

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

	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	0.975	0.000 (0.958 - 0.985)	0.924	0.000	0.776	0.000	0.938	0.000 (0.898 - 0.963)	0.853	0.000	0.676	0.000
Responsiveness	0.965	0.000 (0.941 - 0.979)	0.857	0.000	0.674	0.000	0.937	0.000 (0.896 - 0.962)	0.865	0.000	0.694	0.000
Readability	0.906	0.000 (0.847 - 0.943)	0.604	0.000	0.446	0.000	0.868	0.000 (0.787 - 0.920)	0.494	0.000	0.360	0.000

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH1, 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	0.951	0.000 (0.915 - 0.972)	0.900	0.000	0.753	0.000	0.811	0.000 (0.690 - 0.888)	0.820	0.000	0.639	0.000
Responsiveness	0.915	0.000 (0.856 - 0.951)	0.797	0.000	0.620	0.000	0.836	0.000 (0.728 - 0.903)	0.838	0.000	0.662	0.000
Readability	0.777	0.000 (0.638 - 0.867)	0.419	0.002	0.304	0.002	0.617	0.000 (0.411 - 0.763)	0.331	0.018	0.248	0.011

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH1, 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	0.566	0.683	0.602	0.388	0.470	0.407
Responsiveness	0.446	0.583	0.504	0.339	0.405	0.380
Readability	0.417	0.498	0.466	0.242	0.348	0.318

Table 3: Summary-level correlations (Pearson’s *r*) for C.S.IIITH1, initial and update summaries, no models.

		A>B	A=B	A<B
		C.S.IIITH1	A>B	408
	A=B	0	0	
	A>B			0

Table 4: Discriminative power for C.S.IIITH1 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH1	A>B	408
	A=B	0	0	
	A>B			0

Table 5: Discriminative power for C.S.IIITH1 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH1	A>B	408
	A=B	0	0	
	A>B			0

Table 6: Discriminative power for C.S.IIITH1 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH1	A>B	408
	A=B	0	0	
	A>B			0

Table 7: Discriminative power for C.S.IIITH1 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH1	A>B	407
	A=B	0	0	
	A>B			0

Table 8: Discriminative power for C.S.IIITH1 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH1	A>B	408
	A=B	0	0	
	A>B			0

Table 9: Discriminative power for C.S.IIITH1 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
C.S_IITH1	A>B	231	120	
	A=B	8	916	
	A>B			0

Table 10: Discriminative power for C.S\_IITH1 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S_IITH1	A>B	150	135	
	A=B	37	953	
	A>B			0

Table 11: Discriminative power for C.S\_IITH1 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S_IITH1	A>B	208	143	
	A=B	13	911	
	A>B			0

Table 12: Discriminative power for C.S\_IITH1 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S_IITH1	A>B	122	163	
	A=B	6	984	
	A>B			0

Table 13: Discriminative power for C.S\_IITH1 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S_IITH1	A>B	230	120	
	A=B	183	741	
	A>B			1

Table 14: Discriminative power for C.S\_IITH1 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S_IITH1	A>B	157	120	
	A=B	160	830	
	A>B			8

Table 15: Discriminative power for C.S\_IITH1 compared with Readability. Update summaries, non-model summaries.

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

	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	0.956	0.000 (0.927 - 0.974)	0.901	0.000	0.760	0.000	0.882	0.000 (0.809 - 0.929)	0.720	0.000	0.546	0.000
Responsiveness	0.937	0.000 (0.896 - 0.962)	0.845	0.000	0.678	0.000	0.880	0.000 (0.806 - 0.927)	0.787	0.000	0.596	0.000
Readability	0.894	0.000 (0.827 - 0.936)	0.663	0.000	0.490	0.000	0.837	0.000 (0.739 - 0.900)	0.501	0.000	0.367	0.000

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH2, 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	0.950	0.000 (0.914 - 0.971)	0.897	0.000	0.755	0.000	0.768	0.000 (0.625 - 0.861)	0.715	0.000	0.550	0.000
Responsiveness	0.915	0.000 (0.856 - 0.951)	0.787	0.000	0.610	0.000	0.802	0.000 (0.676 - 0.883)	0.766	0.000	0.586	0.000
Readability	0.776	0.000 (0.637 - 0.867)	0.396	0.004	0.290	0.003	0.617	0.000 (0.411 - 0.763)	0.286	0.042	0.203	0.037

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH2, 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	0.513	0.645	0.554	0.341	0.415	0.344
Responsiveness	0.418	0.541	0.468	0.312	0.367	0.334
Readability	0.389	0.469	0.440	0.184	0.308	0.248

Table 3: Summary-level correlations (Pearson’s *r*) for C.S.IIITH2, initial and update summaries, no models.

		A>B	A=B	A<B
		C.S.IIITH2	A>B	408
	A=B	0	0	
	A>B			0

Table 4: Discriminative power for C.S.IIITH2 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH2	A>B	408
	A=B	0	0	
	A>B			0

Table 5: Discriminative power for C.S.IIITH2 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH2	A>B	408
	A=B	0	0	
	A>B			0

Table 6: Discriminative power for C.S.IIITH2 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH2	A>B	408
	A=B	0	0	
	A>B			0

Table 7: Discriminative power for C.S.IIITH2 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH2	A>B	407
	A=B	0	0	
	A>B			0

Table 8: Discriminative power for C.S.IIITH2 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		C.S.IIITH2	A>B	408
	A=B	0	0	
	A>B			0

Table 9: Discriminative power for C.S.IIITH2 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
C.S.IIITH2	A>B	223	69	
	A=B	16	967	
	A>B			0

Table 10: Discriminative power for C.S.IIITH2 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH2	A>B	106	75	
	A=B	81	1013	
	A>B			0

Table 11: Discriminative power for C.S.IIITH2 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH2	A>B	205	87	
	A=B	16	967	
	A>B			0

Table 12: Discriminative power for C.S.IIITH2 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH2	A>B	94	87	
	A=B	34	1060	
	A>B			0

Table 13: Discriminative power for C.S.IIITH2 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH2	A>B	218	74	
	A=B	196	787	
	A>B			0

Table 14: Discriminative power for C.S.IIITH2 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH2	A>B	119	59	
	A=B	203	891	
	A>B			3

Table 15: Discriminative power for C.S.IIITH2 compared with Readability. Update summaries, non-model summaries.

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

	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	0.786	0.000 (0.663 - 0.867)	0.932	0.000	0.799	0.000	0.608	0.000 (0.416 - 0.747)	0.827	0.000	0.654	0.000
Responsiveness	0.747	0.000 (0.607 - 0.842)	0.860	0.000	0.693	0.000	0.539	0.000 (0.329 - 0.699)	0.835	0.000	0.666	0.000
Readability	0.711	0.000 (0.557 - 0.818)	0.614	0.000	0.460	0.000	0.560	0.000 (0.354 - 0.713)	0.479	0.000	0.353	0.000

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH3, 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	0.965	0.000 (0.939 - 0.980)	0.903	0.000	0.758	0.000	0.884	0.000 (0.805 - 0.933)	0.796	0.000	0.628	0.000
Responsiveness	0.933	0.000 (0.885 - 0.961)	0.781	0.000	0.596	0.000	0.885	0.000 (0.807 - 0.933)	0.812	0.000	0.638	0.000
Readability	0.731	0.000 (0.570 - 0.838)	0.358	0.010	0.242	0.013	0.552	0.000 (0.326 - 0.718)	0.257	0.068	0.180	0.066

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH3, 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	0.498	0.632	0.573	0.417	0.480	0.396
Responsiveness	0.363	0.532	0.468	0.356	0.397	0.364
Readability	0.277	0.407	0.364	0.167	0.261	0.207

Table 3: Summary-level correlations (Pearson’s *r*) for C.S.IIITH3, initial and update summaries, no models.

C.S.IIITH3		A>B	A=B	A<B
	A>B	178	0	
A=B	230	0		
A>B			0	

C.S.IIITH3		A>B	A=B	A<B
	A>B	99	0	
A=B	309	0		
A>B			0	

Table 4: Discriminative power for C.S.IIITH3 compared with Pyramid. Initial summaries, models vs. non-models.

Table 5: Discriminative power for C.S.IIITH3 compared with Pyramid. Update summaries, models vs. non-models.

C.S.IIITH3		A>B	A=B	A<B
	A>B	178	0	
A=B	230	0		
A>B			0	

C.S.IIITH3		A>B	A=B	A<B
	A>B	99	0	
A=B	309	0		
A>B			0	

Table 6: Discriminative power for C.S.IIITH3 compared with Responsiveness. Initial summaries, models vs. non-models.

Table 7: Discriminative power for C.S.IIITH3 compared with Responsiveness. Update summaries, models vs. non-models.

C.S.IIITH3		A>B	A=B	A<B
	A>B	178	0	
A=B	229	1		
A>B			0	

C.S.IIITH3		A>B	A=B	A<B
	A>B	99	0	
A=B	309	0		
A>B			0	

Table 8: Discriminative power for C.S.IIITH3 compared with Readability. Initial summaries, models vs. non-models.

Table 9: Discriminative power for C.S.IIITH3 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
C.S.IIITH3	A>B	215	126	
	A=B	24	910	
	A>B			0

Table 10: Discriminative power for C.S.IIITH3 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH3	A>B	133	69	
	A=B	54	1019	
	A>B			0

Table 11: Discriminative power for C.S.IIITH3 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH3	A>B	194	147	
	A=B	27	907	
	A>B			0

Table 12: Discriminative power for C.S.IIITH3 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH3	A>B	104	98	
	A=B	24	1049	
	A>B			0

Table 13: Discriminative power for C.S.IIITH3 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH3	A>B	207	134	
	A=B	207	727	
	A>B			0

Table 14: Discriminative power for C.S.IIITH3 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IIITH3	A>B	115	85	
	A=B	208	865	
	A>B			2

Table 15: Discriminative power for C.S.IIITH3 compared with Readability. Update summaries, non-model summaries.

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

	Initial summaries						Update summaries							
	<i>r</i>	p (95% CI)		<i>rho</i>	p	<i>tau</i>	p	<i>r</i>	p (95% CI)		<i>rho</i>	p	<i>tau</i>	p
Pyramid	0.950	0.000 (0.917 - 0.970)		0.894	0.000	0.752	0.000	0.865	0.000 (0.782 - 0.918)		0.719	0.000	0.545	0.000
Responsiveness	0.929	0.000 (0.883 - 0.957)		0.834	0.000	0.665	0.000	0.859	0.000 (0.773 - 0.914)		0.778	0.000	0.589	0.000
Readability	0.884	0.000 (0.812 - 0.930)		0.647	0.000	0.479	0.000	0.822	0.000 (0.717 - 0.891)		0.486	0.000	0.359	0.000

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH4, initial and update summaries, all peers.

	Initial summaries						Update summaries							
	<i>r</i>	p (95% CI)		<i>rho</i>	p	<i>tau</i>	p	<i>r</i>	p (95% CI)		<i>rho</i>	p	<i>tau</i>	p
Pyramid	0.946	0.000 (0.906 - 0.969)		0.879	0.000	0.725	0.000	0.770	0.000 (0.627 - 0.862)		0.722	0.000	0.549	0.000
Responsiveness	0.910	0.000 (0.847 - 0.948)		0.767	0.000	0.591	0.000	0.804	0.000 (0.679 - 0.884)		0.769	0.000	0.590	0.000
Readability	0.773	0.000 (0.632 - 0.864)		0.389	0.005	0.282	0.004	0.620	0.000 (0.415 - 0.765)		0.307	0.028	0.210	0.032

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for C.S.IIITH4, 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	0.507	0.639	0.542	0.354	0.407	0.363
Responsiveness	0.409	0.539	0.449	0.320	0.365	0.338
Readability	0.388	0.471	0.436	0.211	0.309	0.256

Table 3: Summary-level correlations (Pearson’s *r*) for C.S.IIITH4, initial and update summaries, no models.

C.S.IIITH4		A>B	A=B	A<B
		A>B	408	0
A=B		0	0	
A>B				0

C.S.IIITH4		A>B	A=B	A<B
		A>B	406	0
A=B		2	0	
A>B				0

Table 4: Discriminative power for C.S.IIITH4 compared with Pyramid. Initial summaries, models vs. non-models.

Table 5: Discriminative power for C.S.IIITH4 compared with Pyramid. Update summaries, models vs. non-models.

C.S.IIITH4		A>B	A=B	A<B
		A>B	408	0
A=B		0	0	
A>B				0

C.S.IIITH4		A>B	A=B	A<B
		A>B	406	0
A=B		2	0	
A>B				0

Table 6: Discriminative power for C.S.IIITH4 compared with Responsiveness. Initial summaries, models vs. non-models.

Table 7: Discriminative power for C.S.IIITH4 compared with Responsiveness. Update summaries, models vs. non-models.

C.S.IIITH4		A>B	A=B	A<B
		A>B	407	1
A=B		0	0	
A>B				0

C.S.IIITH4		A>B	A=B	A<B
		A>B	406	0
A=B		2	0	
A>B				0

Table 8: Discriminative power for C.S.IIITH4 compared with Readability. Initial summaries, models vs. non-models.

Table 9: Discriminative power for C.S.IIITH4 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
C.S.IITH4	A>B	218	53	
	A=B	21	983	
	A>B			0

Table 10: Discriminative power for C\_S\_IITH4 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IITH4	A>B	109	76	
	A=B	78	1012	
	A>B			0

Table 11: Discriminative power for C\_S\_IITH4 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IITH4	A>B	201	70	
	A=B	20	984	
	A>B			0

Table 12: Discriminative power for C\_S\_IITH4 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IITH4	A>B	96	89	
	A=B	32	1058	
	A>B			0

Table 13: Discriminative power for C\_S\_IITH4 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IITH4	A>B	211	60	
	A=B	203	801	
	A>B			0

Table 14: Discriminative power for C\_S\_IITH4 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
C.S.IITH4	A>B	126	57	
	A=B	197	893	
	A>B			2

Table 15: Discriminative power for C\_S\_IITH4 compared with Readability. Update summaries, non-model summaries.