

Assessing Credibility in Online Sexual Health Information
based on the level of Readability, Tone of Speech, and the use
of Dutch Swearing Words

~

Applied Data Science Master Thesis

Esther Luna Sernández
0169765
e.lunasernandez@students.uu.nl

First Supervisor
Prof. Dr. Remco Veltkamp

Second Supervisors
Drs. Mirjam Visscher
Dr. Annemarie van Oosten



Utrecht University

Department of Information and Computing Sciences

Utrecht University

The Netherlands

July 2022

Abstract

Today, for most adolescents the use of social networking sites plays a major role in their lives. However, the great level of accessibility to these networking sites makes it very easy for anyone to publish information related to sexual health information, risking spreading misinformation and low-quality facts. Thus, to improve the quality of sexual health information shared online and help young people to differentiate it from deceptive information, the present research analyses markers that are linked to the degree of information credibility such as the level of readability, the tone of speech, and the use of swearing words. The information concerning sexual health issues has been retrieved from Fok.nl Forum – accounting for a total of 9150 comments previously divided into five different sexual topics. Besides, seven categories of Dutch swearing words have been created.

First, with the proportions of swearing categories per topic, a one-way ANOVA was implemented to see whether there are statistically significant differences across all topics. Furthermore, the readability measure of Gobbledygook was applied to test the level of readability per topic. Lastly, several statistical tests such as Chi-square of Independence, Cramer's V, Goodman and Kruskal Lambda, and the Uncertainty Coefficient were implemented to see whether the topics are statistically independent of one another regarding the tone of speech. This research has found that comments that talk about sexual experiences have a significantly lower readability score compared to the rest of the topics. Along with this, swearing words related to illnesses, feces, sexuality, and homophobic and sexist slurs have a statistically significant different distribution in occurrence among all topics. Lastly, it has been found that there is no statistical difference in terms of the sensational and negative tone between the sexual health topics. Regarding further research, the level of toxicity and the analysis of slang concerning sexual health information are recommended to consider.

Acknowledgments

I would first like to thank my supervisors; Mirjam, Annemarie, and Remco. They have played an essential role in the accomplishment of this research. They have always been encouraging, willing, and highly dedicated throughout the whole process. Thank you for the honest and constructive feedback and for all the time you have always devoted to us. I would also like to express my gratitude to my fellow thesis students Jo and Jaimy. They have never hesitated in helping me when I felt stuck or did not know in which way I should focus my research. In particular, thank you for your original ideas, and open-mindedness and for sharing your data science knowledge with me.

Table of Contents

1. Introduction	5
2. Literature Review	6
2.1. Readability, Swearing Words, and Sensational Tone as Indicators of Credibility 6	
2.2. Differences in Credibility between Sexual Health Topics	8
3. Data	9
3.1. The Fok Forum	9
3.2. Dutch Swearing Words	11
4. Data pre-processing and exploration	13
5. Methods and Implementation	17
5.1. One-way ANOVA	18
5.2. Simple Measure of Gobbledygook (SMOG)	18
5.3. Conditional Probabilities in Crosstabulation Tables	20
5.4. Chi-Square Test of Independence	21
5.5. Cramer's V	21
5.6. Goodman and Kruskal's Lambda	21
5.7. Uncertainty Coefficient	22
6. Results	22
6.1. One-way ANOVA	22
6.2. Simple Measure of Gobbledygook (SMOG)	23
6.3. Conditional Probabilities in Crosstabulation Tables	25
6.4. Statistical Tests	26
7. Discussion and Conclusion	27
8. Limitations	29
9. Future Work	30
10. Bibliography	31

1. Introduction

Today, for most adolescents the use of social networking sites plays a major role in their lives. The usage of the internet for social purposes has dramatically increased over recent years – as a clear example, 95% of US adolescents between the ages of 12 and 17 ‘go’ online regularly, with 80% of the total participating in some type of social media website (Cookingham and Ryan, 2019). Adolescents recur to the internet incessantly looking for entertainment, interacting with friends and colleagues, but to find answers to any question or concern. In particular, according to Suzuki & Calzo (2004, as cited by Albury et al., 2013), online bulletin boards are especially popular among young people. They use these to ask about intimate and sensitive issues such as sexual health, romance, dating, puberty, body issues, sexual identity, and safer sex practices. This provides the opportunity for the youth to open up about certain topics, get involved in all kinds of conversations as well as help each other, which leads them to feel part of a community and a sense of belonging. Cookingham and Ryan (2019) claim that social media provides a very attractive outlet for adolescents during a time in development where self-expression and validation are important, especially when it comes to the dissemination of sensitive information that adolescents may not feel comfortable discussing in face-to-face interactions. According to Keys et al. (2013), young people prefer to access the media to find out about sexual health topics rather than using other sources such as friends, family, or health professionals, even though they show a lower level of trust towards the former.

Nonetheless, the great level of accessibility to these networking sites makes it very easy for anyone to publish information related to sexual health information - whether it entails a suggestion, feedback, or a personal experience -, risking spreading misinformation and low-quality facts (Song et al., 2021) and so decreasing the overall social and sexual wellness in adolescents (Cookingham and Ryan, 2019). Therefore, the freedom that the internet provides when sharing one's own beliefs and experiences also shows a downside when these ideas or thoughts are misleading, false, and even dangerous. Olteanu et al. (2013) state that users generally lack evidence about the factors that characterize credibility such as the author's expertise and trustworthiness in delivering credible information. This reality is particularly harmful when it comes to the youth since they ought to be considered a less informed and more vulnerable and inexperienced part of the population. Even though some authors such as Liu et al. (2021) approach the evaluation of online credibility as a difficult task for many people, it becomes crucial where sexual information is concerned. The online sexual advice that adolescents take from other peers may therefore impact the way they approach any sexual-related topic over the long term and so negatively affect their lives.

Consequently, to improve the quality of sexual health information shared online and help young people to differentiate it from deceptive information, the present research will analyse several markers, such as the level of readability, the tone of speech, and the use of swearing words, markers that have been previously found to be linked to the degree of information credibility (Olteanu et al., 2013). The information concerning sexual health issues used to analyse these markers will be retrieved from Fok.nl – a Dutch website and virtual community that contains active content like news, reviews and polls targeted to a young audience –, particularly from the Fok Forum. A total of 9150 comments previously divided into five different sexual topics from the forum will be analysed regarding these markers of credibility.

2. Literature Review

In the section that follows, some Theoretical Research Questions will be answered to understand the academic background behind the subject of this research. These questions will lead us to the Empirical Research Questions that will be exposed at the end of the section. The Theoretical Research Questions are:

[TRQ. 1] What is credibility?

[TRQ. 2] What is readability?

[TRQ. 3] What are Dutch swearing words? Are there different categories of Dutch swearing words?

[TRQ. 4] What is sensational information? Is sensational information credible?

2.1. Readability, Swearing Words, and Sensational Tone as Indicators of Credibility

The term ‘credibility’ per se can be defined as believability. Simply put, credible information is believable information (Fogg et al., 2001), or in other words, credible information is the knowledge that can be taken seriously and so be applied to individual circumstances for one’s benefit. Besides, the vast majority of researchers identify trustworthiness and expertise as key components of credibility. *Trustworthiness* ‘captures the perceived goodness or morality of the source’ whilst *expertise* ‘captures the knowledge and skill of the source’ (Fogg et al., 2001, p.62). To assess online credibility, previous research has focused on several content features present on websites such as the level of readability, the tone of speech, and the use of swearing words, among

many other markers. Text comprehension, or *readability*, measures how difficult to understand the text on a website is. According to Olteanu et al. (2013), a factor that may affect users' perception of credibility online is their level of education. In accordance to Mc Laughing (1969), a readability formula is an algorithm that first, measures the difficulty experienced by people reading a given text and second, the linguistic characteristics of that text – in other words, how difficult to understand a text in a webpage is. Additionally, other authors such as Morris et al. (2012, as cited in Olteanu et al., 2013, p.561) explain that 'non-standard use of grammar and punctuation is a good indicator of low-quality content and low credibility perceptions'.

Along with this, the tone of speech relates to the style of the sentences (Liu et al., 2021, p.4). The same authors have categorised the tone of speech into sensational and not sensational. Sensational tones often use exaggerating words and frequent use of punctuations and are associated with a perceived attempt to sell a specific point of view or belief or to persuade the other person with certain information (Liu et al., 2021). Besides, 'some studies have found that consumers prefer an objective tone and avoid bias information' (Sun et al. 2019, as cited in Liu et al. 2021, p.6). This suggests that texts that are clear, well-written, and objective are most likely to be considered trustworthy.

In addition, the use of swearing words in online sexual health information will also be taken into account. Gupta and Kumaraguru (2012) state that the presence of swearing words within a comment indicates that it contains the opinion or reaction of the user but rarely provides any information about the subject per se. Ross (2015) agrees with this statement by adding that a text containing swearing words tends to be relevant to the topic, but it is often a personal reaction that does not add much knowledge to the matter. Other authors such as Cruse (1986, as cited in Ruetter, 2019) explain that taboo concepts are often used for swearing, but, 'if a swearword is not taboo enough, the cathartic effect of swearing will not be obtained' (Ruetter, 2019, p.228). Nevertheless, Bednarek (2019) states that some words are considered more taboo than others. For instance, in the English language, the words 'fuck', 'shit', or 'cunt' are frequently rated as more offensive than 'hell' or 'damn', which are seen as mild. However, when it comes to Dutch society, Ruetter (2019) explains that there is a taboo on being in poor health and that swearing using diseases is a prominent part of the contemporary Dutch culture.

Therefore, it is expected to find differences depending on the type of swearing word employed. In his research, Ruetter (2019) splits swearing words into different categories such as 'genitals', 'feces', 'sexuality', 'religion', and 'diseases'. By looking at these sets of words, it could be argued that religious swearing words such as 'god', 'jesus' or 'heilig' may be used to express emotions whilst illness swearing words such as 'kanker', 'tering' or 'tyfus' may be used to be insulting and

hurtful. Therefore, it may be wrong to use all categories of Dutch swearing words together, as swearing words per se, since they could have a different purpose depending on the word.

2.2. Differences in Credibility between Sexual Health Topics

Considering that the present study will work with several different topics in sexual health information, previously retrieved from Fok.nl, it may be appropriate to make a preliminary distinction between them. Topics such as ‘porn, sexual performance, libido and fantasy’ and ‘different sexual positions and experiences’ may be categorised as more luscious and so exhibit a greater number of overdramatised comments. According to Liu et al. (2021), overdramatised information is perceived as biased and lacked in objectivity, as well as it has a negative effect on the trustworthiness of the content – this sort of information usually employs exaggerating words and a tone of persuasion instead of presenting facts (Liu et al., 2021). Besides, it has been found that pornography extensively uses verbal sexual degradation, with name-calling being one of the most common forms of verbal aggression (Bridges et al., 2010, as cited in Tranchese and Sugiura, 2021) as well as misogynistic hate speech Langton (2012, as cited in Tranchese and Sugiura, 2021). Likewise, ‘the imagery of abuse and discourses of masculinity used by some popular porn sites to advertise themselves disturbingly resemble the violent descriptions of sexual intercourse’ (Tranchese and Sugiura, 2021, p.2728). Thus, considering the extreme use of language used when approaching these topic debates, it would not be too surprising to find a larger number of swearing words, particularly chauvinist slurs. Besides, considering these topics to be more overdramatised, it is also expected to find a larger among of sensational tone and negative comments. Besides, in terms of readability, this does not seem as clear. Nonetheless, more swearing words and sensational tone may also imply that the complexity of the text is not as high and so, it may entail a lower level of readability.

On the other hand, topics such as ‘birth control pill and pregnancy’ and ‘STD testing and condom use’ could be categorised as a more serious health information matter, or likewise, less sensational. Besides, since both of these topics refer to using (or not using) protection during sexual intercourse and the potential consequences, it may be appropriate to gather them together. Farrugia et al. (2021) explain that when seeking information about relationships or sexual practices, adolescents show a preference for personal ‘experiences’ and not so much for facts. Along with this, when talking about sexual experiences, young people tend to highlight the emotional and social aspects rather than the physical aspect per se (Cense, 2019). In addition, in a study implemented by Mcpherson et al. (2014) on ‘Exploring the Quality and Perceived Trustworthiness of Online Information About Chronic Conditions Aimed at Children and Young People’, it has been found that participants strongly connected with first-person narratives and found them a powerful source of information. None of these findings imply that the comments on

these topics will be more reliable. However, since more luscious and ostentatious topics are expected to show less reliability, it is therefore foreseen to find a higher level of trustworthiness on topics characterised as more serious in terms of health information. Therefore, a lower amount of profanity, as well as a lesser amount of sensational tone, is expected to be found in these kinds of sexual health topics. By following the same logic, less use of slurs and less sensational and negative tone may imply that the comments hold a higher level of complexity, and so a higher readability score. Lastly, categorising the topic ‘shape of genitals’ into either ‘overdramatic’ or ‘serious health information’ seems to be less obvious.

All in all, and going back to the Theoretical Questions, it could be said that the term credibility should be understood as believability. Information is credible when is reliable, conscientious, and knowledgeable. Besides, any information shows a greater level of readability when is well-written, clear, and objective. Along with this, and regarding the third theoretical question, there have been found to be several categories of Dutch swearing words as well as differences concerning the degree of vulgarity depending on the word. Lastly, it has been found that sensational information is less objective and so less reliable. Taking this into account, the Empirical Research Questions are as follows:

[ERQ. 1] Are there differences in readability in sexual health information depending on the topic?

[ERQ. 2] Are different categories of Dutch swearing words represented to a greater or lesser extent in some sexual health topics than others?

[ERQ. 3] Are some sexual health topics more or less sensational/negative than others?

3. Data

The data employed to assess credibility regarding the use of swearing words, the readability level, and the tone of speech in online sexual health information has been previously web scraped from the forum section of the Fok.nl website by the University of Amsterdam in May 2021. This section will therefore explain the different datasets that have been used to proceed with the analysis as well as which variables will be taken accountable to do so. Besides, since one of the purposes of the report is to analyse the presence, distribution, and use of Dutch swearing words in online sexual information, a new classification of Dutch swearing words has been created.

3.1. The Fok Forum

Fok.nl is a Dutch website and virtual community in which its users can enjoy a wide variety of content including news, reviews, polls, and forums. The site’s content is mostly aimed at a

younger audience, and it entails one of the largest internet communities in the Netherlands accounting for over 500.000 accounts and more than 200 million posts. The comments from the forum were previously web scraped by the University of Amsterdam. Besides, a Latent Dirichlet Allocation, better known as an LDA, was implemented to retrieve the most characteristic ‘topics’ within the forum. What the LDA Topic Modelling does is to discover the abstract topics that are ‘hidden’ within a collection of texts. Therefore, the name ‘Latent’. The resulting LDA was applied on a total of 150 threads.

Among a total of 20 topics found based on the parameters of the LDA, only five topics were significantly distinctive based on the likelihood of word co-occurrence, and so meaningfully interpreted – accounting for a total of 9150 comments based on 30 threads per topic. These are: Topic 2: ‘shape of genitals’, Topic 7: ‘birth control pill and pregnancy’, Topic 10: ‘STD testing and condom use’, Topic 11: ‘porn, sexual performance, libido, and fantasy’ and Topic 15: ‘different sexual positions and experiences’.

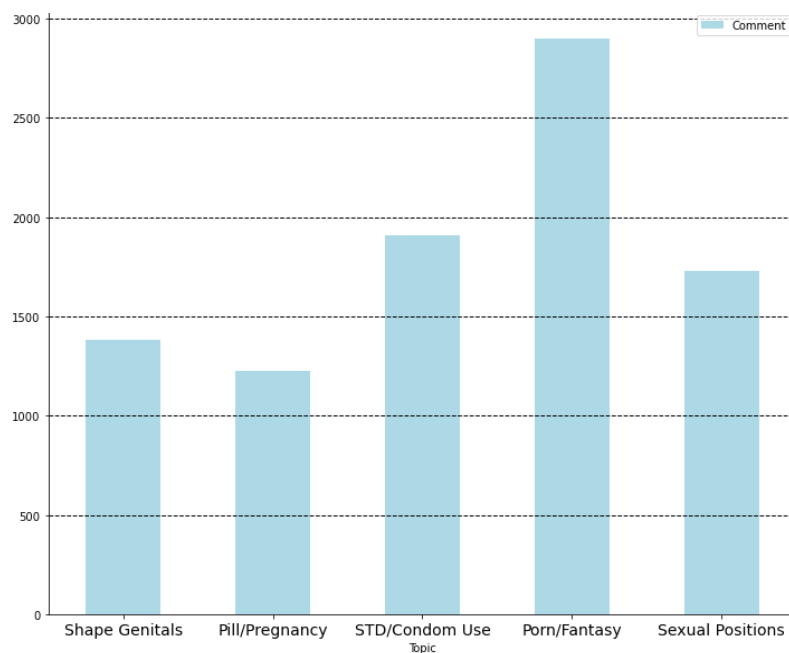


Figure 1. Number of Comments per Topic.

Taking the top 30 threads for each of these five topics account for a total of 9150 comments - 2900 comments for Topic 11, 1911 comments for Topic 10, 1728 comments for Topic 15, 1385 comments for Topic 2 and 1227 comments for Topic 7. Figure 1 displays comment distribution per sexual health topic. Along with this, a variable labeled as ‘sensational/negative tone’ will also be taken into account. This variable was derived by a previous manual coding of the comments and pertains to the sensational/negative tone in a comment, tagged in increasing order from 1 to

5, with 1 standing for ‘not sensational/negative at all’ and 5 for ‘totally sensational’. Table 1 pertains to the three variables we will work with along with their original manual coding.

Variable	Label	Value
Topic	Assigned Topic	2, 7, 10, 11, 15
Comment	‘Comment’	Set of words per row
V4	Negative or sensational tone of comment	1 = ‘not sensational/negative at all’ 2 = ‘not sensational/negative’ 3 = ‘partly sensational/negative’ 4 = ‘sensational/negative’ 5 = ‘totally sensational/negative’

Table 1. Manual Coding for variables Topic, Comment, and V4.

Regarding the length of the comments in terms of the number of words, these show a minimum of 1 word to a maximum of 372 words per comment. Besides, the average and median of the comment’s length account for 25.85 and 3.0 respectively. Nonetheless, taking into account the great difference between the minimum and maximum length, a trimmed mean of 10% will be considered a more robust and realistic mean estimator. The trimmed mean for the comments is equal to 8.21. This entails that most comments are fairly short. As shown in the density plot of Figure 2, it can be seen that the majority of comments barely surpass 50 words.

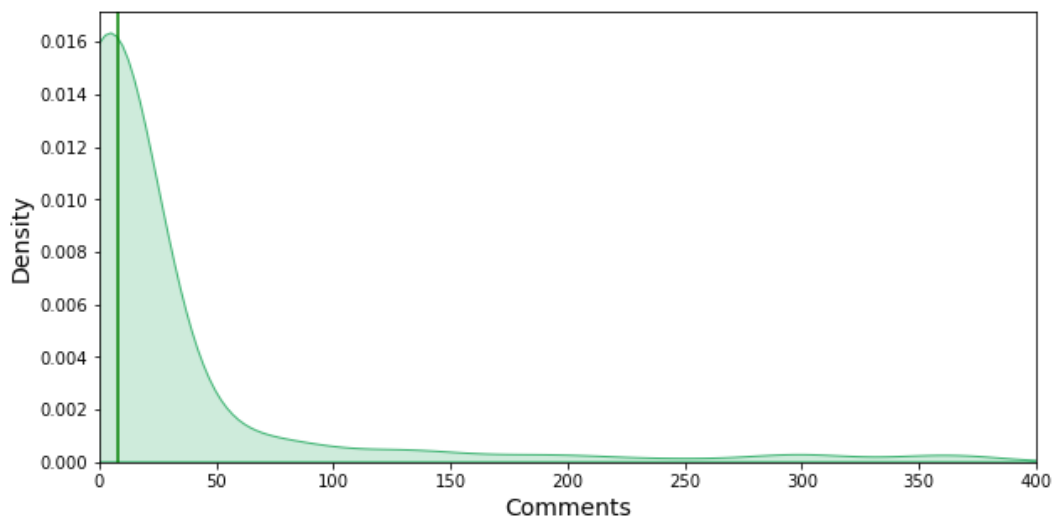


Figure 2. Distribution of Words per Comment.

3.2. Dutch Swearing Words

The Dutch swearing words used in the present research have been retrieved from two different sources. Firstly, Ruetten (2019) made a distinction with lexical items for swearing in which five

categories were taken into account: ‘genitals’, ‘feces’, ‘sexuality’, ‘religion’, and ‘diseases’. However, the number of Dutch swearing words within each category is considered to be limited. Nonetheless, Wikipedia (2022) offers a more solid categorization of swearing words with a wider number of words per category. The words, in this case, are also split into five categories – ‘profanity related to illness and disability’, ‘profanity related to religion and death’, ‘profanity related to sexuality, the human body and animals’, ‘ethnic and social slurs’ and ‘miscellaneous profanity’ – as shown in Table 2.

Ruette (2019)	Wikipedia (2022)
Sexuality	Sexuality, the Human Body, Animals
Religion	Religion and Death
Diseases	Illness and Disability
Genitals	Ethnic and Social Slurs
Feces	Miscellaneous

Table 2. Differences between the Dutch word categories according to Ruette (2019) and Wikipedia (2022).

By a combination of these two groups of swearing words, seven different categories have been created (see Table 3). Besides, ‘Sexists and Homophobic Slurs’ was made out of several swearing words from the original group ‘Sexuality and the Human Body’. This group originally gathered far too many sexual slurs and seemed somewhat ambiguous. Plus, there is a clear difference between words that relate to sexuality and the human body – ‘kut’, ‘kloot’, ‘lul’ – and purely sexist words – ‘slet’, ‘trut’, ‘hoer’ – and homophobic - ‘flikker’, ‘nicht’, ‘kringspiermusketier’. This category may be particularly interesting given that some of the topics are expected to show a larger amount of chauvinist profanity.

Illness and Disability	achterlijk, cholera, ebola, debiel, downie, idioot, jicht, kanker, kankerlijer, klerel, klerelijer, kolere, krijg de, lazarus, lepra, lijer, mazelen, mongool, minkukel, pest, pestkop, pleur, pleuris, pleurlijer, polio, pokken, pokkenlijer, syfilis, stuipen, stom, takke, tering, tyfus, tyfuslijer, vinkentering, ziek, ziekte
Religion and Death	Jezus, godver, gadverdamme, getverderrie, godverdomme, godskolere, godverork, graftak, hel, Jezus Christus, verdomme, verdorie, God, duivel, heilig, hemel, hel, dieu, gatver, dju
Sexuality and the Human Body	snotneus, fucking, aars, bek, dick, dikzak, doos, eikel, emmeren, fok, fuck, hondenlul, huppelkut, kattenkop, kloot, klootzak, klote, kont, kop, kut, lamzak, lul, lullen, manwijf, matennaaijer, mierenneuken, muil, neuk, neuken, neukertje, paardereet, pijp, pik, poes, poot, putain, reet, reetkever, rot, klootviool, rotzak, slijmbal, slons, swaffelen, zak

Ethnic	bamivreter, bosneger, fransoos, franse slag, geitenneuker, hangjongere, Indo, kaaskop, kakker, koelie, lijp, medelander, mocro, mof, nikker, neger, pauper, pinda, plebejer, poepchinesees, proleet, rapalje, rapaille, roetmop, spaghettivreter, loempiavouwer, spanjool, spleetoog, tatta, tuig, tokkie, zandneger, zwartnek, zwartzak, zwartjoekel
Miscellaneous Profanity	knuppel, koekert, koekwaus, NSB'er, aso, oelewapper, pannenkoek, smeerlap, sukkel, sul, bastaard, ouwehoeren, pissig, uilskuiken, sodemieter, ezel, zoi
Feces	kak, shit, crap, schijt, stront, keutel, merde, scheisse, diarree, poep, pis, zeik, chips, kak, flapdrol, drol, zeiken
Sexist and Homophobic Slurs	anaalgeneraal, anusridder, bitch, flikker, hoer, kreng, kringpiersmusketier, muts, nicht, pot, slet, stoephoer, teef, trut, tut, wijf, kenau, kutwijf, homo, gay

Table 3. Final classification of Dutch Swearing Words categories.

Last but not least, the classification of Dutch swearing words has been validated by five Dutch native speakers. The reason behind this is that it is quite convenient, as well as useful, to acknowledge that some words may not be seen as swearing words per se, as well as to verify whether the categories are consistent. Swearing words such as ‘beroerte’, ‘influenza’ and ‘griep’ have therefore been removed. Nonetheless, words such as ‘fucking’, ‘flapdrol’, ‘kutwijf’, and ‘aso’ have been included.

4. Data pre-processing and exploration

The first step in the data pre-processing is to look for null or missing values. For the ‘Comment’ column, no missing values have been found. However, the variable ‘sensational/negative tone’ shows a total of 5090 missing values. The reason behind it is that out of the 9150 comments, only comments that were about sexual health topics (manually coded by coders) were used for this variable. Therefore, comments with a missing value for the variable ‘sensational/negative tone’ will not be considered to analyse whether some sexual health topics are more or less sensational. As the manual coding took into account whether or not a comment was about sexual topics or not, we could filter out all comments that were not about sexual health topics, resulting in a sample of 4055 comments for the analyses on sensational/negative tone. However, these comments will exclusively be deleted to see whether there are differences concerning sensational/negative tone per topic. For the rest of the methods, no comments will be removed and so no information will be lost.

Moreover, in this new filtered dataset, the variable ‘sensational/negative tone’ will be dichotomised, going from 5 different labels to two. By looking at the distribution of this variable

(see Table 4), it can be observed that ‘not sensational/negative at all’ coupled with ‘not sensational/negative’ represent approximately 75% of the total amount of comments. This leaves us with only 25% of the values being partial to totally sensational/negative. Therefore, due to this polarization, a distinction between being sensational and not will be applied. Thus ‘not sensational/negative at all’ and ‘not sensational/negative’ will now be labeled as 0, ‘not sensational/negative’. And so ‘partly sensational/negative’, ‘sensational/negative’, and ‘totally sensational/negative’ will be then labeled as 1, ‘sensational/negative’. The reason behind dichotomising this variable is to see whether some topics are significantly more or less sensational and negative than others.

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
not sensational/negative at all	2103	23,0	51,8	51,8
not sensational/negative	934	10,2	23,0	74,8
partly sensational/negative	662	7,2	16,3	91,1
sensational/negative	324	3,5	8,0	99,1
totally sensational/negative	38	,4	,9	100,0
Total	4061	44,4	100,0	
Missing System	5090	55,6		
Total	9151	100,0		

Table 4. *Sensational/Negative Tone of Comment Distribution in variable V4.*

Next, from the Dutch Swearing Words classification, each group of swearing words was put together into a list, resulting in 7 different lists, one per category. Thereafter, a lambda function was implemented to get any word from each of the seven lists that matched a word in every comment, and so add them into a new column. This resulted in seven columns, the same as the number of swearing word categories (see Figure 3). For instance, ‘Sexuality’ swearing words represent a new column within the data frame that shows which comments have a ‘Sexuality’ swearing word and the word, by showing either the word or an empty square. The same applies to the rest of the categories.

Topic	Comment	V2	V4	Illness	Religion	Sexuality	Ethnic	Miscellaneous	Feces	Homo_Sexist
2	Tijdens heb hebben van seks doet mijn toompje ...	3.0	4.0			eikel				
2	Als je een scheurtje in je toompje hebt moet j...	3.0	2.0							
2	Als je goed hard neukt en je toompje scheurt d...	3.0	2.0							
2	Al eens vaker een maand gewacht. Gelijk 1e kee...	5.0	3.0							
2	Je laten besnijden	3.0	1.0							

Figure 3. Data frame after applying a lambda function.

Subsequently, we have computed the fractions of the number of swearing words in each comment. To do so, we have tokenized each comment and swearing word column with the help of Spacy (see Figure 4). Spacy is a free open-source library for Natural Language Processing that supports more than 66 languages, including the Dutch language. It is also known for being a robust way of tokenizing with high evaluated accuracy. For instance, by tokenizing each comment, Spacy will remove the stop words and return the tokens (words) that are meaningful within each comment. Stop words provide little or no information at all since they occur in abundance, that is why they are often removed from the text beforehand. Besides, in the tokenization process, the comments and swearing words have been lowered cased and punctuations removed. It is important to mention that, because of the stop words removal, the fractions are done per number of non-stop words and not on the original number of words that originally formed each comment. Figure 4 shows how the comments look after being tokenised, under 'Tokenised_Comments'. It also shows the sum of every tokenised comment under 'Token_Count'. Besides, it can also be seen that existing swearing words are also tokenised. As an example, Figure 4 also shows Sexuality tokenised swearing words and the totals under 'sex_count'. The same has been done for every swearing word category.

Topic	Comment	Tokenized_Comments	token_sex	Token_Count	sex_count
2	Tijdens heb hebben van seks doet mijn toompje ...	[seks, toompje, pijn, jaren, seks, last, 6tal,...	[eikel]	67	1
2	Als je een scheurtje in je toompje hebt moet j...	[scheurtje, toompje, bepaalde, tijd, seks, her...	[]	14	0
2	Als je goed hard neukt en je toompje scheurt d...	[goed, hard, neukt, toompje, scheurt, oplossin...	[]	12	0
2	Al eens vaker een maand gewacht. Gelijk 1e kee...	[vaker, maand, gewacht, 1e, keer, mis, doorges...	[]	8	0
2	Je laten besnijden	[laten, besnijden]	[]	2	0

Figure 4. Section of the data frame that shows the tokenization of Comment and Sexuality column and the totals of both columns per comment.

Now, to get the proportions of the swearing words per comment, we can divide each tokenised swearing word by the number of tokenised words of every comment, resulting in 7 columns that contain the fractions. However, after doing so, we have found 129 null values spread across every

topic. The reason behind this may be that these comments only accounted for stop words in the first place and so, after the tokenization process, the 129 tokenized comments only gathered null values. Nonetheless, 129 comments account for only 1.4% of the entire dataset. Therefore, these rows have been removed. We can also remove the columns that are not necessary anymore and keep the relevant ones. Thus, the resulting data frame now includes the seven columns with the fractions per swearing word category (see Figure 5).

	Comment	Topic	V2	V4	ILLNESS	RELIGION	SEXUALITY	ETHNIC	MISCELLANEOUS	FECES	HOMO_SEXIST
0	Tijdens heb hebben van seks doet mijn toompje ...	2	3.0	4.0	0.0	0.0	0.014925	0.0	0.0	0.0	0.0
1	Als je een scheurtje in je toompje hebt moet j...	2	3.0	2.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
2	Als je goed hard neukt en je toompje scheurt d...	2	3.0	2.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
3	Al eens vaker een maand gewacht. Gelijk 1e kee...	2	5.0	3.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0
4	Je laten besnijden	2	3.0	1.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0

Figure 5. Tokenization of the Comment column (section of the data frame).

Along with this, we have calculated the mean of the fraction per category of Dutch swearing words and topic. Figure 6 shows this.

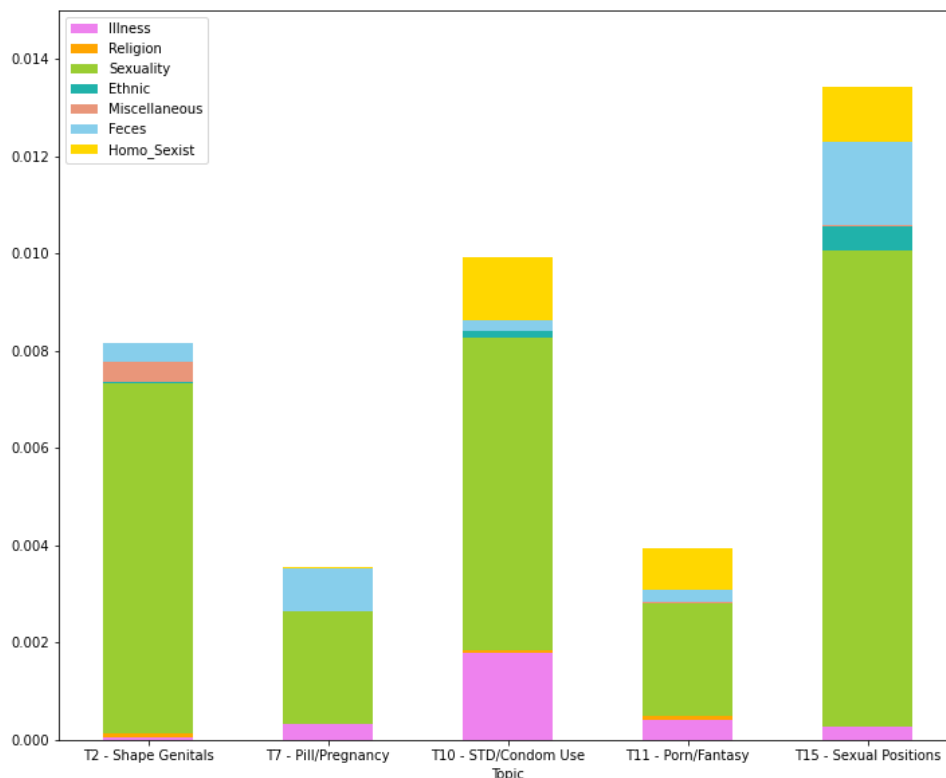


Figure 6. Average of the Fraction of Swearing Words per Topic and Category of Swearing Words.

Besides, Table 5 shows the average of the fraction per topic and swearing word category multiplied by 1000 for interpretation considerations. In general, comments show a high amount

of ‘Sexuality’ swearing words. It can also be observed that ‘Illness’ swearing words are particularly used in comments related to ‘STD testing and condom use’. The great majority of these words are ‘kanker’, ‘ziekte’ and ‘ziek’. Besides, ‘homophobic and sexist’ slurs are especially used in comments related to STD’s, porn, fantasy and sexual positions – with the most used ones being ‘gay’, ‘homo’, ‘wif’ and ‘hoer’. The category ‘feces’ is roughly stable among topics, nonetheless, it shows to be higher for ‘Birth Control pill and Pregnancy’ and ‘Sexual Positions’ with the majority of these words being ‘schijt’, ‘shit’, ‘stront’ and ‘poep’. As for ‘Religion’, ‘Ethnic’, and ‘Miscellaneous’ swearing words, the proportions are very low. Nonetheless, topics that are considered more overdramatised also show stronger slurs such as ‘flikker’, ‘hoer’, ‘slet’, ‘bitch’ or ‘pot’.

	Illness	Religion	Sexuality	Ethnic	Miscell.	Feces	H/Sexist
Shape Genitals	0.01	0.10	0.72	0.00	0.04	0.04	0.00
Pregnancy	0.03	0.00	0.23	0.00	0.00	0.09	0.00
STD’s, condom use	0.18	0.01	0.64	0.01	0.00	0.02	0.13
Porn, Fantasy	0.04	0.01	0.23	0.00	0.00	0.03	0.09
Sexual positions	0.03	0.00	0.98	0.05	0.00	0.17	0.11

Table 5. Average of the fractions per swearing word categorie and topics multiplied by 1000.

In addition, even though words related to ‘Sexuality’ are the most commonly used among all topics, there is slightly different kind of words used in this category depending on the topic. Some of the most utilised ones are shown in Table 6. It can for instance be seen that porn, fantasy, and sexual positions and experiences comments also use stronger slurs such as ‘fuck’ or ‘kont’.

Shape of Genitals	eikel, lul, pik, neuken
Birth Control Pill and Pregnancy	zak, kut, neuken, eikel, doos, lul
STD Testing and Condom Use	neuken, lul, neuk, pik, kut, kop
Porn, Sexual Performance, Libido and Fantasy	neuken, kut, kop, fuck, lul, pik
Different Sexual Positions and Experiences	neuken, kont, kut, lul, pik, reet

Table 6. Most Commonly Used Sexuality Swearing Words per Topic.

5. Methods and Implementation

This chapter describes the methods that will be used to answer the Empirical Research Questions. First, a one-way ANOVA will be implemented to see whether there are statistically significant differences between the swearing words categories among the different topics. Consequently, the level of readability of the comments on each sexual health topic will be analysed by applying the

Simple Measure of Gobbledygook or SMOG. Along with this, a cross-tabulation table will be created with variables ‘topic’ and ‘sensational/negative tone’. Besides, conditional probabilities for these two categorical variables will be implemented too to see the differences in frequencies per topic. Along with this, several statistical tests for categorical variables will be applied to study the relationship between these two categorical variables. These tests are the Chi-square of Independence, Cramer’s V, Goodman and Kruskal’s Lambda, and the Uncertainty Coefficient.

5.1. One-way ANOVA

A one-way ANOVA will be implemented to see whether the frequencies of swearing words are significantly different in terms of their means between the different topics. A one-way ANOVA, or analysis of the variance, compares the means of two or more independent groups to determine whether these groups are statistically different. The ANOVA assumes that all groups have the same population variances, and so if the null hypothesis is true, it implies that the population means are equal for all groups (Herzog et al., 2019, p. 69). Nonetheless, there are several assumptions to take into account before applying a one-way ANOVA. Herzog et al. (2019, p. 72) state the following ones:

- Independent samples.
- The independent variable is discrete, while the dependent variable is continuous.
- Gaussian distributed populations.
- Homogeneity of variance: All groups have the same variance.
- The sample size needs to be determined before the experiment.

Moreover, the F-value will also be taken into consideration, which accounts for the variability between groups and within groups. A larger F implies a larger difference in means (Herzog et al., 2019). The F-value formula can be expressed as the following:

$$F = \frac{\text{Variance estimate based on variability between groups}}{\text{Variance estimate based on variability within groups}} \quad (1)$$

5.2. Simple Measure of Gobbledygook (SMOG)

A readability measure or SMOG score will be applied to the sexual health comments from the Fok.nl forum to discover whether some topics account for a significantly higher or lower level of readability than others, and therefore to see whether people use more or less complex words when approaching different sexual health topics.

The acronym SMOG stands for ‘Simple Measure of Gobbledygook’, but it is better understood as a measure of readability that estimates the years of education a person needs to understand a

piece of writing. According to Ley and Florio (1996, p.9) readability formulas have been extensively used to measure the understandability of health-related information, including the SMOG score. Apart from the SMOG score, other formulas such as the Flesch-Kincaid and the Automatic Readability Index (ARI) have also been considered. Nonetheless, a 2010 study published in the *Journal of the Royal College of Physicians of Edinburgh* by Fitzsimmons et al. (2010, p.295) stated that “SMOG should be the preferred measure of readability when evaluating consumer-oriented healthcare material”. The same study also pointed out that the Flesch-Kincaid formula significantly underestimated the reading difficulty compared to the SMOG formula.

Subsequently, authors such as Mc Laughing (1969) explain that a readability formula is a mathematical equation derived by a regression analysis. This procedure finds the equation that best expresses the relationship between the difficulty of reading a certain text and the measure of the linguistic characteristics of that text. It has been found that word and sentence length are the measures with the highest predictive capacity. This seems reasonable since longer sentences and longer words may gather a more complex grammatical structure than short ones, forcing the reader to make a bigger effort to comprehend the information they are reading. The formula implemented by Mc Laughing (1969) is shown in (2), where p is the number of polysyllables in samples of 30 sentences.

$$SMOG\ Grade = 1.0430 \sqrt{p} + 3.1291 \quad (2)$$

The SMOG formula counts a total of 30 sentences - 10 at the beginning, 10 in the middle, and 10 at the end of the entire text. Then, it counts every word of three or more syllables, including repetitions, within the 30 selected sentences. Subsequently, it estimates the square root of the total number of polysyllable words counted (Mc Laughlin, 1969). A perhaps clearer formula than the one above has been implemented by Amos (2020), shown in (3). This formula explicitly shows that uses 30 sentences and that the polysyllables are counted within these 30 sentences.

$$SMOG\ Grade = 1.0430 \sqrt{30 * \frac{\text{number of polysyllables}}{\text{number of sentences}}} + 3.1291 \quad (3)$$

Therefore, a larger number of polysyllables within a piece of text accounts for a higher score. This does not mean that a text with a higher readability score is easier to read, but it implies that it gathers more complex words and so it requires a higher level of comprehension from the reader. Table 7 illustrates the conversion table for the SMOG score. It can be seen that the total number of polysyllables accounted for within a group of texts correspond to a grade level, and so to a particular age (e.g., an approximate amount of 32 polysyllables would imply that someone in 9th grade, a 14-year-old, should be capable enough to understand the given texts).

Polysyllables Word Count	Grade Level	Age
0-2	4 th	9 - 10 years old
3-6	5 th	10 - 11 years old
7-12	6 th	11 - 12 years old
13-20	7 th	12 - 13 years old
21-30	8 th	13 - 14 years old
31-42	9 th	14 - 15 years old
43-56	10 th	15 - 16 years old
57-72	11 th	16 - 17 years old
73-90	12 th	17 - 18 years old
> 90	> 12 th grade	Older than 18 years old

Table 7. Conversion table for SMOG score.

In this research, since the number of comments per topic is much larger than 30, the SMOG formula will be generalised to the maximum number of comments per topic by using the argument ‘all_sentences = True’ (PyPI.org, 2020). And so, all comments per topic will be considered. Consequently, once we have a score per topic, a single sample t-test will be executed to see whether there are significant differences between the grade of each topic compared to all other grades.

5.3. Conditional Probabilities in Crosstabulation Tables

Variables ‘topic’ and ‘tone of speech’ will be used to create a crosstabulation table and so implement conditional probabilities per sexual health topic. A crosstabulation or contingency table displays the frequencies for the combinations of categorical variables where the values at the row and column intersections are the frequencies for each unique combination of the two variables (Simpson, 1951). These types of tables are implemented to understand the relationship between categorical variables, and, in this case, they will be used to see the relationship between the five different topics and the tone of speech.

Moreover, a conditional probability is the probability of an event happening given that another has occurred too. That is why the name ‘conditional’. Conditional probabilities will be implemented to see the proportions of ‘sensational/negative’ and ‘not sensational/negative’ tone comments among different topics. The formula for a conditional probability is shown in (4).

$$\text{Cond. } P(\text{Row } i, \text{Col } j) = \text{Count cell } (i, j) / \text{Total } (\text{Row } i, \text{Col } j) \quad (4)$$

5.4. Chi-Square Test of Independence

A chi-square test of independence will be applied to see whether there is a statistically significant difference between the topics and having a sensational/negative tone. The Chi-square test of independence determines whether there is a statistically significant relationship between categorical variables (Frankie et al., 2012). This test is used to determine whether there is a statistically significant difference between the observed frequencies and the expected ones within the crosstabulation table. The null hypothesis for this test state that there are no differences between the categorical variables, and so by knowing the value of one variable, it is not possible to predict the value of another variable. We reject this hypothesis when the p-value is lower than .05, meaning that, within the sample, one variable of interest is associated with the second variable of interest (Frankie et al., 2012). The chi-square formula is displayed in (5).

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

5.5. Cramer's V

A Cramer's V will be applied to see how strong the relationship between the categorical variables topic and tone of speech is. Cramer's V is based on the Chi-square Pearson statistic. Cramer's V determines the strength of a relation between two categorical variables, and it retrieves a value from 0 to 1, where 1 means that the association between these two variables is perfect. That is, one variable is completely determined by another variable. This test is stronger and more realistic since it does not increase the value with a larger number of cells like the chi-square does (White and Korotayev, 2004). Cramer's V equation is shown in (6).

$$V = \frac{\sqrt{(X^2/n)}}{\min(k-1, r-1)} \quad (6)$$

5.6. Goodman and Kruskal's Lambda

A lambda measure of the association will be computed to test whether there is a relationship between variables topic and tone of speech by looking at the proportional reduction error. Lambda is a measure of association that reflects the proportional reduction in error (PRE) when the values of the independent variable (topics) are used to predict values of the dependent one (sensational/negative tone) (IBM, 2021b) and it is used in cross-tabulation analyses. Lambda ranges, as well as Cramer's V, from 0 to 1. A value of 1 implies that the independent variable (topics) perfectly predicts the dependent variable (sensational/negative tone). Conversely, a value

of 0 means that the former does not predict the latter in any way whatsoever. The lambda equation is defined in (7):

$$\lambda = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1} \quad (7)$$

Where ε_1 accounts for the error of prediction when the independent variable (topics) is ignored, and ε_2 accounts for the error when the prediction is made on the independent variable (topics).

5.7. Uncertainty Coefficient

Similarly, the uncertainty coefficient will be used to measure whether variables topic and tone of speech are related to one another by looking at the proportional reduction error. This coefficient is a measure of nominal association and indicates the proportional reduction in error (PRE) when the values of one variable are used to predict the values in another variable (IBM, 2021b). For instance, a value of 0.94 would indicate that by knowing one variable we reduce the error when predicting the values for another variable by 94%. It, therefore, determines the degree of association between two variables. The uncertainty coefficient is shown in (8).

$$U(X|Y) = \frac{H(X) - H(X|Y)}{H(X)} \quad (8)$$

6. Results

6.1. One-way ANOVA

Before applying a one-way ANOVA for the swearing word categories and the topics variable, we will first check that our data follows the analysis assumptions. All variables – swearing word categories - are independent of one another and numerical, as well as the independent variable. Topics, in this case, are discrete. Besides, every swearing word category contains different words and so the observations are independent of each other. Nonetheless, we will check that the data follows a Gaussian distribution and that all groups have the same variance.

To check whether the data follows a normal distribution, we imputed a Kolmogorov-Smirnov normality test on every variable, which examines the goodness of fit of a given set of data to a theoretical distribution (Berger and Zhou, 2014). With a p-value $<.05$ for every swearing word variable, we reject the null hypothesis of the variables coming from a normal population distribution. Consequently, we have normalised the values by subtracting the population mean of

every variable and then dividing this difference by the standard deviation. Nonetheless, when checking for homogeneity, our variables do not show to have a homogeneous variance and so we fail to meet the fourth assumption.

Nonetheless, when applying a one-way ANOVA (see Table 8), we can observe that variables ‘illness’, ‘sexuality’, ‘feces’, and ‘homophobic and sexist’ show a p-value $<.05$. We then reject the null hypothesis that states that ‘the means of all variables are equal’. Besides, for p-values $<.05$, the F-value is rather higher than for the rest of the variables, meaning that their differences in means across the topics are significantly different.

	F	Sig.
z_illness	7.409	.000
z_religion	1.310	.263
z_sexuality	17.282	.000
z_ethnic	1.599	.171
z_miscellaneous	1.522	.193
z_feces	3.244	.012
z_homosex	3.909	.004

Table 8. One-way ANOVA for Topics and Swearing Words Variables.

Therefore, according to this result, swearing words related to illnesses, feces, sexuality, and homophobic and sexist slurs have a significantly different distribution among all the topics. Conversely, ethnic and miscellaneous slurs have shown no significant difference at all. We already saw the distributions of the swearing word categories in the data exploration section and found that, for instance, illness swearing words are more predominant in comments that talk about STD testing and condom use, while feces slurs appear more in topics that concern sexual positions and pregnancy. Homophobic and sexist slurs were particularly seen in STDs, porn, fantasy, and sexual positions topics. And finally, even though ‘sexuality’ slurs were generally very predominant on every topic, it could be said that sexual positions, STDs, and the shape of genitals-related information significantly make more use of these than the rest of the topics.

6.2. Simple Measure of Gobbledygook (SMOG)

The measure of Gobbledygook (SMOG) has been implemented using the Natural Language Processing Toolkit NLTK. Natural Language Processing, or NLP, is a wide-ranging field that covers manipulation and computer understanding of human language (Millstein, 2020, p.4). In this research, the SMOG readability metric has been applied to all sexual health comments per topic. Nonetheless, it is important to state that, for each topic, there are a different number of

comments and so the data is imbalanced, as explained in the data section. Besides, the grade level that a person would need to understand the texts has also been imputed, along with the age.

Topic	Name	SMOG	Grade Level	Age
2	Shape of Genitals	10.163	10 th	15 – 16 years
7	Birth Control Pill and Pregnancy	10.403	10 th	15 – 16 years
10	STD Testing and Condom Use	10.649	11 th	16 – 17 years
11	Porn, Sexual Performance, Libido, and Fantasy	10.533	11 th	16 – 17 years
15	Sexual Positions and Experiences	9.97	10 th	15 – 16 years

Table 9. SMOG score, Grade Level, and Age per Sexual Health Topic.

As shown in Table 9, ‘Sexual Positions and Experiences’ shows the lowest SMOG score compared to the rest of the topics, accounting for 9.97, followed by ‘Shape of Genitals’ with a score of 10.162. Conversely, comments in ‘STD Testing and Condom Use’ account for the highest readability score, making this topic use more complex sentences. Besides, topics that are considered to be more serious related to sexual health issues, such as ‘Birth control pill and Pregnancy’ and ‘STD testing and condom use’, stand between the topics with the lowest and highest scores.

By looking at the Grade Level in Table 9, comments in Topics 2, 7, and 15 would need a person to be at least in 10th Grade to comprehend the comments, whilst for Topics 10 and 11, the same applies for 11th Grade. That is to say, according to these scores, topics related to sexual transmission diseases, sexual protection, porn, and fantasy are, according to this score, show a larger number of polysyllables than topics that talk about the shape of the genitals, birth control pill, pregnancy as well as sexual positions, and would need someone older to understand them.

Nonetheless, the difference between the highest and lowest SMOG score among all topics accounts for 0.679, implying that there is not even one year of difference between them. To find out whether there are significant differences between the readability score of every topic to every other score, a one-sample t-test (single sample t-test) will be applied per score. A single sample t-test is used to compare the sample (the score) with the mean of the population (mean of all the scores). The null hypothesis for this test implies that the sample mean equals the population means. When applying this test to every score, only ‘Sexual Positions and Experiences’ shows to be significantly different from the other four remaining topics, with a p-value of 0.038 (< .05). Therefore, we reject the null hypothesis, which in this case implies that the SMOG score for Topic 15 equals the mean of all other scores. And so, this score is significantly different from the rest of the topics, which means that ‘Sexual Positions and Experiences’ has a significantly lower score

Table 11. Conditional Probabilities per Topic.

Nonetheless, several statistical tests will be implemented to know whether there are significant differences between the topics regarding the tone of speech.

6.4. Statistical Tests

As mentioned in the methods section, several statistical tests will be executed to see whether there is a statistically significant difference between the two categorical variables, ‘Topic’ and ‘Sensational/Negative Tone’, in our crosstabulation table. These tests are the Chi-square, Cramer’s V, Goodman and Kruskal Lambda, and the Uncertainty coefficient. Table 12 displays the statistics for the Chi-square test. With a p-value very approximate to zero and $<.05$ we reject the null hypothesis of ‘there are no differences between the categorical variables’, and so we reject that by having the variable ‘topic’, we can predict a ‘sensational/negative’ tone or vice versa.

Chi-square Statistics	Degrees of Freedom	p-value
54.38813732707344	10	4.107821437767782e-08

Table 12. Chi-square test.

Nonetheless, Cramer’s V test is also implemented since it shows how strong the relationship of variables ‘topic’ and ‘sensational/negative tone’ are in cross-tabulation tables larger than 2x2. However, even if the result is statistically significant, a value lower than 0.2 implies that the two categorical variables are weakly associated (IBM, 2021a). With an estimate of .116, it may then be argued that the relationship is fairly weak even though, with a p-value $<.05$, the result is statistically significant.

Cramer’s V	p-value
.116	.000

Table 13. Cramer’s V test.

Along with this, Table 14 shows the lambda and uncertainty coefficient measures. By looking at Lambda’s values, taking ‘sensational/negative tone’ as the dependent variable, this value equals 0. This suggests that there is almost no relationship between both variables, and that variable Topic cannot predict ‘sensational/negative tone’. Lastly, the uncertainty coefficient value when ‘sensational/negative tone’ works as the dependent variable is equal to .012. This translates into, by knowing the variable Topic (independent variable) we reduce the error by 1.2 % when predicting the sensational or negative tone in a comment. This value is indeed very small and close to 0, especially if we take into account that this coefficient ranges from 0 to 1.

		Value	Std. Error	Approximate T	p-value
Lambda	Symmetric	.008	.006	1.324	.186
	Dependent Sensational/neg Tone	.000	.000	-	-
	Dependent Topic	.010	.008	1.324	.186
Uncertainty	Symmetric	.006	.002	3.701	.000
Coefficient	Dependent Sensational/neg Tone	.012	.003	3.701	.000
	Dependent Topic	.004	.001	3.701	.000

Table 14. Directional Measures. Lambda and Uncertainty Coefficient

To sum it up, all these statistical tests agree that both categorical variables are independent of one another and that by knowing the variable ‘Topic’ it is almost impossible to predict the sensational/negative tone. Therefore, the differences in sensational/negative tone among the topics are not significantly different at all.

7. Discussion and Conclusion

The present research has aimed to answer three Empirical Research Questions related to sexual health information sharing. The purpose of the research has been to assess the level of credibility in online sexual health information as well as to understand the way adolescents approach several sexual health topics. This has been done by analysing whether there are differences between these topics in terms of the level of readability, the use of swearing words, and the tone of speech. The sexual topics we have worked with are ‘Shape of Genitals’, ‘Birth Control Pill and Pregnancy’, ‘STD testing and Condom Use’, ‘Porn, Sexual Performance, Libido, and Fantasy’, and ‘Sexual Positions and Experiences’. The first research question aimed to find whether there are differences between sexual health topics in terms of readability or how difficult a piece of text is to comprehend. The second research question seeks to discover whether different categories of Dutch swearing words are represented to a greater or lesser extent in some sexual health topics than others. And finally, the third research question’s goal was to find whether some sexual health topics were more or less sensational and negative than others.

Regarding our first research question, we have found that comments that talk about sexual experiences have a significantly lower readability score compared to the rest of the topics that, statistically, show no difference between them. One of our hypotheses after researching past literature was that topics categorised as being more ostentatious and overdramatised, such as those related to porn, fantasies, and sexual experiences, would show a lower readability score. However,

and according to our findings, this is only true for sexual experiences but not for comments concerning fantasies and pornographic content - implying that when people approach or talk about sexual positions use fewer complex sentences, and so it requires a lower level of comprehension for the reader to understand. It could be said that they approach this topic in a more colloquial and informal style. Therefore, comments that are written in a more colloquial and informal style may not be as skilfully written and so they may show a lower level of credibility to the reader. It is also relevant to note that, according to the readability score, people who are younger than 15 years old may have difficulties understanding the context of comments related to any sexual health matter in the Fok.nl forum.

Concerning the second research question, we have first found that every topic has shown a very different proportion of swearing words. Comments related to sexual positions, STDs, and the shape of the genitals have shown the largest proportions. Besides, swearing words related to illnesses, feces, sexuality, and homophobic and sexist slurs have a statistically significant different distribution among all the topics. We first hypothesised that sexual topics that seemed to be related to a more ostentatious sexual matter would use a larger amount of swearing words than topics categorised as more serious. Surprisingly, when people talk about porn and fantasy swear less than when they talk about sexual positions, even though we related these two topics in the same kind of context in the first place. Besides, people use a lot more profanity when approaching sexually transmitted diseases and condom use than when they talk about pregnancy, topics that we also linked together in the first place. In the case of STDs and condom use issues, people particularly use swearing words related to illnesses. Nonetheless, it is important to note that when people use illness slurs, they may not use it with the purpose of swearing but only to mention a certain disease – especially when talking about sexually transmitted diseases. Along with this, we also hypothesised that when people approach sexual health issues related to porn, fantasies and sexual positions would use more chauvinist slurs due to the extreme and violent language used in pornography. We indeed found that these topics have a larger proportion of sexist swearing words along with comments related to STDs and condom use. But not only that, these comments have shown a stronger use of profanity in general. It could be said that showing strong profanity within a comment may therefore decrease its level of credibility since it may show a lower level of expertise as well as a lower level of morality. Besides, using a large proportion of swearing words within a comment may also cause the same effect. Nonetheless, as mentioned before, some swearing words may not have the purpose of swearing and may be only used as a normal way of talking or as a reaction to a certain opinion. However, even if the purpose of the comment has goodwill and it is truthful, it does not mean that the employment of profanity makes a comment more credible. On the contrary, it may make the reader feel doubtful about the information they are reading.

Lastly, concerning our third research question, we have found that, in general, the great majority of comments are not sensational and negative at all. We firstly hypothesised that we would find a higher level of sensationalism and negativity when people talk about sexual positions, porn, and sexual fantasies, and less negativity when they approach more serious sexual health matters. However, we have found that the differences in sensational/negative tone across the topics are not significantly different from one another. Therefore, if sensational information has a negative impact on the credibility of the content, it could be said that when it comes to this marker, all the sexual topics we have worked with show a low level of sensational tone and so they may be seen as objective and unbiased. Perhaps, the way people talk about sexual matters such as porn and fantasies may not have anything to do with the way porn is produced per se and neither to the kind of language it uses. And the same applies for sexual positions and experiences. This finding could be seen as reassuring, especially when considering that teenagers are the ones involved in these kinds of online conversations.

8. Limitations

One of the limitations found in this research attains to the SMOG score. The score lacks scientific validity when applied to other languages (Contreras et al., 1999). Something that has not been taken into account in this research. Besides, research carried out by Brysbaert et al. (2021) has found that the Dutch language is very similar to English, but that Dutch words are on average half a letter longer – 5.1 letters per word instead of 4.6 (in non-fiction).

The second limitation found within the research is that, on the one hand, the number of comments per topic is not equal among topics. This may disrupt our findings in terms of the number of swearing words, level of readability, and tone of speech since for some topics we gather a lot more information than for other topics. Nonetheless, it is rather complicated to have a balanced number of comments for every sexual health subject given that people may get more involved in some topics than others. It may also be a mistake to get rid of comments to have a balance between the topics since we would therefore be losing information. On the other hand, many comments, a great majority, have shown to be very short. This may particularly bias the level of readability since some very short comments may therefore get a very low score.

Another important limitation we have found relates to the swearing words. First, swearing words are context dependent and they may be used as a profanity in some contexts and as a normal word in others. For instance, swearing words related to body parts may be used to simply name a specific part of the body and not with the purpose of swearing, or insulting someone or something. Coupled with this, the list of swearing words that have been created in this research is by no

means exhaustive. The type of swearing words and the way of swearing may evolve with time, and so words that are seen as profanity today may hold a different meaning in the future. Besides, only swearing words have been taken into account, yet not ways of swearing.

Lastly, when applying a one-way ANOVA test, we failed to meet the homogeneity assumption. In ANOVA, when homogeneity of variance is violated, there is a higher chance of falsely rejecting the null hypothesis which in our case was that 'the means of all swearing words frequencies are equal'. This may therefore imply that some categories are indeed not as significantly different from one another as shown in our analysis.

9. Future Work

Some further research that may be interesting to take into consideration may be to study the level of toxicity related to credibility in online sexual health information. Similar to the analysis of sensational and negative comments, in this case, the goal is to find rude, disrespectful, or unreasonable behaviours that will most likely make the user quit the conversation (Obadimu et al., 2019). The same authors state the relevance of finding these toxic comments online since toxicity, in its various forms, oftentimes disrupts constructive discussions in an online community. This research would be particularly interesting across different sexual health topics and so to find whether there are actual differences between them.

Another potential and intriguing research may be the analysis of slang regarding the level of credibility. According to a study conducted by Mazer and Hunt (2008) on 'affective learning and classroom climate among students', the use of positive slang correlates to higher levels of effective learning as well as to a more positive classroom climate. Some words understood as positive slang are 'cool', 'awesome', or 'sweet'. Besides, the opposite is true for negative slang. Several scholars suggest that the negative use of slang can be detrimental to the student learning outcome (Martin, Weber, and Burant, 1997, as cited in Mazer and Hunt, 2008). The same study refers to negative slang as an informal, obscene, and offensive use of language. According to this, analysing the use of slang may add some relevance when assessing the level of credibility in online sexual health information. More particularly when this is done on websites such as Fok.nl where younger people get involved in these kinds of conversations.

10. Bibliography

- Amos, R. (2020). *Readability Metrics*. GitHub Repository. <https://github.com/cdimascio/py-readability-metrics/blob/master/readability/scorers/smog.py>
- Bednarek, M. (2019). ‘Don ’ t say crap . Don ’ t use swear words .’ – Negotiating the use of swear / taboo words in the narrative mass media. *Discourse, Context & Media*, 29. <https://doi.org/10.1016/j.dcm.2019.02.002>
- Berger, V. W., and Zhou, Y. (2014). Kolmogorov–smirnov test: Overview. *Wiley Statsref: Statistics Reference Online*.
- Brysbaert, M., Sui, L., Duyck, W., and Dirix, N. (2021). Improving reading rate prediction with word length information: Evidence from Dutch. *Quarterly Journal of Experimental Psychology*, 74(11). <https://doi.org/10.1177/17470218211017100>
- Cense, M. (2019). Rethinking sexual agency : proposing a multicomponent model based on young people ’ s life stories Rethinking sexual agency : proposing a multicomponent model based on young people ’ s life stories. *Sex Education*, 19(3), 247–262. <https://doi.org/10.1080/14681811.2018.1535968>
- Contreras, A., García-Alonso, R., Echenique, M., and Daye-Contreras, F. (1999). The SOL formulas for converting SMOG readability scores between health education materials written in Spanish, English, and French. *J Health Commun*, 4(1). <https://doi.org/10.1080/108107399127066>
- Farrugia, A., Waling, A., Pienaar, K., and Fraser, S. (2021). *The “ Be All and End All ”? Young People , Online Sexual Health Information , Science and Skepticism*. <https://doi.org/10.1177/10497323211003543>
- Fitzsimmons, P., Michael, B., Hulley, J., and Scott, G. (2010). A readability assessment of online Parkinson ’ s disease information. *J R Coll Physicians Edinb.*, 40(4), 292–296. <https://doi.org/10.4997/JRCPE.2010.401>
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M., and Hall, C. (2001). *What Makes Web Sites Credible ? A Report on a Large Quantitative Study CHI 2001*. 61–68.
- Frankie, T. M., Ho, T., and Christie, C. A. (2012). The chi-square test: Often used and more often misinterpreted. *American Journal of Evaluation*, 33(3), 448–458.

- Gupta, A., and Kumaraguru, P. (2012). *Credibility Ranking of Tweets during High Impact Events*.
- Herzog, M. H., Francis, G., and Clarke, A. (2019). ANOVA. In *Understanding Statistics and Experimental Design* (pp. 67–82). Springer, Cham.
- IBM. (2021a). *Cramer's V*. <https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-cramers-v>
- IBM. (2021b). *Crosstabs Statistics*. <https://www.ibm.com/docs/en/spss-statistics/25.0.0?topic=crosstabs-statistics>
- Ley, P., and Florio, T. (1996). The use of readability formulas in health care. *Psychology, Health & Medicine*, 1(February 1996), 7–28. <https://doi.org/10.1080/13548509608400003>
- Liu, J., Song, S., and Zhang, Y. (2021). *Linguistic features and consumer credibility judgment of online health information*. 3, 1–9.
- Mazer, J. P., and Hunt, S. K. (2008). The Effects of Instructor Use of Positive and Negative Slang on Student Motivation , Affective Learning , and Classroom Climate The Effects of Instructor Use of Positive and Negative Slang on Student Motivation , Affective Learning , and Classroom Climate. *Communication Research Reports*, 25:1, 44–55. <https://doi.org/10.1080/08824090701831792>
- Mc Laughlin, H. (1969). SMOG Grading- a New Readability Formula. *Journal of Reading*, 12(8), 639–646.
- Mcperson, A. C., Gofine, M. L., Stinson, J., Mcpherson, A. C., Gofine, M. L., Seeing, J. S., Mcpherson, A. C., and Stinson, J. (2014). Seeing Is Believing ? A Mixed-Methods Study Exploring the Quality and Perceived Trustworthiness of Online Information About Chronic Conditions Aimed at Children and Young People. *Health Communication*, 29(5), 473–482. <https://doi.org/10.1080/10410236.2013.768325>
- Millstein, F. (2020). *Natural language processing with Python: natural language processing using NLTK*. Frank Millstein.
- Obadimu, A., Mead, E., Hussain, M. N., and Agarwal, N. (2019). Identifying toxicity within youtube video comment. *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 214–223.

- Olteanu, A., Peshterliev, S., Liu, X., and Aberer, K. (2013). *Web Credibility : Features Exploration and Credibility Prediction*. 557–568.
- PyPI.org. (2020). *py-readability-metrics*. <https://pypi.org/project/py-readability-metrics/#contributors->
- Ross, J. W. (2015). *Features for Ranking Tweets Based on Credibility and Newsworthiness*. *Features for Ranking Tweets Based on Credibility and Newsworthiness A thesis submitted in partial fulfillment by*.
- Ruette, T. (2019). *Why do the Dutch swear with diseases ?*
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2), 238–241.
- Tranchese, A., and Sugiura, L. (2021). “ I Don ’ t Hate All Women , Just Those Stuck-Up Bitches ”: How Incels and Mainstream Pornography Speak the Same Extreme Language of Misogyny. *Violence Against Women.*, 27(14), 2709–2734.
<https://doi.org/10.1177/1077801221996453>
- White, D., and Korotayev, A. (2004). Statistical Analysis of cross-tabs. *Anthrosciences.Org*.
- Wikipedia. (2022). *Dutch Profanity*. https://en.wikipedia.org/wiki/Dutch_profanity