1.7 Alternative Quality Score Weighting Performance Assessment

Our quality score (qs_g) currently consists of 5 dimensions that are weighted equally as depicted in formula (1) in the main manuscript (i.e., the 5 dimensions are just multiplied to get qs_g). To evaluate the current weighting as depicted in formula (1) in the main manuscript, we consider an alternative weighting scheme: maximize the products of s_a, s_k, s_d, while having minimum thresholds (i.e., cutoff-values) for s_n and s_r. Table W1.7 compares the scores of the current against the alternative weighting scheme for all content produced for our field experiments in the IT service industry and the education sector when applying a 50% (i.e., keep 50% of the top scoring pieces of content) and a 25% (i.e., keep 25% of the top scoring pieces of content) cutoff value for s_n and s_r. A positive (negative) value in Table W1.7 means the quality score weighting scheme presented in the article according to formula (1) performs better (worse) than the alternative quality score weighting scheme. For example, using a cutoff value of 50% in the IT service industry sector experiment, the proposed quality score weighting scheme is superior for s_a (.022**), s_k (.042**), s_d (.042**), and qs_g (.008**), which is consistent across both the used cutoff values and experimental contexts. Thus, the alternative quality score weighting we considered does not result in any improvement, as cutting off content with lower s_n and s_r often results in discarding content that performs well in terms of s_a, s_k, s_d, which ultimately translates into a lower score for the overall quality metric qs_g.

Table W1.7: Comparison of Current vs. Alternative Quality Score Weighting Scheme

Field experiment	**s_n & s_r cut-off value**[1]	s_a[2]	s_k[2]	s_d[2]	s_n[2]	s_r[2]	qs_g[2]
IT service	50%	.022**	.042**	.042**	-.083**	-.021**	.008**
IT service	25%	.022**	.046**	.178**	-.167**	-.043**	.022**
Education	50%	.030**	.045**	.013**	-.083**	-.043**	.003**
Education	25%	.045**	.059**	.021**	-.083**	-.021**	.014**

[1]The cut-off value specifies how many top-scoring data points to maintain, i.e., 50% means keep 50% of the top scoring data-points in s_n & s_r, 25% means keep 25% of the top scoring data points in s_n & s_r (25% is thus more conservative)

[2]Reported values are median difference values, i.e., median qs_g score of old quality score scheme minus median qs_g score of new (as suggested by the reviewer) quality score scheme. A positive value means the old scheme is superior, a negative value means the new scheme is superior. Significance codes come from two-tailed Wilcoxon rank sum 2-group comparison tests: *0.05 level, **0.01 level;

Appendix References

Baayen RH, Shafaei-Bajestan E (2019) Analyzing linguistic data: A practical introduction to statistics. Package 'languageR'. Version 1.5.0. *CRAN*. Accessed May 20, 2019, https://cran.r-project.org/web/packages/languageR/languageR.pdf

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A, (2018) "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software*. 3(30). https://doi.org/10.21105/joss.00774

Berger J, Sherman G, Ungar L (2020b) TextAnalyzer. Accessed November 11, 2020, http://textanalyzer.org

Bronnenberg BJ, Kim JB, Mela CF (2016) Zooming in on choice: How do consumers search for cameras online? *Marketing Science*. 35(5):693-712.

Danaher PJ, Mullarkey GW, Essegaier S (2006) Factors affecting website visit duration: A cross-domain analysis. *Journal of Marketing Research*. 43(2):182-194.

Edelman B, Zhenyu L (2016) Design of search engine services: Channel interdependence in search engine results. *Journal of Marketing Research*. 53(6):881-900.

Flanigan, AJ, Metzger, MJ (2007) The role of site features, user attribtues, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*. 9(2):319-342. https://doi.org/10.1177/1461444807075015

Jerath K, Ma L, Park YH (2014) Consumer click behavior at a search engine: The role of keyword popularity. *Journal of Marketing Research*. 51(4):480-486.

Kamoen N, Holleman B, Bergh H (2013) Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*. 7(3):181-189.

Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science*. 37(6):930-952.

Maechler M, Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Conceicao ELT, Palma MA (2020) Basic robust statistics. Package 'robustbase'. Version 0.93-6. *CRAN*. Accessed May 20, 2020, https://cran.r-project.org/web/packages/robustbase/robustbase.pdf

Pennebaker JW, Booth RJ, Boyd RL, Francis ME (2015) Linguistic inquiry and word count: LIWC2015. Austin, TX: Pennebaker Conglomerates. Accessed November 1, 2020, www.LIWC.net.

Pitler E, Nenkova A (2008) Revisiting Readability: A unified framework for predicting text quality. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. 186-195.

Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. OpenAI.

Roberts C (2010) Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist*. 54(1):43-56.

Rocklage MD, Rucker DD, Nordgren LF (2018) Persuasion, emotion and language: the intent to persuade transforms language via emotionality. *Psychological Science*. 29(5):749-760.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomze AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*. 1-15.