

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	0.842	0.000 (0.746 - 0.903)	0.932	0.000	0.798	0.000	0.824	0.000 (0.720 - 0.892)	0.885	0.000	0.723	0.000
Responsiveness	0.815	0.000 (0.706 - 0.886)	0.899	0.000	0.748	0.000	0.774	0.000 (0.646 - 0.860)	0.908	0.000	0.753	0.000
Readability	0.770	0.000 (0.640 - 0.857)	0.674	0.000	0.513	0.000	0.761	0.000 (0.628 - 0.851)	0.609	0.000	0.465	0.000

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for DemokritosGR1, 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	0.974	0.000 (0.955 - 0.985)	0.897	0.000	0.747	0.000	0.856	0.000 (0.759 - 0.916)	0.825	0.000	0.640	0.000
Responsiveness	0.947	0.000 (0.908 - 0.969)	0.845	0.000	0.675	0.000	0.873	0.000 (0.786 - 0.926)	0.861	0.000	0.686	0.000
Readability	0.794	0.000 (0.664 - 0.878)	0.497	0.000	0.359	0.000	0.644	0.000 (0.448 - 0.781)	0.399	0.004	0.306	0.002

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for DemokritosGR1, 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	0.674	0.734	0.752	0.515	0.601	0.520
Responsiveness	0.521	0.635	0.632	0.474	0.492	0.476
Readability	0.407	0.531	0.496	0.273	0.354	0.321

Table 3: Summary-level correlations (Pearson’s *r*) for DemokritosGR1, initial and update summaries, no models.

		A>B	A=B	A<B
		DemokritosGR1	A>B	195
	A=B	213	0	
	A>B			0

Table 4: Discriminative power for DemokritosGR1 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR1	A>B	253
	A=B	155	0	
	A>B			0

Table 5: Discriminative power for DemokritosGR1 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR1	A>B	195
	A=B	213	0	
	A>B			0

Table 6: Discriminative power for DemokritosGR1 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR1	A>B	253
	A=B	155	0	
	A>B			0

Table 7: Discriminative power for DemokritosGR1 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR1	A>B	195
	A=B	212	1	
	A>B			0

Table 8: Discriminative power for DemokritosGR1 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR1	A>B	253
	A=B	155	0	
	A>B			0

Table 9: Discriminative power for DemokritosGR1 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
DemokritosGR1	A>B	236	211	
	A=B	3	825	
	A>B			0

Table 10: Discriminative power for DemokritosGR1 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR1	A>B	141	68	
	A=B	46	1020	
	A>B			0

Table 11: Discriminative power for DemokritosGR1 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR1	A>B	216	231	
	A=B	5	823	
	A>B			0

Table 12: Discriminative power for DemokritosGR1 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR1	A>B	113	96	
	A=B	15	1051	
	A>B			0

Table 13: Discriminative power for DemokritosGR1 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR1	A>B	252	193	
	A=B	160	668	
	A>B			2

Table 14: Discriminative power for DemokritosGR1 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR1	A>B	137	71	
	A=B	187	879	
	A>B			1

Table 15: Discriminative power for DemokritosGR1 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	0.951	0.000 (0.918 - 0.971)	0.924	0.000	0.776	0.000	0.933	0.000 (0.889 - 0.960)	0.888	0.000	0.723	0.000
Responsiveness	0.963	0.000 (0.938 - 0.978)	0.893	0.000	0.727	0.000	0.975	0.000 (0.958 - 0.985)	0.889	0.000	0.726	0.000
Readability	0.906	0.000 (0.846 - 0.943)	0.651	0.000	0.494	0.000	0.895	0.000 (0.829 - 0.936)	0.595	0.000	0.432	0.000

Table 1: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for DemokritosGR2, 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	0.964	0.000 (0.938 - 0.980)	0.888	0.000	0.723	0.000	0.885	0.000 (0.807 - 0.933)	0.824	0.000	0.640	0.000
Responsiveness	0.948	0.000 (0.910 - 0.970)	0.845	0.000	0.671	0.000	0.891	0.000 (0.816 - 0.937)	0.838	0.000	0.668	0.000
Readability	0.791	0.000 (0.660 - 0.876)	0.475	0.000	0.351	0.000	0.670	0.000 (0.483 - 0.798)	0.394	0.004	0.286	0.003

Table 2: System-level correlations (Pearson’s *r*, Spearman’s *rho*, Kendall’s *tau*) for DemokritosGR2, 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	0.631	0.716	0.704	0.496	0.589	0.505
Responsiveness	0.506	0.620	0.625	0.460	0.483	0.470
Readability	0.409	0.515	0.497	0.256	0.338	0.320

Table 3: Summary-level correlations (Pearson’s *r*) for DemokritosGR2, initial and update summaries, no models.

		A>B	A=B	A<B
		DemokritosGR2	A>B	408
	A=B	0	0	
	A>B			0

Table 4: Discriminative power for DemokritosGR2 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR2	A>B	408
	A=B	0	0	
	A>B			0

Table 5: Discriminative power for DemokritosGR2 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR2	A>B	408
	A=B	0	0	
	A>B			0

Table 6: Discriminative power for DemokritosGR2 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR2	A>B	408
	A=B	0	0	
	A>B			0

Table 7: Discriminative power for DemokritosGR2 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR2	A>B	407
	A=B	0	0	
	A>B			0

Table 8: Discriminative power for DemokritosGR2 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		DemokritosGR2	A>B	408
	A=B	0	0	
	A>B			0

Table 9: Discriminative power for DemokritosGR2 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
DemokritosGR2	A>B	235	246	
	A=B	4	790	
	A>B			0

Table 10: Discriminative power for DemokritosGR2 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR2	A>B	144	60	
	A=B	43	1028	
	A>B			0

Table 11: Discriminative power for DemokritosGR2 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR2	A>B	218	263	
	A=B	3	791	
	A>B			0

Table 12: Discriminative power for DemokritosGR2 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR2	A>B	120	84	
	A=B	8	1063	
	A>B			0

Table 13: Discriminative power for DemokritosGR2 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR2	A>B	259	220	
	A=B	153	641	
	A>B			2

Table 14: Discriminative power for DemokritosGR2 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
DemokritosGR2	A>B	143	61	
	A=B	182	889	
	A>B			0

Table 15: Discriminative power for DemokritosGR2 compared with Readability. Update summaries, non-model summaries.