

Supplemental Table S5. Summary of test-retest results for the Cinderella task (retelling a fictional narrative).

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 (<i>p</i> -value)	Systematic difference	SEM / MDC90
Lexical and informativeness	%CIU	NBD	0.64	0.34, 0.83	Moderate (Poor – Good)	0.64 (<i>p</i> = .001)*^	<i>V</i> = 77, <i>p</i> = .04*	0.03
		Aphasia	0.87	0.63, 0.95	Good (Moderate – Exc.)	0.70 (<i>p</i> = .0002)*^	<i>V</i> = 40, <i>p</i> = .002*^	0.09 / 0.21
	PI Density	NBD	0.25 (0.24)	-0.18, 0.59 (-0.17, 0.58)	Poor (Poor – Good) CCC remains poor	0.12 (<i>p</i> = .57)	<i>V</i> = 167.6, <i>p</i> = .63	0.02
		Aphasia	0.79	0.57, 0.90	Moderate (Moderate – Exc.)	0.79 (<i>p</i> < .0001)*^	<i>V</i> = 108.5, <i>p</i> = .38	0.06
	TTR	NBD	0.84	0.62, 0.93	Good (Moderate – Exc.)	0.85 (<i>p</i> < .0001)*^	<i>V</i> = 235, <i>p</i> = .01*	0.03
		Aphasia	0.88	0.64, 0.95	Good (Moderate – Exc.)	0.90 (<i>p</i> < .0001)*^	<i>V</i> = 206, <i>p</i> = .01*	0.05 / 0.11
	Tokens	NBD	0.83	0.60, 0.92	Good (Moderate – Exc.)	0.89 (<i>p</i> < .0001)*^	<i>V</i> = 68, <i>p</i> = .02*	121.26
		Aphasia	0.86	0.54, 0.95	Good (Moderate – Exc.)	0.89 (<i>p</i> < .0001)*^	<i>V</i> = 30.5, <i>p</i> = .001*^	99.36 / 231.85
Fluency / efficiency	CIUs / min	NBD	0.64	0.32, 0.82	Moderate (Poor – Good)	0.68 (<i>p</i> = .0003)*^	<i>V</i> = 160, <i>p</i> = .79	12.77
		Aphasia	0.66	0.35, 0.84	Moderate (Poor – Good)	0.66 (<i>p</i> = .0006)*^	<i>V</i> = 93, <i>p</i> = .18	19.13 / 44.65
	SpeakingSecs	NBD	0.80	0.56, 0.91	Good (Moderate – Exc.)	0.81 (<i>p</i> < .0001)*^	<i>V</i> = 74.5, <i>p</i> = .03*	61.47
		Aphasia	0.81	0.58, 0.92	Good (Moderate – Exc.)	0.81 (<i>p</i> < .0001)*^	<i>V</i> = 76, <i>p</i> = .06	75.37 / 175.87
	WPM	NBD	0.58	0.24, 0.80	Moderate (Poor – Good)	0.67 (<i>p</i> = .0005)*^	<i>V</i> = 138, <i>p</i> = .75	14.56
		Aphasia	0.95	0.82, 0.98	Excellent (Good – Exc.)	0.94 (<i>p</i> < .0001)*^	<i>V</i> = 53, <i>p</i> = .008*^	7.93 / 18.50
Syntactic	MLU	NBD	0.58	0.23, 0.80	Moderate (Poor – Good)	0.55 (<i>p</i> = .007)*^	<i>V</i> = 154, <i>p</i> = .92	1.26
		Aphasia	0.92	0.83, 0.97	Excellent (Good – Exc.)	0.80 (<i>p</i> < .0001)*^	<i>V</i> = 109, <i>p</i> = .39	0.96 / 2.25
	Noun/verb	NBD	0.45 (0.44)	0.07, 0.72 (0.07, 0.71)	Poor (Poor – Moderate) CCC remains poor	0.52 (<i>p</i> = .009)*^	<i>V</i> = 102, <i>p</i> = .18	0.14
		Aphasia	0.70	0.41, 0.86	Moderate (Poor – Good)	0.70 (<i>p</i> = .0003)*^	<i>V</i> = 110, <i>p</i> = .61	0.41 / 0.96

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	0.40 (0.39)	0, 0.69 (-0.003, 0.67)	Poor (Poor – Moderate) CCC remains poor	0.45 (<i>p</i> = .03)*	V = 124.5, <i>p</i> = .48	0.04
		Aphasia	0.60	0.24, 0.81	Moderate (Poor – Good)	0.50 (<i>p</i> = .02)*	V = 120, <i>p</i> = .85	0.18 / 0.41
	VerbUtt	NBD	0.53	0.16, 0.77	Moderate (Poor – Good)	0.46 (<i>p</i> = .02)*	V = 187, <i>p</i> = .30	0.22
		Aphasia	0.89	0.75, 0.95	Good (Good – Excellent)	0.66 (<i>p</i> = .0009)*^	V = 125, <i>p</i> = .71	0.24 / 0.55

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.

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