

## 2.5 Consumers' Content Perceptions

We provide details for the MTurk study presented in the article in which we examine differences in consumer perceptions between the semi-automated and human content. The instructions with which survey participants were presented are shown in Table W2.5.1.

Table W2.5.1: Participants' Survey Instructions

Survey Instructions
<b>Dear study participant</b>
Thank you for participating in our study on SEO & text writing. Your input is vital for us. In the following, besides answering some demographic questions, <b>we will ask you to read and assess 1 text.</b>
<b>It will take you 5 minutes at most to finish the survey.</b>
Please <b>read all questions and the text mindfully and completely</b> , and <b>answer all questions as honestly and spontaneously as possible</b> . Follow your intuition, there are <b>no right or wrong answers</b> .
All information that you provide to us will be <b>strictly treated as anonymous</b> . Thank you for your kind support.
Sincerely, [...]
[New survey page]
<b>Imagine, you are looking for an IT service for your company, and you come across a website with the text below. Please take a look at it.</b>
[Randomized piece of content]
[Questions to assess content]

To assure data quality in our survey-based content perception experiment, we implemented honeypots (for antispam), attention and honesty checks (i.e., reverse coded items

and same questions worded a bit differently), and excluded all surveys with a completion time lower than 1.50 minutes, leaving us with 551 surveys for our analyses. We performed scale reliability checks using Cronbach's Alpha including deleting offset items. Using a series of Kruskal Wallis tests, we assured that participants' characteristics did not differ substantially between the experimental conditions in terms of the time to finish the survey ( $\chi^2(3)=3.38$ ,  $\eta^2=.01$ ,  $p=.337$ ), the participants' gender ( $\chi^2(3)=2.00$ ,  $\eta^2=.00$ ,  $p=.572$ ), the highest completed level of education ( $\chi^2(3)=3.08$ ,  $\eta^2=.01$ ,  $p=.380$ ), age ( $\chi^2(3)=.25$ ,  $\eta^2=.00$ ,  $p=.969$ ), and English reading proficiency ( $\chi^2(3)=.41$ ,  $\eta^2=.00$ ,  $p=.939$ ).

Table W2.5.2 reports operationalizations, references, and scale reliability metrics for the content perception study we conducted.

Table W2.5.2: Operationalizations &amp; Measures of Main Variables for Survey

Variable	Items	Source	Scale Reliability <sup>1</sup>
Readability	Bipolar 5-point scale with following items: “Please indicate whether you perceive the text above as ... <ul style="list-style-type: none"> <li>● poorly written – well written</li> <li>● poorly readable – well readable</li> <li>● not fitting together well – fitting together well</li> <li>● not understandable – understandable</li> <li>● not interesting – interesting”</li> </ul>	Pitler and Nenkova 2008	.91
Understandability	Bipolar 5-point scale with following items: “Please indicate whether you perceive the text above as ... <ul style="list-style-type: none"> <li>● complicated – simple</li> <li>● unclear – clear</li> <li>● chaotic – orderly</li> <li>● illogically arranged – logically arranged</li> <li>● wordy – concise</li> <li>● difficult – easy“</li> </ul>	Kamoen et al. 2013	.88
Credibility	Bipolar 5-point scale with following items: “Please indicate whether the text above is ... <ul style="list-style-type: none"> <li>● unbelievable – believable</li> <li>● inaccurate – accurate</li> <li>● not trustworthy – trustworthy</li> <li>● biased – not biased</li> <li>● incomplete – complete”</li> </ul>	Roberts 2010, Flanigan and Metzger 2000	.87
Attitude toward the content	Bipolar 5-point scale with following items: “Please indicate whether you feel that the text above is ... <ul style="list-style-type: none"> <li>● distant – appealing</li> <li>● reluctant – inviting</li> <li>● boring – fascinating</li> <li>● impersonal – personal</li> <li>● monotonous – varied</li> <li>● interesting – uninteresting”</li> </ul>	Kamoen et al. 2013	.89

<sup>1</sup>Cronbach’s Alpha with optimized number of items

In addition to the scale items included in Table W2.5.2, we measure perceived content naturality using two items. On bipolar five-point scales, we ask respondents to indicate whether they believe that the content feels artificial vs. feels natural, and machine-made vs. human-made. We also ask two questions to assess future intent. To gauge willingness to further inform, we use a slider from 0 to 100 and ask respondents to indicate how they agree with the statement: “I want to further inform myself about the company providing the service.” To measure willingness to buy, we use a slider from 0 to 100 and ask respondents to indicate how much they agree with the statement: “I am willing to buy the described service.”

In Table W2.5.3, we report pairwise correlations between user perception variables using Kendall’s tau b, illustrating high correlations between these items.

Table W2.5.3: Consumer Content Perception: Dimensions’ Inter-Correlations

Dimension	Kendall’s tau b ( $\tau_b$ )						
	Readability	Understandability	Credibility	Attitude Toward the Content	Content Naturality	Willingness to Further Inform	Willingness to Buy
Readability	1.00**	.59**	.57**	.50**	.52**	.41**	.42**
Understandability		1.00**	.43**	.58**	.57**	.44**	.46**
Credibility			1.00**	.40**	.44**	.33**	.37**
Attitude Toward the Content				1.00**	.58**	.52**	.53**
Content Naturality					1.00**	.44**	.49**
Willingness to Further Inform						1.00**	.69**
Willingness to Buy							1.00**

<sup>1</sup>Statistical significance codes: \*0.05 level, \*\*0.01 level, one-tailed; n=551;

We complement the MTurk Study with a computational linguistic analysis. Using LIWC (Pennebaker et al. 2015), the evaluative lexicon (Rocklage et al. 2018), and the text analyzer

(Berger et al. 2020b) software packages that apply various lexica, analyses and scales, we assess the linguistic properties along psychological dimensions including concreteness, familiarity, and emotionality. The analysis presented in Table W2.5.4 reveals that differences between the semi-automated and human content are minor along most dimensions.

Table W2.5.4: Consumer Content Perception (Computational Analysis)

Dimension	Descriptives (Mean, SD) <sup>1</sup>				Kruskal Wallis <sup>2</sup>			
	Revised Machine	Real SEO Experts	Quasi Experts	Novices	$\chi^2$	$\eta^2$	df	p
Concreteness	323.10 (7.45)	326.00 (5.37)	321.30 (7.48)	318.60 (4.28)	9.67	.15	3	.021*
Familiarity	574.14 (7.95)	578.14 (12.73)	579.22 (9.33)	581.47 (9.14)	7.14	.11	3	.067
Emotionality	3.28 (.66)	3.33 (.38)	3.47 (.55)	3.53 (.47)	3.07	.05	3	.380
Emotional Valence	6.15 (.89)	6.23 (.86)	6.45 (.77)	6.69 (.72)	3.70	.06	3	.296
Negations	.004 (.005)	.005 (.003)	.006 (.003)	.007 (.006)	3.28	.05	3	.351
Interrogatives	.011 (.006)	.009 (.004)	.013 (.006)	.013 (.008)	2.31	.04	3	.509
Causation	.028 (.009)	.030 (.013)	.032 (.015)	.026 (.009)	2.07	.03	3	.558
Certainty	.011 (.005)	.013 (.005)	.021 (.009)	.019 (.009)	16.24	.25	3	.001**
Tentativeness	.022 (.010)	.027 (.014)	.022 (.010)	.022 (.009)	1.44	.02	3	.697
Differentiation	.020 (.009)	.026 (.014)	.021 (.009)	.021 (.011)	1.25	.02	3	.740
Focus on future	.009 (.006)	.013 (.006)	.011 (.006)	.015 (.007)	8.54	.13	3	.036*

<sup>1</sup>Dimension scales: for concreteness, familiarity scale range: 100 (abstract, unfamiliar) to 700 (concrete, familiar), emotionality scale range: 0 (no emotion) to 9 (high emotion), emotional valence scale range: 0 (highly negative) to 9 (highly positive); other dimensions like negations, interrogatives, etc., represent percentages of total words in the text;

<sup>2</sup>Statistical significance codes: \*0.05 level, \*\*0.01 level; n=66;

## 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, Essegai 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 attributes, 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](http://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, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*. 1-15.