

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>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Responsiveness	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Readability	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM1, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	p (95% CI)		<i>rho</i>	p		<i>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	0.968	0.000 (0.944 - 0.982)		0.855	0.000		0.696	0.000 (0.521 - 0.815)		0.537	0.000	
Responsiveness	0.936	0.000 (0.889 - 0.963)		0.743	0.000		0.672	0.000 (0.486 - 0.799)		0.531	0.000	
Readability	0.767	0.000 (0.624 - 0.861)		0.344	0.013		0.380	0.006 (0.116 - 0.593)		0.238	0.093	

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM1, 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	0.628	0.726	0.732	0.109	0.221	0.157
Responsiveness	0.463	0.618	0.607	0.089	0.195	0.125
Readability	0.348	0.502	0.451	0.040	0.108	0.111

Table 3: Summary-level correlations (Pearson's *r*) for PKUTM1, initial and update summaries, no models.

		A>B	A=B	A<B
		PKUTM1	A>B	0
	A=B	408	0	
	A>B			0

Table 4: Discriminative power for PKUTM1 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM1	A>B	0
	A=B	408	0	
	A>B			0

Table 5: Discriminative power for PKUTM1 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM1	A>B	0
	A=B	408	0	
	A>B			0

Table 6: Discriminative power for PKUTM1 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM1	A>B	0
	A=B	408	0	
	A>B			0

Table 7: Discriminative power for PKUTM1 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM1	A>B	0
	A=B	407	1	
	A>B			0

Table 8: Discriminative power for PKUTM1 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM1	A>B	0
	A=B	408	0	
	A>B			0

Table 9: Discriminative power for PKUTM1 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
PKUTM1	A>B	230	151	
	A=B	9	885	
	A>B			0

Table 10: Discriminative power for PKUTM1 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM1	A>B	89	103	
	A=B	98	985	
	A>B			0

Table 11: Discriminative power for PKUTM1 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM1	A>B	212	169	
	A=B	9	885	
	A>B			0

Table 12: Discriminative power for PKUTM1 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM1	A>B	67	125	
	A=B	61	1022	
	A>B			0

Table 13: Discriminative power for PKUTM1 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM1	A>B	242	134	
	A=B	167	727	
	A>B			5

Table 14: Discriminative power for PKUTM1 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM1	A>B	90	102	
	A=B	235	848	
	A>B			0

Table 15: Discriminative power for PKUTM1 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>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Responsiveness	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Readability	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM2, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	p (95% CI)		<i>rho</i>	p		<i>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	0.947	0.000 (0.909 - 0.970)		0.806	0.000		0.820	0.000 (0.703 - 0.894)		0.641	0.000	
Responsiveness	0.916	0.000 (0.856 - 0.951)		0.692	0.000		0.819	0.000 (0.702 - 0.893)		0.651	0.000	
Readability	0.762	0.000 (0.616 - 0.858)		0.309	0.027		0.529	0.000 (0.297 - 0.702)		0.215	0.129	

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM2, 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	0.615	0.711	0.726	0.166	0.305	0.218
Responsiveness	0.444	0.600	0.593	0.126	0.265	0.170
Readability	0.328	0.494	0.440	0.068	0.168	0.149

Table 3: Summary-level correlations (Pearson's *r*) for PKUTM2, initial and update summaries, no models.

		A>B	A=B	A<B
		PKUTM2	A>B	0
	A=B	408	0	
	A>B			0

Table 4: Discriminative power for PKUTM2 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM2	A>B	0
	A=B	408	0	
	A>B			0

Table 5: Discriminative power for PKUTM2 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM2	A>B	0
	A=B	408	0	
	A>B			0

Table 6: Discriminative power for PKUTM2 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM2	A>B	0
	A=B	408	0	
	A>B			0

Table 7: Discriminative power for PKUTM2 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM2	A>B	0
	A=B	407	1	
	A>B			0

Table 8: Discriminative power for PKUTM2 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM2	A>B	0
	A=B	408	0	
	A>B			0

Table 9: Discriminative power for PKUTM2 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
PKUTM2	A>B	227	142	
	A=B	12	894	
	A>B			0

Table 10: Discriminative power for PKUTM2 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM2	A>B	98	23	
	A=B	89	1065	
	A>B			0

Table 11: Discriminative power for PKUTM2 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM2	A>B	212	157	
	A=B	9	897	
	A>B			0

Table 12: Discriminative power for PKUTM2 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM2	A>B	87	34	
	A=B	41	1113	
	A>B			0

Table 13: Discriminative power for PKUTM2 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM2	A>B	243	118	
	A=B	163	743	
	A>B			8

Table 14: Discriminative power for PKUTM2 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM2	A>B	97	24	
	A=B	228	926	
	A>B			0

Table 15: Discriminative power for PKUTM2 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>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Responsiveness	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Readability	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM3, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	p (95% CI)		<i>rho</i>	p		<i>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	0.962	0.000 (0.934 - 0.978)		0.894	0.000		0.904	0.000 (0.836 - 0.944)		0.802	0.000	
Responsiveness	0.943	0.000 (0.901 - 0.967)		0.787	0.000		0.919	0.000 (0.862 - 0.953)		0.843	0.000	
Readability	0.755	0.000 (0.605 - 0.853)		0.396	0.004		0.662	0.000 (0.473 - 0.793)		0.374	0.007	

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM3, 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	0.595	0.692	0.710	0.468	0.567	0.472
Responsiveness	0.432	0.595	0.579	0.402	0.478	0.406
Readability	0.303	0.464	0.417	0.209	0.320	0.268

Table 3: Summary-level correlations (Pearson's *r*) for PKUTM3, initial and update summaries, no models.

		A>B	A=B	A<B
		PKUTM3	A>B	0
	A=B	408	0	
	A>B			0

Table 4: Discriminative power for PKUTM3 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM3	A>B	0
	A=B	408	0	
	A>B			0

Table 5: Discriminative power for PKUTM3 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM3	A>B	0
	A=B	408	0	
	A>B			0

Table 6: Discriminative power for PKUTM3 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM3	A>B	0
	A=B	408	0	
	A>B			0

Table 7: Discriminative power for PKUTM3 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM3	A>B	0
	A=B	407	1	
	A>B			0

Table 8: Discriminative power for PKUTM3 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM3	A>B	0
	A=B	408	0	
	A>B			0

Table 9: Discriminative power for PKUTM3 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
PKUTM3	A>B	235	199	
	A=B	4	837	
	A>B			0

Table 10: Discriminative power for PKUTM3 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM3	A>B	151	95	
	A=B	36	993	
	A>B			0

Table 11: Discriminative power for PKUTM3 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM3	A>B	214	220	
	A=B	7	834	
	A>B			0

Table 12: Discriminative power for PKUTM3 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM3	A>B	121	125	
	A=B	7	1022	
	A>B			0

Table 13: Discriminative power for PKUTM3 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM3	A>B	249	185	
	A=B	165	676	
	A>B			0

Table 14: Discriminative power for PKUTM3 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM3	A>B	164	82	
	A=B	161	868	
	A>B			0

Table 15: Discriminative power for PKUTM3 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>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Responsiveness	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	
Readability	NaN	NaN (NaN - NaN)		NaN	NaN		NaN	NaN (NaN - NaN)		NaN	NaN	

Table 1: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM4, initial and update summaries, all peers.

	Initial summaries						Update summaries					
	<i>r</i>	p (95% CI)		<i>rho</i>	p		<i>r</i>	p (95% CI)		<i>rho</i>	p	
Pyramid	0.962	0.000 (0.935 - 0.978)		0.842	0.000		0.794	0.000 (0.663 - 0.877)		0.595	0.000	
Responsiveness	0.930	0.000 (0.881 - 0.960)		0.726	0.000		0.782	0.000 (0.645 - 0.870)		0.591	0.000	
Readability	0.767	0.000 (0.623 - 0.860)		0.329	0.019		0.485	0.000 (0.242 - 0.671)		0.235	0.096	

Table 2: System-level correlations (Pearson's *r*, Spearman's *rho*, Kendall's *tau*) for PKUTM4, 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	0.626	0.723	0.732	0.117	0.263	0.173
Responsiveness	0.458	0.613	0.604	0.088	0.226	0.131
Readability	0.342	0.500	0.448	0.051	0.140	0.125

Table 3: Summary-level correlations (Pearson's *r*) for PKUTM4, initial and update summaries, no models.

		A>B	A=B	A<B
		PKUTM4	A>B	0
	A=B	408	0	
	A>B			0

Table 4: Discriminative power for PKUTM4 compared with Pyramid. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM4	A>B	0
	A=B	408	0	
	A>B			0

Table 5: Discriminative power for PKUTM4 compared with Pyramid. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM4	A>B	0
	A=B	408	0	
	A>B			0

Table 6: Discriminative power for PKUTM4 compared with Responsiveness. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM4	A>B	0
	A=B	408	0	
	A>B			0

Table 7: Discriminative power for PKUTM4 compared with Responsiveness. Update summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM4	A>B	0
	A=B	407	1	
	A>B			0

Table 8: Discriminative power for PKUTM4 compared with Readability. Initial summaries, models vs. non-models.

		A>B	A=B	A<B
		PKUTM4	A>B	0
	A=B	408	0	
	A>B			0

Table 9: Discriminative power for PKUTM4 compared with Readability. Update summaries, models vs. non-models.

		A>B	A=B	A<B
PKUTM4	A>B	230	149	
	A=B	9	887	
	A>B			0

Table 10: Discriminative power for PKUTM4 compared with Pyramid. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM4	A>B	90	19	
	A=B	97	1069	
	A>B			0

Table 11: Discriminative power for PKUTM4 compared with Pyramid. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM4	A>B	212	167	
	A=B	9	887	
	A>B			0

Table 12: Discriminative power for PKUTM4 compared with Responsiveness. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM4	A>B	76	33	
	A=B	52	1114	
	A>B			0

Table 13: Discriminative power for PKUTM4 compared with Responsiveness. Update summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM4	A>B	243	130	
	A=B	165	731	
	A>B			6

Table 14: Discriminative power for PKUTM4 compared with Readability. Initial summaries, non-model summaries.

		A>B	A=B	A<B
PKUTM4	A>B	84	25	
	A=B	241	925	
	A>B			0

Table 15: Discriminative power for PKUTM4 compared with Readability. Update summaries, non-model summaries.