

Supplemental Table S6. Summary of test-retest results for the Sandwich task (procedural "how to" 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	0.10 (0.10)	-0.33, 0.49 (-0.28, 0.45)	Poor (Poor) CCC remains poor	-0.10 (<i>p</i> = .64)	V = 149, <i>p</i> = .99	0.03
		Aphasia	0.71	0.43, 0.87	Moderate (Poor – Good)	0.74 (<i>p</i> < .0001)*^	V = 140, <i>p</i> = .96	0.12 / 0.28
	PI Density	NBD	0.11 (0.11)	-0.31, 0.49 (-0.30, 0.48)	Poor (Poor) CCC remains poor	0.08 (<i>p</i> = .70)	V = 138, <i>p</i> = .74	0.02
		Aphasia	0.84	0.66, 0.93	Good (Moderate – Exc.)	0.69 (<i>p</i> = .0002)*^	V = 145, <i>p</i> = .84	0.03 / 0.08
	TTR	NBD	0.70	0.41, 0.86	Moderate (Poor – Good)	0.64 (<i>p</i> = .0008)*^	V = 188, <i>p</i> = .28	0.03
		Aphasia	0.69	0.39, 0.86	Moderate (Poor – Good)	0.75 (<i>p</i> < .0001)*^	V = 150, <i>p</i> = .73	0.08 / 0.18
	Tokens	NBD	0.80	0.59, 0.91	Good (Moderate – Exc.)	0.82 (<i>p</i> < .0001)*^	V = 69.5, <i>p</i> = .02*	121.26
		Aphasia	0.73	0.47, 0.88	Moderate (Poor – Good)	0.89 (<i>p</i> < .0001)*^	V = 92, <i>p</i> = .27	28.52 / 66.55
Fluency / efficiency	CIUs / min	NBD	0.41 (0.40)	0.03, 0.69 (0.02, 0.68)	Poor (Poor – Moderate) CCC remains poor	0.42 (<i>p</i> = .04)*	V = 188, <i>p</i> = .29	12.77
		Aphasia	0.90	0.78, 0.96	Excellent (Good – Exc.)	0.87 (<i>p</i> < .0001)*^	V = 137, <i>p</i> = .99	13.49 / 31.48
	SpeakingSecs	NBD	0.79	0.55, 0.91	Moderate (Good – Exc.)	0.83 (<i>p</i> < .0001)*^	V = 69, <i>p</i> = .02*	61.47
		Aphasia	0.73	0.47, 0.87	Moderate (Poor – Good)	0.81 (<i>p</i> < .0001)*^	V = 120, <i>p</i> = .60	22.89 / 53.41
	WPM	NBD	0.55	0.19, 0.77	Moderate (Poor – Good)	0.57 (<i>p</i> = .004)*^	V = 175, <i>p</i> = .49	14.56
		Aphasia	0.94	0.87, 0.98	Excellent (Moderate – Exc.)	0.95 (<i>p</i> < .0001)*^	V = 87, <i>p</i> = .13	11.15 / 26.01
Syntactic	MLU	NBD	0.29 (0.28)	-0.13, 0.62 (-0.12, 0.60)	Poor (Poor – Moderate) CCC remains poor	0.23 (<i>p</i> = .27)	V = 114, <i>p</i> = .48	1.26
		Aphasia	0.81	0.60, 0.91	Good (Moderate – Exc.)	0.78 (<i>p</i> < .0001)*^	V = 145.5, <i>p</i> = .83	1.44 / 3.37
	Noun/verb	NBD	0.52	0.16, 0.76	Moderate (Poor – Good)	0.69 (<i>p</i> = .0003)*^	V = 183, <i>p</i> = .36	0.14
		Aphasia	0.22 (0.21)	-0.24, 0.59 (-0.22, 0.57)	Poor (Poor – Moderate) CCC remains poor	0.56 (<i>p</i> = .007)*^	V = 110, <i>p</i> = .61	2.01 / 4.70

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	0.33 (0.32)	-0.09, 0.64 (-0.05, 0.61)	Poor (Poor – Moderate) CCC remains poor	0.14 (<i>p</i> = .52)	V = 144, <i>p</i> = .87	0.04
		Aphasia	0.77	0.53, 0.90	Good (Moderate – Exc.)	0.70 (<i>p</i> = .0002)*^	V = 138, <i>p</i> > .99	0.30 / 0.70
	VerbUtt	NBD	0.11 (0.10)	-0.28, 0.47 (-0.27, 0.45)	Poor (Poor) CCC remains poor	-0.08 (<i>p</i> = .70)	V = 105, <i>p</i> = .21	0.22
		Aphasia	0.69	0.40, 0.86	Moderate (Poor – Good)	0.73 (<i>p</i> < .0001)*^	V = 134, <i>p</i> = .53	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).