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

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.945	0.000 (0.910 - 0.967)	0.878	0.000	0.710	0.000	0.936	0.000 (0.894 - 0.962)	0.828	0.000	0.655	0.000
Responsiveness	0.948	0.000 (0.913 - 0.969)	0.857	0.000	0.680	0.000	0.965	0.000 (0.942 - 0.979)	0.861	0.000	0.687	0.000
Readability	0.903	0.000 (0.842 - 0.942)	0.658	0.000	0.503	0.000	0.915	0.000 (0.860 - 0.949)	0.657	0.000	0.496	0.000

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY1, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.956	0.000 (0.923 - 0.974)	0.869	0.000	0.702	0.000	0.898	0.000 (0.827 - 0.941)	0.825	0.000	0.661	0.000
Responsiveness	0.949	0.000 (0.911 - 0.970)	0.789	0.000	0.608	0.000	0.903	0.000 (0.836 - 0.944)	0.848	0.000	0.675	0.000
Readability	0.778	0.000 (0.640 - 0.868)	0.412	0.003	0.308	0.002	0.619	0.000 (0.414 - 0.764)	0.370	0.007	0.256	0.009

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY1, 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	0.586	0.672	0.702	0.379	0.488	0.353
Responsiveness	0.449	0.590	0.594	0.333	0.395	0.304
Readability	0.325	0.471	0.434	0.162	0.213	0.167

Table 3: Summary-level correlations (Pearson's *r*) for CLASSY1, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY1	A>B	327	0	
	A=B	81	0	
	A>B			0

Table 4: Discriminative power for CLASSY1 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	400	0	
	A=B	8	0	
	A>B			0

Table 5: Discriminative power for CLASSY1 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	327	0	
	A=B	81	0	
	A>B			0

Table 6: Discriminative power for CLASSY1 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	400	0	
	A=B	8	0	
	A>B			0

Table 7: Discriminative power for CLASSY1 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	327	0	
	A=B	80	1	
	A>B			0

Table 8: Discriminative power for CLASSY1 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	400	0	
	A=B	8	0	
	A>B			0

Table 9: Discriminative power for CLASSY1 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY1	A>B	232	219	
	A=B	7	817	
	A>B			0

Table 10: Discriminative power for CLASSY1 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	54	1	
	A=B	133	1087	
	A>B			0

Table 11: Discriminative power for CLASSY1 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	215	236	
	A=B	6	818	
	A>B			0

Table 12: Discriminative power for CLASSY1 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	50	5	
	A=B	78	1142	
	A>B			0

Table 13: Discriminative power for CLASSY1 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	264	187	
	A=B	150	674	
	A>B			0

Table 14: Discriminative power for CLASSY1 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY1	A>B	47	8	
	A=B	278	942	
	A>B			0

Table 15: Discriminative power for CLASSY1 compared with Readability. Update summaries, non-model summaries.

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

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.945	0.000 (0.910 - 0.967)	0.878	0.000	0.710	0.000	0.944	0.000 (0.907 - 0.966)	0.857	0.000	0.689	0.000
Responsiveness	0.948	0.000 (0.913 - 0.969)	0.857	0.000	0.680	0.000	0.963	0.000 (0.938 - 0.978)	0.889	0.000	0.728	0.000
Readability	0.903	0.000 (0.842 - 0.942)	0.658	0.000	0.503	0.000	0.915	0.000 (0.861 - 0.949)	0.653	0.000	0.482	0.000

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY2, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.967	0.000 (0.942 - 0.981)	0.881	0.000	0.721	0.000	0.900	0.000 (0.831 - 0.942)	0.797	0.000	0.617	0.000
Responsiveness	0.951	0.000 (0.916 - 0.972)	0.790	0.000	0.603	0.000	0.903	0.000 (0.836 - 0.944)	0.814	0.000	0.635	0.000
Readability	0.774	0.000 (0.634 - 0.865)	0.408	0.003	0.302	0.002	0.620	0.000 (0.416 - 0.765)	0.322	0.021	0.225	0.021

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY2, 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	0.623	0.709	0.721	0.370	0.482	0.388
Responsiveness	0.468	0.617	0.606	0.358	0.416	0.349
Readability	0.336	0.503	0.445	0.201	0.260	0.225

Table 3: Summary-level correlations (Pearson's *r*) for CLASSY2, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY2	A>B	327	0	
	A=B	81	0	
	A>B			0

Table 4: Discriminative power for CLASSY2 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	401	0	
	A=B	7	0	
	A>B			0

Table 5: Discriminative power for CLASSY2 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	327	0	
	A=B	81	0	
	A>B			0

Table 6: Discriminative power for CLASSY2 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	401	0	
	A=B	7	0	
	A>B			0

Table 7: Discriminative power for CLASSY2 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	327	0	
	A=B	80	1	
	A>B			0

Table 8: Discriminative power for CLASSY2 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	401	0	
	A=B	7	0	
	A>B			0

Table 9: Discriminative power for CLASSY2 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY2	A>B	236	274	
	A=B	3	762	
	A>B			0

Table 10: Discriminative power for CLASSY2 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY2	A>B	117	27	
	A=B	70	1061	
	A>B			0

Table 11: Discriminative power for CLASSY2 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY2	A>B	216	294	
	A=B	5	760	
	A>B			0

Table 12: Discriminative power for CLASSY2 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY2	A>B	99	45	
	A=B	29	1102	
	A>B			0

Table 13: Discriminative power for CLASSY2 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY2	A>B	270	240	
	A=B	144	621	
	A>B			0

Table 14: Discriminative power for CLASSY2 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY2	A>B	108	36	
	A=B	217	914	
	A>B			0

Table 15: Discriminative power for CLASSY2 compared with Readability. Update summaries, non-model summaries.

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

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.909	0.000 (0.852 - 0.945)	0.906	0.000	0.752	0.000	0.953	0.000 (0.921 - 0.972)	0.849	0.000	0.685	0.000
Responsiveness	0.899	0.000 (0.836 - 0.939)	0.847	0.000	0.669	0.000	0.961	0.000 (0.935 - 0.977)	0.872	0.000	0.725	0.000
Readability	0.844	0.000 (0.750 - 0.905)	0.600	0.000	0.452	0.000	0.907	0.000 (0.849 - 0.944)	0.660	0.000	0.503	0.000

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY3, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.937	0.000 (0.891 - 0.964)	0.871	0.000	0.696	0.000	0.890	0.000 (0.813 - 0.936)	0.819	0.000	0.634	0.000
Responsiveness	0.932	0.000 (0.883 - 0.961)	0.795	0.000	0.608	0.000	0.919	0.000 (0.862 - 0.953)	0.866	0.000	0.695	0.000
Readability	0.768	0.000 (0.624 - 0.861)	0.434	0.001	0.316	0.001	0.705	0.000 (0.533 - 0.821)	0.422	0.002	0.301	0.002

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY3, 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	0.599	0.685	0.705	0.406	0.509	0.420
Responsiveness	0.441	0.598	0.589	0.387	0.443	0.387
Readability	0.307	0.484	0.422	0.221	0.310	0.263

Table 3: Summary-level correlations (Pearson's *r*) for CLASSY3, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY3	A>B	225	0	
	A=B	183	0	
	A>B			0

Table 4: Discriminative power for CLASSY3 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	406	0	
	A=B	2	0	
	A>B			0

Table 5: Discriminative power for CLASSY3 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	225	0	
	A=B	183	0	
	A>B			0

Table 6: Discriminative power for CLASSY3 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	406	0	
	A=B	2	0	
	A>B			0

Table 7: Discriminative power for CLASSY3 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	225	0	
	A=B	182	1	
	A>B			0

Table 8: Discriminative power for CLASSY3 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	406	0	
	A=B	2	0	
	A>B			0

Table 9: Discriminative power for CLASSY3 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY3	A>B	228	263	
	A=B	11	773	
	A>B			0

Table 10: Discriminative power for CLASSY3 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	130	60	
	A=B	57	1028	
	A>B			0

Table 11: Discriminative power for CLASSY3 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	214	277	
	A=B	7	777	
	A>B			0

Table 12: Discriminative power for CLASSY3 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	111	79	
	A=B	17	1068	
	A>B			0

Table 13: Discriminative power for CLASSY3 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	279	212	
	A=B	135	649	
	A>B			0

Table 14: Discriminative power for CLASSY3 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY3	A>B	154	36	
	A=B	171	914	
	A>B			0

Table 15: Discriminative power for CLASSY3 compared with Readability. Update summaries, non-model summaries.

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

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.853	0.000 (0.764 - 0.910)	0.926	0.000	0.783	0.000	0.953	0.000 (0.921 - 0.972)	0.891	0.000	0.731	0.000
Responsiveness	0.830	0.000 (0.729 - 0.896)	0.860	0.000	0.689	0.000	0.949	0.000 (0.916 - 0.970)	0.912	0.000	0.764	0.000
Readability	0.774	0.000 (0.645 - 0.859)	0.606	0.000	0.464	0.000	0.887	0.000 (0.817 - 0.932)	0.620	0.000	0.460	0.000

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY4, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	r	p (95% CI)	*rho*	p	*tau*	p	*r*	p (95% CI)	*rho*	p	*tau*	p
Pyramid	0.968	0.000 (0.944 - 0.982)	0.882	0.000	0.723	0.000	0.911	0.000 (0.848 - 0.948)	0.837	0.000	0.661	0.000
Responsiveness	0.951	0.000 (0.916 - 0.972)	0.796	0.000	0.611	0.000	0.927	0.000 (0.875 - 0.958)	0.877	0.000	0.716	0.000
Readability	0.784	0.000 (0.649 - 0.871)	0.407	0.003	0.304	0.002	0.683	0.000 (0.502 - 0.807)	0.426	0.002	0.314	0.001

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for CLASSY4, 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	0.630	0.712	0.721	0.453	0.553	0.449
Responsiveness	0.480	0.619	0.611	0.399	0.455	0.395
Readability	0.364	0.518	0.467	0.195	0.290	0.245

Table 3: Summary-level correlations (Pearson's *r*) for CLASSY4, initial and update summaries, no models.

		A>B	A=B	A<B
CLASSY4	A>B	180	0	
	A=B	228	0	
	A>B			0

Table 4: Discriminative power for CLASSY4 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	386	0	
	A=B	22	0	
	A>B			0

Table 5: Discriminative power for CLASSY4 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	180	0	
	A=B	228	0	
	A>B			0

Table 6: Discriminative power for CLASSY4 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	386	0	
	A=B	22	0	
	A>B			0

Table 7: Discriminative power for CLASSY4 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	180	0	
	A=B	227	1	
	A>B			0

Table 8: Discriminative power for CLASSY4 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	386	0	
	A=B	22	0	
	A>B			0

Table 9: Discriminative power for CLASSY4 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
CLASSY4	A>B	236	284	
	A=B	3	752	
	A>B			0

Table 10: Discriminative power for CLASSY4 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	135	52	
	A=B	52	1036	
	A>B			0

Table 11: Discriminative power for CLASSY4 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	216	304	
	A=B	5	750	
	A>B			0

Table 12: Discriminative power for CLASSY4 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	115	72	
	A=B	13	1075	
	A>B			0

Table 13: Discriminative power for CLASSY4 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	273	247	
	A=B	141	614	
	A>B			0

Table 14: Discriminative power for CLASSY4 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
CLASSY4	A>B	143	44	
	A=B	182	906	
	A>B			0

Table 15: Discriminative power for CLASSY4 compared with Readability. Update summaries, non-model summaries.