EasyChair Preprint

№ 1311

Research on the Readability of Food Safety Online Health Education Information

Qin Qin, Ke Qing, Wang Chuanlei and Ding Songyun

July 19, 2019

Track Number：_____9_____ Track Title：___智慧健康与智慧养老___

Research on the Readability of Food Safety Online Health Education Information

Qin Qin[1]，Ke Qing[1]，Wang Chuanlei[2]，Ding Songyun[1]
(1. Nanjing University, School of Information Management，210023
2. Anhui University, School of Business，230601)

Abstract

Food safety is a major issue of people's livelihood, the rapid development of the economy and society and the improvement of material living standards have led to an increasing demand for public access to food safety health knowledge. In the Internet age, online health information websites have become the main source of public access to health education information. The food safety health education materials in Chinese Health Education Network were selected for readability calculation and evaluation, the status quo of food safety health education information readability in China was analyzed and the suggestions were proposed in this study.

Keywords: Food safety, health education, online health information, readability

Introduction

With the rapid development of mobile internet information technology, health websites have become the main way for the public to obtain health information. Promote the integration of the internet with food safety, health and medical care, smart elderly care, and develop health services based on internet actively, which will help to promote new health industries, businesses, and models. Food safety concerns public health and safety, constructing the environment of "Internet + Food Safety" and establishing a food safety information platform based on big data analysis will help improve food safety risk management capabilities, improve food safety situation, and enhance public safety and satisfaction.

As a public health education communication medium, health website is becoming more and more popular. Many research institutions and scholars have begun to study the information services evaluation of health websites. Readability means the extent to which information is easily discovered, accepted, and understood. The broad readability refers to all the elements of the text material that affect the effective use of specific readers and their interactions, reflecting the consistency between the text and potential readers; the narrow readability means the influence on the difficulty of text reading and understanding from the text material's internal language, external typesetting and illustration[1]. As an important indicator to measure the reader's difficulty in understanding the information, readability is of great significance in measuring the quality of health information. The evaluation of health information readability allows health information users to understand what information is suitable for themselves more intuitively, so that they can browse health information and make health decisions more easily and effectively. It can also help health website builders understand the website's positioning and do better. This study has selected the food safety health education materials in Chinese Health Education Network (CHEN) to calculate and evaluate the readability, and analyzed the status quo of food safety health education information readability in China.

Materials and Methods

Materials

This study selected CHEN (http://www.nihe.org.cn/) as the experimental carrier website. CHEN is the official website of Chinese Health Education Center (CHEC), it's a comprehensive website including health education materials, health news, health knowledge and science information, health education dynamics in each province, and professional guidance information. As the government's portal site, CHEN has authoritative and authentic educational materials. A total of 230 food safety health education materials published between January 1, 2013 and April 1, 2019 were selected from the 'health knowledge' part of CHEN as the text source for readability calculation. These materials are mainly from Health News Network(HNN,32%), Xinhua News Network(XNN,25%), National Health and Family Planning Commission Network (NHFPCN,9%), People's Network (PN,8%), China News Network (CNN,7%), Economic Information Daily Network (EIDN,3%), China Food and Drug Administration Network (CFDAN,2%), China Network (CN,2%), Lifetimes Network (LN,1%), China Ministry of Health Network (CMHN,1%). The materials with a distribution ratio of 1% or more were selected as the data source for readability calculate, a total of 208 articles including HNN(74), XNN(58), NHFPCN(20), PN(17), CNN(17), EIDN(6), CFDAN(5), CN(5), LN(3), CMHN(3), among them, NHFPCN, CFDAN and CMHN belong to government agency website, XNN, PN, CNN and CN belong to comprehensive network media, HNN, EIDN and LN belong to professional network media.

Methods

Gilliland (1972) proposed that the readability study consists of three aspects, firstly, the "ease of reading" obtained by measuring the recognition speed of words, the error rate, and the number of eye stops per second; secondly, the "interest or compellingness" obtained by measuring the interest of the person and the propositional density and problem style of the text; thirdly, the "ease of understanding" obtained by examining the characteristics of a word or sentence, to find difficulty of the text for some readers[2]. At present, foreign readability calculation methods mainly include Flesch Reading Ease[3], Flesch-Kincaid Formula[4], Gunning Fog Index[5], SMOG Grading[6], Fry Graph[7], Coleman-Liau Index[8], New Dale-Chall Formula[9], etc., the SMOG Grading is a more accurate and convenient method to analyze the readability of health education information. Chinese readability calculation methods are mainly proposed by Yang (1971)[10], Sun Hanyin (1992)[11], Jing Xiyu (1995)[12], Guo Wanghao (2010)[13], Li Ping (2018)[14] and so on, the readability calculation method proposed by Jing Xiyu is most authoritative, Jing analyzed the Chinese textbooks of 12 grades in Taiwan primary and secondary schools, selected three variables to formulate the readability calculation formula, R(Readability)=17.5255+0.0024*X1+0.4415*X2-18.3344*X3, "R" represents the readability value, X1 represents the total number of vocabularies in the article, X2 represents the total number of sentences in the article, and X3 represents the number of familiar vocabularies [14].

On the basis of Jing's calculation formula, Li Ping adjusted the three indicators of article length, average sentence length and accounting terminology in the annual report as the measure indicators of the company's annual report readability, Li proposed that R =17.5255+0.0024*X1+0.4415*X2-18.3344*(1-X3), X1 represents the total number of vocabularies in the article, X2 represents the average sentence length of the article (total number of words/number of sentences), and X3 represents accounting terminology proportion (total number of accounting terminology/total number of vocabularies) 16. The commonly used words in the selected materials account for a large proportion and the familiar words are difficult to define, so Jing's readability calculation method is not suitable for this study, Li's calculation method is selected, R=17.5255+0.0024*X1+0.4415*X2-18.3344*(1-X3) (Formula 1), X1 represents the total number of vocabularies, X2 represents the average sentence length (total number of words/number of sentences), and X3 represents the medical terminology proportion (total number of food safety medical terminology/total number of vocabularies). The smaller the R, the easier the material is to read; the larger the R, the harder it is to read. According to the readability calculation formula, the readability value of the health information is calculated from the three indicators of the total number of words, the average sentence length and the terminology proportion.

Jieba is a Chinese word segmentation component of Python, it's easy to use and the word segmentation result is precise. Create stop words and segment the word of food safety health education materials by using jieba, count the total number of vocabularies (X1) in each article. Count the total number of words and sentences of each article, and calculate the average sentence length(X2).

The definition of terminology is based on the medical health terminology dictionary provided by the LetPub website, the LetPub medical English dictionary has collected 16229 medical terminology of 46 taxonomic disciplines for example nutrition and food hygiene, psychology, microbiology, psychiatry, internal medicine, surgery, pediatrics, obstetrics and gynecology. The LetPub medical English dictionary was selected as a medical terminology lexicon, the excel vlookup was used to compare the word segmentation results of food safety health education materials with the lexicon, find the medical terminology, and count the proportion of medical terminology in food safety health education materials(X3).

Results

Table1. Food Safety Health Education Materials Readability Descriptive Calculation Results

Website	Articles	X1	X2	X3	R
		M(SD)	M(SD)	M(SD)	M(SD)
HNN	74	198.55(161.20)	55.01(11.37)	7.43%(4.05%)	25.32(4.89)
XNN	58	292.64(232.95)	50.69(12.49)	6.22%(3.72%)	23.41(5.47)
NHFPCN	20	237.30(154.78)	58.42(8.76)	8.19%(2.80%)	27.05(3.92)
PN	17	290.71(213.49)	40.90(10.36)	7.31%(2.68%)	19.29(4.63)
CNN	17	243.65(170.85)	49.31(7.87)	7.29%(2.29%)	22.88(3.51)
EIDN	6	302.83(130.78)	49.86(9.04)	10.96%(1.35%)	23.94(4.11)
CFDAN	5	444.40(223.52)	43.42(10.12)	8.82%(2.79%)	21.04(4.23)
CN	5	319.20(212.11)	61.33(19.90)	4.65%(1.64%)	27.89(8.20)
LN	3	417.33(171.04)	32.45(2.33)	11.19%(0.63%)	16.57(0.64)
CMHN	3	144.33(72.58)	56.12(9.96)	12.00%(3.58%)	26.51(4.80)
CHEN	208	253.92(198.29)	51.93(12.43)	7.33%(3.68%)	24.07(5.38)

According to Table1, the R mean of HNN, NHFPCN, CN and CMHN food safety health education materials is higher than the R mean of all materials, the R mean of other sites is lower than the R mean of all materials, indicating that compared with other sites, HNN, NHFPCN, CN and CMHN have higher requirements for public reading, with CN being the highest. About the total number of vocabularies, the average total number of vocabularies in CFDAN is the highest, the average total number of vocabularies and the standard deviation (SD) in CMHN is the lowest, indicating that the average length of education materials and the gap between each other in CMHN is small. About the average sentence length, the average length and the SD of CN is the largest, the average length and the SD of LN is the smallest, indicating that the education materials of CN are more verbose and the articles are widely different with each other, the education materials of LN are concise and consistent. About the terminology proportion, CMHN have the highest proportion of medical terminology, and CN have the lowest proportion of medical terminology, indicating that the educational materials of CMHN are more professional than CN.

Discussion

Import the readability values of all materials into SPSS 25.0 to test the normality, the result is shown in Table 2.

Table 2. Readability Results Normality Test

	Kolmogorov-Smirnov[a]			**Shapiro-Wilk**		
	Stats	df	Sig.	Stats	df	Sig.
R Results	0.046	208	0.200	0.989	208	0.096

According to Table 2, the readability calculation results show that the Sig value is greater than 0.05, so the R value is normally distributed and has good statistical characteristics.

Analyze the readability value of food safety health education materials from each website, the result is shown in Table 3.

Table 3. Food Safety Health Education Material Readability Characteristic Value of Each Website

Website	R Maximum	R Minimum	R Mean	R Variance
HNN	38.79	15.32	25.32	24.21
XNN	42.51	10.85	23.41	30.44
NHFPCN	34.22	20.37	27.05	16.13
PN	27.42	9.83	19.29	22.81
CNN	28.62	15.58	22.88	13.06
EIDN	31.85	19.74	23.94	20.28
CFDAN	28.10	16.52	21.04	22.33
CN	37.36	17.45	27.89	83.94
LN	17.27	15.73	16.57	0.61
CMHN	33.06	21.65	26.51	34.67
CHEN	42.51	9.83	24.07	29.06

According to Table 3, R average value of the food safety health education materials on each website is mainly between 20-30. The article with the largest R value is published by XNN, and the article with the smallest R value is published by PN. The R mean of HNN, NHFPCN, CN and CMHN are higher than the R mean of all materials, the R mean of XNN, PN, CNN, EIDN, CFDAN, and LN are lower than the R mean of all materials. Among them, CN has the highest R mean and the R variance is much higher than others' variance and overall variance, LN has the lowest R mean and the R variance is much lower than others' variance and overall variance, there is a wade gap between them. It can be seen that the readability of LN food safety health education materials are generally high, and it's more inclusive for users, while the CN food safety education materials are less readable and the readability between the materials is quite different, and it's harsher for users.

Calculate the readability of food safety health education materials according to the character of each website, the result is shown in Table 4.

Table 4. Readability of Food Safety Health Education Materials for Websites of Different Characters

Character of Websites	Websites	R Maximum	R Minimum	R Mean	R Variance
Government Agency Website	NHFPCN	34.22	16.52	25.92	22.63
	CFDAN				
	CMHN				
Comprehensive Network Media	XNN	42.51	9.83	22.83	31.31
	PN				
	CNN				
	CN				
Professional Network Media	HNN	38.79	15.32	24.90	25.57
	EIDN				
	LN				
	CHEN	42.51	9.83	24.07	29.06

According to Table 4, the R maximum and R minimum of all materials are distributed in the

comprehensive network media. The R mean of the government agency website is highest, the professional network media is the second, and the comprehensive network media is lowest, and the R variance of them is converse, indicating that among the three types of websites, the comprehensive network media have the strongest readability but the most readability difference, and the government agency website have the weakest readability and the smallest difference in readability values.

The stacked column chart was used to display the readability distribution of all materials and each website, the result is shown in Figure 1.

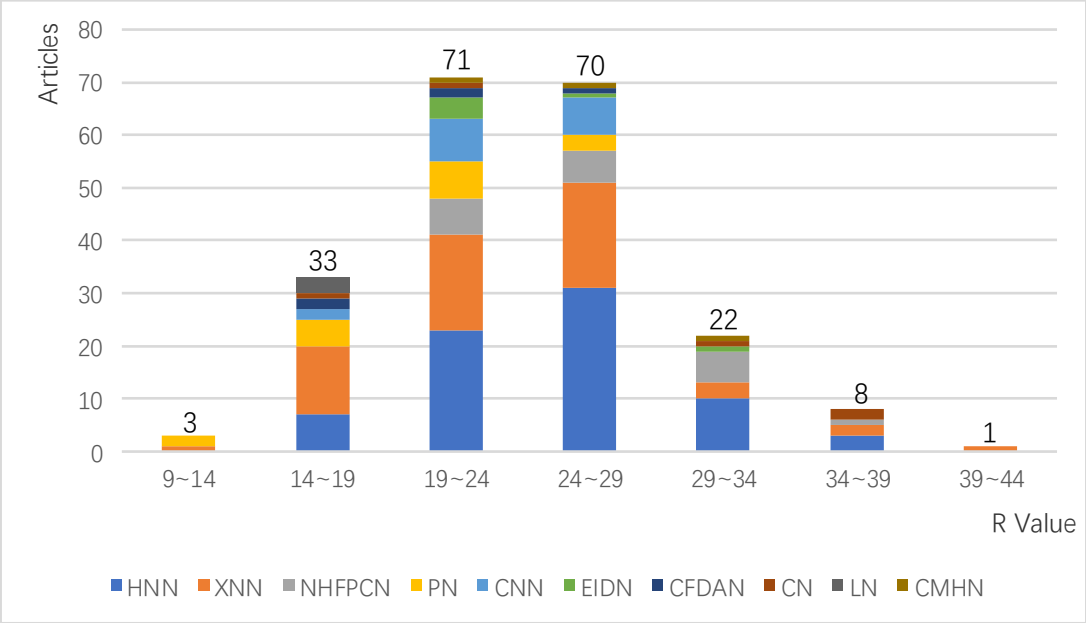

Figure 1. Readability Value Distribution of All Materials and Each Website's Materials

According to Figure 1, the readability values of all food safety health education materials are mainly distributed between 19 and 29, materials with R values greater than 19 account for 83%, indicating that health information users must can read health information with more than 19 readability value to understand the content of most educational materials. Among the websites, the distribution characteristics of HNN and XNN are basically consistent with overall distribution characteristic. The readability values of PN, CNN, LN and CFDAN are mainly distributed below 29, indicating that the education materials published on these websites are highly readable. The readability values of CN are relatively scattered and most of them distributed between 29 and 39, indicating that the food safety health education materials of CN are lowly readable and the read difficulty is uneven.

Status quo of domestic food safety online health education information readability

As the official website of CHEC, CHEN includes health education materials from government portal websites and some commercial health websites, the published food safety health information has certain representativeness and authority, so the readability status quo of domestic food safety online health education information is basically the same. For all of materials, the average number of vocabularies is 253.92, the average sentence length is 51.93, the average terminology proportion is 7.33%, and the education materials with more than 10% terminology account for more than 50, reflecting that the domestic food safety online health education information capacity is large, the sentence is lengthy, the material statement is not appropriate enough, which confuses the sentence logic, and causes the sentences difficult to understand. Many education materials contain a large number of medical terminologies, health information is generally less readable, health information users need to have a certain level of reading and medical knowledge reserve to understand health information and meet their own health needs better.

For websites of different characters, the food safety health education materials published by the comprehensive network media are the most readable, the professional network media is the second, and the government agency website is the weakest. For each website, LN is far more readable than

others, PN, CFDAN and CNN are more readable than CN, NHFPCN, CMHN and HNN, and the readability of CN is the weakest.

Suggestions to improve domestic food safety online health education information readability

Although Formula 1 is appropriate to this study, it lacks an authoritative reading level grading standard, so we only calculate and evaluate the absolute value of readability. In this study, suggestions are proposed to improve the readability of domestic food safety health education information from internal factors, regardless of the subjective factors and external environmental factors of health information users.

From the Formula 1 and readability calculation results, the readability of health information is affected by the total number of vocabularies, the average sentence length and the medical terminology proportion, and is most affected by the average sentence length. Therefore, about the total number of vocabularies, authors and network editors of health education materials should minimize the text length and avoid disparity between texts to reduce vocabularies. About the average sentence length, try to reduce long sentences with complicated structure, avoid using lengthy paragraphs, remove unnecessary content, and make articles more concise. About the terminology proportion, on the basis of ensuring the professionalism of the article, reduce the frequency of unfamiliar and awkward medical terminology by using common vocabularies and express health information by an accessible way.

Conclusion

We have selected the food safety health education materials in CHEN for readability calculation, evaluated the calculation results and analyzed the readability status quo in this study. Founding that the domestic food safety online health education information has low readability and high requirements for health information users. Proposing suggestions for improving domestic food safety online health education information from the internal factors of health education information content.

References

1. Xiao, Y. Y. 2012. *Textbook Readability Evaluation Research*, Shandong Normal University.
2. Gilliland J. 1972. *Readability*, London: Hodder and Stoughton.
3. Flesch, R. F. 1948. “A New Readability Yardstick,” J Appl Psychol (32), pp. 221-233.
4. Kincaid, J. P., Fishburne, R. P., and Rodgers, R. L. 1975. “Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel,” Naval Technical Training Command (1), pp. 8-75.
5. Gunning, R. 1968. *The Technique of Clear Writing*, New York: McGraw-Hill Book Company.
6. McLaughlin, G. H. 1969. “SMOG Grading: A New Readability Formula,” J Read (12), pp. 639-646.
7. Fry, E. 1968. “A Readability Formula That Saves Time,” J Reading (11), pp. 513-516.
8. Coleman, M., Liau, T. L. 1975. “A Computer Readability Formula Designed for Machine Scoring,” Journal of Applied Psychology (60:2), pp. 283-284.
9. Dale, E., Chall, J. S. 1948. *A Formula for Predicting Readability*, Columbus: Bureau of Educational Research.
10. Sun, G. 2015. *Research on Chinese Text Readability Prediction Method Based on Linear Regression*, Nanjing University.
11. Sun, H. Y. 1992. *Chinese Comprehensibility Formula*, Beijing Normal University.
12. Jing, X. Y. 1995. “Study on the Readability of Chinese Textbooks: Estimation of Applicable Grades,” Educational Research Information (5), pp. 113-127.
13. Guo, W. H. 2010. *Study on the Readability Formula of Chinese Text as a Foreign Language*, Shanghai Jiao Tong University.
14. LI, P. 2018. “Readability Analysis of ‘Management Discussion and Analysis’ in the Annual Report of China’s Listed Companies,” Times Finance (9), pp. 225-226.