

Supplemental Table S8. Summary of test-retest results for the Refused Umbrella task (describing a picture sequence).

Koo and Li (2016) gives the following suggestion for interpreting ICC: below 0.50 = poor; between 0.50 and 0.75 = moderate; between 0.75 and 0.90 = good; and above 0.90 = excellent. Lin's concordance correlation coefficient (CCC) is given in cases where ICC is poor, to identify if this improves the estimate. If it does improve the estimate, it suggests that test-retest the low ICC is due to lack of spread (i.e., lack of true intra-group variability).

Primary Proxy	Measure	Group	ICC (CCC)	95% ICC CI (95% CCC CI)	Koo & Li (2016) ICC Quality (CI Quality)	Spearman's rho (p -value)	Systematic difference	SEM / MDC90
Lexical and informativeness	%CIU	NBD	0.79	0.57, 0.90	Good (Moderate – Exc.)	0.08 ($p = .72$)	$V = 125, p = .49$	0.06
		Aphasia	0.82	0.62, 0.92	Good (Moderate – Exc.)	0.67 ($p = .001$) [^]	$V = 109, p = .39$	0.10 / 0.23
	PI Density	NBD	0.35 (0.34)	-0.05, 0.65 (-0.05, 0.64)	Poor (Poor – Moderate)	0.27 ($p = .20$)	$V = 111, p = .27$	0.04
		Aphasia	0.90	0.77, 0.95	Excellent (Good – Exc.)	0.73 ($p < .0001$) [^]	$V = 105, p = .33$	0.04 / 0.10
	TTR	NBD	0.72	0.45, 0.87	Moderate (Poor – Good)	0.74 ($p < .0001$) [^]	$V = 151, p = .70$	0.05
		Aphasia	0.72	0.45, 0.87	Moderate (Poor – Good)	0.63 ($p = .001$) [^]	$V = 105, p = .32$	0.06 / 0.15
	Tokens	NBD	0.78	0.50, 0.91	Good (Moderate – Exc.)	0.84 ($p < .0001$) [^]	$V = 56.5, p = .008^*$	26.81
		Aphasia	0.69	0.38, 0.86	Moderate (Poor – Good)	0.87 ($p < .0001$) [^]	$V = 58.5, p = .02^*$	37.24 / 86.91
	CIUs / min	NBD	0.74	0.49, 0.88	Moderate (Poor – Good)	0.57 ($p = .004$) [^]	$V = 173, p = .53$	18.95
		Aphasia	0.90	0.79, 0.96	Excellent (Good – Exc.)	0.89 ($p < .0001$) [^]	$V = 82, p = .09$	12.52 / 29.21
Fluency / efficiency	SpeakingSecs	NBD	0.76	0.42, 0.90	Good (Poor – Exc.)	0.82 ($p < .0001$) [^]	$V = 46.5, p = .006^*$	10.92
		Aphasia	0.56	0.21, 0.78	Moderate (Poor – Good)	0.79 ($p < .0001$) [^]	$V = 90, p = .39$	29.39 / 68.57
	WPM	NBD	0.65	0.35, 0.83	Moderate (Poor – Good)	0.58 ($p = .003$) [^]	$V = 183, p = .36$	21.15
		Aphasia	0.89	0.76, 0.95	Good (Good – Exc.)	0.84 ($p < .0001$) [^]	$V = 83, p = .10$	13.68 / 31.91
Syntactic	MLU	NBD	0.46	0.08, 0.72	Poor (Poor – Moderate)	0.56 ($p = .005$) [^]	$V = 67.5, p = .02^*$	1.26
		Aphasia	0.81	0.62, 0.92	Good (Moderate – Exc.)	0.65 ($p = .001$) [^]	$V = 104, p = .32$	1.11 / 2.58
	Noun/verb	NBD	0.25	-0.14, 0.58	Poor (Poor – Moderate)	0.18 ($p = .41$)	$V = 107, p = .23$	0.13

Primary Proxy	Measure	Group	ICC (CCC)	95% ICC CI (95% CCC CI)	Koo & Li (2016) ICC Quality (CI Quality)	Spearman's rho (<i>p</i> -value)	Systematic difference	SEM / MDC90
			(0.24)	(-0.14, 0.56)	CCC remains poor			
		Aphasia	0.07	-0.37, 0.49	Poor (Poor)	0.04 (<i>p</i> = .87)	V = 113, <i>p</i> = .95	0.67 /1.57
			(0.07)	(-0.30, 0.42)	CCC remains poor			
	Open/closed	NBD	0.08	-0.35, 0.47	Poor (Poor)	0.36 (<i>p</i> = .08)	V = 169, <i>p</i> = .60	0.22
			(0.07)	(-0.30, 0.43)	CCC remains poor			
		Aphasia	0.18	-0.24, 0.55	Poor (Poor – Moderate)	0.45 (<i>p</i> = .04)*	V = 128, <i>p</i> = .97	0.51 / 1.20
			(0.17)	(-0.14, 0.45)	CCC remains poor			
	VerbUtt	NBD	0.29	-0.10, 0.61	Poor (Poor – Moderate)	0.32 (<i>p</i> = .13)	V = 91, <i>p</i> = .16	0.25
			(0.28)	(-0.09, 0.59)	CCC remains poor			
		Aphasia	0.75	0.51, 0.89	Good (Moderate – Good)	0.65 (<i>p</i> = .001)*^	V = 98.5, <i>p</i> = .37	0.33 / 0.77

CI = confidence interval; %CIU = Percentage of correct information units; CIUs/min = correct information units per minute; MLU = mean length of utterance (in words); VerbUtt = verbs per utterance; Noun/verb = noun-to-verb ratio; Open/closed = open-to-closed class word ratio; SpeakingSecs = speaking duration in seconds; PI Density = propositional idea density; TTR = type-token ratio; WPM = words per minute.

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