

Appendix A.

Because there are essentially three topic assigners (Reviewer 1, Reviewer 2, and the Arbitrator), we calculated a generalized kappa to measure inter-reviewer agreement.

Fleiss' Generalized Kappa

Number of Raters:	3
Number of Items:	564
Number of Rating Categories:	10
p_a	0.90543735

Overall Rater Agreement

kappa	0.89155393
s^2	0.00003620
z	148.18948247
p	0.00000000
LCI	0.87976196
UCI	0.90334590

Inter-reviewer Agreement by Category

κ_1	0.9258	κ_2	0.9163	κ_3	0.9491	κ_4	0.9002	κ_5	0.8347
Var	0.0037		0.0036		0.0036		0.0045		0.0042
SE	0.0605		0.0598		0.0600		0.0671		0.0650
z	15.3141		15.3166		15.8308		13.4100		12.8347
p	0.0000		0.0000		0.0000		0.0000		0.0000
Upper CI	1.0442		1.0335		1.0666		1.0318		0.9621
Lower CI	0.8073		0.7990		0.8316		0.7687		0.7072
κ_1	0.7834	κ_7	0.8756	κ_8	0.9023	κ_9	0.7522	κ_{10}	0.9085
Var	0.0047		0.0039		0.0054		0.0048		0.0037
SE	0.0685		0.0628		0.0737		0.0693		0.0606
z	11.4280		13.9381		12.2504		10.8596		14.9902
p	0.0000		0.0000		0.0000		0.0000		0.0000
Upper CI	0.9177		0.9988		1.0467		0.8879		1.0273
Lower CI	0.6490		0.7525		0.7580		0.6164		0.7897

We used software available at

<http://www.ccitonline.org/jking/homepage/interrater.html>
to calculate kappa.

Data are available at <https://idash-data.ucsd.edu/community/27>