



CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022

Audio delivery of health information: An NLP study of information difficulty and bias in listeners

^aArif Ahmed*, ^aGondy Leroy, ^aHan Yu Lu, ^bDavid Kauchak, ^aJeff Stone, ^aPhilip Harber, ^aStephen A. Rains, ^aPrashant Mishra, ^aBhumi Chitroda

^aThe University of Arizona, Tucson 85721, U.S.A
^bPomona College, 333N College Way, Claremont 91711, U.S.A

Abstract

Health literacy is the ability to understand, process, and obtain health information and make suitable decisions about health care [3]. Traditionally, text has been the main medium for delivering health information. However, virtual assistants are gaining popularity in this digital era; and people increasingly rely on audio and smart speakers for health information. We aim to identify audio/text features that contribute to the difficulty of the information delivered over audio. We are creating a health-related audio corpus. We selected text snippets and calculated seven text features. Then, we converted the text snippets to audio snippets. In a pilot study with Amazon Mechanical Turk (AMT) workers, we measured the perceived and actual difficulty of the audio using the response of multiple choice and free recall questions. We collected demographic information as well as bias about doctors' gender, task preference, and health information preference. Thirteen workers completed thirty audio snippets and related questions. We found a strong correlation between text features lexical chain, and the dependent variables, and multiple choice response, percentage of matching word, percentage of similar word, cosine similarity, and time taken (in seconds). In addition, doctors were generally perceived to be more competent than warm. How warm workers perceive male doctors correlated significantly with perceived difficulty.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2022

Keywords: Health information; Health literacy; audio information delivery; Text features; Natural Language Processing; NLP

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: arifahmed@arizona.edu

1. Introduction

Improving health literacy is a major national goal in the U.S. Poor health literacy can cost patients thousands of dollars [5]. Text is a cost-effective and efficient medium for delivering healthcare information and has been the most common practice for decades. In comparison, interactive videos and other multimedia tools are relatively new. In recent years, audio information delivery is becoming increasingly popular with patients. The use of virtual assistants and smart speakers for health-related queries is increasing. By 2021, nearly 91M smart speakers (i.e., Siri, Alexa) will be used in the U.S. [2, 3], and the adoption rate has increased annually by 30-40%. Hospitals are incorporating smart speakers into their systems, and patients are using the technology to communicate with clinicians and make queries [4]. Generally, audio is generated by recording speech and then stored as an audio file [11]. Alternatively, audio can also be generated by converting text audio, e.g., Microsoft Azure, Amazon Web Services (AWS)) have support to automatically generate audio from text. The audio can be generated as a male/female voice, accents can be applied, and emotions can also be added.

In addition to simplified audio information, several other features may influence audio information. Research focusing on in-person communication provides guidance, e.g., the perception of credibility plays a critical role in information comprehension [12]. Source credibility encompasses perceptions that the person is an expert and trustworthy [13]. Factors that foster those perceptions enhance the effectiveness, and factors that reduce them diminish the impact of information [14]. These principles extend to the dissemination of health information [15]. Further critical factors of stereotypes are gender, race/ethnicity, and age. According to the stereotype content model [16], the content of most stereotypes describes a group as warm to cold and as competent to incompetent. Regarding source credibility, warmth conveys trust, and competence conveys expertise [17, 18]. When patients view professionals as less of an expert, they may pay attention which can inhibit retention of health information.

One implication of biases is that during an audio delivery of a message, a cue may activate a stereotype – overgeneralized belief about the traits and attributes of members of social and ethnic groups, leading the target to be judged and treated in a stereotype-consistent manner [19]. Specifically, when a patient "hears" that the speaker is female, elderly, or with a regional or international accent, this can activate negative stereotypes, which, if it biases perceptions of the speaker's expertise or trustworthiness, will reduce the effectiveness of the information [20]. Thus, the biases that patients hold due to stereotypes moderate the impact of audio delivery of health information.

The main goal of this paper is to examine the first set of features that contribute to the perceived and actual difficulty of audio health information. We will also examine how patient biases towards doctors play a role in the perception of health information.

2. Literature Review

The U.S. National Action Plan is to enhance health literacy which specifies the clarity of health-related information, and the Plain Writing Act urges lucidity in communications. There are many existing guidelines for text, but none seem to include audio. For example, in the Plain Language Summit in 2019, they showed that plain language can be easily understood when people read it for the first time [8], but audio was not mentioned. With the growing popularity of smart speakers and virtual assistants, this promising method must be studied.

Researchers have been trying to figure out how to measure the difficulty of a text. Readability formulas (e.g., Flesch-Kincaid) are frequently used to assess the readability of a given text [26]. However, they do not include the linguistic structure that also contribute to text difficulty [27]. Recent studies have shown that Flesch-Kincaid is not useful for text simplification evaluation metrics. [28] Research focused on text simplification shows semantic analysis to reduce the text difficulty [9], but they do not incorporate audio. Another study showed the association of text difficulty with the number of nouns, text lengths, and the number of verbs [10], while another suggested that the lexical chain feature can be used to differentiate between easy and difficult texts [26].

While it is important to examine the text and audio features that contribute to the difficulty of health information, it is equally important to examine human factors that influence the retention of audio information. One factor includes biases of listeners and how it affects the credibility of the audio source. Bias occurring during an audio delivery of health information may activate negative stereotypes toward certain groups of people. For instance, when a listener hears health information delivered by someone who is the subject of negative stereotypes, it may bias the listener's perception of the speaker's expertise or trustworthiness and thus reduce the effectiveness of the health information.

The stereotype content model posits that higher-status groups, such as doctors, are generally perceived to be high in competence and lower in warmth, while lower-status groups are perceived to be lower in competence and higher in warmth [21]. In most cultures, men possess greater structural power than women and belong to high-status groups across economic, political, and social institutions. They are perceived as higher in competence but lower in warmth than women [22]. The stereotypes toward men and women can be defined as paternalistic and envious stereotypes, respectively [17]. Paternalistic stereotypes typically describe a group that is pitied and not respected and carry overtones of compassion, sympathy, and tenderness. This stereotype is evident in gender, age, race, and accent prejudice. Women, the elderly, racial minorities, and those with a foreign accent are often the targets of paternalistic stereotypes. Interestingly, non-traditional women, such as career women, are targets of envious stereotypes and are perceived as competent but not warm. As career women, female doctors are seen as highly competent and hardworking, are envied and seen as too ambitious, but simultaneously characterized as cold and aloof.

Given the biases that patients hold due to the stereotypes that laypeople hold towards certain groups, it is important to understand its influence and impact on the delivery of health information.

3. Methods

3.1 Study Design

Demographic and Bias Questions

Standard questions assessed respondents' age, gender, race, educational attainment, and amount of English spoken at home. We also included a six-item measure of need for cognition to evaluate participants' baseline tendency to engage in information [23]. Ratings were made on a 7-point scale, with higher scores indicating a greater propensity to think about information without being prompted.

The competency and warmth scales were taken from a previous study [24]. Participants were also asked to rate the competence and warmth of physicians on a scale of 1 to 7. There were four groups of physicians: male, female, racial minority, and those with a foreign accent. Seven attributes related to competency (four) and warmth (three) were captured. Specifically, attributes related to competency include competent, unintelligent, unskilled, and not trained well. Attributes related to warmth include friendly, warm, and snobby. In this pilot, the snobby attribute was not measured for doctors with a foreign accent.

Audio/Text Generation

We collected health-related text on common diseases from the Mayo Clinic website. We used ICD-10 (International Statistical Classification of Diseases and Related Health Problems) codes to list 147 diseases. Subsequently, we randomly selected ten text snippets from these different topics and generated audio snippets using those text snippets. We used the Microsoft Azure text-to-audio tool to generate the audio snippets. Microsoft Azure has a wide variety of features for audio generation from texts. It supports various accents and male/female voices for audio generation. In this pilot study, we used the U.S. male voice.

For all text used in our study, we calculated text features that we have found to influence actual and perceived difficulty in prior work [10], i.e., Google frequency, percentage of nouns, percentage of verbs, percentage of adverbs, percentage of function words, percentage of adjectives, and several lexical chain features. Table 1 provides the details of the text features that were calculated.

Table 1. Text features for pilot study

No	Topic	Google frequency	Percentage of Nouns	Percentage of Verbs	Percentage of Adverbs	Percentage of Function Words	Percentage of Adjective	Number of Lexical Chains
1	Acne	5426733903	60	10	0	10	20	1
2	Microcephaly	5389831840	32.10	16.24	2.54	13.20	17.259	14
3	Brucellosis	8930577709	51.85	11.11	3.70	11.11	14.81	0
4	Diverticulitis	5992601531	43.10	9.48	1.72	17.24	18.10	8
5	Measles	6225998400	24.63	15.94	4.35	13.04	21.73	5
6	Cirrhosis	6421494525	42.85	9.183	4.08	20.41	15.30	4

7	Scabies	5617736519	29.63	14.81	7.40	22.22	18.51	1
8	Emphysema	6069903857	36.67	28.33	0	18.33	6.66	4
9	Dystonia	6563083195	43.54	15.65	1.36	12.93	14.29	4
10	Pseudogout	6040037769	36.36	18.18	2.72	15.45	18.18	7

Question generation

We chose ten topics (diseases) randomly from the text corpus. Then we selected text snippets from these topics and created four types of multiple-choice questions about the cause, risk, signs or symptoms, and medication of the disease.

Perceived and Actual Difficulty Measures

Participants were asked to rate on a scale of 1 to 5 how difficult they perceived each audio to be. A score of 1 indicates very difficult, and a score of 5 indicates very easy. An average perceived difficulty score was calculated for each participant, with higher scores indicating lower perceived difficulty, meaning audio was easier to understand.

For measuring the actual difficulty of the given audio information, we used the response of the multiple choice and a free recall question asking participants to write what they remembered from the text/audio. For free recall responses, we measured the percentage of similar words (worker's response similar to the actual information), the percentage of matching words (exact match of the worker's response to the actual information), and cosine similarity between the worker's response and the actual information.

3.2 Study Procedure

Participants were recruited from Amazon Mechanical Turk. They were first asked demographic questions and bias-related questions. We used Qualtrics for these questions. After successfully submitting the demographic and bias questions, they received a six-digit code to enter on AMT to continue the study. They were then invited to complete as many human intelligence tasks (HITs) as were available (the common approach on AMT). The workers received \$0.25 per HIT. Each HIT was presented to three different AMT workers and consisted of one audio snippet played, and then workers answered the perceived and actual difficulty questions.

4. Results

4.1 Demographic and bias result

Table 2 below shows participants' demographic information. Thirteen workers completed the thirty HITs. 54% of the participants were female. Participants were given the option to choose multiple races, and the majority identified as white (77%), with the second largest group being Asian (15%). The education level of participants was skewed towards most having a high school diploma as their highest degree (62%), followed by a bachelor's degree (23%).

Table 2: Participant Demographic Information

Variable	Choice	Count (%)
Sex	Female	7 (54)
	Male	5 (38)
	Other	1 (8)
Race (Multiple options possible)	American Indian or Alaska Native	1 (8)
	Asian	2 (15)
	Black or African American	1 (8)
	White	10 (77)
Education Level	High school diploma	8 (62)
	Associate degree	1 (8)
	Bachelor's degree	3 (23)
	Master's degree	1 (8)

Age	Younger than 30 years old	4 (31)
	31-40 years old	4 (31)
	41-50 years old	2 (15)
	51-60 years old	2 (15)
	71 years old or better	1 (8)
Language Spoken at Home	Mostly English	10 (77)
	Only English	3 (23)
Need for Cognition	High vs. Low	
	Low (1) vs. High (7)	5.31

Participants’ age levels were well represented, with the majority being either younger than 30 years old (31%) or between 31 to 40 years old (31%). All participants spoke mostly English at home (77%).

The mean level of need for cognition ($M = 5.31, SD = 1.25$) was above the scale midpoint, indicating that participants generally had a modest baseline tendency to think about information without prompting.

4.2 Participant Biases

Tables 3 and 4 below show the competence and warmth responses by participant demographics. Overall, doctors

Table 3: Competency responses by participants demographics

Participants by	Doctors who are				
	Male	Female	Racial Minority	Foreign Accent	
Gender					
	Male	5.95	6.10	6.15	6.15
	Female	5.00	6.18	6.11	6.29
	Other	5.00	4.75	6.00	4.75
Age					
	Younger than 30 years	5.63	6.00	6.00	6.00
	31 – 40 years	5.94	5.94	6.00	6.19
	41 – 50 years	6.00	5.88	6.50	5.88
	51 – 60 years	6.00	6.00	6.00	6.00
	71 years or better	7.00	7.00	6.50	7.00
Race					
	Asian	5.50	6.25	6.25	6.25
	Asian/White	6.00	6.00	6.00	6.00
	American	5.75	5.75	6.00	6.00
	Indian/Native Alaskan				
	Black or African American	6.00	6.75	6.75	6.75
	White	6.00	5.97	6.06	6.06
Education					
	High School diploma	6.31	6.38	6.34	6.72
	Associate degree	6.00	6.00	6.00	6.00
	Bachelor’s degree	5.00	4.92	5.33	4.92
	Master’s degree	5.75	6.75	6.75	5.00
Language					
	Mostly English	5.08	5.33	5.42	5.42
	Only English	6.20	6.25	6.33	6.33

were judged as more competent than warm. For male doctors, there was a significantly higher competence score ($M = 5.94, SD = 0.84$) than warmth score ($(M = 4.59, SD = 1.33), t(12) = 4.35, p < .05$). For female doctors, results also indicated a significantly higher competence score ($M = 6.04, SD = 0.93$) than warmth score ($(M = 5.23, SD = 1.11), t(12) = 2.83, p < .05$). Racial minority doctors also had higher competence score ($M = 6.12, SD = 0.81$) than warmth score ($(M = 5.05, SD = 1.08), t(12) = 3.62, p < .05$); and doctors with a foreign accent as well [competence ($M = 6.12, SD = 0.98$), warmth ($(M = 5.00, SD = 1.17), t(12) = 3.37, p < .05$)].

4.3 Perceived and Actual Difficulty

There were no significant correlations found between the perceived difficulty measures and the text features. The average accuracy for multiple choice questions was 84%. The average percentage of matching words was 17%, the average percentage of similar words was 22%, and the average percentage of cosine similarity was 16%.

The number of lexical chains strongly correlates with actual difficult measures. The number of lexical chains positively correlate with the percentage of matching words and time taken and negatively correlates with multiple choice response, percentage of similar words, and cosine similarity, which can be considered a factor that can be used to measure text difficulty. Table 5 contains the correlation results of the text features with the dependent variables.

Table 4: Warmth responses by participants' demographics

Participants by	Doctors who are				
	Male	Female	Racial Minority	Foreign Accent	
Gender					
	Male	4.67	5.60	5.00	5.40
	Female	4.62	5.05	5.14	4.71
	Other	4.00	4.67	4.67	5.00
Age					
	Younger than 30 years	4.50	4.67	4.50	4.75
	31 – 40 years	3.83	5.58	5.17	4.63
	41 – 50 years	5.50	5.83	5.83	6.00
	51 – 60 years	4.17	4.17	4.17	4.25
	71 years or better	7.00	7.00	7.00	7.00
Race					
	Asian	4.00	3.67	4.00	4.00
	Asian/White	4.00	4.00	4.00	4.00
	American	2.33	6.00	4.00	5.00
	Indian/Native Alaskan				
	Black or African American	5.00	6.00	5.00	6.00
	White	4.93	5.37	5.41	5.11
Education					
	High School diploma	4.79	5.75	5.46	5.31
	Associate degree	4.33	4.33	4.33	4.50
	Bachelor's degree	4.00	4.22	4.22	4.33
	Master's degree	5.00	5.00	5.00	5.00
Language					
	Mostly English	3.44	4.56	4.00	4.33
	Only English	4.93	5.43	5.37	5.20

4.4 Bias with Perceived Difficulty

We calculated a Pearson correlation coefficient to assess the linear relationship between average perceived difficulty and warmth scores for male doctors. There was a moderately positive correlation between the two variables, ($r(11) = .56, p = .048$). The warmth of male doctors correlates with the perceived difficulty of the text. Specifically, the warmer participants found the doctors, the less difficult they perceived the information.

Table 5. Correlation of text features with dependent variables

	Google frequency	Percentage Noun	of	Percentage Verb	of	Percentage Adverb	of	Percentage of Function Word	Percentage of adjective	Number of Lexical Chains
Google frequency	1.00									
Percentage Noun	of 0.32	1.00								
Percentage Verb	of -0.19	-0.47		1.00						
Percentage Adverb	of 0.13	-0.52		-0.25		1.00				
Percentage Function Word	of -0.28	-0.46		0.17		0.48		1.00	-	
Percentage adjective	of -0.25	-0.09		-0.63		0.40		-0.25	1.00	
Number of Lexical Chains	of -0.44	-0.41		0.15		-0.17		-0.03	0.09	1.00
Perceived difficulty score	0.04	-0.08		0.00		0.07		0.18	0.09	-0.12
Multiple choice response	0.18	0.17		0.04		-0.17		-0.12	0.00	-0.30 **
Percentage of Matching word	of 0.23	0.10		-0.21		0.40 **		0.05	0.16	0.43 **
Percentage similar word	of 0.11	0.15		-0.15		0.25		-0.06	0.16	-0.39 **
Cosine Similarity	0.57 **	0.54**		-0.16		-0.02		-0.35	-0.11	-0.59 **
Time taken (in Seconds)	(in -0.02	-0.06		-0.06		-0.15		-0.06	0.01	0.53 **

* Correlation is significant at the 0.05 level (1-tailed).

** Correlation is significant at the 0.01 level (1-tailed).

5. Conclusion

We are developing an audio corpus for healthcare-related text with information about diseases' cause, risk, signs, and medication. We found a significant correlation between one text feature (lexical chain) and actual difficulty. The results are in line with earlier work focusing on text simplification. As we have limited data points for this study, we expect to find additional interesting findings in our next full-length study.

Our study results also support the stereotype content model that higher-status groups are perceived as higher in competence than warmth. Our findings suggest that doctors, generally seen as a group of high status, were rated more competent than warm. This is consistent across all doctor groups, including females, racial minorities, and those with a foreign accent.

We looked at the correlation and perceived difficulty of the audio. Male doctors' competence and warmth scores were correlated with the participants' perceived difficulty with the audio. We found a moderate correlation where the colder one finds the doctor, the more difficult one perceives the audio to be. When participants found the voice warm, they tended to find it easier to understand the audio.

6. Future Work

This is the first audio corpus with measurements for the perceive and actual difficulty of health information. In future work, we will introduce additional questions. i.e., cloze test, select all that apply, true-false, etc., to get a robust result from which we can identify the helpful text features which will solidify our current findings. In this study, we only use the text collected from the Mayo Clinic. In our next full-scale study, we will use text from

various sources i.e., Wikipedia, Cochrane, PubMed, and other health-related websites. Additionally, we intend to develop a metric to measure the difficulty of information in the future.

On the bias findings, we hope to replicate the findings we have in this study with a bigger dataset as well as to expand and include audios of different genders, ethnicities, and accents in the future. An interesting direction would be to add emotion and intonation to the audio, making it even more natural sounding.

Acknowledgments

Research reported in this paper was supported by the National Library of Medicine of the National Institutes of Health under Award Number 2R01LM011975. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We also acknowledge the participation of the AMT workers, who are the main subjects of this study.

References

- [1] Health literacy. Official web site of the U.S. Health Resources & Services Administration. (2019, August 13). Retrieved July 20, 2022, from <https://www.hrsa.gov/about/organization/bureaus/ohe/healthliteracy/index.html#:~:text=Health%20literacy%20is%20the%20degree,who%20have%20low%20socioeconomic%20status>
- [2] Kinsella B. Global Smart Speaker Growth Cools in Q1 as Pandemic Leads to Declining China Sales, Amazon Retains Top Spot Says Strategy Analytics. <https://voicebot.ai/2020/05/25/global-smart-speaker-growth-cools-in-q1-as-pandemic-leads-to-declining-china-sales-amazon-retains-top-spot-says-strategy-analytics/2020>.
- [3] Laricchia, F. (2022, March 29). *US: Smart speaker installed base 2018-2022*. Statista. Retrieved July 30, 2022, from <https://www.statista.com/statistics/967402/united-states-smart-speakers-in-households/#:~:text=The%20installed%20base%20of%20smart,nearly%2091%20million%20in%202021>.
- [4] Soshea.Leibler@cshs.org. Cedars-Sinai Taps Alexa for Smart Hospital Room Pilot. <https://www.cedars-sinai.org/newsroom/cedars-sinai-taps-alexa-for-smart-hospital-room-pilot/2019>.
- [5] Eichler K, Wieser S, Brügger U. The costs of limited health literacy: a systematic review. *Int J Public Health*. 2009;54(5):313-24 (PMC3785182 PMID: 19644651).
- [6] United States. Plain Writing Act of 2010. <https://www.plainlanguage.gov/law/2010>.
- [7] Koh HK, Berwick DM, Clancy CM, Baur C, Brach C, Harris LM, et al. New Federal Policy Initiatives To Boost Health Literacy Can Help The Nation Move Beyond The Cycle Of Costly ‘Crisis Care’. *Health Affairs*. 2012;31(3):434-43 PMID: 22262723.
- [8] digital.gov, plainlanguage.gov. Plain Language Summit 2019: A Day of talks and conversations on how to advance plain language in government communications. <https://digital.gov/event/2019/09/05/plain-language-summit-2019/20198>
- [9] Sulem, E., Abend, O., & Rappoport, A. (n.d.). Semantic structural evaluation for text simplification. ACL Anthology. Retrieved July 21, 2022, from <https://aclanthology.org/N18-1063/>
- [10] D. Kauchak, O. Mouradi, C. Pentoney, and G. Leroy, “Text simplification tools: Using machine learning to discover features that identify difficult text,” 2014 47th Hawaii International Conference on System Sciences, 2014.
- [11] *What is audio?* Computer Hope. (2021, April 12). Retrieved July 28, 2022, from <https://www.computerhope.com/jargon/a/audio.htm>
- [12] Paulsel, M. L., McCroskey, J. C., & Richmond, V. P. (2006). Perceptions of health care professionals' credibility as a predictor of patients' satisfaction with their medical care and physician. *Communication Research Reports*, 23(2), 69–76. <https://doi.org/10.1080/08824090600668832>
- [13] Hovland, C. I., Kelley, H. H., & Janis, I. L. (1963). *Communication and persuasion: Psychological studies of opinion change*. Yale University Press.
- [14] Paulsel, M. L., Richmond, V. P., McCroskey, J. C., & Cayanus, J. L. (2005). The relationships of perceived health professionals' communication traits and credibility with perceived patient confidentiality. *Communication Research Reports*, 22(2), 129–142. <https://doi.org/10.1080/00036810500130588>
- [15] Kareklas, I., Muehling, D. D., & Weber, T. J. (2015). Reexamining health messages in the digital age: A fresh look at source credibility effects. *Journal of Advertising*, 44(2), 88–104. <https://doi.org/10.1080/00913367.2015.1018461>
- [16] Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- [17] Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the bias map. *Advances in Experimental Social Psychology*, 61–149. [https://doi.org/10.1016/s0065-2601\(07\)00002-0](https://doi.org/10.1016/s0065-2601(07)00002-0)
- [18] Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences*, 3, 45–50. <https://doi.org/10.1016/j.cobeha.2015.01.010>

- [19] Bogart, L. M., Bird, S. T., Walt, L. C., Delahanty, D. L., & Figler, J. L. (2004). Association of stereotypes about physicians to health care satisfaction, help-seeking behavior, and adherence to treatment. *Social Science & Medicine*, 58(6), 1049–1058. [https://doi.org/10.1016/s0277-9536\(03\)00277-6](https://doi.org/10.1016/s0277-9536(03)00277-6)
- [20] Baquiran, C. L., & Nicoladis, E. (2019). A doctor's foreign accent affects perceptions of competence. *Health Communication*, 35(6), 726–730. <https://doi.org/10.1080/10410236.2019.1584779>
- [21] Fiske, S. T., Xu, J., Cuddy, A., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, 55, 473–491.
- [22] Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles*, 47, 99–114. <https://doi.org/10.1023/A:1021020920715>
- [23] Coelho, G., Hanel, P., & Wolf, L. (2018). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, 27(8), 1870–1885. <https://doi.org/10.1177/1073191118793208>
- [24] Bogart, L. (2001). Relationship of stereotypic beliefs about physicians to health care-relevant behaviors and cognitions among african american women. *Journal of Behavioral Medicine*, 24(6), 573–586. <https://doi.org/10.1023/A:1012991327297>
- [25] Mukherjee P, Leroy G, Kauchak D. Using lexical chains to identify text difficulty: A corpus statistics and classification study. *IEEE J Biomed Health Inform*. 2019 Sep;23(5):2164-2173. doi: 10.1109/JBHI.2018.2885465. Epub 2018 Dec 6. PMID: 30530380; PMCID: PMC6551329.
- [26] *Flesch reading ease and the Flesch Kincaid grade level*. Readable. (2021, July 9). Retrieved July 29, 2022, from <https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/>
- [27] Leroy G, Helmreich S, Cowie JR, Miller T, Zheng W. Evaluating online health information: beyond readability formulas. *AMIA Annu Symp Proc*. 2008 Nov 6;2008:394-8. PMID: 18998902; PMCID: PMC2656067.
- [28] Tanprasert, T., & Kauchak, D. (2021). Flesch-Kincaid is not a text simplification evaluation metric. *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*. <https://doi.org/10.18653/v1/2021.gem-1.1>