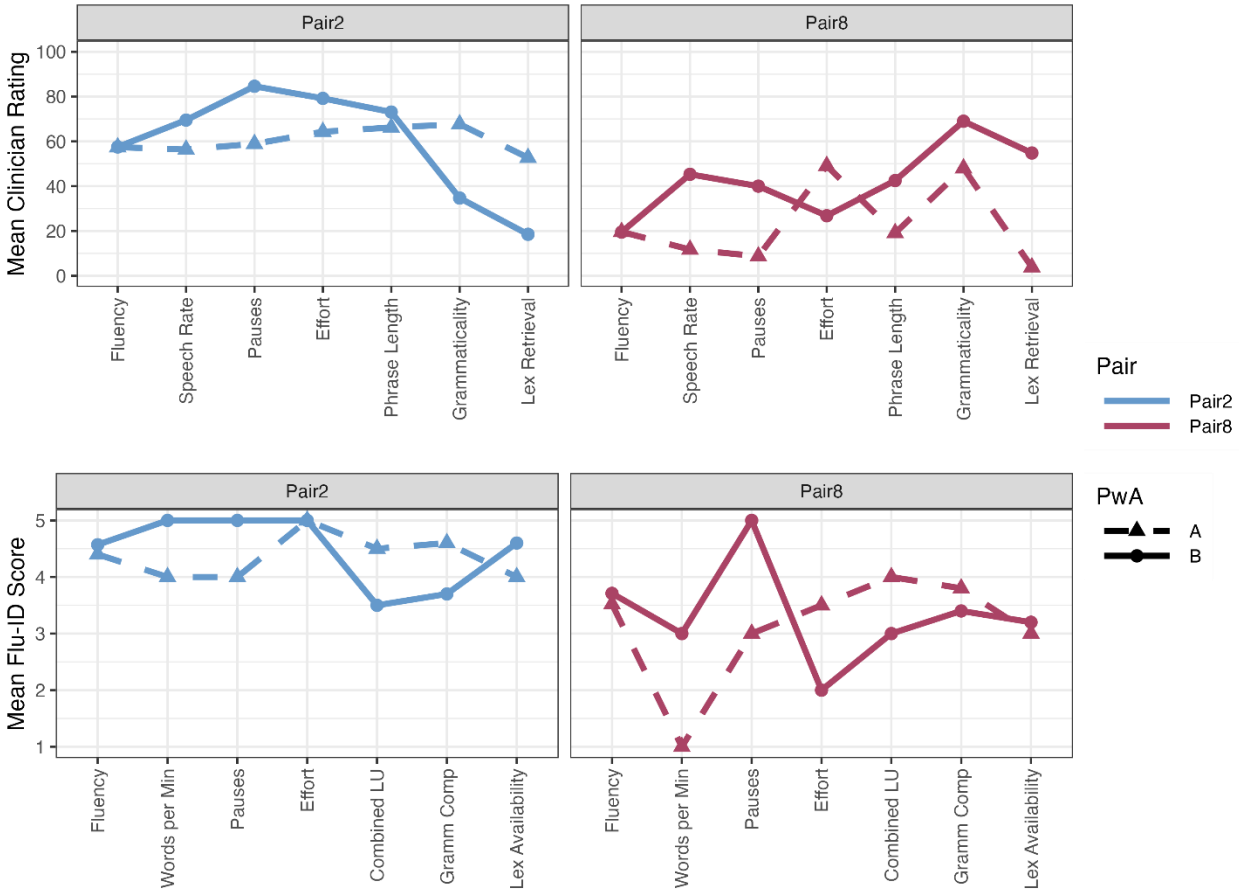
Supplemental Material S8:
Comparison of Flu-ID scores with clinical ratings for two sample pairs of PwA

To further illustrate the similarities and differences between the subjective clinical ratings and the objective measures of the Flu-ID, we compared the profiles generated by each approach for two pairs of speakers, Pair 2 and Pair 8, illustrated below.



The two pairs were selected to represent high and low overall dimension scores, respectively, and for their relatively large divergences in some of the fluency dimensions. For example, both PwA in Pair 2 were rated highly on the OVERALL FLUENCY rating and obtained high Overall Fluency domain scores above, left). Clinician ratings for PwA 2A and 2B diverged on SPEECH RATE and PAUSING, with PwA 2B rated higher, and these differences were also reflected in the corresponding Flu-ID dimension scores. However, the Flu-ID scores did not diverge on the Effort dimension as they did for clinician ratings,

which may reflect either a ceiling effect in the Flu-ID scores, or the influence of other perceived dimensions, such as SPEECH RATE and PAUSING, on the EFFORT rating by clinicians. The ratings of LEXICAL RETRIEVAL by clinicians also did not correspond to the domain scores for this pair of PwA; clinicians rated PwA 2B much lower than PwA 2A, but the Lexical Availability domain score for PwA 2B is slightly higher than PwA 2A. Examining the individual dimension scores revealed that PwA 2B produced a lot of empty speech, accounting for an appropriately low LEXICAL RETRIEVAL rating. However, in the Flu-ID, this dimension does not contribute to the LA domain score, because the purpose of the Flu-ID is to reflect not simply lexical availability, but the extent to which lexical availability contributes to fluency.

Both PwA in Pair 8 (above, right) were rated quite low on overall fluency by clinicians relative to their dimension ratings; however, their Overall Fluency domain scores more closely reflect the corresponding dimension scores. This may indicate that clinicians were more strongly influenced by the lower ratings of SPEECH RATE and PAUSING for PwA 8B, and EFFORT for PwA 8A, whereas the Overall Fluency domain is calculated as a weighted average in the Flu-ID. As with PwA 2A and 2B, the divergence in SPEECH RATE and PAUSING in clinician ratings was borne out by the dimension scores on the Flu-ID. However, the relative performance of the 2 PwA on ratings of PHRASE LENGTH, GRAMMATICALITY, and LEXICAL RETRIEVAL did not match their corresponding dimension/domain scores. The output of PwA 8A was characterized by frequent repairs, while PwA 8B paused more frequently and spoke more slowly. It appears that frequent pausing created the impression of shorter utterances and reduced grammatical competence in PwA 8B, whereas the Flu-ID relies primarily on syntactic criteria to measure utterance length. PwA 8A made more grammatical errors but this was offset by a faster speech rate, resulting in similar Grammatical Competence scores. The lower LEXICAL RETRIEVAL rating for Sample PwA 8B appears to be influenced by the high number of repairs. Overall, ratings and measured dimensions show some correspondence, but also important differences—this is discussed further in the main paper's Discussion section.

Supplemental material, Gordon & Clough, "The Flu-ID: A New Evidence-Based Method of Assessing Fluency in Aphasia," ***AJSLP***, https://doi.org/10.1044/2024_AJSLP-23-00424