

Supplemental Table S7. Summary of test-retest results for the Broken Window 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	0.87	0.72, 0.94	Good (Moderate – Exc.)	0.79 ($p < .0001$)*^	$V = 137, p = .73$	0.07
		Aphasia	0.65	0.32, 0.83	Moderate (Poor – Good)	0.57 ($p = .006$)*^	$V = 124, p = .69$	0.14 / 0.32
	PI Density	NBD	0.15 (0.15)	-0.25, 0.51 (-0.23, 0.49)	Poor (Poor – Moderate) CCC remains poor	0.18 ($p = .40$)	$V = 177.5, p = .44$	0.04
		Aphasia	0.67	0.37, 0.84	Moderate (Poor – Good)	0.77 ($p < .0001$)*^	$V = 97, p = .35$	0.06 / 0.14
	TTR	NBD	0.50	0.13, 0.75	Moderate (Poor – Good)	0.60 ($p = .003$)*^	$V = 221, p = .04^*$	0.06
		Aphasia	0.84	0.66, 0.93	Good (Moderate – Exc.)	0.81 ($p < .0001$)*^	$V = 157, p = .33$	0.06 / 0.13
	Tokens	NBD	0.67	0.37, 0.84	Moderate (Poor – Good)	0.68 ($p = .0002$)*^	$V = 70, p = .04^*$	26.39
		Aphasia	0.78	0.54, 0.90	Good (Moderate – Exc.)	0.95 ($p < .0001$)*^	$V = 49, p = .02^*$	32.39 / 75.58
Fluency / efficiency	CIUs / min	NBD	0.77	0.54, 0.89	Good (Moderate – Good)	0.80 ($p < .0001$)*^	$V = 123, p = .46$	18.19
		Aphasia	0.89	0.75, 0.95	Good (Moderate – Exc.)	0.90 ($p < .0001$)*^	$V = 97, p = .22$	12.19 / 28.44
	SpeakingSecs	NBD	0.73	0.47, 0.87	Moderate (Poor – Good)	0.71 ($p = .0001$)*^	$V = 97, p = .13$	11.21
		Aphasia	0.69	0.41, 0.86	Moderate (Poor – Good)	0.75 ($p < .0001$)*^	$V = 105, p = .50$	24.49 / 57.15
	WPM	NBD	0.77	0.54, 0.89	Good (Moderate – Good)	0.63 ($p = .001$)*^	$V = 110, p = .26$	14.36
		Aphasia	0.91	0.81, 0.96	Excellent (Good – Exc.)	0.91 ($p < .0001$)*^	$V = 108, p = .38$	10.80 / 25.20
	MLU	NBD	0.22 (0.21)	-0.21, 0.57 (-0.19, 0.56)	Poor (Poor – Moderate) CCC remains poor	0.17 ($p = .43$)	$V = 137, p = .72$	2.17
		Aphasia	0.77	0.54, 0.90	Good (Moderate – Exc.)	0.72 ($p = .0001$)*^	$V = 130, p = .82$	1.63 / 3.81
Syntactic	Noun/verb	NBD	0.11 (0.10)	-0.30, 0.48 (-0.29, 0.46)	Poor (Poor) CCC remains poor	0.03 ($p = .90$)	$V = 180, p = .39$	0.34
		Aphasia	0.74	0.47, 0.88	Moderate (Poor – Good)	0.83 ($p < .0001$)*^	$V = 119, p = .92$	0.59 / 1.37
	Open/closed	NBD	0.49	0.10, 0.74	Poor (Poor – Moderate)	0.40 ($p = .05$)	$V = 233, p = .02^*$	0.15
		Aphasia	0.08	-0.32, 0.46	Poor (Poor)	0.37 ($p = .08$)	$V = 169, p = .36$	0.38 / 0.90

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
			(0.08)	(-0.22, 0.36)	CCC remains poor			
	VerbUtt	NBD	-0.002	-0.42, 0.40	Poor (Poor)	0.06 (<i>p</i> = .77)	V = 104, <i>p</i> = .31	0.34
			(-0.002)	(-0.36, 0.36)	CCC remains poor			
		Aphasia	0.76	0.51, 0.89	Good (Moderate – Good)	0.71 (<i>p</i> = .0001)*^	V = 104, <i>p</i> = .48	0.31 / 0.72

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).