ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

Curvilinear Relationship Between Confrontation Naming and Monologic Narrative Discourse in People With Aphasia: A Partial Replication With Expanded Analysis of Discourse

Christie Carroll-Duhigg,[a] Sarah Grace Dalton,[b] Yu-Yu Hsiao,[c] Honey I. Hubbard,[a] and Jessica D. Richardson[a]

[a] Department of Speech and Hearing Sciences, The University of New Mexico, Albuquerque [b] Department of Communication Sciences and Special Education, University of Georgia, Athens [c] Department of Educational Psychology, The University of New Mexico, Albuquerque

ARTICLE INFO

Article History:
Received September 15, 2025
Revision received February 10, 2026
Accepted July 27, 2026

Editor-in-Chief: Catherine A. Off
Editor: Suma R. Devanga

https://doi.org/10.1044/2026_AJSLP-25-00419

ABSTRACT

Purpose: The purpose of this study was to examine the relationship between latent naming and narrative discourse ability in people with aphasia, extending prior research with a revised discourse measurement model focused on narrative production.
Method: Data from 223 participants from AphasiaBank were analyzed (final n = 220). Latent naming ability was indicated by the Boston Naming Test–Short Form, Verb Naming Test, and Western Aphasia Battery–Revised Naming Subtest (WABN). Latent narrative discourse ability, in the context of AphasiaBank's Cinderella story-telling task, was indicated by main concept analysis (MCA) subscores Accurate and Complete (AC), and Attempts (ATT), Sequencing (SEQ), Story Grammar Episodic Complexity (SGEC), and Core Lexicon (CoreLex). Confirmatory factor analysis (CFA) and structural equation modeling were used to evaluate the narrative discourse measurement model and its association with latent naming ability, respectively. A scatter plot with locally estimated scatterplot smoothing revealed a curvilinear relationship between latent constructs, showing a weaker relationship at lower ranges of naming ability, and a disproportionately stronger relationship at higher levels. Linear (a priori) and quadratic (post hoc) paths were tested using robust maximum likelihood estimation to account for nonnormal distribution of AC and WABN.
Results: Narrative discourse CFA demonstrated a good fit (e.g., comparative fit index [CFI] = .998, root-mean-square error of approximation [RMSEA] = .031, standardized root-mean-square residual [SRMR] = .004), with high loadings across indicators. The linear model also fit the data well (e.g., CFI = .990, RMSEA = .054, SRMR = .024), showing a strong positive association (β = .66; R^2 = .44). However, the curvilinear model improved fit (Δ Akaike Information Criterion = −12) and explained more variance (R^2 = .61).
Conclusions: Differences in elicited narrative discourse performance appear to be minimal among those with lower naming ability. However, larger differences in narrative discourse for those with higher level naming abilities suggest that even mild naming impairment may correspond with noticeably impaired narrative discourse.
Supplemental Material: https://doi.org/10.23641/asha.34113318

Correspondence to Christie Carroll-Duhigg: ccduhigg@unm.edu.
Publisher Note: This article is part of the Special Issue: Select Papers from the 54th Clinical Aphasiology Conference. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

People with aphasia experience difficulties in performing everyday communicative tasks (e.g., ordering food at a restaurant), socially interacting with others, maintaining relationships, and fulfilling roles and responsibilities (Baker et al., 2020; Hilari et al., 2012; Worrall et al., 2011). In line with these reported challenges, there is broad agreement

 Copyright © 2026 American Speech-Language-Hearing Association

in the literature that functional communication outcomes should be prioritized over isolated impairment-focused outcomes (e.g., confrontation naming) (Armstrong & Hersh, 2024; Brady et al., 2014, 2016; Kagan et al., 2008; Worrall et al., 2011). This emphasis is consistent with the expressed preferences of people with aphasia, who have repeatedly called for therapeutic outcomes that directly relate to real-life functioning (Hilari et al., 2012; Wallace et al., 2017; Worrall et al., 2011).

Following the World Health Organization's (2002) International Classification of Functioning, Disability, and Health model, a wide range of measures have been developed to capture different aspects of functional communication, such as communicative capacity (i.e., the ability to communicate effectively in every day contexts), communicative performance in everyday tasks (i.e., how communicative tasks are actually performed in real life), participation in meaningful life activities, and personal and environmental factors that may impede or facilitate engagement (e.g., communication confidence, depressive symptoms, care-partner attitudes, social support, clinical practice trends) (World Health Organization, 2002). Collectively, these measures reflect the complexity and multifaceted nature of functional communication.

Encouragingly, in a recent survey of practicing clinicians, approximately 69% of clinicians reported assessing functional communication in > 80% of patients, although mostly through informal assessment of connected speech (Tierney-Hendricks et al., 2022). However, standardized naming and impairment-based measures such as the Boston Naming Test (BNT; Lansing et al., 1999; Mack et al., 1992) and Western Aphasia Battery–Revised (WAB-R; Kertesz, 2006) continue to be mainstays in clinical care (RELEASE Collaboration, 2020; Tierney-Hendricks et al., 2022). In a review of studies and data sets (dating from 1973 to 2017), spoken language impairment was identified as the most common intervention target, with 73% of spoken language interventions focusing specifically on naming (RELEASE Collaboration, 2020). Similarly, confrontation naming and overall language impairment continue to serve as the primary outcome in research investigations (Dignam et al., 2024; Godecke et al., 2021; Jacobs et al., 2021, 2023; Kristinsson et al., 2023; Low et al., 2024; Simic et al., 2022), including those evaluating discourse-level interventions (Dipper, Marshall, Boyle, Botting, et al., 2021). The ubiquitous use of isolated naming outcomes warrants further inquiry into their relationship with overall communicative functioning.

Relationship Between Naming Ability and Discourse

Despite the widespread use of confrontation naming outcomes in both research and clinical settings, evidence for robust generalization to connected speech remains limited. Prior research examining the relationship between single-word retrieval and discourse production suggests that confrontation naming may be associated with word-level outcomes in discourse (e.g., number of words, substantive words, informative words) but with some variability across discourse contexts. For example, the literature shows that confrontation naming may be associated with retrieval of specific word classes (e.g., number of nouns, content words) in conversation (Herbert et al., 2008); verb–noun differences during monologic discourse tasks (R. Li & Kiran, 2024); and presence of informative words in picture description (Boucher et al., 2022), free speech, and narrative storytelling (Fergadiotis et al., 2019; Fergadiotis & Wright, 2016). Confrontation naming has also been found to be significantly associated with proposition-level outcomes (i.e., number of accurate and complete propositions, per main concept analysis [MCA], during elicited monologic tasks) but only for those with Broca's and Wernicke's aphasia (Richardson et al., 2018). However, studies have also reported inconsistencies between naming and discourse performance, such as greater difficulty retrieving words during confrontation naming compared to video narration (Pashek & Tompkins, 2002), picture description, and conversation (Mayer & Murray, 2003); more self-corrections during connected speech than during confrontation naming (Mayer & Murray, 2003); and a weak correlation between naming and substantive conversational contribution (i.e., likelihood of producing substantive turns in conversation; Herbert et al., 2008).

Across much of aphasia literature, language processing is described in terms of distinct levels of analyses, including phonological (sound level), morphological, lexical (word level), semantic (meaning level), syntactic (sentence level), and discourse (i.e., connected speech, consisting of more than one sentence, including both micro- and macrolinguistic features). Within language processing, word retrieval is often modeled as a multistage process involving lexical–semantic access to a concept and its associated lemma (i.e., an abstract representation of the concept), followed by retrieval of the word's phonological representation (e.g., sequence of sounds, syllable structure, and stress patterns; Maher & Raymer, 2004). These distinct units of analysis (e.g., phonological, lexical, semantic, syntactic) are typically used to characterize language impairment profiles and guide impairment-focused aphasia assessment and intervention (American Speech-Language-Hearing Association, n.d.). However, a recent review of neuroimaging studies reported evidence of spatial overlap and functional interdependence between different levels of language processing (e.g., between phonological and syntactic processing; Fedorenko et al., 2024). This may help to explain some of the associations reported between naming and connected speech in previous studies. However, the review highlighted a distinct feature of discourse-level processing, in that it also recruited

cognitive systems involved in attention, memory integration, and situation modeling (i.e., maintaining a coherent mental representation of discourse), which are believed to support macrostructural aspects of discourse, such as global coherence and cohesion (Fedorenko et al., 2024). These findings suggest that, while word-level and discourse-level abilities may be functionally integrated to some degree, discourse production is uniquely complex and reflects more than the accumulation of lower level linguistic skills.

Monologic Discourse as a Functional Communication Outcome

Discourse is a broad, complex construct that includes both conversational dialogue and monologue (e.g., describing a picture, giving instructions, narrative storytelling, personal narratives; Armstrong et al., 2011). Dominant theoretical frameworks of functional communication (Clark, 2012; Doedens & Meteyard, 2022) emphasize "interactiveness" as a key feature. As such, conversational discourse is typically considered to be more representative of functional communication than monologic discourse (Armstrong & Ferguson, 2010; Armstrong et al., 2011; Armstrong & Hersh, 2024; Leaman & Edmonds, 2021). However, it is arguable that noninteractive communicative skills, such as describing a scene to someone over the phone, providing stepwise instructions, and telling stories, are also important aspects of everyday communication (Dipper, Marshall, Boyle, Hersh, et al., 2021; Pritchard et al., 2018). Additionally, a growing body of work has demonstrated differences in linguistic and informational output between monologic discourse and conversation (Armstrong et al., 2011; Beeke et al., 2008; Leaman & Edmonds, 2021; Pashek & Tompkins, 2002), suggesting that monologic assessments may provide information about communicative functioning that is distinct from dialogue assessment. For example, monologic discourse has been associated with greater syntactic productivity and increased errors (Armstrong et al., 2011), as well as greater informativeness and efficiency (Boyle et al., 2023), in comparison to conversational discourse. Differences in word class production have also been observed, with studies showing context-dependent variability in noun and verb production between monologue and dialogue tasks (Armstrong et al., 2011; Mayer & Murray, 2003). While such output differences have been interpreted as evidence of a lack of generalizability of monologic discourse performance to conversation (Brady et al., 2014; Leaman & Edmonds, 2021), these findings also highlight the potential value of monologues as a functionally relevant communicative context that places unique demands on the speaker, which may not necessarily be reflected in conversational dialogue. These differences emphasize the value of evaluating connected speech across different contexts (Armstrong et al., 2011).

Monologic discourse skills can be examined in naturalistic contexts and/or through structured or semistructured elicited tasks. Although naturalistic contexts better reflect real-world functional monologic skills, elicited monologue tasks, such as those embedded in standardized language assessments (e.g., "Cookie Theft" picture description in the Boston Diagnostic Aphasia Examination; Goodglass & Kaplan, 1983), continue to be widely used in research and clinical settings (Bryant et al., 2016; Tierney-Hendricks et al., 2022). Elicited monologic discourse tasks also facilitate the use of standardized, normed checklists and scoring procedures (e.g., MCA; Nicholas & Brookshire, 1995) that support reliability, comparison across participants, and longitudinal measurement (Dalton & Richardson, 2019; Richardson & Dalton, 2016) and offer insights into multiple levels of processing (i.e., word-level, proposition-level, and macrolinguistics outcomes; Greenslade et al., 2020; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016; Richardson et al., 2021) across various monologic contexts (MacWhinney et al., 2011). While elicited monologic discourse tasks may not be a direct proxy for naturalistic monologues, successful performance on tasks such as the Cinderella storytelling task (see AphasiaBank; MacWhinney et al., 2011) engages core narrative and language processes that are also relevant to real-world monologic production, such as lexical selection, organization of causally linked event chains, maintenance of global coherence, inclusion of essential story elements, and continued monitoring of language output and listener needs (Dutta et al., 2025; Richardson & Dalton, 2016).

Furthermore, several advancements now make it possible to investigate monologic discourse skills in large samples of people with aphasia. The first major advancement is the development and ongoing expansion of AphasiaBank (http://talkbank.org/AphasiaBank), an online repository of standardized test results and transcribed discourse samples collected from people with aphasia and control participants across a range of standardized discourse tasks (Forbes et al., 2012; MacWhinney et al., 2011). AphasiaBank has facilitated two additional advancements—the expansion of normed MCA checklists (Dalton et al., 2020; Richardson & Dalton, 2016) into a more comprehensive, multilevel framework (i.e., main concept [MC], sequencing, and story grammar [MSSG]; Greenslade et al., 2020; Richardson et al., 2021) for evaluating discourse informativeness and macrostructure, as well as the continued development of Core Lexicon (CoreLex) checklists (Dalton et al., 2020, 2022; Dalton & Richardson, 2015), which offer an efficient, quantifiable way to assess the typicality of lexical selection within discourse. Lastly, the development of a unified theoretical framework, the Linguistic Underpinnings of Narrative in Aphasia (LUNA) treatment research, proposed by Dipper, Marshall, Boyle, Hersh, et al. (2021), provides a broad theoretical basis for evaluating narrative discourse across studies.

By leveraging these recent advancements, the present study examined the relationship between latent confrontation naming and latent narrative discourse production, as indicated by a theory-driven measurement model informed by the LUNA framework. Earlier work by Fergadiotis et al. (2019) tested a similar path model examining the relationship between confrontation naming and monologic discourse informativeness, as indicated by correct information units (CIUs; Nicholas & Brookshire, 1993) across three different monologic discourse contexts (free speech, picture description, and narrative storytelling), in which a relatively strong predictive relationship between confrontation naming and monologic discourse informativeness was reported.

Modeling Narrative Discourse Using the LUNA Framework

A comprehensive, underlying "theory" of overall discourse has been elusive, as much of the discourse literature has focused on evaluating discrete aspects of discourse analysis (Armstrong, 2000; Linnik et al., 2016; Wright, 2011) such as informativeness (Doyle et al., 1995; Fergadiotis et al., 2019; Leaman & Edmonds, 2019; Nicholas & Brookshire, 1993), cohesion and coherence (Alyahya et al., 2022; Kong et al., 2018; Linnik et al., 2022; Zhang et al., 2021), lexical retrieval/linguistic analyses (Boyle, 2011; Bryant et al., 2016; Dalton & Richardson, 2015; Kim et al., 2019), and multimodal/nonverbal features of discourse (Caute et al., 2013; Dutta & Mohapatra, 2024; Pritchard et al., 2015; Stark & Cofoid, 2022). While the LUNA framework does not provide a model for evaluating discourse production as a whole, Dipper, Marshall, Boyle, Hersh, et al. (2021) propose a comprehensive model for evaluating spoken narratives. The LUNA model consists of four core categories: (a) pragmatics, (b) macrostructure planning, (c) propositional planning and selection, and (d) linguistic features. To facilitate systematic evaluation of the theory-driven measurement model, a structured narrative discourse task (Cinderella storytelling) was selected to provide a controlled and replicable assessment context supported by prespecified checklists and standardized scoring procedures (e.g., MCA, CoreLex) reflecting each of the four LUNA categories (see Supplemental Material S1 for conceptual mapping of MCA-derived indicators of narrative discourse onto LUNA categories).

Within LUNA, pragmatics exerts an overarching influence on narrative discourse production, shaping decisions based on various contextual, interpersonal, and interactional demands. Pragmatic processing involves consideration of communicative purpose (e.g., to inform, persuade), the informational needs of the listener, appropriate tone based on setting, listener familiarity, and content being shared, as well as overall social awareness. Although pragmatic skills were not formally assessed in the current study, pragmatic demands were likely reduced by the use of a structured narrative task (i.e., "Cinderella storytelling"), including standardized stimuli and elicitation instructions, administered in a controlled research environment. Additionally, pragmatic skills such as attending to the informational needs of the listener may be indirectly reflected in MCA—Accurate and Complete (AC), and Attempts (ATTs), which captures the speaker's ability to relay "essential" (i.e., informative and relevant) concepts of the story.

The macrostructure planning component of LUNA reflects the global organization of narrative discourse. This includes constructing an overall narrative framework, selecting and arranging major structural elements (e.g., setting, initiating events, attempts, consequences, and conclusion), conveying them in a logical sequence, and keeping track of structural elements as the story is being told. In the present study, macrostructure planning was primarily operationalized using Story Grammar Episodic Complexity (SGEC) and Sequencing (SEQ; see Greenslade et al., 2020, and Richardson et al., 2021, for scoring details). SGEC evaluates the presence and configuration of core story grammar components across episodes, thereby measuring the speaker's ability to organize narrative discourse at the level of macrostructure rather than local linguistic form. SEQ evaluates how propositions are ordered to meet macrostructural requirements of the story (e.g., propositions describing an initiating event precede propositions conveying attempts).

The propositional component of LUNA refers to prelinguistic planning. At this level, speakers decide which details are necessary to fulfill macrostructural plans, as well as the order in which they should be shared to convey informational dependency (e.g., passage of time, cause and effect). Propositional processing does not require linguistic precision but rather reflects decisions about meaning and content. From this perspective, MCA AC, ATTs, and SEQ all reflect aspects of propositional processing to varying degrees. As MCA centers on prespecified propositions, propositional planning is best captured by SEQ, which is the logical ordering of propositions, and ATTs, which accounts for all attempted propositions, regardless of linguistic errors. AC also reflects production of essential elements of the story but excludes propositions with errors.

Finally, the linguistic component of LUNA represents the microlinguistic realization of propositions. This includes lexical selection and phonological, grammatical, and syntactic structuring. Linguistic processing determines how precisely and effectively selected propositions are expressed. In the present study, linguistic aspects of narrative discourse were captured primarily by CoreLex and AC scores. CoreLex provides information about lexical specificity and typicality, capturing the extent to which

speakers access and produce task-relevant vocabulary rather than relying on vague or underspecified terms. AC reflects the successful syntactic and lexical realization of predefined propositions.

Taken together, this mapping situates MCA-derived measures within a theoretically grounded model of monologic narrative discourse (see Supplemental Material S1 for conceptual mapping). The LUNA framework integrates distinct aspects of narrative discourse analysis (e.g., pragmatics, structural organization, content selection, and linguistic realization) to examine monologic narrative discourse as a single, complex construct.

Study Aims

Given the inherent complexity of narrative discourse, establishing a valid, theory-driven measurement model is an essential prerequisite for examining the relationship between confrontation naming and narrative discourse production. The present study has two primary aims.

- Aim 1: To evaluate the adequacy of a theory-driven narrative discourse measurement model using confirmatory factor analysis (CFA).

  Hypothesis 1: Observed variables (AC, ATTs, SGEC, SEQ, and CoreLex) are indicators of latent narrative discourse ability.

- Aim 2: To examine the relationship between latent confrontation naming ability and latent narrative discourse performance.

  Hypothesis 2: Latent confrontation naming ability demonstrates a significant positive association with latent narrative discourse ability.

Method

Participants

Data from a total of 370 participants, across 23 sites in the United States, were initially retrieved from AphasiaBank (MacWhinney et al., 2011). The initial AphasiaBank sample included individuals with aphasia resulting from stroke and other brain injuries (e.g., tumor resection, infection, traumatic brain injury). Individuals with aphasia due to neurodegenerative etiology were excluded. Because contributing sites did not consistently utilize the same measures, sites that were missing more than one third of data for any of the key variables of the current study were excluded from the final data set. This resulted in a total of 223 participants, all of whom had stroke-induced aphasia. See Table 1 for participant details.

Table 1. Participants' description (n = 220).

Characteristic	Total
Gender	127 M; 96 F
Age	M = 61.36 (12.90)
	Range: 25.6–90.7
Years of education	M = 15.47 (2.82)
	Range: 11–25
Some high school (< 13 years)	n = 52 (24%)
High school (13 years)	n = 8 (3.6%)
Some college (14–16 years)	n = 94 (43%)
Bachelor's or higher (> 17)	n = 58 (26%)
Unknown (missing data)	n = 8 (3.6%)
Aphasia severity (per WAB-R Aphasia Quotient)	M = 70.80 (19.44)
	Range: 16.1–99.6
Aphasia type	
Not aphasia by WAB-R (> 93.7)	n = 20 (9%)
Anomic	n = 71 (32%)
Conduction	n = 42 (19%)
Wernicke	n = 15 (7%)
Broca	n = 57 (26%)
Global	n = 4 (2%)
Transcortical motor	n = 9 (4%)
Transcortical sensory	n = 2 (< 1%)

Note. Sites: Aphasia Center of West Texas (n = 10), Adler Aphasia Center (n = 25), Boston University (n = 12), Aphasia Center of California (n = 14), University of South Carolina (n = 10), University of Kansas (n = 21), Emerson College (n = 3), Montclair State University (n = 7), Snyder Center for Aphasia Life Enhancement (n = 27), Stroke Aphasia Recovery Program (n = 1), Triangle Aphasia Project (n = 18), Texas Christian University (n = 6), Northwestern University (n = 13), Aphasia House (n = 20), Stroke Comeback Center (n = 22), Intensive Residential Aphasia Community Therapy (n = 7), and East Carolina University (n = 7). M = male; F = female; WAB-R = Western Aphasia Battery–Revised.

Data Collection

This study used de-identified data from AphasiaBank (TalkBank), including standardized naming and aphasia severity scores, as well as Codes for Human Analysis of Transcripts (CHAT) of the Cinderella narrative task. As investigators did not have access to participant identifying information, the study did not meet the criteria for human subjects research under U.S. federal regulations. The University of New Mexico Institutional Review Board (IRB) determined that use of this de-identified data was exempt from additional review. All contributing studies to AphasiaBank obtained local IRB approval and used informed consent procedures that included additional permissions for data sharing.

Measurement

Confrontation Naming Measures

Replicating Fergadiotis et al. (2019), naming ability was indicated by the BNT–Short Form (BNT-SF; Lansing

et al., 1999; Mack et al., 1992), Verb Naming Test (VNT) of the Northwestern Assessment of Verbs and Sentences (Cho-Reyes & Thompson, 2012), and WAB-R (Kertesz, 2006) Naming Subtest (WABN). For WABN, participants are shown 20 real tangible objects and asked to name them; responses are scored from 0 to 3, depending on accuracy and whether cues are needed (i.e., 3 points if named correctly with no cues; 2 points if response is recognizable but includes a phonemic paraphasia, with no cues; 1 point if a cue is needed; 0 points if incorrect or no response), resulting in a possible score range of 0–60. Similar to the WABN, the 15-item BNT-SF evaluates confrontation naming of nouns but with line drawings rather than real objects. It is derived from the longer 60-item BNT (Kaplan et al., 2001). Correct responses receive 1 point each, for a total of 15 maximum points. The VNT is administered using 22 line drawings representing action words, with 1 point given for each correct response, for a total 22 possible points.

Monologic Narrative Discourse Measures

Indicators of latent narrative discourse ability, operationalized within the context of AphasiaBank's Cinderella storytelling task, included MCA-derived (Nicholas & Brookshire, 1995; Richardson & Dalton, 2016) scores for the total number of ACs and the total number of ATTs, as well as SEQ, SGEC (Greenslade et al., 2020; Richardson et al., 2021), and CoreLex (Dalton et al., 2020). All narrative discourse samples were elicited using the AphasiaBank protocol and procedures, in which participants were asked to review an illustrated (i.e., wordless) copy of Cinderella (provided by AphasiaBank) and then asked to tell as much of the story as possible in their own words. The AphasiaBank protocol provides detailed instructions for administering discourse tasks, including scripts for examiners to follow. The Cinderella story portion of the AphasiaBank CHAT was extracted to be coded and scored for the variables in this study.

MCA AC and ATTs. MCA uses a normed checklist and standardized scoring procedures to evaluate the accuracy and completeness of a narrative (Kong, 2009; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016). Unlike CIUs, which provide credit for numerous alternative productions and circumlocution attempts (as long as the words are accurate, not repeated, meaningful, and relevant), MCs consist of prespecified essential elements, which typically include syntactic and lexical criteria, such as a main verb, related nouns, and prepositional phrases and/or subordinate clauses (e.g., "The prince[1] falls in love[2] with Cinderella[3]" has three essential elements). Acceptable alternative productions are provided in the MCA checklist (Nicholas & Brookshire, 1995; Richardson & Dalton, 2016).

To streamline coding and scoring calculations, all MCs, including alternative productions, were structured within a Microsoft Excel worksheet. Each MC was scored as either "accurate and complete" (all essential elements are accurate and present), "accurate and incomplete" (AI; content is accurate, but one or two essential elements are missing; "The prince[1] falls in love[2]"), "inaccurate and complete" (IC; some of the content is inaccurate, but all essential elements are present; "The king[1] falls in love[2] with Cinderella[3]"), "inaccurate and incomplete" (II; some of the content is inaccurate, and one of more essential elements are missing; "The king[1] falls in love[2]"), or "absent" (AB; all essential elements are missing). Each predefined MC was evaluated based on the speaker's first identifiable production of that concept. Subsequent self-corrections, repetitions, elaborations, or returns to the same concept were not rescored or considered additional attempts. Pronouns or reduced forms were accepted when referents were identifiable, consistent with MCA scoring conventions and checklists (see Richardson & Dalton, 2016). Accordingly, ATTs reflect the total number of MCs that were coded as AC, AI, IC, or II and excludes concepts coded as AB.

The Cinderella storytelling task has a total of 34 MCs, and thus, a maximum score of 34 ACs and 34 ATTs. Although a weighted composite score (i.e., [no. of AC × 3] + [no. of AI × 2] + [no. of IC × 2] + [no. of II × 1]) may be calculated for MCA, yielding a maximum score of 102, the composite score was not included in this study. This decision was made to minimize the risk of extreme collinearity typically shared between subtest and composite scores (Kline, 2023). Additionally, AC and ATT may be more sensitive to reflecting nuanced differences between speakers than composite score. For example, someone who produces only 30% of a story with high accuracy could potentially have the same composite score as someone who tells 65% of the story but with less accuracy (e.g., says "king" instead of "prince"). Such subtle differences in microlinguistic accuracy and macrolinguistic narrative completeness would be missed by composite score alone. All scorers were trained following a predetermined training protocol, as described in Richardson et al. (2021). Twenty percent of transcripts were rescored by a coauthor, with an interrater reliability of 95.4%.

SGEC. Story grammar reflects the macrostructural conventions used to tell a story, commonly used to orient listeners, evoke emotions, connect disparate elements, and denote meaning and relevance to key characters and events. MSSG scores are based on the presence and configuration of six story grammar "components," including (a) setting (i.e., characters, location, time), (b) initiating events, (c) attempts, (d) direct consequences, (e) mental states (i.e., internal/emotional responses), and (f) conclusion (Greenslade et al., 2020). The developers of MSSG preassigned each MC with one of six story grammar codes, although some (i.e., five

MCs) were assigned with more than one coding option (e.g., could be an initiating event or a direct consequence, depending on the order of production, neighboring MCs, and other formulation details) (Greenslade et al., 2020; Richardson et al., 2021). MSSG developers then identified story grammar "episodes" by preassigning each MC with an episode number (1–5); two MCs were assigned multiple coding options. Five episodes (each including multiple MCs) were identified based on their inclusion of at least three story grammar components (i.e., an "initiating event," "attempt," and "direct consequence"; Greenslade et al., 2020).

Each episode (five total) was evaluated—those with at least two out of the three required story grammar components received 1 point, and episodes with one or no components received 0 points, for a total possible SGEC score of 5. Interrater reliability for SGEC was 97.6%. Although other scoring options are available with MSSG (e.g., total number of episodic components), SGEC has been recommended as a more meaningful measure of macrolinguistic processing (i.e., narrative complexity) (Greenslade et al., 2020; Richardson et al., 2021).

SEQ. Meaningful conditional relationships between propositions (e.g., where one event or component necessitates another) are required for narratives to be coherent (van Dijk, 1980). To determine whether MCs were produced sequentially, line numbers were added to orthographic transcripts. To be considered as "logically sequenced," each MC had to come after the MC listed before it in the checklist, with exceptions for when the sequence was interchangeable (Greenslade et al., 2020). Three points were given if MCs were produced in the correct sequence, 2 points if the participant self-corrected, 1 point if MCs were out of sequence and no self-correction attempts were made, and 0 points if MCs were absent, for a maximum score of 102. Interrater reliability for SEQ was 98.7%.

CoreLex. CoreLex is a measure of lexical typicality (i.e., the production of specific and commonly used words) during structured discourse tasks. The Cinderella CoreLex list consists of 94 single words previously identified as typically produced by controls when describing the Cinderella story (Dalton et al., 2020). These words are less frequently used by people with aphasia. Compared to controls, people with aphasia were more likely to use vague terms (e.g., "girl," "lady") over more specific terms, such as "Cinderella" and "Godmother," and/or grammatical function words such as articles (e.g., "a," "an") and prepositions (e.g., "after"). Participants were given 1 point for producing each word on the CoreLex checklist, with credit given for any inflected version of a CoreLex item but not for production of synonyms, for a total possible score of 94 points. CoreLex scoring was conducted using AphasiaBank discourse analysis tools, CHAT transcription, and Computerized Language Analysis scoring.

Preliminary Analysis

Model assumptions were evaluated using statistical software R (Version 4.4.0; The R Foundation for Statistical Computing). A scatter plot with locally estimated scatterplot smoothing (LOESS) line was used to assess linearity. Outliers were identified using Mahalanobis distance (i.e., cases exceeding a χ^2 significance threshold of $p < .001$; Kline, 2023). Descriptive statistics, including correlations among observed variables, were evaluated (see Table 2). Most variables were relatively normal in distribution (i.e., per skewness > 0.7 and kurtosis > 3.5; Lei & Lomax, 2005). However, AC showed a strong positive skew (1.68) with a median score of 1.0 (out of 34 possible points), suggesting a floor effect, and WABN demonstrated a negative skew (−1.12), indicating a potential ceiling effect.

Statistical Analysis

MPlus statistical software (Version 8.7; Muthén & Muthén, 1998) was used to conduct a CFA of narrative discourse ability, indicated by ATT, AC, SGEC, SEQ, and CoreLex. Structural equation modeling (SEM) was employed to evaluate the linear relationship between exogenous latent variable naming and endogenous latent variable narrative discourse production. Robust maximum likelihood estimation was used to account for nonnormality of AC and WABN (C.-H. Li, 2021). Common factor units were established using marker variables, with BNT-SF and AC loadings fixed at a value of 1. Model fit was evaluated using a χ^2 significance threshold of $p > .05$ (Barrett, 2007), root-mean-square error of approximation (RMSEA) of < .06 (Hu & Bentler, 1999), comparative fit index (CFI) of ≥ .95 (Hu & Bentler, 1999), Tucker–Lewis Index (TLI) of ≥ .95 (Hu & Bentler, 1999), and standardized root-mean-square residual (SRMR) of < .08 (Hu & Bentler, 1999). Standardized model results are reported. Post hoc sensitivity analyses were conducted to assess the impact of model assumption violations on model fit and parameter estimates.

Results

Participants

The final aggregate sample included 223 individuals with stroke-induced aphasia. Three participants had missing data points (i.e., two were missing VNT scores, and one was missing VNT and WAB-R Aphasia Quotient [WAB-R AQ]). Although reasons for missingness are unknown, no notable similarities in demographics, aphasia severity, naming ability, or narrative discourse ability (per available measures) were found across the three

Table 2. Descriptive statistics and correlations (n = 220).

Variable	BNT-SF	VNT	WABN	AC	ATTs	SGEC	SEQ	CoreLex
BNT-SF	1							
VNT	.762	1						
WABN	.772	.776	1					
AC	.555	.560	.471	1				
ATTs	.560	.581	.511	.803	1			
SGEC	.511	.534	.463	.719	.914	1		
SEQ	.590	.598	.528	.822	.975	.920	1	
CoreLex	.485	.520	.402	.750	.779	.756	.808	1
M	7.15	14.62	44.52	3.54	8.43	1.67	25.00	36.20
SD	4.64	6.73	16.35	4.90	6.73	1.63	19.76	19.20
Mdn	7.0	16.0	50.50	1.0	7.0	1.0	21.0	35.0
Range	0–15	0–22	0–60	0–26	0–30	0–5	0–90	0–83
Skew	0.08	–0.71	–1.12	1.68	0.54	0.46	0.55	0.12
Kurtosis	–1.21	–0.68	0.28	2.62	–0.72	–1.14	–0.66	–0.77

Note. BNT-SF = Boston Naming Test–Short Form; VNT = Verb Naming Test; WABN = Western Aphasia Battery–Revised Naming Subtest; AC = Accurate and Complete; ATTs = Attempts; SGEC = Story Grammar Episodic Complexity; SEQ = Sequencing; CoreLex = Core Lexicon.

participants, suggesting that the data are missing completely at random (MCAR). Given the likelihood of MCAR and very small amount of missing data in comparison to total sample size, listwise deletion was used. The final sample (n = 220) consisted of 127 men (58%) and 93 women (42%), with a mean age of 61.36 years (range: 25.6–90.7) and aphasia severity (per WAB-R AQ) ranging from nonaphasic by WAB (i.e., individuals who report aphasia symptoms but score above the WAB cutoff score of ≥ 93.8; n = 20), mild (76–93.7; n = 77), moderate (51–75; n = 94), severe (26–50; n = 21), and very severe aphasia (≤ 25; n = 8). Aphasia types represented in the sample (in order of most frequent to least) included anomic (n = 71), Broca's (n = 58), conduction (n = 42), nonaphasic by WAB (n = 20), Wernicke's (n = 15), transcortical motor (n = 9), global (n = 4), and transcortical sensory (n = 2). See Table 1 for participant characteristics.

CFA for Narrative Discourse

The narrative discourse measurement model showed a good fit to the data, $\chi^2(3) = 3.636$, $p = .304$, RMSEA = .031 (90% confidence interval [CI] [.000, .122]), CFI = .998, TLI = .994, and SRMR = .004. All five indicators, AC (λ = .824), ATT (λ = .981), SGEC (λ = .927), SEQ (λ = .994), and CoreLex (λ = .809), showed strong associations with latent narrative discourse ability. However, the loading of SEQ (λ = .994) suggested potential redundancy, which was also reflected in high correlations with other narrative discourse indicators (r > .90; Kline, 2023). In post hoc analysis, an alternative CFA excluding SEQ was tested. Loadings remained strong for AC (β = .824), ATT (β = .971), SGEC (β = .942), and CoreLex (β = .803) under the alternative CFA model. These findings support a well-defined latent narrative discourse construct. See Table 3 for CFA results.

Hypothesized SEM Model and Modifications

The original hypothesized model, in which latent naming ability defined by BNT, VNT, and WABN was specified to predict narrative discourse ability, as defined by AC, ATT, SGEC, SEQ, and CoreLex, did not fit the data adequately (see Table 4). Model modifications were made by correlating residuals between (a) BNT and WABN, (b) AC and SGEC, and (c) AC and CoreLex. These modifications are theoretically justified, as both the BNT and WABN measure the ability to name nouns, and CoreLex and SGEC represent micro- and macrolinguistic features, respectively, which may also be captured by AC to varying degrees. The modified model demonstrated a good fit to the data, $\chi^2(16) = 26.12$, $p = .05$, RMSEA = .054 (90% CI [.000, .090]), CFI = .990, TLI = .982, and SRMR = .024. Model results showed a significant positive association between naming and narrative discourse ability (β = .660, SE = .037, p < .001; R^2 = .436), with naming ability accounting for approximately 44% of the variance in narrative discourse. See Table 4 for the model fit comparison and Table 5 for the modified model results.

Model Assumptions

Based on Mahalanobis distance, five outliers were identified. These cases were retained in the primary linear and quadratic models and were removed only for post hoc sensitivity analyses. In both the linear and quadratic models, re-estimation with outliers excluded resulted in only minimal

Table 3. Confirmatory factor analysis of monologic narrative discourse indicators (n = 220).

Model fit					
χ^2(df), p	**CFI**	**TLI**	**RMSEA [90% CI]**	**SRMR**	**AIC**
6.02(3), .111	.998	.994	.068 [.000, .147]	.004	5,926.26
Factor loadings					
Indicator	**Standardized λ (*SE*)**			**R^2 (computed as λ^2)**	
AC	0.824 (0.022)***			.679	
ATTs	0.981 (0.003)***			.962	
SGEC	0.927 (0.010)***			.859	
SEQ	0.994 (0.002)***			.988	
CoreLex	0.809 (0.024)***			.655	

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximate; CI = confidence interval; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; AC = Accurate and Complete; ATTs = Attempts; SGEC = Story Grammar Episodic Complexity; SEQ = Sequencing; CoreLex = Core Lexicon.

***p < .001.

changes in standardized coefficients (maximum $\Delta\beta$ = .004), and all model fit indices remained stable, indicating that these cases were not influential. Visual inspection of demographic and clinical characteristics did not reveal a consistent pattern distinguishing these cases from the remainder of the sample. As such, the assumption of no influential outliers was met, and results are reported for the full sample. Despite the modified model showing a good fit to the data, visual inspection of the LOESS-smoothed scatter plot revealed a curvilinear relationship between naming ability and narrative discourse, showing a more modest relationship with narrative discourse at low-to-moderate levels of naming ability and a stronger association at higher levels (see Figure 1).

Post Hoc Curvilinear Model Results

To investigate this further, a latent quadratic term was introduced using the XWITH function in Mplus (Klein & Moosbrugger, 2000; Moosbrugger et al., 2009; Muthén & Muthén, 1998). Adding the quadratic term reduced AIC by 12, indicating a better fit to the data compared to the linear model (see Table 4). Detailed model fit differences are not reported, as local fit indices are not available when using XWITH in Mplus. The quadratic model explained greater variance in narrative discourse (R^2 = .610) compared to the linear model (R^2 = .438). See Table 5 for nonlinear model results and Figure 2 for nonlinear model path diagram.

Additional Post Hoc Sensitivity Analyses

To evaluate whether potential floor effects of AC contributed to the observed curvature, a scatter plot with LOESS line was generated after removing AC from the model. The resulting curve showed no meaningful change in shape, suggesting that AC floor effects did not have a meaningful impact on the shape of the model. Lastly, to evaluate the potential impact of extreme correlations with SEQ, the quadratic model was re-estimated excluding SEQ as an indicator. Model fit improved considerably, with AIC reduced by 1,426, and the structural relationship between naming ability and narrative discourse was nearly unchanged, suggesting that SEQ provided minimal unique explanatory value.

Discussion

Summary of Findings

A theory-driven measurement model of narrative discourse and SEM was estimated to examine the relationship between latent naming ability and monologic

Table 4. Model fit comparison (n = 220).

Model	**AIC**	**χ^2(df), p**	**RMSEA [90% CI]**	**CFI**	**TLI**	**SRMR**
Hypothesized linear	8,998.85	51.74(19), p = .000	.089 [.060, .118]	.961	.942	.034
Modified linear	10,011.47	26.12(16), p = .052	.054 [.000, .090]	.990	.982	.024
Nonlinear	9,999.56	—	—	—	—	—

Note. Em dashes indicate data not reported. AIC = Akaike information criterion; RMSEA = root-mean-square error of approximate; CI = confidence interval; CFI = comparative fit index; TLI = Tucker–Lewis index; SRMR = standardized root-mean-square residual.

Table 5. Modified linear and nonlinear model results (n = 220).

Latent construct	Indicators	Linear λ (*SE*)	Nonlinear λ (*SE*)
Naming	BNT-SF	0.846 (0.028)	0.831 (0.023)
	VNT	0.909 (0.031)	0.910 (0.025)
	WABN	0.846 (0.027)	0.846 (0.027)
Discourse	AC	0.825 (0.021)	0.849 (0.023)
	ATTs	0.980 (0.013)	0.984 (0.011)
	SGEC	0.926 (0.009)	0.938 (0.009)
	SEQ	0.995 (0.005)	0.996 (0.004)
	CoreLex	0.809 (0.024)	0.834 (0.024)
Latent path		**Linear, β (*SE*) (R^2 = .436)**	**Nonlinear, β (*SE*) (R^2 = .610)**
Naming → Discourse (linear)		.660 (0.037)*	.736 (0.038)*
Naming^2 → Discourse (quadratic)			.186 (0.035)*

Note. The nonlinear path model includes both a linear term and a quadratic term for latent naming ability to test for curvilinear effects on latent discourse ability. Standardized (STDYX) estimates reported. BNT-SF = Boston Naming Test–Short Form; VNT = Verb Naming Test; WABN = Western Aphasia Battery–Revised Naming Subtest; AC = Accurate and Complete; ATTs = Attempts; SGEC = Story Grammar Episodic Complexity; SEQ = sequencing; CoreLex = Core Lexicon.

*p < .05.

narrative discourse in people with aphasia. Two primary aims guided this investigation: (a) to evaluate the adequacy of a narrative discourse measurement model grounded in the LUNA framework (Dipper, Marshall, Boyle, Hersh, et al., 2021) and (b) to examine the relationship between latent naming and latent narrative discourse ability. CFA results suggest that latent narrative discourse was well defined by indicators, although SEQ demonstrated near saturation and potential redundancy with other narrative discourse indicators; however, post hoc sensitivity analyses showed that latent narrative discourse remained stable without SEQ. Although naming ability, as measured by the BNT-SF, VNT, and WABN, was a strong predictor of narrative discourse ability, it was better explained by a curvilinear relationship. The curvilinear model provided a superior fit to the data. Findings were consistent with a threshold pattern, in which differences in narrative discourse ability appeared to be more modest across individuals with relatively weaker naming ability and larger for those with stronger

Figure 1. Multivariate scatter plot with locally estimated smoothing (LOESS) line illustrating the relationship between latent confrontation naming ability and latent narrative discourse (n = 220). The LOESS curve suggests potential nonlinearity in the association between constructs.

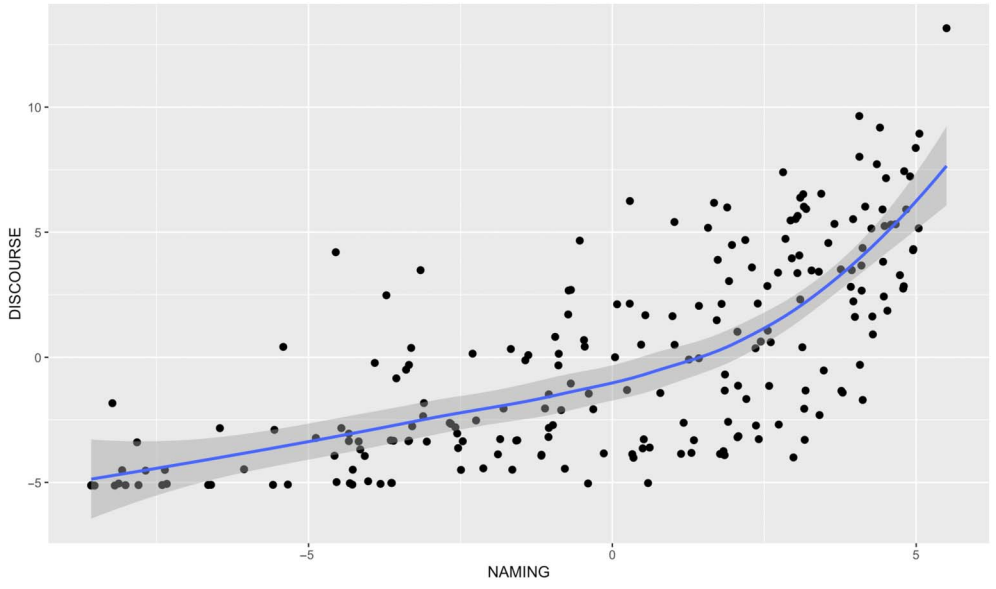
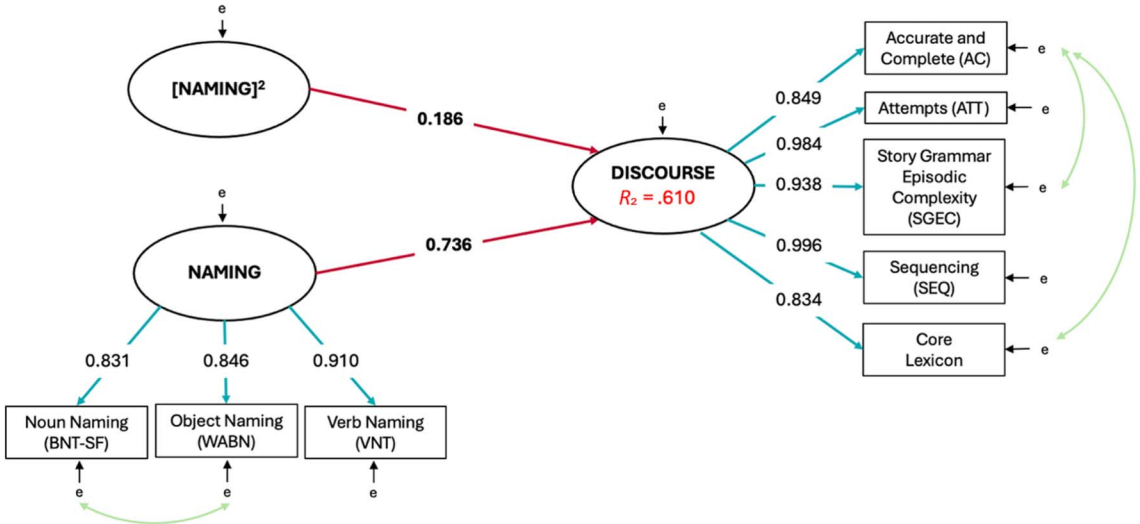


Figure 2. Structural equation path diagram illustrating the nonlinear predictive relationship between latent naming ability and latent narrative discourse (n = 220). Standardized parameter estimates displayed. BNT-SF = Boston Naming Test–Short Form; WABN = Western Aphasia Battery–Revised Naming Subtest; VNT = Verb Naming Test.



naming ability (e.g., "moderately strong" vs. "strong" vs. "very strong").

Measurement Model Specification

All narrative discourse indicators loaded strongly on the latent construct; however, SEQ approached saturation and demonstrated extreme correlations with other indicators. These findings likely reflect substantial conceptual overlap between SEQ and measures aligning with macrostructural organization and propositional planning (see Supplemental Material S1). Within the LUNA framework, such overlap is expected, as decisions about proposition order are intrinsically tied to story grammar in a nested manner (i.e., sequencing decisions are guided by macrostructural plans). Importantly, post hoc sensitivity analyses excluding SEQ yielded stable loadings and did not meaningfully alter the structural relationship between naming ability and narrative discourse. This pattern suggests that the latent narrative discourse construct is robust while also highlighting the challenges of achieving parsimony when modeling highly complex, interdependent narrative discourse components.

Additionally, model fit improved after correlating residuals between (a) BNT-SF and WABN, (b) AC and SGEC, and (c) AC and CoreLex, acknowledging the potential conceptual overlap between these measures. Although AC and SGEC have different units of measurement (i.e., word–phrase level for AC and macrostructure level for SGEC), both capture aspects of narrative completeness. Similarly, as AC and CoreLex are scored according to the presence of prespecified words (and/or phrases for AC), both capture microlinguistic accuracy. Residuals were not correlated between (a) AC and SEQ or between (b) AC and ATT, as correct sequencing is not required for AC (i.e., MCs, as well as words/phrases within an MC, may be produced in any order to be scored as AC) and ATTs do not have to be accurate or complete.

Linear Relationship Between Naming and Narrative Discourse

SEM revealed a significant positive association between latent naming ability and latent narrative discourse ability, with naming accounting for approximately 44% of the variance in latent narrative discourse ability. This finding aligns with existing literature showing positive associations between confrontation naming and discourse outcomes, such as retrieval of word classes (e.g., proportion of nouns; Herbert et al., 2008), informativeness (Fergadiotis et al., 2019; Fergadiotis & Wright, 2016), and MCs (i.e., for individuals with Broca's and Wernicke's aphasia; Richardson et al., 2018). Additionally, from a language network perspective, the observed association is consistent with neurocognitive models that propose functional integration across levels of language processing (Fedorenko et al., 2024). Similarly, Fergadiotis et al. (2019) reported a relatively strong linear association between latent naming ability and monologic discourse informativeness, as measured by CIUs across multiple discourse contexts (Fergadiotis et al., 2019), with naming accounting for approximately 63% of the variance in discourse informativeness across contexts. Although discourse constructs differed between the current study and

Fergadiotis et al. (2019), both findings support the presence of a relationship between naming ability and discourse-level outcomes.

Interpretation of the Curvilinear Relationship

Beyond establishing a linear association, this study provides evidence that the relationship between naming and narrative discourse may be better described as curvilinear. Visual inspection of the data and latent quadratic modeling indicated a weaker association at lower levels of naming ability, with a stronger relationship with narrative performance at higher levels of naming. Additionally, the quadratic model accounted for more variance in narrative discourse (R^2 = .61) than the linear model. This pattern remained stable across multiple sensitivity analyses, including removal of AC scores to evaluate potential floor effects, as well as SEQ to examine potential model redundancy.

The curvilinear relationship suggests that once naming ability drops below a certain functional range, narrative discourse performance may remain uniformly weaker despite differences in naming ability. Conversely, at higher levels of naming ability, above this threshold, even relatively mild deficits in naming may correspond to disproportionately larger deficits in narrative discourse. This interpretation does not imply that individuals with lower naming ability lack macrostructural knowledge and/or propositional intent. Rather, it may suggest that impaired lexical retrieval may play a dominant role by constraining how narrative discourse plans are fully realized and reflected in observable narrative performance.

This pattern is consistent with the LUNA framework's characterization of narrative discourse production as a nested process, where macrostructural plans guide the selection and ordering of propositions, and linguistic processes translate propositions into fully realized utterances (Dipper, Marshall, Boyle, Hersh, et al., 2021). Under this framework, macrostructural and propositional planning may be intact but inadequately externalized when linguistic processes are compromised. However, people with aphasia may also experience deficits in macrostructural and propositional conceptualization (Hameister & Nickels, 2018). Therefore, when lexical retrieval is impaired, gesturing and other multimodal communication behaviors (e.g., writing, drawing) may serve an important compensatory role, allowing propositional content to be externalized through nonverbal modalities. Examining multimodal output may provide greater insight into where breakdowns in narrative discourse are occurring. Consistent with this expectation, Dutta and Mohapatra (2024) reported that multimodal scoring yielded higher macrolinguistic and MC scores than verbal-only scoring, particularly for those with nonfluent aphasia (Dutta & Mohapatra, 2024). Additionally, gestures have been found to convey semantically rich information (Pritchard et al., 2015) and enhance overall communicative success (Caute et al., 2013).

The curvilinear pattern is also consistent with reports of severity-dependent relationships between word retrieval and discourse. Prior work has shown a weak or inconsistent association between naming and discourse among individuals with more severe aphasia and a stronger association for those with milder impairment (Mayer & Murray, 2003). Similarly, Cherney (2010) found that discourse-level benefits following oral reading intervention were observed primarily in individuals with mild and moderate aphasia. These findings align with the present results and suggest that naming ability may be a more informative predictor of narrative discourse performance for those with milder naming deficits.

Implications

From a clinical perspective, these findings suggest that naming outcomes may underestimate differences in discourse-level functioning (in the context of elicited narratives) for individuals with relatively mild naming impairment. However, the observed threshold effect illustrated by the curvilinear model suggests that narrative discourse measures may be less sensitive to detecting smaller incremental differences in narrative discourse ability among individuals with naming abilities below this threshold. When interpreted alongside the LUNA model, the shared variance observed between confrontation naming and elicited narrative discourse performance as a whole illustrates the role of lexical retrieval processes in narrative production. As such, naming assessment may complement narrative assessment by detecting smaller, more nuanced gains among individuals with relatively similar deficits in narrative discourse performance. These findings suggest that impairment-level measures such as naming may still provide clinically meaningful information when interpreted within the broader context of narrative discourse performance.

Limitations

While the study contributes novel insights into the relationship between naming ability and narrative discourse performance, several limitations should be acknowledged. First, the study focused exclusively on a single narrative storytelling task. Structured narrative storytelling tasks such as the Cinderella task (MacWhinney et al., 2011) differ in important ways from narrative production embedded within more naturalistic conversational interactions. Unlike conversational storytelling, the Cinderella task relies on a fixed stimulus, constrained narrative content, and standardized elicitation procedures, and performance may be impacted by

prior familiarity and cultural awareness of story elements, which may vary across individuals. Therefore, findings may not be generalizable across narrative contexts. However, successful performance on elicited narrative tasks relies on cognitive–linguistic processes that are also relevant to broader communicative functioning, including lexical selection, cohesion, coherence, macrostructural organization, and monitoring of language output and listener needs.

Second, this study relied heavily on MC checklist-dependent measures. The MCA framework provides a structured, theory-driven, standardized method for measuring discourse ability, which may be particularly useful in clinical and research settings due to its protocol-driven administration/scoring methods and psychometric strengths (Kong, 2009, 2011; Kong et al., 2016; Nicholas & Brookshire, 1995; Pritchard et al., 2017; Richardson & Dalton, 2016; Richardson et al., 2018, 2021). MC checklists streamline analysis by focusing on prespecified meaningful concepts (with their constituent elements), enabling efficient scoring for diagnostic use and differentiating aphasia severity and subtypes (Dalton & Richardson, 2015; Kong, 2009; Kong et al., 2016). In doing so, it excludes utterances that do not correspond with listed MCs, potentially omitting important aspects of discourse production. Therefore, in the context of CFA and SEM, the reliance on only MC-dependent measures may have introduced statistical redundancy into the measurement model.

Relatedly, MCA intentionally does not account for the effective use of total communication and/or circumlocution strategies, both of which can reveal meaningful differences in discourse effectiveness among those with varying aphasia severity and subtypes; therefore, other measures may serve as valuable complements to future discourse models. For example, CIUs are capable of capturing both lexical accuracy and circumlocution attempts; however, it does not capture structural aspects of discourse (i.e., macrolinguistic planning) or whether circumlocutory attempts were successful in conveying the message. Importantly, because both confrontation naming tasks and CIUs are scored at the word level, analyses relying exclusively on CIUs may overemphasize word-level processes when estimating the relationship between naming and discourse. This shared unit of analysis may artificially strengthen or linearize observed associations, without capturing the broader organizational (i.e., macrostructural) demands of narrative discourse. Finally, beyond task constraints, the present study did not examine a range of individual and contextual factors that may influence narrative discourse production, including aphasia severity, subtype, cultural differences in narrative conventions, familiarity with the stimulus (e.g., the Cinderella story), comorbid cognitive deficits, fatigue, and psychological well-being.

Future Directions

Further refinement of the narrative discourse measurement model may be warranted to address concerns regarding overlapping measurement content (e.g., SEQ). Alternate sequencing measures that explicitly evaluate "informational dependency," independent of MCA checklists, may offer a theory-driven way to address redundancy concerns. MCA-derived measures may also be fully retained, by using statistical remedies (e.g., adding a latent method factor) to address potential method variance—although there is some debate about which remedies are most appropriate (Kock, 2017; Manata & Boster, 2024; Tehseen et al., 2017). Future SEM studies of discourse should test other theory-driven measurement models to examine the relationship between naming and discourse ability across multiple monologic discourse tasks (e.g., picture description, procedural, conversation) and conversational discourse.

Additionally, incorporating complementary word-level measures of informativeness (e.g., CIUs) could strengthen models by accounting for meaningful communicative words that may not be captured in prespecified checklists. Similarly, adapting and validating existing frameworks to account for gesture use (e.g., Dutta & Mohapatra, 2024) may provide a more comprehensive view of communication, particularly for those with greater deficits in lexical retrieval. Emerging work on cospeech gesture use may also offer promising avenues for capturing total/multimodal communication more comprehensively (Caute et al., 2021; Dresang et al., 2026; Hameister & Nickels, 2018; Kong et al., 2017; Rose et al., 2016; Stark & Cofoid, 2022). These developments have the potential to improve tools for assessing both micro- and macrolinguistic elements of discourse and lead to more accurate and inclusive (i.e., nuanced) models of functional communication in people with aphasia. However, it is also important to balance nuanced precision with clinical feasibility, as there may be notable trade-offs between precision-based model refinement and clinical applicability of more elaborate measurement models.

Further research is also needed to examine the impact of pragmatic skills on successful discourse production (i.e., how people with aphasia accommodate differences in setting, familiarity with listeners and/or conversational partners, fluctuating topics, and communicative purpose, across different contexts). Additionally, comparative studies examining how contextual factors and/or potential moderators (e.g., working memory, psychological well-being) influence the relationship between naming and monologic discourse ability are needed to improve the generalizability of these findings to a broader range of people with aphasia. There is still much to be learned about how personal and environmental factors contribute

to nuanced differences in discourse skills across individuals and contexts.

Given the diversity of discourse analysis tools and methods in use, some researchers have called for a consensus-based core outcomes set for evaluating discourse ability in people with aphasia (Dietz & Boyle, 2018), while others have advocated for prioritizing measurement standardization efforts (Wallace et al., 2018). Despite much recent progress, consensus on a core set of discourse measures has not yet been reached (Wallace et al., 2019), and many existing tools still lack comprehensive psychometric validation (Pritchard et al., 2017). Additionally, findings from the current study suggests that some measures of discourse may be less sensitive to detecting nuanced differences in narrative discourse ability across participants. As such, future work is needed to examine the sensitivity of discourse measures to detect meaningful differences and treatment-related change across individuals with differing language profiles and impairment levels.

Conclusions

This study examined the relationship between latent naming ability and latent narrative discourse production and tested a theory-driven measurement model of narrative discourse. Findings revealed a curvilinear relationship between latent constructs, in which small differences in naming were associated with large differences in narrative performance for those with higher naming ability but limited differences for those with lower naming ability. These findings suggest that lexical retrieval is a meaningful contributor to narrative discourse production while also illustrating the limitations of this relationship. Collectively, these findings reinforce the importance of theoretically grounded measures that capture the complex, multilevel processes involved in communicative functioning while also highlighting current challenges in discourse measurement, including overlapping measurement content and the inherent difficulty of modeling highly interdependent cognitive–linguistic processes. Addressing these challenges will require stronger theoretical consensus across contexts and continued refinement of measurement tools that capture functionally meaningful aspects of communication while also balancing methodological rigor with clinical feasibility.

Data Availability Statement

The data set for the current study was derived from primary data retrieved from the AphasiaBank database (MacWhinney et al., 2011), which receives support from National Institute on Deafness and Other Communication Disorders (NIDCD) Grant R01-DC008524 for 2022–2028. To protect participants' privacy, the AphasiaBank database is not publicly available. However, researchers, clinicians, educators, and students, who are interested in accessing the data, may formally request access at https://talkbank.org/aphasia/. For questions regarding the final data set, please contact Christie Carroll-Duhigg at ccduhigg@unm.edu.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders (Grant R01DC018282-02S1, awarded to principal investigator Jessica D. Richardson), the National Institute of General Medical Sciences (Grant P20GM109089; PI: Claude W. Shuttleworth), and the National Institute on Aging (Grant P30AG086404; PI: Gary Rosenberg). The authors gratefully acknowledge the individuals with aphasia and their care partners who generously gave their time and effort to participate in aphasia research and allowed their assessment data to be used in future projects. The authors also thank AphasiaBank and the many sites that have contributed to its database.

References

Alyahya, R. S., Lambon Ralph, M. A., Halai, A., & Hoffman, P. (2022). The cognitive and neural underpinnings of discourse coherence in post-stroke aphasia. *Brain Communications, 4*(3), Article fcac147. https://doi.org/10.1093/braincomms/fcac147

American Speech-Language-Hearing Association. (n.d.). *Aphasia*. Retrieved September 1, 2025, from https://www.asha.org/practice-portal/clinical-topics/aphasia/

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

Armstrong, E., Ciccone, N., Godecke, E., & Kok, B. (2011). Monologues and dialogues in aphasia: Some initial comparisons. *Aphasiology, 25*(11), 1347–1371. https://doi.org/10.1080/02687038.2011.577204

Armstrong, E., & Ferguson, A. (2010). Language, meaning, context, and functional communication. *Aphasiology, 24*(4), 480–496. https://doi.org/10.1080/02687030902775157

Armstrong, E., & Hersh, D. (2024). Speaking up and being heard: The importance of functional communication and discourse principles in aphasia intervention. *Seminars in Speech and Language, 45*(4), 356–367. https://doi.org/10.1055/s-0044-1788981

Baker, C., Worrall, L., Rose, M., & Ryan, B. (2020). 'It was really dark': The experiences and preferences of people with aphasia to manage mood changes and depression. *Aphasiology, 34*(1), 19–46. https://doi.org/10.1080/02687038.2019.1673304

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815–824. https://doi.org/10.1016/j.paid.2006.09.018

Beeke, S., Maxim, J., & Wilkinson, R. (2008). Rethinking agrammatism: Factors affecting the form of language elicited via clinical test procedures. *Clinical Linguistics & Phonetics, 22*(4–5), 317–323. https://doi.org/10.1080/02699200801918911

Boucher, J., Marcotte, K., Brisebois, A., Courson, M., Houzé, B., Desautels, A., Léonard, C., Rochon, E., & Brambati, S. M. (2022). Word-finding in confrontation naming and picture descriptions produced by individuals with early post-stroke aphasia. *The Clinical Neuropsychologist, 36*(6), 1422–1437. https://doi.org/10.1080/13854046.2020.1817563

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

Boyle, M., Akers, C. M., Cavanaugh, R., Hula, W. D., Swiderski, A. M., & Elman, R. J. (2023). Changes in discourse informativeness and efficiency following communication-based group treatment for chronic aphasia. *Aphasiology, 37*(3), 563–597. https://doi.org/10.1080/02687038.2022.2032586

Brady, M. C., Ali, M., Fyndanis, C., Kambanaros, M., Grohmann, K. K., Laska, A. C., Hernández-Sacristán, C., & Varlokosta, S. (2014). Time for a step change? Improving the efficiency, relevance, reliability, validity and transparency of aphasia rehabilitation research through core outcome measures, a common data set and improved reporting criteria. *Aphasiology, 28*(11), 1385–1392. https://doi.org/10.1080/02687038.2014.930261

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

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

Caute, A., Dipper, L., & Roper, A. (2021). The City Gesture Checklist: The development of a novel gesture assessment. *International Journal of Language & Communication Disorders, 56*(1), 20–35. https://doi.org/10.1111/1460-6984.12579

Caute, A., Pring, T., Cocks, N., Cruice, M., Best, W., & Marshall, J. (2013). Enhancing communication through gesture and naming therapy. *Journal of Speech, Language, and Hearing Research, 56*(1), 337–351. https://doi.org/10.1044/1092-4388(2012/11-0232)

Cherney, L. R. (2010). Oral reading for language in aphasia: Impact of aphasia severity on cross-modal outcomes in chronic nonfluent aphasia. *Seminars in Speech and Language, 31*(1), 42–51. https://doi.org/10.1055/s-0029-1244952

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Clark, H. H. (2012). *Using language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology, 24*(4), S923–S938. https://doi.org/10.1044/2015_AJSLP-14-0161

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

Dalton, S. G., Stark, B. C., Fromm, D., Apple, K., MacWhinney, B., Rensch, A., & Rowedder, M. (2022). Validation of an automated procedure for calculating core lexicon from transcripts. *Journal of Speech, Language, and Hearing Research, 65*(8), 2996–3003. https://doi.org/10.1044/2022_JSLHR-21-00473

Dalton, S. G. H., Kim, H., Richardson, J. D., & Wright, H. H. (2020). A compendium of core lexicon checklists. *Seminars in Speech and Language, 41*(1), 45–60. https://doi.org/10.1055/s-0039-3400972

Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: Have we reached the tipping point? *Aphasiology, 32*(4), 459–464. https://doi.org/10.1080/02687038.2017.1398803

Dignam, J., Rodriguez, A. D., O'Brien, K., Burfein, P., & Copland, D. A. (2024). Early within therapy naming probes as a clinically-feasible predictor of anomia treatment response. *Neuropsychological Rehabilitation, 34*(2), 196–219. https://doi.org/10.1080/09602011.2023.2177312

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

Dipper, L., Marshall, J., Boyle, M., Hersh, D., Botting, N., & Cruice, M. (2021). Creating a theoretical framework to underpin discourse assessment and intervention in aphasia. *Brain Sciences, 11*(2), Article 183. https://doi.org/10.3390/brainsci11020183

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

Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology, 4*(4), 130–134. https://doi.org/10.1044/1058-0360.0404.130

Dresang, H. C., Dalton, S. G., McCreary, E., Pittman, E., & Stark, B. C. (2026). Iconic gesture use during discourse production in latent aphasia. *Aphasiology, 40*(8), 1485–1508. https://doi.org/10.1080/02687038.2025.2511215

Dutta, M., & Mohapatra, B. (2024). Expanding the scope: Multimodal dimensions in aphasia discourse analysis—Preliminary findings. *Frontiers in Human Neuroscience, 18,* Article 1419311. https://doi.org/10.3389/fnhum.2024.1419311

Dutta, M., Murray, L. L., Park, H., Burklow, E., Bose, A., Kim, H., Greenslade, K., Ramage, A. E., Combs, C., Balasubramanian, A., & Casilio, M. (2025). Let's chat about spoken discourse: A tutorial to support use of spoken discourse analysis when providing aphasia clinical services. *American Journal of Speech-Language Pathology, 34*(4), 1931–1966. https://doi.org/10.1044/2025_ajslp-24-00534

Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience, 25*(5), 289–312. https://doi.org/10.1038/s41583-024-00802-4

Fergadiotis, G., Kapantzoglou, M., Kintz, S., & Wright, H. H. (2019). Modeling confrontation naming and discourse informativeness using structural equation modeling. *Aphasiology, 33*(5), 544–560. https://doi.org/10.1080/02687038.2018.1482404

Fergadiotis, G., & Wright, H. H. (2016). Modelling confrontation naming and discourse performance in aphasia. *Aphasiology, 30*(4), 364–380. https://doi.org/10.1080/02687038.2015.1067288

Forbes, M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language, 33*(3), 217–222. https://doi.org/10.1055/s-0032-1320041

Godecke, E., Armstrong, E., Rai, T., Ciccone, N., Rose, M. L., Middleton, S., Whitworth, A., Holland, A., Ellery, F., Hankey,

G. J., Cadilhac, D. A., Bernhardt, J., & VERSE Collaborative Group. (2021). A randomized control trial of intensive aphasia therapy after acute stroke: The Very Early Rehabilitation for SpEech (VERSE) study. *International Journal of Stroke, 16*(5), 556–572. https://doi.org/10.1177/1747493020961926

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Lea & Febiger.

Greenslade, K. J., Stuart, J. E., Richardson, J. D., Dalton, S. G., & Ramage, A. E. (2020). Macrostructural analyses of Cinderella narratives in a large nonclinical sample. *American Journal of Speech-Language Pathology, 29*(4), 1923–1936. https://doi.org/10.1044/2020_AJSLP-19-00151

Hameister, I., & Nickels, L. (2018). The cat in the tree—Using picture descriptions to inform our understanding of conceptualisation in aphasia. *Language, Cognition and Neuroscience, 33*(10), 1296–1314. https://doi.org/10.1080/23273798.2018.1497801

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology, 22*(2), 184–203. https://doi.org/10.1080/02687030701262613

Hilari, K., Needle, J. J., & Harrison, K. L. (2012). What are the important factors in health-related quality of life for people with aphasia? A systematic review. *Archives of Physical Medicine and Rehabilitation, 93*(1), S86–S95.e4. https://doi.org/10.1016/j.apmr.2011.05.028

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jacobs, M., Briley, P. M., Fang, X., & Ellis, C. (2021). Telepractice treatment for aphasia: Association between clinical outcomes and client satisfaction. *Telemedicine Reports, 2*(1), 118–124. https://doi.org/10.1089/tmr.2020.0024

Jacobs, M., Briley, P. M., Wright, H. H., & Ellis, C. (2023). Marginal assessment of the cost and benefits of aphasia treatment: Evidence from community-based telerehabilitation treatment for aphasia. *Journal of Telemedicine and Telecare, 29*(4), 271–281. https://doi.org/10.1177/1357633X20982773

Kagan, A., Simmons-Mackie, N., Rowland, A., Huijbregts, M., Shumway, E., McEwen, S., Threats, T., & Sharp, S. (2008). Counting what counts: A framework for capturing real-life outcomes of aphasia intervention. *Aphasiology, 22*(3), 258–280. https://doi.org/10.1080/02687030701282595

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test–Second Edition*. Lippincott Williams & Wilkins. https://proedinc.com/products-11870.html

Kertesz, A. (2006). *The Western Aphasia Battery–Revised (WAB-R)*. Pearson. https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Speech-%26-Language/Western-Aphasia-Battery-Revised/p/100000194

Kim, H., Kintz, S., Zelnosky, K., & Wright, H. H. (2019). Measuring word retrieval in narrative discourse: Core lexicon in aphasia. *International Journal of Language & Communication Disorders, 54*(1), 62–78. https://doi.org/10.1111/1460-6984.12432

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*(4), 457–474. https://doi.org/10.1007/BF02296338

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press. https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462551910

Kock, N. (2017). Common method bias: A full collinearity assessment method for PLS-SEM. In H. Latan & R. Noonan (Eds.), *Partial least squares path modeling* (pp. 245–257). Springer. https://doi.org/10.1007/978-3-319-64069-3_11

Kong, A. P.-H. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders, 42*(6), 442–464. https://doi.org/10.1016/j.jcomdis.2009.06.002

Kong, A. P.-H. (2011). The main concept analysis in Cantonese aphasic oral discourse: External validation and monitoring chronic aphasia. *Journal of Speech, Language, and Hearing Research, 54*(1), 148–159. https://doi.org/10.1044/1092-4388(2010/09-0240)

Kong, A. P.-H., Law, S.-P., & Chak, G. W.-C. (2017). A comparison of coverbal gesture use in oral discourse among speakers with fluent and nonfluent aphasia. *Journal of Speech, Language, and Hearing Research, 60*(7), 2031–2046. https://doi.org/10.1044/2017_JSLHR-L-16-0093

Kong, A. P.-H., Linnik, A., Law, S.-P., & Shum, W. W.-M. (2018). Measuring discourse coherence in anomic aphasia using rhetorical structure theory. *International Journal of Speech-Language Pathology, 20*(4), 406–421. https://doi.org/10.1080/17549507.2017.1293158

Kong, A. P. H., Whiteside, J., & Bargmann, P. (2016). The main concept analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. *Logopedics Phoniatrics Vocology, 41*(3), 129–141. https://doi.org/10.3109/14015439.2015.1041551

Kristinsson, S., Basilakos, A., den Ouden, D. B., Cassarly, C., Spell, L. A., Bonilha, L., Rorden, C., Hillis, A. E., Hickok, G., Johnson, L., Busby, N., Walker, G. M., McLain, A., & Fridriksson, J. (2023). Predicting outcomes of language rehabilitation: Prognostic factors for immediate and long-term outcomes after aphasia therapy. *Journal of Speech, Language, and Hearing Research, 66*(3), 1068–1084. https://doi.org/10.1044/2022_JSLHR-22-00347

Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Randolph, C. (1999). An empirically derived short form of the Boston Naming Test. *Archives of Clinical Neuropsychology, 14*(6), 481–487. https://doi.org/10.1093/arclin/14.6.481

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

Leaman, M. C., & Edmonds, L. A. (2021). Assessing language in unstructured conversation in people with aphasia: Methods, psychometric integrity, normative data, and comparison to a structured narrative task. *Journal of Speech, Language, and Hearing Research, 64*(11), 4344–4365. https://doi.org/10.1044/2021_JSLHR-20-00641

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling, 12*(1), 1–27. https://doi.org/10.1207/s15328007sem1201_1

Li, C.-H. (2021). Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables. *Behavior Research Methods, 53*(5), 2191–2213. https://doi.org/10.3758/s13428-021-01547-z

Li, R., & Kiran, S. (2024). Noun and verb impairment in single-word naming and discourse production in Mandarin-English bilingual adults with aphasia. *Aphasiology, 38*(2), 337–365. https://doi.org/10.1080/02687038.2023.2189994

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

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

Low, T. A., Chilvers, M. J., Zhu, H., Carlson, H. L., Harris, A. D., Goodyear, B. G., & Dukelow, S. P. (2024). Structural network topology associated with naming improvements following intensive aphasia therapy in post-stroke aphasia. *Journal of the Neurological Sciences, 462,* Article 123065. https://doi.org/10.1016/j.jns.2024.123065

Mack, W., Freed, D., Williams, B. W., & Henderson, V. W. (1992). Boston naming test: Shortened versions for use in Alzheimer's disease. *Journal of Gerontology, 47*(3), P154–P158. https://doi.org/10.1093/geronj/47.3.P154

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

Maher, L. M., & Raymer, A. M. (2004). Management of anomia. *Topics in Stroke Rehabilitation, 11*(1), 10–21. https://doi.org/10.1310/318R-RMD5-055J-PQ40

Manata, B., & Boster, F. J. (2024). Reconsidering the problem of common-method variance in organizational communication research. *Management Communication Quarterly, 38*(3), 677–697. https://doi.org/10.1177/08933189231226242

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology, 17*(5), 481–497. https://doi.org/10.1080/02687030344000148

Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation approaches. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 103–136). Sense.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (*9th ed.*). https://www.statmodel.com/html_ug.shtml

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

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research, 38*(1), 145–156. https://doi.org/10.1044/jshr.3801.145

Pashek, G. V., & Tompkins, C. A. (2002). Context and word class influences on lexical retrieval in aphasia. *Aphasiology, 16*(3), 261–286. https://doi.org/10.1080/02687040143000573

Pritchard, M., Dipper, L., Morgan, G., & Cocks, N. (2015). Language and iconic gesture use in procedural discourse by speakers with aphasia. *Aphasiology, 29*(7), 826–844. https://doi.org/10.1080/02687038.2014.993912

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders, 52*(6), 689–732. https://doi.org/10.1111/1460-6984.12318

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2018). Psychometric properties of discourse measures in aphasia: Acceptability, reliability, and validity. *International Journal of Language & Communication Disorders, 53*(6), 1078–1093. https://doi.org/10.1111/1460-6984.12420

RELEASE Collaboration. (2020). Communicating simply, but not too simply: Reporting of participants and speech and language interventions for aphasia after stroke. *International Journal of Speech-Language Pathology, 22*(3), 302–312. https://doi.org/10.1080/17549507.2020.1762000

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

Richardson, J. D., Dalton, S. G., Greenslade, K. J., Jacks, A., Haley, K. L., & Adams, J. (2021). Main concept, sequencing, and story grammar analyses of Cinderella narratives in a large sample of persons with aphasia. *Brain Sciences, 11*(1), Article 110. https://doi.org/10.3390/brainsci11010110

Richardson, J. D., Hudspeth Dalton, S. G., Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2018). The relationship between confrontation naming and story gist production in aphasia. *American Journal of Speech-Language Pathology, 27*(1S), 406–422. https://doi.org/10.1044/2017_AJSLP-16-0211

Rose, M. L., Mok, Z., Carragher, M., Katthagen, S., & Attard, M. (2016). Comparing multi-modality and constraint-induced treatment for aphasia: A preliminary investigation of generalisation to discourse. *Aphasiology, 30*(6), 678–698. https://doi.org/10.1080/02687038.2015.1100706

Simic, T., Laird, L., Brisson, N., Moretti, K., Théorêt, J.-L., Black, S. E., Eskes, G. A., Leonard, C., & Rochon, E. (2022). Cognitive training to enhance aphasia therapy (Co-TrEAT): A feasibility study. *Frontiers in Rehabilitation Sciences, 3,* Article 815780. https://doi.org/10.3389/fresc.2022.815780

Stark, B. C., & Cofoid, C. (2022). Task-specific iconic gesturing during spoken discourse in aphasia. *American Journal of Speech-Language Pathology, 31*(1), 30–47. https://doi.org/10.1044/2021_AJSLP-20-00271

Tehseen, S., Ramayah, T., & Sajilan, S. (2017). Testing and controlling for common method variance: A review of available methods. *Journal of Management Sciences, 4*(2), 142–168. https://doi.org/10.20547/jms.2014.1704202

Tierney-Hendricks, C., Schliep, M. E., & Vallila-Rohter, S. (2022). Using an implementation framework to survey outcome measurement and treatment practices in aphasia. *American Journal of Speech-Language Pathology, 31*(3), 1133–1162. https://doi.org/10.1044/2021_AJSLP-21-00101

van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition* (1st ed.). Routledge. https://doi.org/10.4324/9780429025532

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A.-C., Marshall, J., . . . Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke, 14*(2), 180–185. https://doi.org/10.1177/1747493018806200

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

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

World Health Organization. (2002, January). *IFC beginner's guide: Towards a common language for functioning, disability and health.*

https://www.who.int/publications/m/item/icf-beginner-s-guide-towards-a-common-language-for-functioning-disability-and-health

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology, 25*(3), 309–322. https://doi.org/10.1080/02687038.2010.508530

Wright, H. H. (2011). Discourse in aphasia: An introduction to current research and future directions. *Aphasiology, 25*(11), 1283–1285. https://doi.org/10.1080/02687038.2011.613452

Zhang, M., Geng, L., Yang, Y., & Ding, H. (2021). Cohesion in the discourse of people with post-stroke aphasia. *Clinical Linguistics & Phonetics, 35*(1), 2–18. https://doi.org/10.1080/02699206.2020.1734864

Copyright of American Journal of Speech-Language Pathology is the property of American Speech-Language-Hearing Association and its content may not be copied or emailed to multiple sites without the copyright holder's express written permission. Additionally, content may not be used with any artificial intelligence tools or machine learning technologies. However, users may print, download, or email articles for individual use.