

Supplemental Table S2. Summary of test-retest results across tasks, stratified by sample length and aphasia severity.

Given that sample length (in both subject groups) and aphasia severity (Aphasia group only) may affect discourse measure reliability, here we stratify ICC, Spearman's rho, and Wilcoxon's signed rank test *p*-value by each factor. Sample length is the average sample length across all discourse tasks (*n* = 12 in each sample length group for NBD group (Med = 437.8 words); *n* = 11 in long sample and *n* = 12 in short sample length group for Aphasia group (Med = 229 words)).

Proxy	Group	Measure	Long sample length	Short sample length	Mild or Latent Severity (<i>n</i> = 14)	Moderate or Severe Severity (<i>n</i> = 9)	Latent Aphasia (<i>n</i> = 6)	Clinical Aphasia (<i>n</i> = 17)
			ICC (95% CI) [Spearman's rho] Wilcoxon <i>p</i> -value	ICC (95% CI) [Spearman's rho] Wilcoxon <i>p</i> -value	ICC (95% CI) [Spearman's rho] Wilcoxon <i>p</i> -value	ICC (95% CI) [Spearman's rho] Wilcoxon <i>p</i> -value	ICC (95% CI) [Spearman's rho] Wilcoxon <i>p</i> -value	ICC (95% CI) [Spearman's rho] Wilcoxon <i>p</i> -value
Lexical and informativeness	NBD	%CIU	0.94 (0.8, 0.98) [0.81], <i>p</i> = .03	0.61 (0.05, 0.87) [0.71], <i>p</i> = .85	N/A	N/A	N/A	N/A
	Aphasia		0.91 (0.72, 0.98) [0.75], <i>p</i> = .32	0.94 (0.73, 0.98) [0.94], <i>p</i> = .02*	0.89 (0.70, 0.96) [0.79], <i>p</i> = .22	0.91 (0.45, 0.98) [0.87], <i>p</i> = .04*	0.62 (-0.29, 0.94) [0.49], <i>p</i> = .56	0.95 (0.83, 0.98) [0.94], <i>p</i> = .02*
	NBD	PI Density	0.28 (-0.35, 0.72) [0.34], <i>p</i> = .53	0.15 (-0.50, 0.66) [0.30], <i>p</i> = .91	N/A	N/A	N/A	N/A
	Aphasia		0.76 (0.29, 0.93) [0.78], <i>p</i> = .07	0.93 (0.78, 0.98) [0.86], <i>p</i> = .30	0.92 (0.76, 0.97) [0.86], <i>p</i> = .19	0.93 (0.74, 0.98) [0.97], <i>p</i> = .20	0.65 (-0.17, 0.94) [0.60], <i>p</i> = .44	0.94 (0.84, 0.98) [0.90], <i>p</i> = .07
	NBD	TTR	0.43 (-0.19, 0.80) [0.20], <i>p</i> = .68	0.58 (-0.05, 0.87) [0.78], <i>p</i> = .01	N/A	N/A	N/A	N/A
	Aphasia		0.72 (0.16, 0.92) [0.66], <i>p</i> = .04*	0.88 (0.62, 0.97) [0.88], <i>p</i> = .13	0.83 (0.56, 0.94) [0.90], <i>p</i> = .36	0.94 (0.78, 0.99) [0.93], <i>p</i> = .30	0.77 (-0.03, 0.97) [0.37], <i>p</i> = .06	0.92 (0.79, 0.97) [0.93], <i>p</i> = .26
	NBD	Tokens	0.57 (0.07, 0.85) [0.29], <i>p</i> = .11	0.70 (-0.01, 0.92) [0.78], <i>p</i> = .005	N/A	N/A	N/A	N/A
	Aphasia		0.57 (-0.06, 0.87) [0.76], <i>p</i> = .002*^	0.93 (0.77, 0.98) [0.92], <i>p</i> = .27	0.72 (0.24, 0.90) [0.79], <i>p</i> = .004*^	0.96 (0.82, 0.99) [0.95], <i>p</i> = .05	0.52 (-0.15, 0.91) [0.66], <i>p</i> = .03*	0.93 (0.77, 0.98) [0.97], <i>p</i> = .03*

Proxy	Group	Measure	Long sample length	Short sample length	Mild or Latent Severity ($n = 14$)	Moderate or Severe Severity ($n = 9$)	Latent Aphasia ($n = 6$)	Clinical Aphasia ($n = 17$)
			ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value
Fluency / efficiency	NBD	CIUs/min	0.91 (0.70, 0.97) [0.90], $p = .21$	0.81 (0.46, 0.94) [0.72], $p = .68$	N/A	N/A	N/A	N/A
	Aphasia		0.90 (0.66, 0.97) [0.87], $p = .83$	0.96 (0.88, 0.99) [0.92], $p = .20$	0.94 (0.83, 0.98) [0.91], $p = .81$	0.97 (0.86, 0.99) [0.95], $p = .50$	0.90 (0.45, 0.99) [0.83], $p = .84$	0.95 (0.88, 0.98) [0.94], $p = .35$
	NBD	SpeakingSecs	0.63 (0.14, 0.88) [0.34], $p = .09$	0.73 (0.22, 0.92) [0.70], $p = .04$	N/A	N/A	N/A	N/A
	Aphasia		0.57 (0.02, 0.86) [0.78], $p = .02^*$	0.91 (0.72, 0.97) [0.84], $p = .47$	0.63 (0.19, 0.86) [0.69], $p = .08$	0.95 (0.79, 0.99) [0.98], $p = .25$	0.51 (-0.16, 0.91) [0.54], $p = .03^*$	0.87 (0.68, 0.95) [0.92], $p = .21$
	NBD	WPM	0.84 (0.49, 0.95) [0.85], $p = .08$	0.78 (0.40, 0.93) [0.78], $p = .68$	N/A	N/A	N/A	N/A
	Aphasia		0.92 (0.73, 0.98) [0.86], $p = .37$	0.98 (0.93, 0.99) [0.95], $p = .18$	0.94 (0.84, 0.98) [0.94], $p = .24$	0.99 (0.97, 0.999) [0.98], $p = .30$	0.90 (0.49, 0.99) [0.77], $p = .69$	0.97 (0.92, 0.99) [0.96], $p = .13$
Syntactic	NBD	Noun/verb	0.50 (-0.11, 0.83) [0.39], $p = .73$	0.27 (-0.36, 0.72) [0.31], $p = .85$	N/A	N/A	N/A	N/A
	Aphasia		0.86 (0.58, 0.96) [0.86], $p = .90$	0.50 (-0.02, 0.82) [0.58], $p = .15$	0.79 (0.48, 0.93) [0.76], $p = .24$	0.49 (-0.19, 0.86) [0.67], $p = .73$	0.67 (-0.29, 0.95) [0.60], $p = .56$	0.59 (0.19, 0.83) [0.82], $p = .33$
	NBD	Open/closed	0.12 (-0.54, 0.65) [0.27], $p = .57$	0.57 (0.01, 0.86) [0.47], $p = .02$	N/A	N/A	N/A	N/A
	Aphasia		0.82 (0.45, 0.95) [0.72], $p = .58$	0.63 (0.1, 0.88) [0.90], $p = .91$	0.97 (0.90, 0.99) [0.84], $p = .67$	0.48 (-0.29, 0.86) [0.70], $p = .65$	0.86 (0.26, 0.98) [0.89], $p > .99$	0.69 (0.32, 0.88) [0.87], $p = .96$
	NBD	MLU	0.62 (0.11, 0.87) [0.66], $p = .42$	0.64 (0.15, 0.88) [0.46], $p = .57$	N/A	N/A	N/A	N/A
	Aphasia		0.71 (0.26, 0.91) [0.39], $p = .32$	0.96 (0.86, 0.99) [0.96], $p = .13$	0.88 (0.67, 0.96) [0.67], $p = .81$	0.92 (0.70, 0.98) [0.93], $p > .99$	0.69 (-0.01, 0.95) [0.09], $p = .56$	0.95 (0.86, 0.98) [0.89], $p = .38$

Proxy	Group	Measure	Long sample length	Short sample length	Mild or Latent Severity ($n = 14$)	Moderate or Severe Severity ($n = 9$)	Latent Aphasia ($n = 6$)	Clinical Aphasia ($n = 17$)
			ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value	ICC (95% CI) [Spearman's rho] Wilcoxon p -value
	NBD	VerbUtt	0.43 (-0.20, 0.80) [0.59], $p = .38$	0.65 (0.19, 0.88) [0.27], $p = .20$	N/A	N/A	N/A	N/A
	Aphasia		0.35 (-0.19, 0.76) [0.06], $p = .12$	0.93 (0.73, 0.98) [0.95], $p = .06$	0.77 (0.42, 0.92) [0.55], $p = .86$	0.93 (0.72, 0.98) [0.93], $p > .99$	0.06 (-0.72, 0.78) [-0.20], $p = .44$	0.92 (0.81, 0.97) [0.88], $p = .43$

Koo and Li (2016) gives the following suggestion for interpreting intraclass correlation coefficient (ICC), including confidence intervals: below 0.50 = poor; between 0.50 and 0.75 = moderate; between 0.75 and 0.90 = good; and above 0.90 = excellent.

* = significant; ^ = significant after Bonferroni correction (11 row-wise within group corrections; new $p < .0045$)