as either (1) message immediately understood or (2) understood after an error or following request for information. For each communicative modality, efficiency of communication was calculated by dividing the number of messages immediately understood by the total number of attempted messages in that modality	Not assessed
Interaction (accuracy)	**Number of misunderstandings** Carlomagno et al. (2005a, 2005b, 2013)	Effectiveness of the referential expression produced by the participant. This is operationalized as the listener's accuracy in understanding the reference. That is, the number of items when the listener incorrectly interpreted or did not understand the reference due to inaccurate or insufficient information	Not assessed
Interaction (communicative ease/effectiveness)	**The amount and type of conversational support provided by the listener** Carlomagno et al. (2013)	The listener's prompts were scored, e.g., encouraging, interpreting, initiating repair, clarifying	Not assessed
Interaction (communicative ease/effectiveness)	**Measure of Skill in Supported Conversation (MSC)** Kagan et al. (2004) Correll et al. (2010) **Adapted Measure of Support in Conversation** (Togher et al., 2010) Tu et al. (2011)	Correll et al. (2010): Includes an Acknowledging Competence (AC) scale that measures interaction and a Revealing Competence (RC) scale that measures transaction. Each scale has a 9-point Likert scale, with five anchor points and half points between them. The scale anchors were adapted for use with conversation partners who are family members Tu et al. (2011): includes two subscales that measure the skills of the partner in acknowledging the competence of the person with TBI and facilitating the exchange thoughts and feelings	Tu et al. (2011): interrater reliability: 22 of the 24 ratings were point-to-point or within 0.5 point difference on a 9-point scale. All ratings were within 1.0 point difference
Interaction (turns/moves)	**Number and % of exchanges initiated by the communication partner** Togher et al. (1996)	Defined as the extent to which the communication partner controlled the exchange. This was based on the number of exchanges initiated by the communication partner	Not assessed
Interaction (turns/moves)	**Form of the message** Flowers and Peizer (1984)	The form of partner's turns were categorised as; yes–no question, wh-question, multiple choice question, indirect question, statement, unfinished/interrupted utterance	Coding for a fraction of data was checked by two trained raters. Percentage agreement was 99%
Interaction (turns/moves)	**Apparent purpose of the message** Flowers & Peizer (1984)	The purpose of the partner's turns were categorised, e.g., obtain new information, request repetition or give feedback	Coding for a fraction of data was checked by two trained raters. Percentage agreement was 94% and 87% for each raters
Interaction (turns/moves)	**Number of interruptions** Simmons-Mackie et al. (2005)	Defined as instances when the partner spoke over the PWA or when the partner did not allow time for the PWA to response to a question	Interrater reliability: agreements averaged 87% across conditions and three behaviours of interest
Interaction (turns/moves)	**Number of convergent questions** Simmons-Mackie et al. (2005)	Defined as questions requiring a one-word response to which the partner already knew the answer	Interrater reliability: agreements averaged 87% across conditions and three behaviours of interest

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (turns/moves)	**Number of negative teaching** Simmons-Mackie et al. (2005)	Defined as instances when the partner corrected a successful communication attempt, e.g., correcting articulation on an intelligible response or instructing the PWA to use spoken language when the message was successfully communicated non-verbally	Interrater reliability: agreements averaged 87% across conditions and three behaviours of interest
Cognitive–linguistic measure (word level)	**Number of paraphasias** Carlomagno et al. (2005a, 2005b)	Includes lexical paraphasia and circumlocutions	Interrater reliability: percentage agreement = 0.78 (Carlomagno et al. 2005b)
Cognitive–linguistic measure (cohesion)	**% complete cohesive ties** Jorgensen and Togher (2009)	Cohesive ties were categorised as complete (the cohesive marker referred to information that was easily found and unambiguous) or incomplete/error (the cohesive marker referred to information that was absent or ambiguous). The measure, % complete cohesive ties, was operationalized by dividing the number of complete cohesive ties by the total number of cohesive ties × 100	Interrater reliability of 20% of discourse transcripts from all participants: percentage agreement was 80%
Cognitive–linguistic measure (productivity)	**Number of words** Carlomagno et al. (2005a, 2005b) Gupta et al. (2011, 2012) Hengst (2003)	Includes fillers, repetition and comments (Carlomagno et al., 2005a, 2005b) Words were not restricted by morphological or syntactic forms and included false starts, fillers, and back-channelling (Gordon & Duff, 2016; Gupta et al., 2011, 2012) Includes neologisms, false starts and placeholders (Hengst, 2003)	Point-to-point agreement on 12% of the data: • interrater reliability: 99% (Gupta et al., 2011), 98% (Gupta et al., 2012), 99% (Gordon & Duff, 2016) • intra-rater reliability: 98% (Gupta et al., 2011), 97% (Gupta et al., 2012), 99% (Gordon & Duff, 2016) reliability assessed on 4% of the data for number of words: • interrater reliability: 99.5% (Hengst, 2003) • intra-rater reliability: 98.5% (Hengst, 2003)
Cognitive–linguistic measure (complexity)	**Conceptual complexity of speaker-initiated turns** Hux et al. (2010)	Utterances were coded for conceptual complexity using a scale with four levels (matching experience, selective analysis of experience, reordering experience and reasoning about experience). An overall conceptual complexity score for each speaker was calculated to give a score ranging from 1.00 to 4.00. This was operationalized by (1) tallying the number of times a speaker produced a comment or oblige corresponding with the conceptual complexity scale: (2) multiplying each tally by 1, 2, 3, or 4 depending on its complexity level; (3) summing the resultant products; and (4) dividing by the total number of utterances. Higher scores indicated conversations in which a speaker produced a large number of conceptually complex utterances	Interrater reliability on one third of the data: percentage agreement was calculated for utterance-by-utterance comparison and yielded: • 94% for classifying comments and obliges • 77% for complexity level assignment • disagreements in complexity differed by more than 1 point on less than 5% of utterances
Interaction (accuracy)	**% essential units of information** Jorgensen and Togher (2009)	Information units were coded as essential (relevant information consistent with major details selected for the task) or non-meaningful (irrelevant, redundant, off-topic or incorrect). The measure was operationalised as the number of essential information units divided by the total number of information units × 100	Interrater reliability on 20% of discourse transcripts from both participant groups: percentage agreement for both story grammar elements and essential units of information was 89%

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Interaction (accuracy)	**Number of salient ideas communicated by the PWAs** Carragher et al. (2015)	Salient ideas were defined as content words that at least 50% of control participants produced in their retelling of the assessment stimuli. The threshold of 50% was used to distil the target story into the essential components or ideas. Transcripts of the participant and partner were analysed to identify instances when the participant verbally or non-verbally produced a main idea that corresponded to the core salient idea	Not assessed
Cognitive–linguistic measure (structural)	**Total no. of C-units** Jorgensen and Togher (2009)	A communication unit (C-unit) is an independent clause plus any associated subordinate clauses. Total number of C-units were calculated across tasks and participant groups	Interrater reliability on 20% of transcripts from both participant groups: percentage agreement for both total number of C-units and words per C-unit was 85%
Cognitive–linguistic measure (structural)	**Words per C-unit** Jorgensen and Togher (2009)	Operationalized as the average length of C-units calculated by dividing the number of words by the number of C-units	Interrater reliability on 20% of transcripts from both participant groups: percentage agreement for both total number of C-units and words per C-unit was 85%
Interaction (accuracy)	**Semantic content of initial descriptions (follow-up analysis)** Gupta et al. (2011)	A director's first attempt to describe the target stimulus was analysed for how they characterised the target, e.g., using biological, nonbiological, geometric characteristics	Not assessed
Interaction (accuracy)	**Content units** Hux et al. (2010)	Defined as a single piece of new information in the conversation. This included any type of information such as labels of objects, names of people or locations, temporal markers, characteristics, and events. Within an utterance, each unique piece of information counted as a separate content unit. For example, 'It was a 1948 Chevy Coupe' = three content units ('1948', 'Chevy', and 'Coupe'). Debriefing interviews with communication partners were transcribed and content units analyses were used to indicate how much information the communication partner had obtained, the accuracy of the information and the modality through which the PWA communicated the information. Content units were coded into four categories: • correct information gained from conversation or visual scene display • information remembered incorrectly by the partner • correct information misinterpreted by the partner • incorrect information stated by the PWA	Interrater reliability on one third of the data: point-by-point agreement of identification and categorization was 84.84%
Interaction (accuracy)	**Crucial information score** Carlomagno et al. (2005a, 2005b, 2013)	Measure of lexical encoding of information (lexical deficit). Operationalized as the number of target words or plausible synonyms (i.e., crucial information) produced by the participant to identify target stimuli	Interrater reliability: percentage agreement = 0.96 Carlomagno et al. (2005b)

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Cognitive–linguistic measure (story grammar)	**% story grammar elements** Jorgensen and Togher (2009)	Three story grammar elements were coded: initiating event, action, and direct consequence. The measure was operationalized as the number of story grammar elements present divided by the number of expected elements ×100	Interrater reliability on 20% of discourse transcripts from both participant groups: percentage agreement for both % story grammar elements and % essential units of information was 89%
Cognitive–linguistic measure (memory)	**Retention of card labels after the trial** Gordon and Duff (2016)	30 min after the last trial, TBI participants were asked to identify cards used in the study from foil cards	Not assessed
Multimodal (partner measure)	**Partner monitoring through eye gaze** Gupta et al. (2011)	Speakers monitor each other's faces to gauge whether they are being understood. Thus, partner monitoring through eye gaze was interpreted as a sign of collaboration and development of common ground. The measure was operationalized by quantifying the timing and duration of the participant looking at the partner's face across three non-consecutive trials	Analysis was completed for the PWA and three control participantsDuration of gaze: reliability ratings were checked for approx. 12% of data: • Interrater reliability: 98% • Intra-rater reliability: 97%
Multimodal expression	**Channel use** Feyereisen et al. (1988)	Measured as the number of picture stimuli described by gestural, oral, or gestural + oral means. Gestures included illustrative, accompaniment or deictic movements and were only recorded when it led to selection of the correct picture. Oral messages were defined as any verbal attempt to communicate about the stimuli (including paraphasias and neologisms). Participants' responses to yes/no questions were ignored	Not assessed
Multimodal expression	**Manner of delivery** Flowers and Peizer (1984)	How participants and their partners delivered a message, i.e., via verbal, gesture, written/graphic or non-verbal means	Coding for a fraction of data was checked by two trained raters. Percentage agreement was 100%
Multimodal expression	**Number of judges out of 10 who identified the target referent from participants' gestures** Feyereisen et al. (1988)	The judges watched video clips of participants communicating a target via gesture, and selected the picture which corresponded to the gesture from a selection. A score per participant and per condition (apraxia assessment versus referential communication task) was calculated by dividing the total number of identifications by the maximum possible number in the condition	Not assessed
Multimodal expression (gesture)	**Number of (informative) gestures** Carlomagno et al. (2005a, 2005b, 2013) Hengst (2003)	Hand/arm gestures produced by participants and judged as contributing to correct identification of the referent (Carlomagno et al., 2005a, 2005b, 2013) Gestures included postures, movements and sound effects were counted only if they acted as a word substitute, called attention to the ongoing activity, described the target card or place and/or were taken up and responded to by the partner (Hengst, 2003)	Interrater reliability for the number of informative gestures: • 78% (Carlomagno et al., 2005b) • 97.9% (Hengst, 2003) Interrater reliability for the number of gestures agreed by two raters divided by the total number of gestures identified: percentage agreement = 92.0–94.7%(Carlomagno et al., 2013)Intra-rater reliability: percentage agreement = 77.4% (Hengst, 2003)
Multimodal expression (gesture)	**Classification of gestures** Carlomagno et al. (2013)	Gestures were classified as iconic, deictic, meta-linguistic, emblem or indefinite	Interrater reliability: percentage agreement = 84.6% (repetition) to 94.8% (emblems)

(Continues)

TABLE B1 (Continued)

Domain	Name	Operational definition	Reliability
Multimodal expression (gesture)	**Number of coverbal gestures** Carlomagno et al. (2013)	Defined as gestures produced alongside speech. Operationalized as the number of coverbal gestures, type (iconic, repetition, emblem, deictic, batonic, indefinite), distribution of iconic gestures, recognizable semantic content, and occurrence of gestures with empty or parapahasic speech	Not assessed
Multimodal expression (including prosody)	**Type of voice** Hengst et al. (2008)	Dimensions of dialogic voice: • social voices • representing others' words/acts • personalized voice	Not assessed
Interaction (time)	**Trial time** Doedens et al. (2021) **Task duration** Flowers and Peizer (1984) Yorkston et al. (1980) **Total time of the interaction** Feyereisen et al. (1988) **Time to completion** Gordon and Duff (2016) Gupta et al. (2011, 2012) **Total time** Guo and Togher (2008) **Duration of each transcript** Togher et al. (2004)	Doedens et al. (2021): timing started when participants began to communicate on a trial and ended when one of the participants pressed the button to indicate they had completed the trial Flowers & Peizer (1984): timing began after the partner asked the initial question, and ended when the partner reported they knew the target. Each picture had a 3-min limit Yorkston et al. (1980): timing started at the end of the partner's first question and stopped when the partner started to write down their answer. Duration was calculated as the mean across 15 exchanges Feyereisen et al. (1988): duration of the interaction was timed for four stimuli Gordon and Duff (2016), Gupta et al. (2011), Gupta et al. (2012): time to complete each trial (s)	Not assessed