Discourse Fluency Disruptions in Latent Aphasia: A Mixed-Effects Analysis Across Elicitation Contexts

Brielle C. Stark, PhD[1]; Kent Meinert, BA[1,2]; Davida Fromm, PhD[3]; Brian MacWhinney, PhD[3]

1 Indiana University Bloomington; 2 University of Queensland; 3 Carnegie Mellon University

Introduction

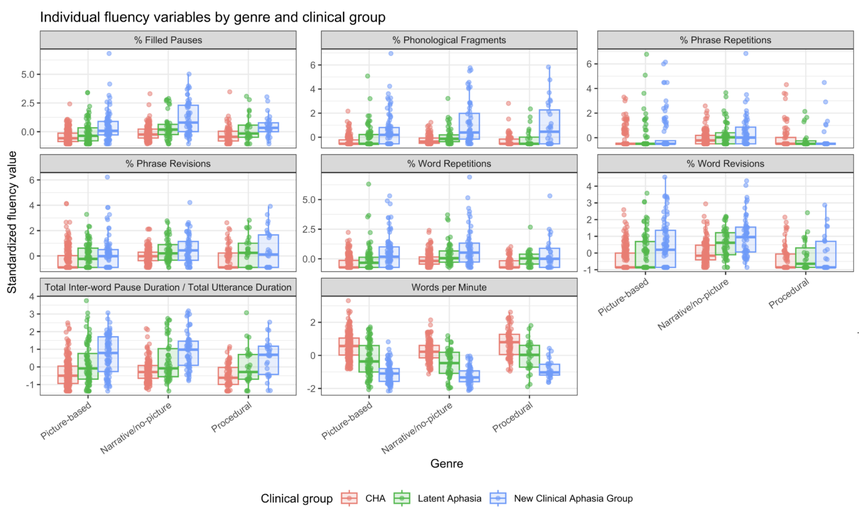
- ✓"Fluency" is a complicated phenomenon → generally refers to how efficiently and smoothly a speaker produces connected speech, reflecting the coordination of speech rate, pausing, repetitions, revisions, fragments, and filled pauses during real-time language formulation
- ✓Individuals with mildest aphasia have long been differentiated from neurologically healthy controls based on efficiency of language access/fluency (e.g., DeDe & Salas, 2020), but also by lexical-semantic access (Lanzi, Dalton & Stark, 2025).
- ✓Further, it is well known that discourse tasks (e.g., picture-based, non-picture-based narratives, procedural tasks) vary in their cognitive-linguistic loads, and this impacts linguistic outcomes (e.g., Stark, 2019; Stark & Fukuyama, 2021), though few have evaluated this impact on multiple fluency variables, especially in mildest aphasia (e.g., Stark et al., 2023). Recently, a new clinical cut-off of aphasia has been suggested for the Western Aphasia Battery-Revised Aphasia Quotient [AQ] of <96.7, raised from <93.8 (Senthilkumar et al., 2026, AJSLP), coupled from another study suggesting a score of AQ of <97 (Garden et al., 2026, Aphasiology)

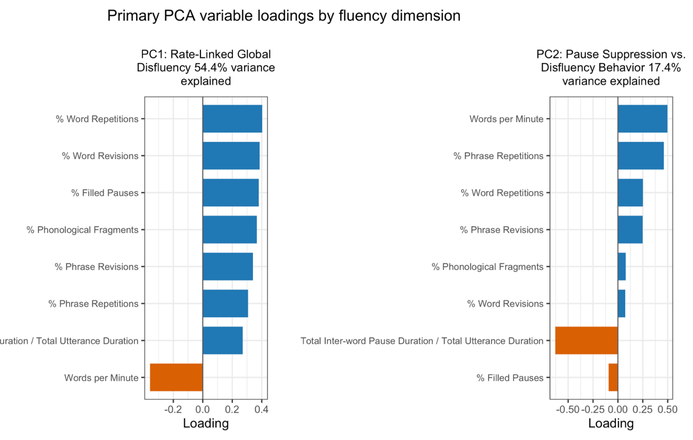
Research Questions

1. **Research Question 1: Do automated measures of discourse fluency detect residual aphasia-related disruption in individuals at the mildest end of the aphasia continuum, taking into account the known differences in discourse tasks?**
2. **Research Question 2: Do individuals who score above the recently proposed WAB-R AQ clinical aphasia cutoff of 96.7 nevertheless continue to show measurable fluency differences from neurologically healthy adults?**

Methodology and Statistical Analysis Plan

1. Test and retest design, tasks always elicited in the same order each day (see QR code), 7 +/- 10 days testing period (explained in NEURAL2 publication: Stark et al., JSLHR, 2025)
2. Ran **FLUCALC command in CLAN** to extract automatically-tagged fluency variables from transcripts that are time-locked at the word level to audio/video files: (1) % filled pauses, (2) % phonological fragments, (3) % phrase repetitions, (4) % phrase revisions, (5) % word repetitions, (6) % word revisions, (7) total inter-word pause duration divided by total utterance duration in seconds, and (8) words per minute (ignoring revisions and repetitions)
3. Evaluated each variable across six different tasks, each divided into three genres: **Picture-based tasks** [Cat Rescue, Broken Window, and Refused Umbrella: CHA: *M* = 122.71 words, SD = 59.95; Latent Aphasia: *M* = 103.83, SD = 67.42]; **Narrative/no-picture tasks** [Cinderella retell and Neutral Cue autobiographical narrative: CHA: *M* = 893.51, SD = 528.16; Latent Aphasia: *M* = 726.38, SD = 667.21]; and the **Procedural task [**how to make a sandwich: CHA: *M* = 127.86, SD = 69.93; Latent Aphasia: *M* = 126.63, SD = 118.56].
4. Examined the extent to which the fluency variables loaded onto unrotated principal components in order to reduce dimensionality of the dataset. **Two PCs were determined to meet Kaiser qualifications for retainment (Eigenvalues >1) and explained >71.8% variance together.**
5. Reliability (using ICC and Minimal Detectable Chance) of the PCs were evaluated for each genre over time, to examine whether the test and retest timepoints tended to associate with vastly different fluency behaviors.
6. Mixed-effects models tested whether PCA-derived fluency dimensions differed by clinical group across discourse genres (secondary: by task), while accounting for repeated samples from each participant, test/retest timepoint, age, education, and total word count.



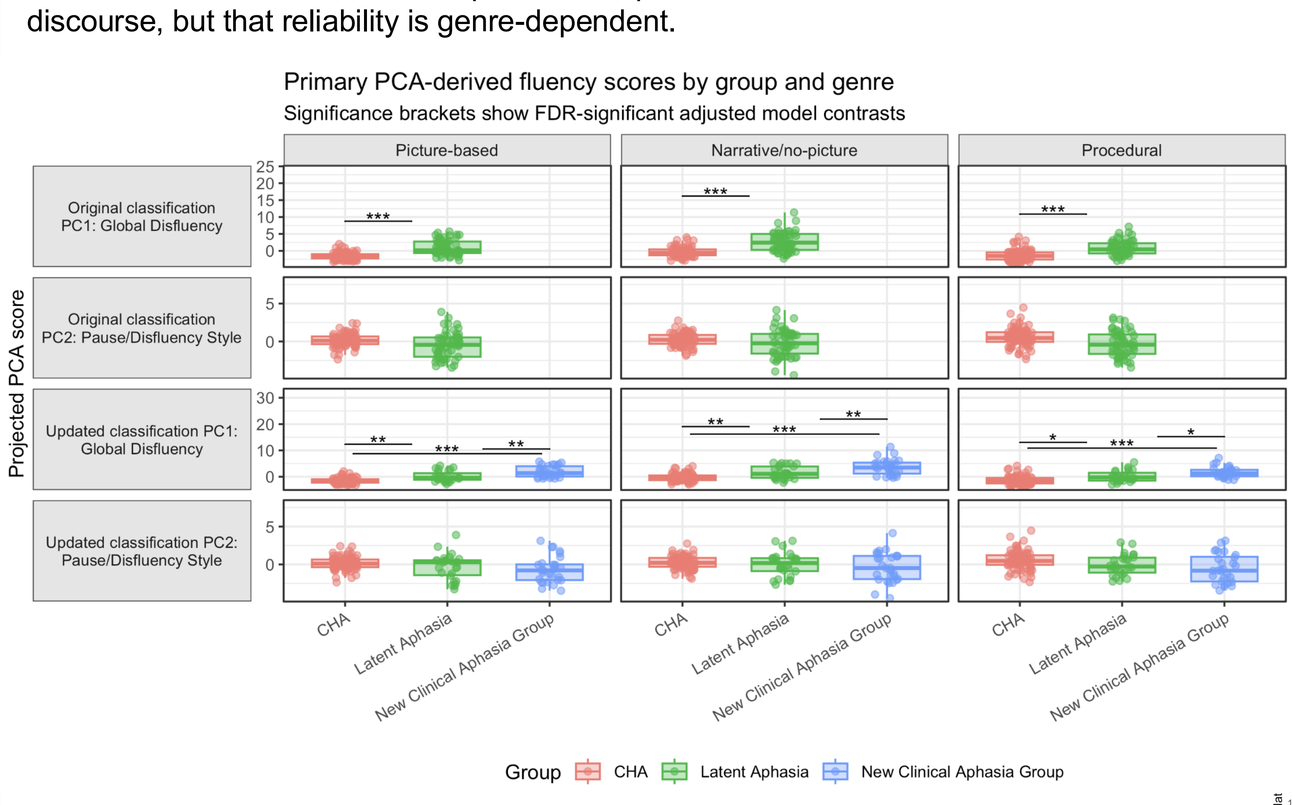


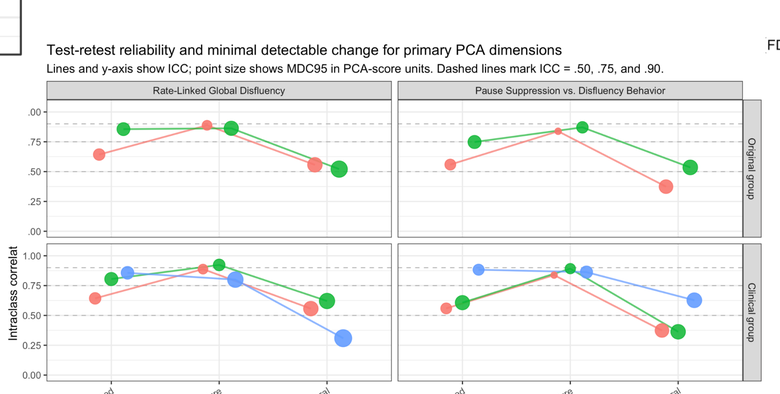
Group	N	Age, mean (SD)	Educ. mean (SD)	Female, n (%)	MoCA, mean (SD)	WAB-R AQ, mean (SD)	WAB-R Fluency, mean (SD)
Cog. Healthy adult (CHA)	38	66.7 (7.9)	16.8 (2.8)	24 (63.2%)	27.4 (2.2)	—	—
Latent Aphasia* ((AQ>96.7)	13	58.3 (10.8)	17.2 (3.2)	8 (61.5%)	—	98.28 (0.98)	9.77 (0.44)
New Clinical Aphasia Group* (AQ>93.8, <96.7)	14	55.3 (13.4)	17.6 (3.1)	7 (50.0%)	—	95.19 (0.89)	9.07 (0.27)

** = combined, these make the "original" latent group shown in results, i.e., n = 27*

Results

- ✓**Research Question 1: Genre mixed-effects models showed robust PC1 group differences across all genres.** In the original two-group model, CHA had significantly lower Rate-Linked Global Disfluency scores than Latent Aphasia in narrative/no-picture discourse, *b* = −2.82, SE = 0.48, *t*(91.37) = −5.87, *p* < .001; picture-based discourse, *b* = −2.37, SE = 0.48, *t*(89.38) = −4.96, *p* < .001; and procedural discourse, *b* = −2.13, SE = 0.48, *t*(89.29) = −4.47, *p* < .001.
- ✓**Research Question 2: Global disfluency is sensitive even with the new cut-off.** In the updated three-group model, PC1 showed a graded pattern across genres, with CHA < Latent Aphasia < New Clinical Aphasia Group; for example, in narrative/no-picture discourse, CHA differed from Latent Aphasia, *b* = −1.83, SE = 0.55, *t*(97.95) = −3.35, *p* = .001, CHA differed from New Clinical Aphasia, *b* = −3.89, SE = 0.56, *t*(98.38) = −6.98, *p* < .001, and Latent Aphasia differed from New Clinical Aphasia, *b* = −2.07, SE = 0.63, *t*(101.53) = −3.29, *p* = .001; by contrast, PC2 contrasts were not reliably significant after correction, indicating that group differentiation was specific to Rate-Linked Global Disfluency rather than the pause-suppression/disfluency-style dimension.
- ✓**FDR-corrected covariate tests indicated that no covariate significantly improved model fit after correction**. Education showed the largest uncorrected contribution for PC1 in the clinical-group model, $\chi^2(1) = 4.81$, p = .028, FDR-adjusted q = .339, and original-group model, $\chi^2(1) = 3.71$, p = .054, q = .324. All other covariate tests were also nonsignificant after FDR correction, including age and total word count for PC1, χ^2s = 0.84–1.45, qs ≥ .599, and all PC2 covariates, χ^2s = 0.07–1.07, qs ≥ .599. Thus, demographic and sample-length covariates did not significantly account for the PCA-derived fluency effects after multiple-comparison correction.
- ✓**Test-retest reliability supported the stability of the PCA-derived fluency dimensions, but reliability varied by genre and group.** At the participant level, PC1, Rate-Linked Global Disfluency, showed excellent reliability, ICC = .91, and PC2, Pause Suppression vs. Disfluency Behavior, showed good reliability, ICC = .84, across 65 participants. Genre-specific estimates showed the strongest reliability for narrative/no-picture discourse: for CHA, PC1 ICC = .89 and PC2 ICC = .84; for Latent Aphasia under the updated clinical grouping, PC1 ICC = .92 and PC2 ICC = .89; and for the New Clinical Aphasia Group, PC1 ICC = .80 and PC2 ICC = .86. Picture-based discourse showed moderate-to-good reliability, with PC1 ICCs ranging from .64 to .86 and PC2 ICCs ranging from .56 to .88 across groups. Procedural discourse was less stable, particularly for PC2 in CHA, ICC = .37, and Latent Aphasia, ICC = .36, and for PC1 in the New Clinical Aphasia Group, ICC = .31. Overall, these results indicate that PCA-derived fluency dimensions are reproducible, especially in narrative/no-picture discourse, but that reliability is genre-dependent.





SCAN ME FOR STATISTICS & SUCH

SCAN ME FOR PROTOCOL & DATA

Discussion

- ✓**Automated fluency metrics detected residual aphasia-related disruption at the mildest end of the aphasia continuum.** Rate-Linked Global Disfluency distinguished Latent Aphasia from CHA across narrative/no-picture, picture-based, and procedural discourse, supporting fluency as a sensitive marker of subtle post-stroke language difficulty.
- ✓**Fluency remained sensitive even after applying the newer WAB-R AQ cutoff.** Individuals above the proposed 96.7 cutoff still showed higher global disfluency than CHA, suggesting that discourse fluency captures residual language vulnerability not fully represented by standardized aphasia classification.
- ✓**The effect was specific to global fluency efficiency, not all fluency dimensions.** PC1 showed robust and graded group differences, whereas PC2 did not reliably distinguish groups after correction, indicating that clinical differentiation was driven by rate-linked disfluency rather than pause/disfluency style.
- ✓**Discourse genre mattered.** Group differences were present across genres, but reliability was strongest for narrative/no-picture discourse, suggesting that longer, internally generated discourse may provide the most stable context for detecting subtle fluency impairment.
- ✓**Findings were not explained by age, education, or sample length.** FDR-corrected covariate tests showed that demographic variables and total word count did not significantly account for the PCA-derived fluency effects.
- ✓**Clinical implication:** People who score above aphasia cutoffs may still show measurable discourse-level impairment; automated fluency analysis may help identify individuals whose communication difficulties are subtle but clinically meaningful.

Supported by NIH R01 R01DC008524 (PI MacWhinney, Co-I Stark)