

Supplemental Table S4. Summary of test-retest results for the Cat Rescue task (describing a single picture).

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 ("exc"). 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 (<i>p</i> -value)	Systematic difference	SEM / MDC90
Lexical and informativeness	%CIU	NBD	0.68	0.39, 0.85	Moderate (Poor – Good)	0.37 (.07)	$V = 178, p = .44$	0.06
		Aphasia	0.76	0.52, 0.89	Good (Moderate – Good)	0.77 ($p < .0001$)*^	$V = 108, p = .38$	0.10 / 0.24
	PI Density	NBD	0.41 (0.40)	0.001, 0.69 (0.002, 0.68)	Poor (Poor – Moderate) CCC remains poor	0.30 ($p = .15$)	$V = 145.5, p = .91$	0.03
		Aphasia	0.68	0.38, 0.85	Moderate (Poor – Good)	0.63 (.001)*^	$V = 108.5, p = .38$	0.04 / 0.10
	TTR	NBD	0.35 (0.34)	-0.06, 0.65 (-0.06, 0.64)	Poor (Poor – Moderate) CCC remains poor	0.32 (.12)	$V = 180, p = .41$ $V = 138, p > .99$	0.07
		Aphasia	0.64	0.32, 0.83	Moderate (Poor – Good)	0.63 (.001)*^		0.08 / 0.19
	Tokens	NBD	0.69	0.41, 0.85	Moderate (Poor – Good)	0.63 (.001)*^	$V = 110, p = .26$	41.51
		Aphasia	0.79	0.54, 0.90	Good (Moderate – Exc.)	0.88 ($p < .0001$)*^	$V = 84.5, p = .11$	33.29 / 77.68
Fluency / efficiency	CIUs / min	NBD	0.82	0.63, 0.92	Good (Moderate – Exc.)	0.81 ($p < .0001$)*^	$V = 170, p = .58$	14.49
		Aphasia	0.90	0.79, 0.96	Good (Good – Exc.)	0.93 ($p < .0001$)*^	$V = 178, p = .23$	11.76 / 27.44
	SpeakingSecs	NBD	0.82	0.63, 0.92	Good (Moderate – Exc.)	0.85 ($p < .0001$)*^	$V = 84, p = .10$	14.49
		Aphasia	0.63	0.30, 0.83	Moderate (Poor – Good)	0.71 ($p = .0001$)*^	$V = 61, p = .04^*$	23.67 / 55.24
	WPM	NBD	0.69	0.40, 0.85	Moderate (Poor – Good)	0.63 (.001)*^	$V = 172, p = .55$	17.92
		Aphasia	0.91	0.79, 0.96	Excellent (Good – Exc.)	0.91 ($p < .0001$)*^	$V = 196, p = .08$	12.28 / 28.67
Syntactic	MLU	NBD	0.66	0.36, 0.84	Moderate (Poor – Good)	0.72 ($p < .0001$)*^	$V = 127, p = .53$	1.74
		Aphasia	0.85	0.68, 0.93	Good (Moderate – Exc.)	0.86 ($p < .0001$)*^	$V = 163, p = .46$	1.35 / 3.15
	Noun/verb	NBD	0.23 (0.23)	-0.20, 0.58 (-0.17, 0.56)	Poor (Poor – Moderate) CCC remains poor	0.37 (.08)	$V = 123, p = .46$	0.24
		Aphasia	0.68	0.37, 0.85	Moderate (Poor – Good)	0.66 (.0006)*^	$V = 78, p = .07$	0.52 / 1.21

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
	Open/closed	NBD	0.20 (0.20)	-0.16, 0.54 (-0.17, 0.51)	Poor (Poor – Moderate) CCC remains poor	0.21 (.32)	$V = 212, p = .08$	0.09
		Aphasia	0.78	0.56, 0.90	Good (Moderate – Exc.)	0.51 ($p = .02$)*	$V = 160, p = .29$	0.18 / 0.43
	VerbUtt	NBD	0.73	0.46, 0.87	Moderate (Poor – Good)	0.57 ($p = .004$)*^	$V = 167, p = .64$	0.35
		Aphasia	0.79	0.57, 0.90	Good (Moderate – Exc.)	0.74 ($p < .0001$)*^	$V = 163, p = .46$	0.31 / 0.71

CCC = Concordance correlation coefficient; 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; MDC90 = Minimal detectable change at 90% confidence, only calculated for aphasia group; SEM = standard error of measurement.

* = significant; ^ = significant after Bonferroni correction (11 row-wise within group corrections; new $p < .004$).