Supplemental Material S1

4-Point GCSconv Interrater Reliability Data for Varied Sampling Methodologies

Raters	SAMPLE SIZE: 18 Conversations								SAMPLE SIZE: 6 Conversations			
	34% Utterances by Time[1]				34% Utterances by # of Utterances[2]				100% of 6 Conversations			
	Total Utt.	# Utt. Analyzed (% of data)	Coefficient (C.I.)	*p*-value	Total Utt.	# Utt. Analyzed (% of data)	Coefficient (C.I.)	*p*-value	Total Utt.	# Utt. Analyzed (% of data)	Coefficient (C.I.)	*p*-value
FA/RA1	601	206 (34.3%)	.983[3] [.921-.997]	.000	601	225 (37.4%)	.879[4] [n. a.]	.001	n. a.	n. a.	n. a.	n. a.
FA/RA2	508	177 (34.8%)	.932[3] [.711-.986]	.000	508	186 (36.6%)	.972[3] [.886-.994]	.000	n. a.	n. a.	n. a.	n. a.
FA/RA	n. a.	n. a.	n. a.	n. a.	n. a.	n. a.	n. a.	n. a.	378	378 (100%)	.912[3] [.486-.987]	.003

Note. 4-point GCSconv = 4-point Global Coherence Scale; [1] reported in manuscript; # = number; [2] to ensure 34% of each transcript was subjected to interrater reliability, RA1 needed to code an additional 19 utterances; RA2 needed to code an extra 9 utterances; utt. = utterances; C.I. = 95% confidence interval FA = first author; RA1 = research assistant 1; [3] intraclass correlation procedure used for parametric data, model description: 2-way random effects with absolute agreement and single rater measures; [4] Spearman's rho correlation procedure for non-parametric data; n. a. = not applicable; RA2 = research assistant 2; RA = research assistants data combined due to small sample size.