

Uses models:	No
Uses documents:	Yes
Uses topic statements:	Yes
Emulates:	Responsiveness

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.113	0.392 (-0.147 - 0.359)	0.179	0.175	0.078	0.404	0.017	0.901 (-0.241 - 0.271)	0.159	0.228	0.103	0.267
Responsiveness	0.093	0.481 (-0.167 - 0.341)	0.187	0.156	0.090	0.333	-0.058	0.661 (-0.310 - 0.201)	0.063	0.637	0.018	0.853
Readability	0.049	0.712 (-0.210 - 0.301)	0.083	0.530	0.036	0.706	-0.043	0.744 (-0.296 - 0.215)	-0.101	0.447	-0.116	0.211

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST1, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.837	0.000 (0.729 - 0.904)	0.401	0.004	0.280	0.004	0.545	0.000 (0.317 - 0.713)	0.384	0.005	0.287	0.003
Responsiveness	0.786	0.000 (0.651 - 0.872)	0.281	0.046	0.188	0.055	0.549	0.000 (0.322 - 0.716)	0.353	0.011	0.265	0.007
Readability	0.642	0.000 (0.445 - 0.780)	0.083	0.561	0.057	0.564	0.371	0.007 (0.106 - 0.586)	-0.037	0.794	-0.038	0.702

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST1, initial and update summaries, no models.

	Initial summaries			Update summaries		
	all	avg per topic	avg per assessor	all	avg per topic	avg per assessor
Pyramid	0.341	0.550	0.489	0.195	0.329	0.233
Responsiveness	0.278	0.455	0.469	0.202	0.271	0.233
Readability	0.265	0.395	0.392	0.056	0.199	0.209

Table 3: Summary-level correlations (Pearson’s *r*) for CIST1, initial and update summaries, no models.

		A>B	A=B	A<B
		CIST1	A>B	68
	A=B	206	0	
	A>B			134

Table 4: Discriminative power for CIST1 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST1	A>B	58
	A=B	218	0	
	A>B			132

Table 5: Discriminative power for CIST1 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST1	A>B	68
	A=B	206	0	
	A>B			134

Table 6: Discriminative power for CIST1 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST1	A>B	58
	A=B	218	0	
	A>B			132

Table 7: Discriminative power for CIST1 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST1	A>B	68
	A=B	205	1	
	A>B			134

Table 8: Discriminative power for CIST1 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST1	A>B	58
	A=B	218	0	
	A>B			132

Table 9: Discriminative power for CIST1 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CIST1	A>B	152	52	
	A=B	84	984	
	A>B			3

Table 10: Discriminative power for CIST1 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST1	A>B	49	35	
	A=B	122	1053	
	A>B			16

Table 11: Discriminative power for CIST1 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST1	A>B	144	60	
	A=B	74	994	
	A>B			3

Table 12: Discriminative power for CIST1 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST1	A>B	49	42	
	A=B	70	1105	
	A>B			9

Table 13: Discriminative power for CIST1 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST1	A>B	151	44	
	A=B	251	817	
	A>B			12

Table 14: Discriminative power for CIST1 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST1	A>B	50	18	
	A=B	243	932	
	A>B			32

Table 15: Discriminative power for CIST1 compared with Readability. Update summaries, non-model summaries.

Uses models:	No
Uses documents:	Yes
Uses topic statements:	Yes
Emulates:	Responsiveness

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.113	0.392 (-0.147 - 0.359)	0.179	0.175	0.078	0.404	0.017	0.901 (-0.241 - 0.271)	0.159	0.228	0.103	0.267
Responsiveness	0.093	0.481 (-0.167 - 0.341)	0.187	0.156	0.090	0.333	-0.058	0.661 (-0.310 - 0.201)	0.063	0.637	0.018	0.853
Readability	0.049	0.712 (-0.210 - 0.301)	0.083	0.530	0.036	0.706	-0.043	0.744 (-0.296 - 0.215)	-0.101	0.447	-0.116	0.211

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST2, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.847	0.000 (0.746 - 0.910)	0.418	0.002	0.301	0.002	0.578	0.000 (0.360 - 0.737)	0.419	0.002	0.309	0.001
Responsiveness	0.794	0.000 (0.663 - 0.877)	0.274	0.052	0.185	0.059	0.582	0.000 (0.365 - 0.739)	0.407	0.003	0.301	0.002
Readability	0.645	0.000 (0.449 - 0.781)	0.076	0.595	0.047	0.632	0.391	0.005 (0.129 - 0.602)	0.002	0.990	-0.014	0.890

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST2, initial and update summaries, no models.

	Initial summaries			Update summaries		
	all	avg per topic	avg per assessor	all	avg per topic	avg per assessor
Pyramid	0.336	0.533	0.491	0.173	0.324	0.236
Responsiveness	0.280	0.442	0.465	0.188	0.264	0.213
Readability	0.267	0.385	0.382	0.058	0.196	0.185

Table 3: Summary-level correlations (Pearson’s *r*) for CIST2, initial and update summaries, no models.

		A>B	A=B	A<B
		CIST2	A>B	68
	A=B	206	0	
	A>B			134

Table 4: Discriminative power for CIST2 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST2	A>B	58
	A=B	218	0	
	A>B			132

Table 5: Discriminative power for CIST2 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST2	A>B	68
	A=B	206	0	
	A>B			134

Table 6: Discriminative power for CIST2 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST2	A>B	58
	A=B	218	0	
	A>B			132

Table 7: Discriminative power for CIST2 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST2	A>B	68
	A=B	205	1	
	A>B			134

Table 8: Discriminative power for CIST2 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST2	A>B	58
	A=B	218	0	
	A>B			132

Table 9: Discriminative power for CIST2 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CIST2	A>B	153	47	
	A=B	85	989	
	A>B			1

Table 10: Discriminative power for CIST2 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST2	A>B	50	21	
	A=B	132	1067	
	A>B			5

Table 11: Discriminative power for CIST2 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST2	A>B	144	56	
	A=B	76	998	
	A>B			1

Table 12: Discriminative power for CIST2 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST2	A>B	49	25	
	A=B	77	1122	
	A>B			2

Table 13: Discriminative power for CIST2 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST2	A>B	152	41	
	A=B	254	820	
	A>B			8

Table 14: Discriminative power for CIST2 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST2	A>B	50	12	
	A=B	261	938	
	A>B			14

Table 15: Discriminative power for CIST2 compared with Readability. Update summaries, non-model summaries.

Uses models:	No
Uses documents:	Yes
Uses topic statements:	Yes
Emulates:	Responsiveness

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.176	0.183 (-0.084 - 0.413)	0.239	0.068	0.131	0.156	0.089	0.502 (-0.171 - 0.337)	0.226	0.085	0.161	0.081
Responsiveness	0.154	0.243 (-0.106 - 0.395)	0.248	0.059	0.144	0.121	0.017	0.896 (-0.240 - 0.272)	0.129	0.331	0.077	0.412
Readability	0.106	0.424 (-0.154 - 0.353)	0.154	0.243	0.090	0.337	0.025	0.849 (-0.232 - 0.280)	-0.032	0.808	-0.057	0.539

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST3, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.837	0.000 (0.729 - 0.904)	0.401	0.004	0.280	0.004	0.545	0.000 (0.317 - 0.713)	0.384	0.005	0.287	0.003
Responsiveness	0.786	0.000 (0.651 - 0.872)	0.281	0.046	0.188	0.055	0.549	0.000 (0.322 - 0.716)	0.353	0.011	0.265	0.007
Readability	0.642	0.000 (0.445 - 0.780)	0.083	0.561	0.057	0.564	0.371	0.007 (0.106 - 0.586)	-0.037	0.794	-0.038	0.702

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST3, initial and update summaries, no models.

	Initial summaries			Update summaries		
	all	avg per topic	avg per assessor	all	avg per topic	avg per assessor
Pyramid	0.341	0.550	0.489	0.195	0.329	0.233
Responsiveness	0.278	0.455	0.469	0.202	0.271	0.233
Readability	0.265	0.395	0.392	0.056	0.199	0.209

Table 3: Summary-level correlations (Pearson’s *r*) for CIST3, initial and update summaries, no models.

		A>B	A=B	A<B
		CIST3	A>B	71
	A=B	209	0	
	A>B			128

Table 4: Discriminative power for CIST3 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST3	A>B	63
	A=B	220	0	
	A>B			125

Table 5: Discriminative power for CIST3 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST3	A>B	71
	A=B	209	0	
	A>B			128

Table 6: Discriminative power for CIST3 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST3	A>B	63
	A=B	220	0	
	A>B			125

Table 7: Discriminative power for CIST3 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST3	A>B	70
	A=B	209	0	
	A>B			128

Table 8: Discriminative power for CIST3 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		CIST3	A>B	63
	A=B	220	0	
	A>B			125

Table 9: Discriminative power for CIST3 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CIST3	A>B	152	52	
	A=B	84	984	
	A>B			3

Table 10: Discriminative power for CIST3 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST3	A>B	49	35	
	A=B	122	1053	
	A>B			16

Table 11: Discriminative power for CIST3 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST3	A>B	144	60	
	A=B	74	994	
	A>B			3

Table 12: Discriminative power for CIST3 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST3	A>B	49	42	
	A=B	70	1105	
	A>B			9

Table 13: Discriminative power for CIST3 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST3	A>B	151	44	
	A=B	251	817	
	A>B			12

Table 14: Discriminative power for CIST3 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST3	A>B	50	18	
	A=B	243	932	
	A>B			32

Table 15: Discriminative power for CIST3 compared with Readability. Update summaries, non-model summaries.

Uses models:	No
Uses documents:	Yes
Uses topic statements:	Yes
Emulates:	Responsiveness

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.176	0.183 (-0.084 - 0.413)	0.239	0.068	0.131	0.156	0.089	0.502 (-0.171 - 0.337)	0.226	0.085	0.161	0.081
Responsiveness	0.154	0.243 (-0.106 - 0.395)	0.248	0.059	0.144	0.121	0.017	0.896 (-0.240 - 0.272)	0.129	0.331	0.077	0.412
Readability	0.106	0.424 (-0.154 - 0.353)	0.154	0.243	0.090	0.337	0.025	0.849 (-0.232 - 0.280)	-0.032	0.808	-0.057	0.539

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST4, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>	<i>r</i>	<i>p</i> (95% CI)	<i>rho</i>	<i>p</i>	<i>tau</i>	<i>p</i>
Pyramid	0.847	0.000 (0.746 - 0.910)	0.418	0.002	0.301	0.002	0.578	0.000 (0.360 - 0.737)	0.419	0.002	0.309	0.001
Responsiveness	0.794	0.000 (0.663 - 0.877)	0.274	0.052	0.185	0.059	0.582	0.000 (0.365 - 0.739)	0.407	0.003	0.301	0.002
Readability	0.645	0.000 (0.449 - 0.781)	0.076	0.595	0.047	0.632	0.391	0.005 (0.129 - 0.602)	0.002	0.990	-0.014	0.890

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for CIST4, initial and update summaries, no models.

	Initial summaries			Update summaries		
	all	avg per topic	avg per assessor	all	avg per topic	avg per assessor
Pyramid	0.336	0.533	0.491	0.173	0.324	0.236
Responsiveness	0.280	0.442	0.465	0.188	0.264	0.213
Readability	0.267	0.385	0.382	0.058	0.196	0.185

Table 3: Summary-level correlations (Pearson’s *r*) for CIST4, initial and update summaries, no models.

		A>B	A=B	A<B
CIST4	A>B	71	0	
	A=B	209	0	
	A>B			128

Table 4: Discriminative power for CIST4 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CIST4	A>B	63	0	
	A=B	220	0	
	A>B			125

Table 5: Discriminative power for CIST4 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CIST4	A>B	71	0	
	A=B	209	0	
	A>B			128

Table 6: Discriminative power for CIST4 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CIST4	A>B	63	0	
	A=B	220	0	
	A>B			125

Table 7: Discriminative power for CIST4 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CIST4	A>B	70	1	
	A=B	209	0	
	A>B			128

Table 8: Discriminative power for CIST4 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CIST4	A>B	63	0	
	A=B	220	0	
	A>B			125

Table 9: Discriminative power for CIST4 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CIST4	A>B	153	47	
	A=B	85	989	
	A>B			1

Table 10: Discriminative power for CIST4 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST4	A>B	50	21	
	A=B	132	1067	
	A>B			5

Table 11: Discriminative power for CIST4 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST4	A>B	144	56	
	A=B	76	998	
	A>B			1

Table 12: Discriminative power for CIST4 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST4	A>B	49	25	
	A=B	77	1122	
	A>B			2

Table 13: Discriminative power for CIST4 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CIST4	A>B	152	41	
	A=B	254	820	
	A>B			8

Table 14: Discriminative power for CIST4 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CIST4	A>B	50	12	
	A=B	261	938	
	A>B			14

Table 15: Discriminative power for CIST4 compared with Readability. Update summaries, non-model summaries.