First Steps Toward Measuring the Readability of Security Advice

Elissa M. Redmiles, Miraida Morales*, Lisa Maszkiewicz, Rock Stevens,
Everest Liu, Dhruv Kuchhal†, and Michelle L. Mazurek
University of Maryland, *Rutgers University, †Maharaja Agrasen Institute of Technology

***Abstract*—Security advice is one key way that consumers learn security behaviors. However, prior work has shown that this advice may not always be helpful and may be less accessible to those with lower internet skill or less education. As a first step toward improving the quality of security advice, we analyzed the readability of 1878 internet security advice documents drawn from crowdsourced search queries and expert recommendations. We measured readability via the commonly used Flesch Reading Ease Score. Our results provide the first characterization, to our knowledge, of the readability of a large corpus of security advice. We find that less than 25% of security advice meets or exceeds the "Standard" (e.g., Reader's Digest) reading level. Preliminary results suggest that security advice is more readable than corporate privacy policies, nearly equally as readable as Wikipedia articles, less readable than health advice, and far less readable than well-known book chapters. Further, we find that ostensibly authoritative advice sources such as those from `.gov` and `.edu` domains score the lowest for readability.**

I. Introduction

Users often rely on security advice – typically, text-based articles or instructions – in order to learn new security behaviors and stay safe online. While there is a vast array of security advice available online, our prior work shows that over 50% of consumers report a serious security or privacy incident, despite reporting taking advice [22]. Further, we have also observed a there is a "digital divide" – a skill- or socioeconomic-based gap – in who takes advice from which sources [21]: lower-skill and lower-education users tend to rely on less authoritative advice sources, and taking advice from those sources is also correlated with a higher rate of security incidents [22]. Thus, security advice may be falling short of effectively imparting consumers with security knowledge, and certain advice sources such as workplace advice and advice distributed in the media may be especially inaccessible for lower-skill or lower-literacy users. One potential reason for this failing may be incomprehensibility (or high reading level) of security advice, relative to the reading level of general consumers. While prior work has examined the readability of warning messages [12] and privacy policies [6], [17], [25], no similar analysis, to our knowledge, has been conducted on security advice.

As a first step at evaluating and potentially improving text-based security advice, we have collected a corpus of 1878 unique security advice articles collected based on user-generated search queries (*search*, 989 documents) as well as recommendations from experts (*expert*, 889 documents). We measured the reading level of this corpus using the Flesch Reading Ease score (FRES), a standard measure of text readability [8].

Our preliminary results indicate that less than a quarter of security advice is at the "Standard" (e.g., similar to Readers Digest) to "Very Easy" (e.g., similar to newspaper comics) reading level. Further, our initial results show that advice from expert sources is very slightly, but statistically significantly easier to read than the collection of search results generated by user queries. Based on our analysis of the reading ease scores of the security advice documents, we find that articles hosted by `.gov` and `.edu` have lower average readability scores than those from other domains (e.g., `.com`, `.org`, `.net`). We also compare the readability of our corpus to the reading level of privacy policies, health advice, and general reading material such as books and Wikipedia articles. Our preliminary results suggest that the reading ease of security advice is significantly higher than that of privacy policies, nearly the same as that of Wikipedia articles, significantly lower than that of health advice, and significantly lower than that of common book chapters.

We also examine how closely the FRES for the security-advice documents correlate with people's self-reported perceptions of the comprehensibility of the documents. We found strong correlation for the search corpus, but not for the expert corpus. We hypothesize that this may be the case due to differences in structure and cohesiveness of official advice documents (e.g., narrative flow, use of bullet points, etc.), which are not well captured by traditional readability metrics [10].

In the rest of this paper, we report on these results in more detail and then discuss our plans for future work building on these preliminary results.

II. Related Work

In this section we provide background on readability metrics and assessments and then review prior work assessing the readability of security- and privacy-related content such as warning messages and privacy policies.

A. Readability Metrics

Readability, broadly defined, is a concept indicating how easy or difficult to read a certain text is for someone [26]. Because reading is a complex phenomenon involving both social [9] and cognitive factors [18], there are many formulas for estimating readability, each focusing on different predictor variables [14]. The most widely used are the Flesch-Kincaid

Grade Level [13], Flesch Reading Ease Formula [8], SMOG formula [16] and the Dale-Chall Readability Formula [4].

These formulas use combinations of sentence length, word length, and word familiarity to predict readability, with an underlying assumption that longer sentences and words — which often co-occur with complex syntax — indicate greater reading difficulty [5], [7]. Additionally, since shorter words tend to be more common than longer ones in English [24], longer words are considered less likely to be familiar to the reader. While these assumptions do not account for individual readers' vocabulary and reading experience, simple metrics such as sentence and word length can provide a useful first step in assessing readability.

In this work, we use a combination of readability formulasand self-reported readability perceptions as a first step toward assessing whether a given security-advice document is easy to read, easy to understand, usable, and ultimately actionable.

B. Security and Privacy Readability

Harbach et al. evaluated the readability of warning message descriptions using multiple traditional measures of text readability including the Flesch-Kincaid readability test and Gunning-Fog Index [12]. They evaluated the accuracy of these metrics using a Cloze test [3], with participants filling in the blank for certain parts of the warning messages. The authors conclude that readability metrics have promise for improving warning message design, but also raise concerns that metrics designed for grade-school readability may not be as accurate for adult populations.

McDonald et al. measured the readability of six companies' privacy policies using multiple metrics including word counts, passive word proportions, and the Flesch-Kincaid test [17]. They then conducted a large-scale survey to evaluate how well this readability measure matched with user comprehension, measured by accuracy and time taken when answering questions about the policy. They found no correlation between readability metrics and observed comprehension. Singh et al. examined readability of privacy policies in mobile environments, using Cloze tests with policies from 10 popular websites. They found that desktop environments are potentially more suitable than mobile for reading and understanding privacy policies. Ermakova et al. used comprehension questions to evaluate participants' understanding of privacy policies and found that those who fully comprehended a policy exhibited more trust toward that company [6].

Finally, Rader and Wash used LDA topic modeling to explore the content of security advice drawn from news articles, websites, and personal stories of security experiences [20]. They found that news articles focus on the consequences of security-related attacks, websites focus on attack methodology, and stories focus on the people who conduct attacks.

We build on this prior work by examining readability for security advice specifically and by evaluating a significantly larger corpus than in previous security and privacy readability studies. This allows us to highlight areas for improvement across a broad swatch of consumer-relevant information sources.

III. METHOD

In this section we describe collecting our corpus of security advice, evaluating and validating the reading level of this corpus, comparing to other corpi, and the limitations of our analysis.

A. Advice Corpus Generation

We used two approaches to collect text-based security advice: (1) We collected search queries for security advice from crowdworkers and scraped the top 20 articles surfaced in response to their queries, and (2) We collected a list of authoritative security-advice sources from computer security experts and librarians and scraped articles provided by those resources.

1) User Search Query Generation: We recruited 50 participants from Amazon Mechanical Turk to write search queries for security advice. To obtain a broad range of queries, we used two different surveys. The first survey asked participants to list three digital security topics they would be interested in learning more about, then write five search queries for each topic. Participants in the second survey were instructed to read a security-related news article, then asked if they were interested in learning more about digital security topics related to the article. If the participant answered yes, they were prompted to provide three associated search queries. Participants who answered no were asked to read additional articles until they reported interest; if no interest was reported after five articles, the survey ended without creating queries. Twenty-five people participated in each survey and were compensated $0.25 (first survey) or $0.50 (second survey). Participants completed these tasks in four minutes or less, and our protocol was approved by the University of Maryland IRB.

From these surveys, we collected 140 security-advice search queries. After manual cleaning by the researchers (removing duplicates and off-topic queries such as "fashion websites"), 110 queries remained. Examples of these queries include, "How safe is my encoded information online," "how to block all windows traffic manually allow," and "common malware."

We then aggregated the top 20 Google search results for each query using parameterized GET requests, yielding a cumulative URL index that preserved rank order of search results. We then used the Diffbot API [1] to parse and sanitize HTML body elements within each identified site, merging all such elements to create one text file per site. Our collection was conducted in September 2017.

In total, the resulting search corpus includes 990 documents. Examples of advice in this corpus include Apple and Facebook help pages and security advice, news articles from Guardian, the New York Times, and other media sources, and advice or sales material from McAfee, Avast, or Norton.

2) Expert Advice List: To collect a corpus of online security advice recommended or consulted by experts, we asked 10 people for a list of websites from which they personally get security advice or which they would recommend that others go to find advice. These included five people holding or pursuing a Ph.D. in computer security, two employees of our university's

IT department who have security-related job responsibilities, and three librarians from our universities libraries.

Two researchers manually visited each recommended website and collected URLs for referenced advice articles. Manual collection was required, as many of these expert sites required hovering, clicking images, and traversing multiple levels of sub-pages to surface relevant advice. (An initial attempt to use an automated crawl of all URLs one link deep from each page missed more than 90% of the provided advice.) As with the search corpus, we used the Diffbot API to parse and sanitize body elements.

The resulting expert corpus includes 894 documents. Exemplar pieces of advice in this corpus include U.S. CERT pages, FBI articles, and articles from Bruce Schneir's blog. Only five documents of advice were in both the expert and search corpi: an article from the FTC on malware, a veracrypt page, an article on passwords from security-in-a-box, an article from `safetynetkids.org`, and an article from `axantum.com`.

3) Data Cleaning: The raw data scraped from our advice sources contained uncommon words and jargon, typos, contractions, embedded URLs, and other elements that are potentially problematic for our automated readability tools (described below). To address this, we tested several cleaning approaches on a 10% random sample of documents: (1) removing any words not recognized by the pycharm dictionary [1]; (2) removing URLs beginning with *http:* or *https:*; and (3) manual cleaning by two members of the research team. We then compared the FRES scores for the raw files and all three cleaning approaches. We found that approach (2) (removing URLs) exhibited strong correlation with manual cleaning ($r = 0.98$), so for efficiency we applied approach (2) to the entire corpus.

Finally, after computing FRES metrics on the corpi (as discussed below), we manually checked the outliers (144 documents that scored below 30 or above 80 on FRES) for quality. As a result we removed 11 irrelevant documents from the search corpus and five from the expert corpus. Examples included text output of a video that contained all numbers and the terms of service for a website. The resulting final corpus includes 1878 pieces of security advice; 989 in the search corpus and 889 in the expert corpus.

B. Readability Measurement and Analysis

We used the `textstat` library [2] to measure the FRES reading level of our corpi. To check that these scores matched relatively well with real users' perceptions of reading ease, we randomly selected three documents from each corpus in each of the following FRES ranges: 0-25, 25-55, and 55+, and asked MTurkers to answer the following item on a 5-point Likert scale for each document: "How easy was it to understand this document?" We also asked them to tell us which of the three documents they read they found the easiest to read. We recruited 400 MTurkers to answer these questions for the selected documents (each MTurker read three of the 18 documents). We used Spearman correlations to assess the relationship between FRES scores and the perceptions of reading ease measured with Likert scales.

We also sought to compare FRES scores for our corpi to each other and to other kinds of documents. To do this, we applied Student's T-tests[2] to make pairwise comparisons between our corpi, and to compare each of our corpi separately to four comparison corpi:

- a corpus of six corporate privacy policies collected by McDonald et al. [17];
- 1083 documents of health advice provided by U.K. hospitals to patients and families in palliative care, collected by Payne et al. [19];
- 1,708 chapters sampled from 3,206 books scraped by the Gutenberg project and cleaned by Lahiri [15];
- and a random sample of 999 Wikipedia articles we drew from the set of 20,000 articles scraped and cleaned by Shaol [23].

We applied Bonferroni-Holm correction to compensate for potential multiple-testing error.

Note that for the palliative care dataset we use a χ^2 Proportion test (instead of a T-test) to compare the distribution of FRES scores in our dataset to the distribution reported by Payne et al. for their health leaflets, as these authors released neither their raw data nor state the mean and standard deviation (SD) of their sample.

C. Limitations

While we attempted to collect a broad corpus of advice (including two methods of using MTurkers to generate advice queries as well as expert recommendations), our techniques were not exhaustive, and may not be entirely representative. Further, there may be errors in our corpus, ranging from documents that do not appropriately fit our criteria of text-based security advice to typos. We attempted to mitigate this via semi-automated data cleaning and manual inspection of outliers, as described in Section III-A3 above; however, there are likely still errors that may affect our results. We believe, however, that our semi-cleaned corpus is sufficiently accurate and representative to support the preliminary results we present here.

Finally, our comparative analysis is limited by the data published by the researchers to whom we compare. For privacy policies, the sample size is relatively small, and for the health leaflets we can compare only by distribution across FRES ranges rather than by means. We present these limited comparisons for context, but focus the majority of our analysis on our own data.

IV. Results

We report our preliminary results about the readability of the advice in our corpi, including comparisons within our corpi as well as comparisons to other kinds of documents.

[1] https://www.jetbrains.com/pycharm/

[2] Visual inspection of quantile-quantile plots indicated normality.

A. Security-Advice Readability

Less Than 25% of Security Advice Readable at the Standard Level. FRES scores range from 0 to 100, with 100 easiest to read. We find that the mean FRES for our corpus of security advice is 48.7 (SD=14.4), between the high school and college readability levels. This is considered difficult text for the regular reader (newspapers and books tend to score in the 60-80 range). Further, only 22% of the advice achieves a "Standard" (middle school) or easier reading level.

Perceptions of Reading Ease Correlate More Strongly with FRES for User-Query Generated Advice. To validate that our FRES measurements aligned with users' perception of how easy it was to read the documents, we first examined the correlation between FRES scores and self-reported five-point Likert perceptions (very easy to very hard) for 18 documents read by an average of 66 MTurk workers per document. Across all documents, participants reported an average rating of 2.97 (neutral, SD = 1.01). These ratings correlate significantly with FRES scores within the search corpus ($r = 0.79$ *(large)*, $p = 0.011$), but not within the expert corpus ($p = 0.34$). However, when removing the perception rating for one outlier, the expert correlation becomes significant and large ($r = 0.56$, $p = 0.023$).

We also asked participants to rate which of the three documents they read was easiest to read. Each participant saw one document each from the three FRES groups 0-25, 25-55, and 55+; these were drawn from three randomly-selected documents from each corpus and FRES category. We similarly find a significant correlation between the frequency that a document is rated easiest to read and FRES for the search corpus ($r = 0.70$ (large), $p = 0.038$) but no significant correlation for the expert corpus ($p = 0.29$). From trend observations, we in fact find no observable relationship between FRES and these comparative ease ratings for the expert corpus. This suggests that FRES may not be a sufficient evaluation metric for the expert advice, but we must sample more documents to further evaluate this possibility. Further, we discuss potential alternatives to FRES measurements in Section V below.

B. Comparing Security-Advice Sources

Within the search corpus, the mean FRES is 47.8 (SD=15.2), while the mean FRES for the expert corpus is 49.6 (SD=12.6). This difference is statistically significant ($p = 0.007$), with the documents generated by user queries scoring slightly more difficult to read: 52% of the advice collected based on user queries scored less than 50 ("Difficult" or worse, equivalent to a corporate report, grade level "in College" or higher) compared to 49% of the advice collected based on expert recommendations. The distribution of FRES for the expert advice seems to be weighted toward the extremes, while the FRES for the search corpus is more evenly distributed.

`.gov`, `.edu` Websites Score Lower Than Others We also examined whether certain sources (e.g., government websites) generated more or less readable advice than others. We first examined whether websites within particular top-level domains

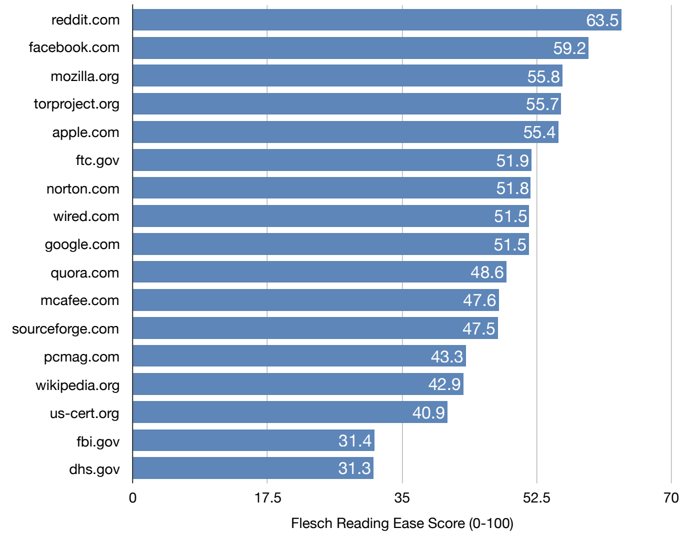

Fig. 1. Mean FRES for a sample of advice providers (specific websites) across both the expert and search corpi.

(TLDs) (e.g., `.net`, `.gov`, `.com`) scored higher or lower than others (Figure 1). We report here on only those TLDs associated with at least 10 documents (1% or more) in our sample. Because of the small sample sizes, we do not conduct hypothesis tests to compare these results, but only report on trends. Within the expert corpus, we find that `.co`, `.com`, and `.org` all have FRES scores above 50, while `.net` averaged 49.4 (SD=12.4) and `.gov` averaged 40.8 (SD=15.4). Within the search corpus, only `.net` documents scored above 50 (mean=53.6, SD=12.2); `.com`, `.co`, `.org`, and `.edu` averaged between 40 and 50, and `.gov` averaged 34.4 (SD=16.7). It is interesting to note that across both of our corpuses, `.gov` sites scored the lowest for readability, more than 5 points lower than the next lowest TLD we observed. `.edu` documents also had low FRES, scoring an average of 41.5 within the search corpus. Within the expert corpus, there were only three `.edu` documents, but these averaged a similarly low 39.5 in FRES.

Wide Variation in Readability by Advice Provider. Briefly, we review the readability of specific websites (we consider only those websites with five or more documents in our corpus). Within the search corpus, we find that `apple.com` and `reddit.com` average above 60 ("Standard" readability equivalent to the middle-school level), while `norton.com`, `wired.com`, `quora.com`, `mcafee.com`, and `google.com` each average within 2 points of 50 ("Fairly Difficult" readability equivalent to the high school level). `pcmag.com` and `wikipedia.org` rate in the 40s ("Difficult," college student).

Among the expert advice, documents from `facebook.com`, `mozilla.org`, and `torproject.org` are on average the most readable (FRES between 55-59, "Fairly Difficult" to "Standard", middle to high school level), while advice from `ftc.gov`, `google.com`, and `sourceforge.com` each averages near 50 ("Fairly Difficult", high school). Finally, advice from `us-cert.org`

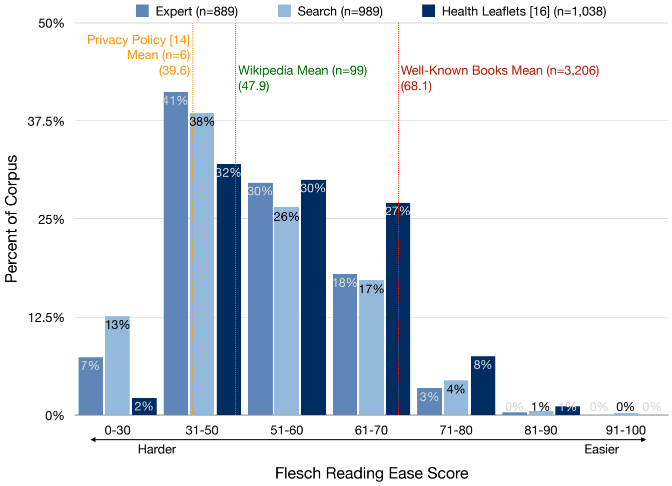

Fig. 2. Comparing the distribution of FRES across security advice from our two security-advice corpi as well as a collection of pallative care leaflets analyzed by Payne et al. [19]. Figure also includes, for reference, the mean FRES for privacy policies collected by McDonald et al. [17], 1000 Wikipedia articles, and 1,708 Gutenberg book chapters.

averages 40.9 ("Difficult," college student), while `fbi.gov` and `dhs.gov` average near 31 ("Very Difficult", readability equivalent to a scientific journal article). These results are illustrated in Figure 1.

C. Security Advice vs. Other Content

Next we detail preliminary findings comparing the FRES for our corpi to four other types of content: privacy policies, wikipedia articles, health advice documents, and chapters of well-known books. Figure 2 illustrates our results.

Better than Privacy Policies. McDonald et al. found that all six privacy policies (from Disney, Microsoft, Nextag, IBM, Walmart, and O'Reilly) that they analyzed had FRES under 50 [17]. The mean of the six policies was 39.6 (SD=5.86), which is significantly lower than the means of our expert ($p = 0.010$) and search documents ($p = 0.020$).

About the Same as Wikipedia. The mean FRES for the 999 Wikipedia articles was 47.3, SD=12.0, which was not significantly different from the mean FRES for the search corpus ($p = 0.402$) but was significantly lower than the expert mean ($p < 0.001$). Nine percent of the Wikipedia articles had an FRES under 30 ('Very Difficult', requires a college degree, example: Legal Contract, Academic Journal) compared to 13% of search advice and 7% of expert advice.

Worse than Health Advice. Compared to the corpus of pallative care leaflets collected by Payne et al. [19], both our corpi are significantly harder to read (χ^2 = 3347, $p < 0.001$ for expert; χ^2 = 3804, $p < 0.001$ for search). Ten percent of our security advice corpus had an FRES under 30 compared to 2% of the health leaflets.

Far Worse than Well-Known Book Chapters. Finally, the FRES mean for the corpus of book chapters is 68.1 (SD=12.5), with 78% of book chapters having an FRES of 60 or above compared to 22% of expert and user-query generated security advice, respectively. Thus, well-known book chapters are significantly easier to read than the security advice recommended by our experts ($p < 0.001$) and generated by the user queries we collected ($p < 0.001$).

V. Discussion and Future Work

Our preliminary results indicate that people's perceptions of the ease of understanding and reading a document correlate strongly with FRES scores for nine documents sampled from the advice obtained using MTurkers' search queries, but not for documents recommended by experts. If these results prove to hold for larger subsamples, we hypothesize it may be because the expert articles are less narrative and have more complex structure, such as bullet points. These attributes may actually make documents easier to read, but are not well accounted for in metrics such as FRES. This suggests that to fully evaluate security advice, additional metrics may be needed. In future work, we hope to develop a new such measure, synthesized from some of the nearly 100 readability, cohesion, and word difficulty metrics provided by the Cohmetrix readability tool [11]. The new measure may also require specialized word-difficulty value for security- and privacy-specific jargon.

Additionally, while we evaluated people's perception of reading ease, we did not assess whether they are able to take action on the supplied advice, or whether doing so would be beneficial in practice; we hope to include this additional evaluation in future work as well. By evaluating the quality of available security advice along these three axes (readability, actionability, usefulness), we hope to both understand the current status of the advice ecosystem and provide guidance for how to improve it, ultimately improving consumers' ability to learn useful security behaviors and protect themselves.

VI. Acknowledgements

We thank Marine Carpuat at the University of Maryland, Danielle Clewley at the Institute for Intelligent Systems in the University of Memphis, and the anonymous reviewers for their assistance and feedback. This material is based upon work supported by a UMIACS contract under the partnership between the University of Maryland and DoD, the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1322106, and a Facebook Fellowship.

References

[1] Diffbot.
[2] S. Bansal. textstat.
[3] J. R. Bormuth. Cloze test readability: Criterion reference scores. *Journal of educational measurement*, 5(3):189–196, 1968.
[4] J. S. Chall and E. Dale. *Readability revisited: The new Dale-Chall readability formula*. Brookline Books, 1995.
[5] E. Dale and R. W. Tyler. A study of the factors influencing the difficulty of reading materials for adults of limited reading ability. *The Library Quarterly*, 4(3):384–412, 1934.
[6] T. Ermakova, A. Baumann, B. Fabian, and H. Krasnova. Privacy policies and users? trust: Does readability matter? 2014.
[7] L. Feng, N. Elhadad, and M. Huenerfauth. Cognitively motivated features for readability assessment. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 229–237. Association for Computational Linguistics, 2009.

[8] R. Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

[9] J. P. Gee. Three paradigms in reading (really literacy) research and digital media. In *Reading at a Crossroads?: Disjunctures and Continuities in Current Conceptions and Practices*. Taylor and Francis Inc., 2015.

[10] A. C. Graesser, D. S. McNamara, and J. M. Kulikowich. Coh-metrix: Providing multilevel analyses of text characteristics. *Educational researcher*, 40(5):223–234, 2011.

[11] A. C. Graesser, D. S. McNamara, M. M. Louwerse, and Z. Cai. Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2):193–202, May 2004.

[12] M. Harbach, S. Fahl, T. Muders, and M. Smith. Towards measuring warning readability. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 989–991, New York, NY, USA, 2012. ACM.

[13] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch, 1975.

[14] W. Kintsch and D. Vipond. Reading comprehension and readability in educational practice and psychological theory. *Perspectives on learning and memory*, pages 329–365, 2014.

[15] S. Lahiri. Complexity of Word Collocation Networks: A Preliminary Structural Analysis. In *Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 96–105, Gothenburg, Sweden, April 2014. Association for Computational Linguistics.

[16] G. H. Mc Laughlin. Smog grading-a new readability formula. *Journal of reading*, 12(8):639–646, 1969.

[17] A. M. McDonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor. *A Comparative Study of Online Privacy Policies and Formats*, pages 37–55. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[18] S. G. Paris and S. A. Stahl. *Children's reading comprehension and assessment*. Routledge, 2005.

[19] S. Payne, S. Large, N. Jarrett, and P. Turner. Written information given to patients and families by palliative care units: a national survey. *The Lancet*, 355(9217):1792, 2000.

[20] E. Rader and R. Wash. Identifying patterns in informal sources of security information. *Journal of Cybersecurity*, 1(1):121–144, 2015.

[21] E. Redmiles, S. Kross, and M. L. Mazurek. How i learned to be secure: a census-representative survey of security advice sources and behavior. In *CCS*, 2016.

[22] E. M. Redmiles, S. Kross, and M. L. Mazurek. Where is the digital divide?: A survey of security, privacy, and socioeconomics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 931–936. ACM, 2017.

[23] C. Shaoul. The westbury lab wikipedia corpus. *Edmonton, AB: University of Alberta*, 2010.

[24] B. Sigurd, M. Eeg-Olofsson, and J. Van Weijer. Word length, sentence length and frequency–zipf revisited. *Studia Linguistica*, 58(1):37–52, 2004.

[25] R. Singh, M. Sumeeth, and J. Miller. Evaluating the readability of privacy policies in mobile environments. *Developments in Technologies for Human-Centric Mobile Computing and Applications*, 56, 2012.

[26] M. Zamanian and P. Heydari. Readability of texts:: State of the art. *Theory and Practice in Language Studies*, 2(1):43, 2012.