Effects of Peer Feedback on Password Strength

Faisal Irfan Yar Khan

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Computer Science & Software Engineering

University of Washington

2017

Reading Committee:

Marc Dupuis, Chair

Mark Kochanski

William Erdly

Program Authorized to Offer Degree:

Computing and Software Systems

University of Washington

Abstract

Effects of Peer Feedback on Password Strength

Faisal Irfan Yar Khan

Chair of the Supervisory Committee:
Dr. Marc Dupuis
Computing and Software Systems

This thesis is a study on the effects of peer-feedback on a user's password strength. Passwords are a common sight in everyday use of an average end user. Text-based passwords are heavily relied upon when it comes to user authentication employed in various account management scenarios. Most users do not pay attention to or understand the importance of creating a secure password. Lack of strong passwords means that it is the single most vulnerable point to gain unauthorized access to the resource as prior studies have uncovered that most passwords are significantly weak and hence, easy to crack. Consequently, exploring mechanisms which improve password security has been the main focus of a significant body of research. To this end, we introduced a peer-feedback password meter which shows how the strength of the user's password compares to the strength of passwords used by other users. To achieve this goal, we

conducted a user study where we asked users to create an account on a hypothetical website. The users were either shown a traditional password meter or a peer-feedback meter. Our findings suggest that when told to create a unique password, the peer-feedback password meter significantly increased the strength of the password as compared to a traditional password meter. This approach could potentially be one of the methods to encourage end users to create a stronger password.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to offer my sincere gratitude to my supervisor and committee chair Dr. Marc J. Dupuis, for his countless hours of guidance, encouragement, support, and patience. Special thanks to Mr. Mark Kochanski and Dr. William Erdly for agreeing to serve on my committee. I would also like to thank the faculty, my peers, and staff at the University of Washington Bothell who motivated me in their unique ways, especially Mrs. Megan Jewell, Mrs. Krista Garg, and Mrs. Rosemary Simmons. I also want to acknowledge and thank various services available at the University of Washington Bothell which helped me reach my goals.

DEDICATION

I dedicate this thesis work to my family and close friends who supported me through this entire process. No words can reflect the amount of gratitude I have for everyone, especially my parents and siblings. I would also like to extend special thanks to Mr. Nadeem Aziz, Mr. Naveed Aziz, and Mr. Zain Aziz for the constant encouragement, support, and contribution throughout my master's program and special thanks to Dr. Sobia Azhar Khan for proof-reading and editing this thesis.

Chapter 1. INTRODUCTION

According to some estimates, online users are expected to reach 4 billion by 2019, generating around 44 zettabytes of data by 2020 [1]. All of this data is valuable for the individual, the institution, and any malicious actor. Recent data breaches suffered by Yahoo!, Dropbox, and LinkedIn had exposed up to 732 million user details [2]. The exposure of personally identifiable information (PII) includes information such as email addresses, passwords, secret questions, and their answers. It has compromised the confidentiality and integrity of information and risked the privacy and security of the user. Institutions suffering these attacks face economic loss as well as loss of reputation. Some studies suggest that in 2015 cyber crime victims spent around $126 billion to deal with the fallout of an attack [3]. Around 62% of the data breaches that occurred in 2016 were as a result of hacking, and out of those a staggering 81% leveraged the use of either prior stolen information or leveraged weak passwords [4].

Passwords have dominated the world of authentication. Their widespread use has made them a prized target for attackers. If the attackers can get hold of a password, they can essentially compromise the security of an entire system. In one study, it was found that “123456” was the most popular password followed by “12345” [5]. Thus, it is not surprising that out of 500 compromises in 15 countries, 28% of those were as a result of weak passwords [6]. The compromised financial accounts (e.g., PayPal) went onto to be sold in the black market for upwards of $60 per account [7]. Yet, other research revealed that around 15% of the passwords were either a partner’s or child’s name, followed by the names of football teams and pets [8]. In one survey conducted it was found that 26.2% of the respondents have on average more than 50 accounts. Additionally, 50.9% of the respondents stated that they often reuse their passwords because they are easy to remember [9]. If one account gets compromised, this could potentially create a chain reaction of other associated accounts of that individual to be compromised as well.

A weak password is defined as phrases that are trivial to guess or those likely to be cracked by using a dictionary attack [10]. Researchers identified passwords as being the weakest link in the security chain at least as early as 1979 [11]. Many reasons have been speculated as to why a user may choose weak passwords to begin with. The reason comes down to human psychology. People prefer convenience, speed, and memorability of a weak password than the complexity of a strong password [12].

Kevin Mitnick, arguably the most famous hacker, pointed out that [8]:

> The human side of computer security is easily exploited and constantly overlooked. Companies spend millions of dollars on firewalls, encryption, and secure access devices, and it's money wasted because none of these measures address the weakest link in the security chain.

Mitnick's statement essentially means that no matter how much money and effort is spent building security around a system to prevent attacks, as long as humans – the weakest link in this security chain – continue to create hackable passwords, computer systems will remain vulnerable.

New and improved ways of authentication have been introduced in the industry, such as two-factor authentication and biometrics, but the use of passwords still dominates the industry. In many cases, users are forced to choose stronger passwords in accordance with strict password policies. However, such policies tend to isolate a user and do not improve password quality and may, in turn, lead to the repetitive use of a single password across websites and services [13]. In such scenarios, a user might tend to use mechanisms (e.g. writing passwords down on paper) which might be even more detrimental to security than the use of weak passwords [14].

Various schemes have been employed to strengthen password security to resist such attacks. Numerous websites and applications use password meters to help users create a stronger password. As shown in Figure 1-1, a password meter is a graphical representation that informs the user how secure the password is on some scale.

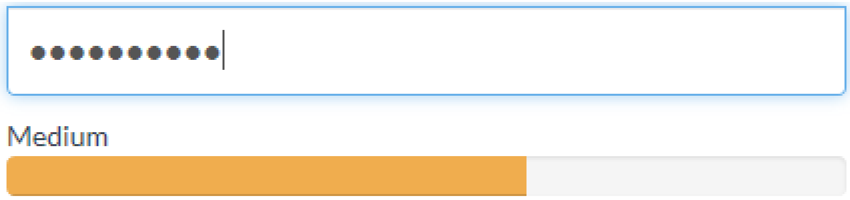

Figure 1-1. Graphical password meter

The objective of having a password meter is to provide visual feedback to the user on their choice of a password by labeling it weak, medium, or strong, for example. The criteria for calculating the strength of a password is set by a developer or an institution, such as checking for minimum length, use of dictionary words, and use of special characters. One drawback of this approach is that there is a multitude of varying implementations of password meters, each having its unique set of password rules, which can yield a different result for the same password phrase. The differing results can cause confusion and even distrust by the user. When conflicting feedback for the same password occurs, it decreases a user's trust and willingness to comply with the system [15].

The evidence suggests the presence of password meters for password change and creation might lead to a more secure password [16]. While implementing such meters may encourage users to create longer and hence more secure passwords, users tend to dislike having to follow strict guidelines for choosing such passwords [17]. Another drawback is that it is hard to remember such passwords since password meters do not necessarily increase their memorability.

In this thesis, we incorporated social influence, which is the effect others have on an individual's attitude and behavior [18]. This social influence, commonly known as peer-feedback, was incorporated in the design of a password meter which we are calling a peer-feedback meter. Our approach is an application of social navigation which states that users navigate through a system based on the inherent design of the system. This method also allows users to understand and interpret the actions of others as well [19]. This modified password meter was used to assess whether it resulted in a stronger password than a traditional password meter, and if so, to what extent.

1.1 THESIS OUTLINE

The remainder of the thesis is organized into the following chapters. Chapter two provides background by examining related work on text-based passwords, password meters, and details on how a user navigates through social influences. Chapter three provides the design and methodology used in our study to investigate the effect of peer-feedback on the strength of the password. Chapter four presents the results of our study. Finally, chapter five provides a summary of the results, the conclusion of our study, its limitations, and the implications of our work.

Chapter 2. LITERATURE REVIEW

Typically, user authentication falls into one of three categories: 1) “what you know” (e.g. text-based passwords or PINs); 2) “what you have” (e.g. a token or a smart card), and 3) “who you are” (e.g. biometrics) [20]. In this thesis, we are going to focus on “what you know,” specifically text-based passwords and how we can improve their effectiveness in security.

2.1 PASSWORDS AND PASSWORD METERS

More than 35 years ago, Moris and Thompson identified text-based passwords as the weak point in a system’s security [11]. In the study, they found that approximately 86% of the users had weak passwords characterized either as short, all lower or uppercase, or could easily be found in the dictionary. Despite their efforts to raise awareness, contemporary password and password meter designs have not evolved at the same pace over the decades as they would have expected and still to this day remain one of the weakest links.

Researchers and industry experts understood that because of poor password choices by the user, the inherent security risks associated with it materialized [21]. Because of this, many researchers proposed that creating good password policies would help increase security [22]. However, later research established that these policies are often misunderstood and even become cumbersome for users, thereby defeating the purpose of having passwords. As a result, users default to insecure practices such as writing their passwords down on post-it notes and attaching it on their workstations. These are the very practices which password policies hoped to mitigate [23].

Beyond password policies, password strength meters have been used as a visual representation of how secure a password is against the rules set by the developer. As shown in Figure 1-1, they advise or force the user to choose a complex password if the password they chose is considered weak. This method helps to prevent the user from choosing easily guessable passwords and may even enforce the password policies. These password meters are also known as proactive password checkers and have been in use for many years [21]. Instead of relying on a user to create a robust and a secure password, they continuously check the typed phrase against the set of rules and provide feedback on it (e.g. if the password is too short or is missing special characters). The evidence suggests that implementing password meters assists in creating stronger passwords [16].

2.1.1 *Inner workings of a password strength meter*

The goal of every password meter is to check if an attacker can easily crack a given password and provide this feedback to the end user. To achieve this, password meters employ the calculation of bit-space or entropy based on the length and the character set (e.g. lowercase, uppercase, numbers, symbols) used in the password. These password checkers often penalize a user for using a dictionary word since a dictionary attack can be used to crack such passwords easily. National Institute of Standards and Technology (NIST) recommend that passwords have a minimum of one lowercase, one uppercase, one number, and one special character with a minimum length of eight characters. Additionally, they should not be a permutation of the username and should not contain phrases from a language dictionary [24].

However, there is no formalized standard across the industry which defines how much entropy is considered weak, strong, or very strong when it comes to password meters. In fact, in a large-scale comparative analysis of various password meters used in popular websites (e.g. Microsoft, Google, Dropbox, Apple, Yahoo!, Skype, Twitter, PayPal), researchers found that these platforms do not explain the logic behind the strength assignment of the password and give varying strength scores for the same password [13].

This is where our work differentiates from previous work. In our study, administrators do not set the values of how strong a given password is; instead, it is left to the potential user of the system. Brown et al. support our view that users should be able to decide which application requires stringent password security (e.g. account with financial and banking information) as opposed to accounts like forums that may not require the same level of password strength [25].

2.2 PSYCHOLOGY AND SOCIAL ASPECTS

Theoretically, using password meters can be used to block all possible weak passwords, but in reality, it is impractical as these passwords are often difficult to memorize [26]. Users tend to avert stringent rules [27], and as a result, indulge in the same insecure practices we mentioned earlier. There is always a trade-off to be made between security (i.e. using strong passwords) and usability (i.e. easy to remember passwords) [28]. This security tradeoff creates a problem for IT professionals. Users are usually aware of what a good password is, but despite this understanding, they are still inclined to take risks and are optimistically biased; they believe a negative event is unlikely to happen to them [29]. In a study of 20,907 people, researchers found that around 76% of the repondents know that they should protect their information online yet they engage in sharing their passwords and other risky behaviors [3]. When their understanding is lacking, it may be due to them overestimating the benefits of a few predictable practices, such as adding digits or phrases, which inadvertently weakens their password [30].

In this study, we incorporate the recommendations of earlier research that password meters should be targeted towards users and be data driven. We provide feedback based on an individual's password proactively, which is unique to that specific individual. Through this study, we seek to develop a different type of password meter in hopes to achieve a better balance between the usability and security trade-off.

Incorporating social influences into cyber security is not new. In one study, researchers found that users shared their information depending on how their peers made decisions [31]. In another study, the researchers showed how the influence of framing impacts user tolerance on security delays [32]. In cognitive psychology, the notion of framing effect states that how a risk is framed can influence people's actions [33]. By manipulating the way information is presented, we can influence and alter decisions and judgments about that information. This theory suggests that people tend to avoid risks if a positive frame is offered and seek risks when a negative frame is presented. The idea of social influence has been adopted in other computing fields apart from security. For instance, YouTube and Netflix recommend content to users based on what other users have liked. Our design of the system reflects such social influence attributes. The feedback the password meter gives is based upon peer-feedback and how the feedback is expressed is influenced by the framing effect.

Considering this past research, the following hypothesis was developed. In this study, we wanted to explore if users can be encouraged to generate strong passwords if they are given the feedback on how their passwords compare with their peers, which can be referred to as subjective norm. Subjective norm is defined as an individual's perception of what people in society think about their behavior [34]. This comparison with their peers includes the strength of passwords used by individuals in their university community. We are interested in finding out if the peer feedback password meter results in stronger passwords compared to a traditional password meter.

> H1: Using peer-feedback password meters increases the password strength as compared to using the traditional password meter.

Chapter 3. DESIGN AND METHODOLOGY

For testing of the hypothesis, we designed a single blind experiment where we masked the true goals of this study so that the participants in this study will not consciously be biased when they saw both password meters. Thus, conducting a laboratory experiment was a logical choice as it gave us more control of the environment and any external influencing factors. We created a hypothetical website for UW students called “Slack @UW.” Slack is a cloud-based set of team collaboration tools and services which allow users to create groups, conduct real-time communication, and sharing of resources for projects.

We developed and tested a functioning prototype of an experimental peer-feedback password meter which displayed the password strength based on how it compared to previously registered users. However, to check the peer-feedback meter and its effect we only had to test how the visual representation of the peer-feedback meter impacted the password strength. Hence, to be consistent, we changed how the score was calculated on the backend of the system. Instead of calculating based on a comparison to previously registered users, we calculated based on the scoring algorithm discussed in section 3.1. Both the traditional and peer-feedback meters used the same color schemes, scoring rules, and evaluation of feedback based on scores. The feedback evaluation is illustrated in Table 3-2. The value of the progress bar was calculated by multiplying the password strength value by two.

In our study, we informed each participant that this system is being developed specifically for UW students to assist them in collaborating on group projects. Hence, we created a perception amongst the participants that this was a real prototype website soon to be launched for them and they had a real stake in it. This was the key in our study as previous research had shown that systems which contain sensitive or critical information (e.g. financial) about users are perceived to require strong passwords as compared to accounts associated with blogs or forums [35]. In our view, our hypothetical website was around the middle to upper-middle level on the scale of sensitive information stored.

During the early part of our study and based in part on the survey they each had completed, we became concerned that participants were simply using an existing password rather than developing a new one. At approximately the midpoint of the study, we decided to investigate if explicitly stating in the oral instructions to users that they need to create a unique password would cause them to instead develop a new password rather than reusing an existing one. The use and subsequent effectiveness of cues have been shown in other research [31].

We created two variations of the same website. The first one was the control with the traditional meter as illustrated in Figure 3-1.

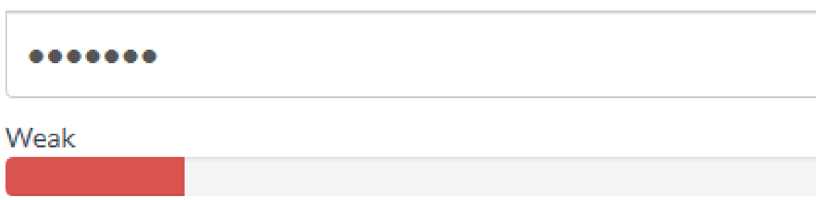

Figure 3-1. Traditional Password Meter

The second one was our experimental variant with the peer-feedback meter as shown in Figure 3-2.

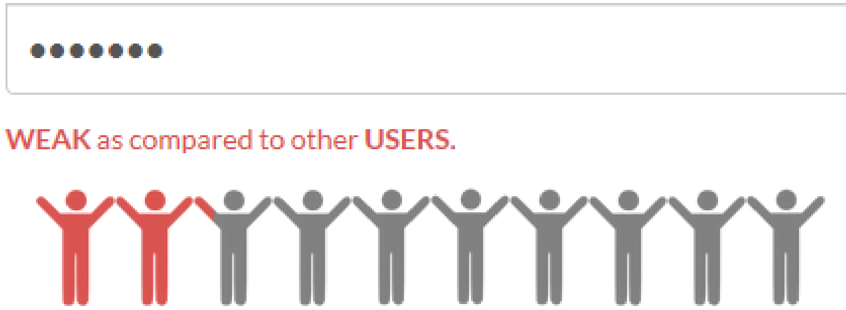

Figure 3-2. Peer-feedback meter

The difference between the two meters was how they were visualized and what feedback was given. The traditional meter simply stated the feedback and set the percentage value on the progress bar. The peer-feedback meter, however, stated the feedback in comparison to other users and instead of displaying a traditional progress bar we created a 10-person series bar which reflected that it is a percentage from a population. Apart from this meter, both the variants had identical content and design. The full web page screenshots are available in Appendix B.

3.1 Password Scoring Algorithm

An open source library for password strength calculation was used called "jQuery Password Strength Meter for Twitter Bootstrap" [36]. This library uses custom rules and settings to calculate the strength of a given password. The library allows a developer to use a scoring system which rewards or penalizes when rules match. Rules can be minimum length, word repetitions, characters used, etc. The default scoring rules are shown in Table 3-1.

Table 3-1. Algorithm rule based scores

Rule	**Score**
wordNotEmail	-100
wordLength	-50
wordSimilarToUsername	-100
wordSequences	-50
wordTwoCharacterClasses	2
wordRepetitions	-25
wordLowercase	1
wordUppercase	3
wordOneNumber	3
wordThreeNumbers	5
wordOneSpecialChar	3
wordTwoSpecialChar	5
wordUpperLowerCombo	2
wordLetterNumberCombo	2
wordLetterNumberCharCombo	2

The total score at the end is raised to the power of 1.4 based on the password length. For the traditional meter, the following score values, shown in Table 3-2, are used to determine the feedback verdict given back to the user.

Table 3-2. Traditional meter scoring

Score	**Feedback verdict**
Score < 0	Very Weak
0 > Score < 14	Weak
14 > Score < 26	Normal
26 > Score < 38	Medium
38 > Score < 50	Strong
Score > 50	Very Strong

We kept the same scoring rules for the peer feedback meter. However, instead of displaying the same feedback verdict, we calculated the score of the password and then calculated the percentage of users that score below the current user's score. Moreover, we used the score results to display the feedback. While a password could still be strong despite being weaker than a significant number of peers, the primary goal in this research was to determine how a user responds when presented with information on how the strength of his password compares with his peers. Thus, it is possible that other factors could be considered in such a comparison, such as the form or structure of the password apart from its strength, but that is not the focus of this research.

3.2 DESIGN OF THE SYSTEM

The website was coded in HTML, CSS, and JavaScript for the front end. The backend server functionality was coded in PHP. A MySQL database was used to store the data on the server. Apart from the website, there was another web page created which acted as an admin panel. It was password protected with credentials so only the research team would have access to the data.

The data stored for the study was the user's email address, degree program, their IP address and location (resolved through IP address lookup), the meter type (traditional or peer-feedback) they were exposed to, their password strength calculated using the algorithm discussed in 3.1, the characteristics of the passwords (e.g. number of digits, lowercases, uppercase and repetitions), hashed password using "bcrypt" algorithm and salted with a random cryptographic salt, and the number of unique password tries along with their score. Each user was assigned a unique identifier or UUID which was based on Mersenne Twister Random Number Generator. The database itself was encrypted and each data column was encrypted with AS 256-bit algorithm for further protection.

Figure 3-3 shows the workflow of how the system operated. The user would interact with the user front end–the "Slack @UW" website. Various web service calls can be made through the frontend for the required functionality. A password protected MySQL database was used to store the necessary data and information for the study. The researchers could interact using a password protected admin frontend to view data in the database.

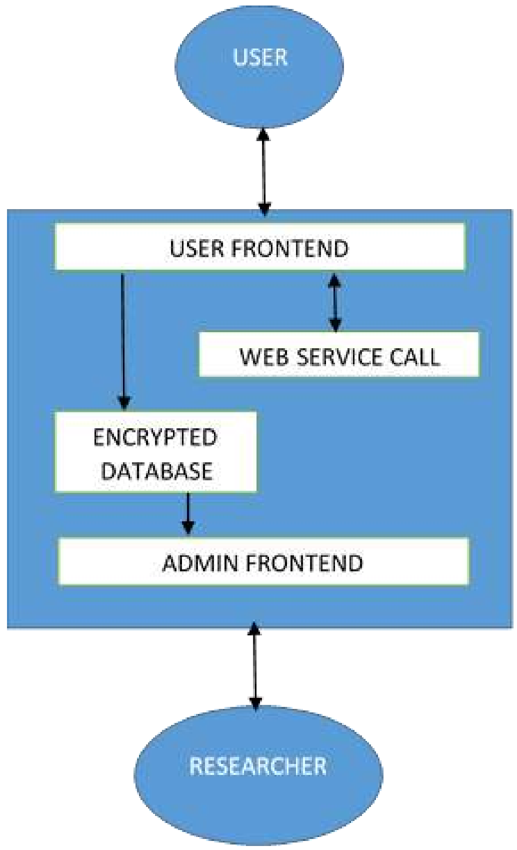

Figure 3-3. System workflow diagram

We administered a survey to participants to collect demographic information along with Internet security behavior questions to gauge the sample population's expertise in online security and how they behave online. The entire survey is in Appendix A of this thesis. The survey also served as a distraction to determine if the participants would be able to recall the password they chose after approximately five minutes.

IRB approval was obtained before recruitment and subsequent participation in the study. The study ran from April 12, 2017 until May 10, 2017. For the recruitment of participants, flyers were posted around various campus buildings. The potential recruits were required to sign up by completing a qualifying survey where if selected they were invited to schedule an appointment to participate in the study. The participants were provided with a $20 Amazon gift card as compensation for their time.

Chapter 4. RESULTS

Forty-eight students were recruited as participants for this study. They ranged in age from 18 to 39 years old with 24 males and 24 females. Table 4-1 shows how the participants were distributed in each age group. One participant declined to identify an age group.

Table 4-1. Age group of participants

Age Group	# of participants
18-19	9
20-24	28
25-29	6
30-34	2
35-39	2

Figure 4-1, shows the distribution of gender across various ethnicities. Approximately 54% of participants identified as White/Caucasian with 42% females and 58% males, followed by 27% who identified as Asian/Pacific Islander with 31% males and 69% females.

Figure 4-1. Ethnicity by gender

Out of the 48 total participants, 14 were from a computing background and the rest were from a non-computing background.

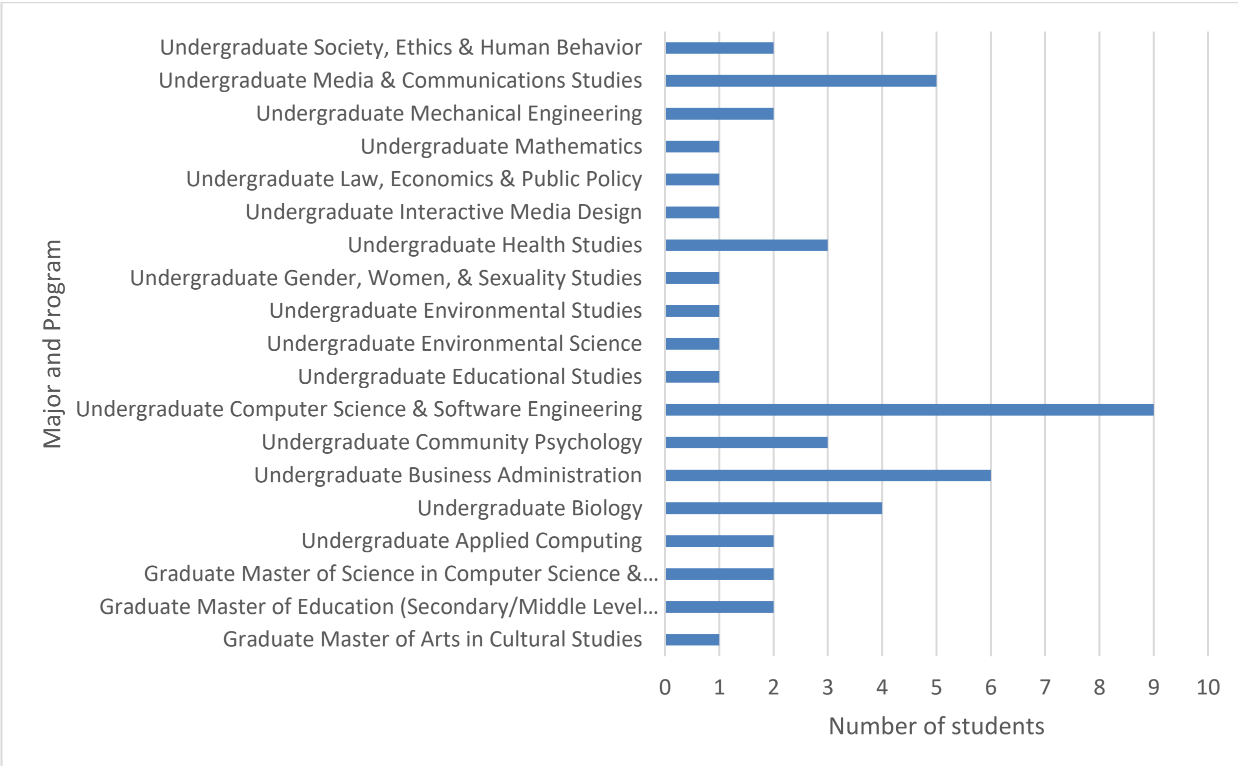

Figure 4-2. Major and program distribution of participants

Figure 4-2 shows participants from each of the degree programs. Nine participants were from the Computer Science & Software Engineering undergraduate program followed by six from the undergraduate Business Administration program.

Twenty-four of the participants were in the control group, while the other 24 were in the treatment group. Half of the participants in each group were given explicit instructions to create an account using a unique password that they had not used anywhere previously. The other half of the participants in each group were not given explicit instructions. The distribution of participants is illustrated in Table 4-2. Both groups were told to explore the website and create an account on the sign-up page. After registering for an account, they were redirected to an online survey. Upon completion of the survey, they were asked to log in using the credentials they had created on the sign-up page.

Table 4-2. Participant distribution among the groups

Explicit instruction group		**Non-explicit instruction group**	
Male Peer Feedback	6	Male Peer Feedback	6
Male Traditional	6	Male Traditional	6
Female Peer Feedback	6	Female Peer Feedback	6
Female Traditional	6	Female Traditional	6
Total Peer Feedback	12	**Total Peer Feedback**	12
Total Traditional	12	**Total Traditional**	12
Total Male	12	**Total Male**	12
Total Female	12	**Total Female**	12
Total	24	**Total**	24

Both groups had identical lab settings. To minimize any external influence, the participant and the researcher met individually in a meeting room. The participants were asked to read and sign a consent form. The researcher then presented the participant with either the control website with the traditional password meter or the experimental website with the peer-feedback password meter. Participants were then given instructions. Care was taken to ensure each group had an equal number of both males and females. The participants were redirected to sign in with their credentials which they had created on the sign-up page after they had completed the online survey. The participants were not told about the sign-in beforehand as this approach ensured that the participants would not make a conscious effort to memorize or write down the password to find out if creating a stronger password had any effect on the memorability of the said password phrase.

For hypothesis testing, IBM SPSS was used to perform independent samples t-test on the results obtained. As mentioned in chapter three, we could not find a statistically significant result without the explicit instructions to create unique password. However, when the cue was explicitly stated during both the traditional and peer-feedback meters, a statistically significant difference between the control and treatment groups were found ($t=1.882$; $p < 0.05$). Statistics on mean, standard deviation and standard error mean are shown in Table 4-3.

Table 4-3. Statistical results

Group Statistics

	Meter Type	N	Mean	Std. Deviation	Std. Error Mean
Password Meter with instructions	Traditional	12	39.8082	15.80699	4.56308
	Peer-Feedback	12	53.3495	19.26955	5.56264

Out of all the participants, only 16.67% (eight participants) had difficulties remembering their entered password. Four of the participants could enter their correct password at most on their second attempt. Table 4-4 shows participants who failed to enter their correct password at any point on multiple tries.

Table 4-4. Number of participants who failed to remember password

	# of participants failed to remember password
Traditional meter with no explicit instruction	1
Peer-Feedback meter with no explicit instruction	1
Traditional meter with explicit instruction	2
Peer-Feedback meter with explicit instruction	0

Further analysis of the data collected revealed that when the peer-feedback meter was presented, it took an average of 6.42 tries until the user could finalize their password selection.

Table 4-5 shows the average number of attempts it took to decide upon the final password for each treatment given.

Table 4-5. Number of attempts to submit final password

	Average # of attempts to submit final password
Traditional meter with no explicit instruction	2.92
Peer-Feedback meter with no explicit instruction	2.83
Traditional meter with explicit instruction	5.50
Peer-Feedback meter with explicit instruction	6.42

Figure 4-3 illustrates how the score of both password meter treatments within the explicit instruction groups influenced the resulting password strength of users. There was an increase in the average password strength of peer-feedback meter group as compared to the traditional meter group. This suggests that the peer-feedback meter influenced the users to create a stronger password when every other condition was kept same between the two treatment groups.

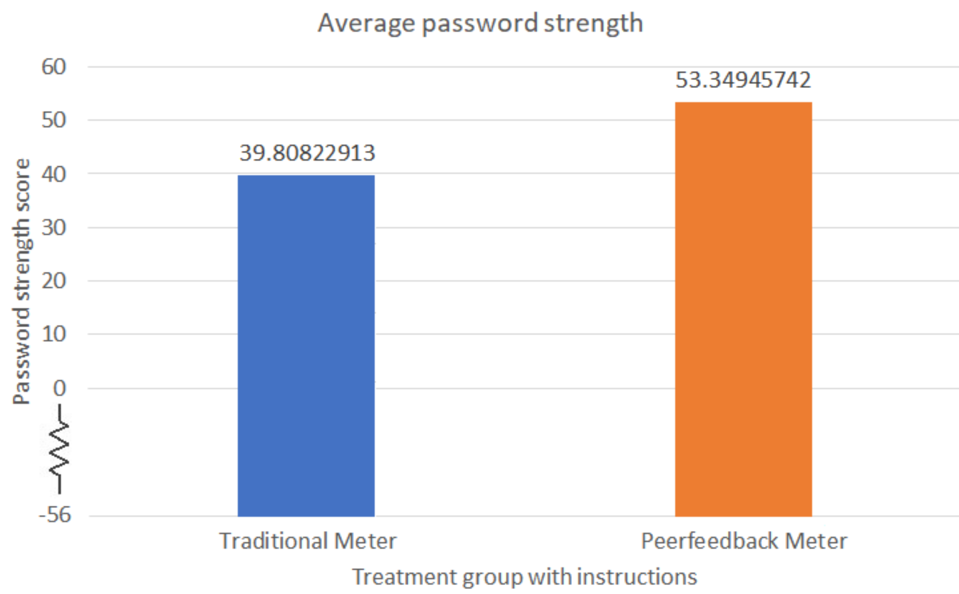

Figure 4-3. Average password strength with explicit instructions

Taking a deeper look into the above results, Figure 4-4 illustrates that males consistently created stronger passwords as compared to females; however, the difference found was not statistically significant. Moreover, as the figure shows, when given explicit instructions the peer-feedback meter increases password strength as compared to traditional password meters.

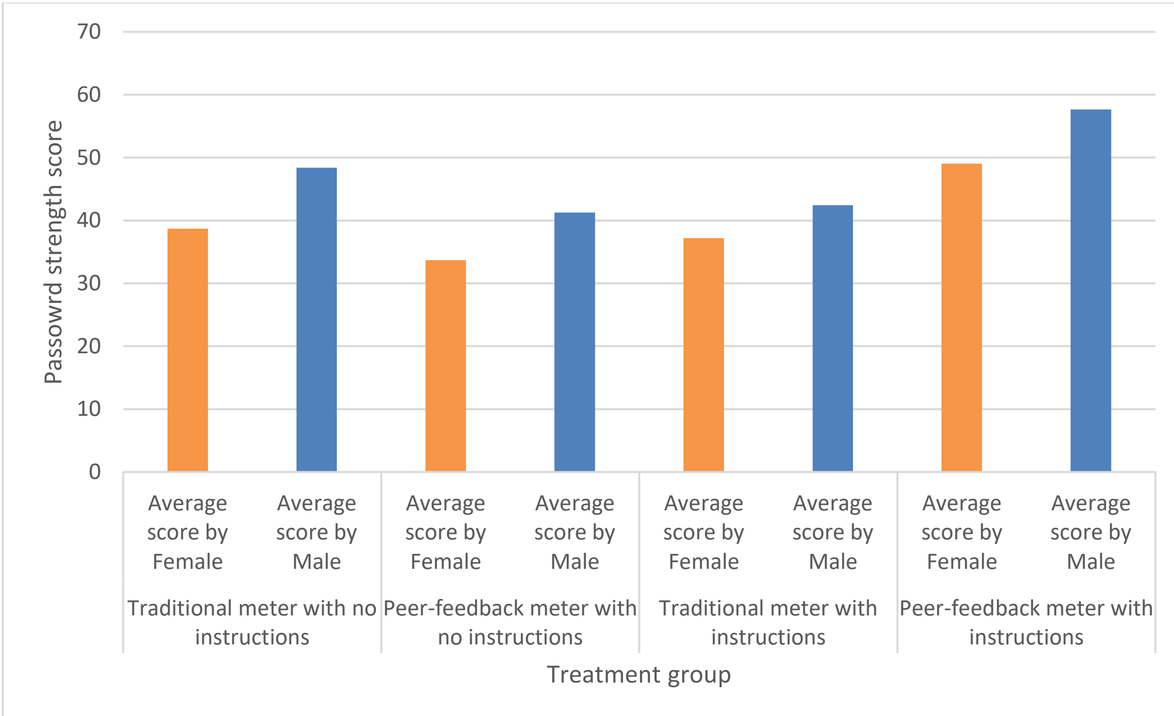

Figure 4-4. Average password strength score by gender and treatment

We analyzed the data collected from the surveys and present some of the results here. Figure 4-5 depicts the frequency in which participants changed their passwords.

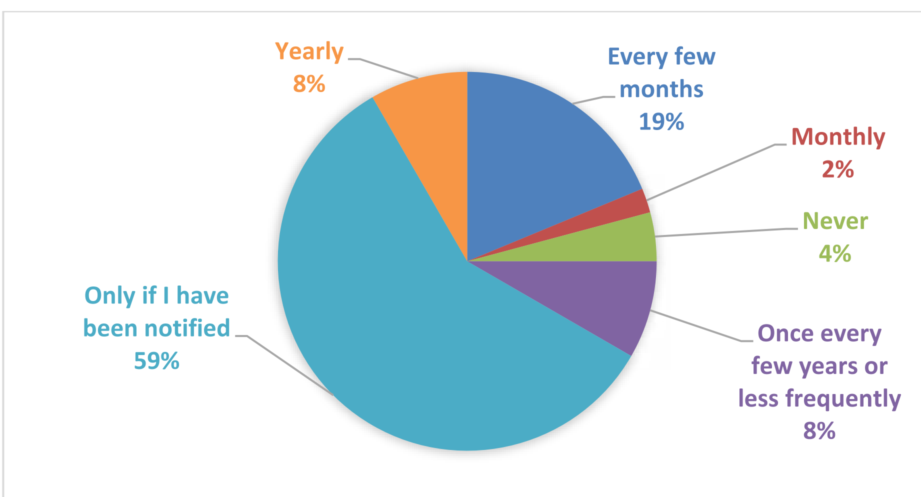

Figure 4-5. Password change frequency

Additionally, it revealed that approximately 59% of the participants only change their passwords if they have been notified of a security breach, as shown in Figure 4-5. This approach to managing passwords is problematic [37].

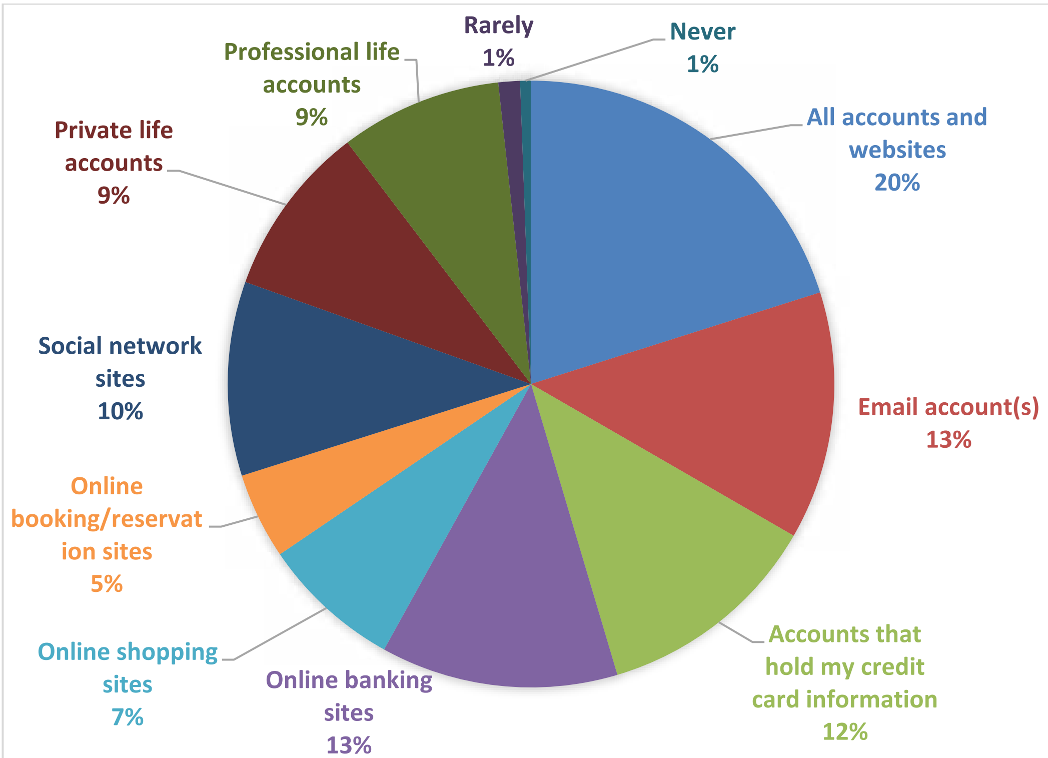

Figure 4-5. Attempts to create secure passwords

Another interesting insight we gathered from the survey results, as shown in Figure 4-5, is that around 20% of the participants stated they create secure passwords for all their accounts, followed by emails and banking websites at 13%.

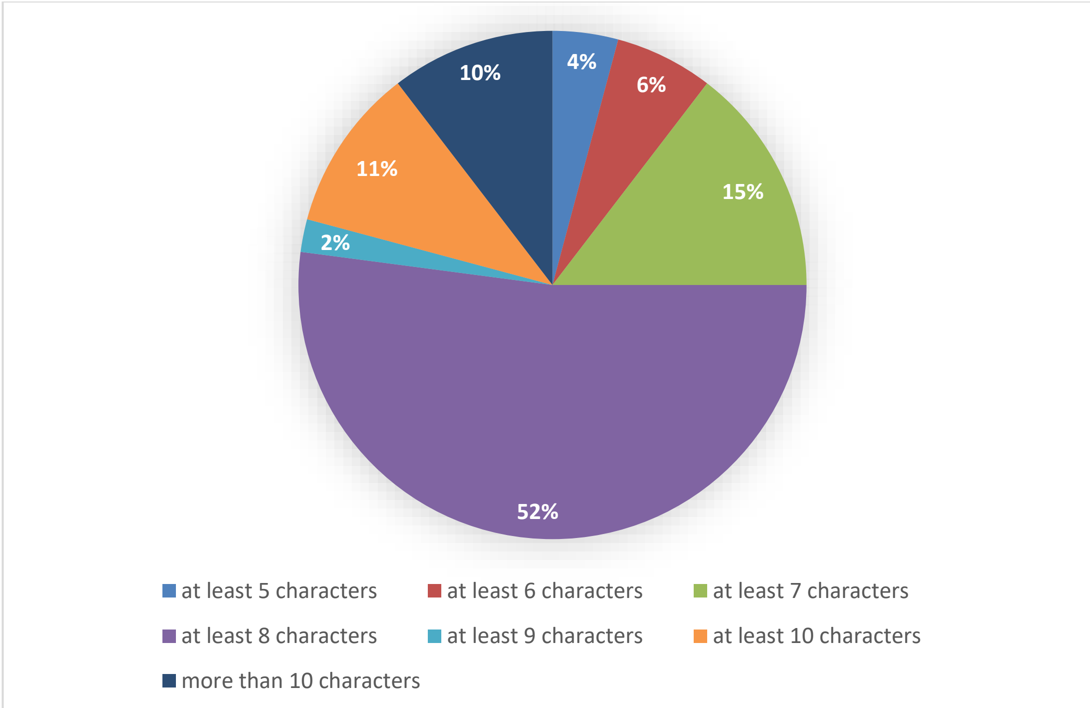

Figure 4-6. Minimum length for the password

Approximately 52% of the users perceived that having a minimum of eight characters in a password constitutes as good, with around 10% of the users believed the password should be more than 10 characters as shown in Figure 4-6.

Additional statistics from the survey are mentioned in Appendix C of this thesis.

Chapter 5. CONCLUSION

The hypothesis under investigation in this study was whether or not peer influence had any effect on a user's password choice. To address this question, we conducted an experiment on a pool of 48 university students. In the design of this experiment, we had two forms of password meters, one was the traditional password meter and the other was a peer-feedback password meter. Additionally, we administered two different treatments. One was without explicit instructions to create a unique password and the other was with these explicit instructions.

Our results suggest that the peer-feedback meter, when administered alongside explicit instructions, would create relatively stronger passwords as compared to traditional password meters. However, when no explicit instructions were given, we did not find a statistically significant difference. We hypothesize that such prototype peer-feedback meters could have the most benefit on platforms which already depend upon social connections between users, alongside a mechanism to prompt the user to create a unique password. Platforms like Facebook, LinkedIn, Google and Outlook, are prime examples.

5.1 LIMITATIONS

Our experimental design had several limitations. First, our sample population was not a true representative sample of the general population. Our participants were university students whose mean age ranged from 20 to 24 years old. This age group is a part of "Net Generation" and is considered more intuitively aware of the digital and IT space around them [38]. Additionally, Western college-educated participants represent a unique cognitive makeup different from the population as a whole [39]. The laboratory experimental design of our study has an inherent limitation in that it has low generalizability and lacks ecological validity. Participants were out of their natural environment and we were able to exert some control over them. Thus, field experiments would need to be done to determine if the results obtained here would also be found in the natural environment [40].

Another limitation is that in our quest to find effects of peer influence in password meters, we did not test different variations of the GUI design. Although it is uncertain if a different design would or would not have any effect, different prototypes could be developed and tested in a pilot study to determine the effectiveness of each. Some of the prototype designs we initially considered include: 1) how would the peer-feedback influence if the users were told their passwords strength compared to users from another geographical region, and 2) showing users a percentage value of how much stronger or weaker their password is when compared to others.

Due to constraints of this thesis, we did not have the opportunity to conduct a post-interview with the participants. This post-interview could have yielded insights into what the participants semantically inferred from the feedback "Weak as compared to other USERS." Did they perceive it as their password strength being similar or different from other users?

Additionally, our study involved compensation worth $20 and thus could influence some individuals to have participated in the study that otherwise would not have participated.

Finally, for new systems traditional password meters would need to be shown to gather password strengths data which would be used to compare against other users. This may introduce a vulnerability of the system, if attackers can breach the system and crack the database encryption, they might find accounts which have weaker passwords and prioritize by attacking them first. Long-term implementation of such a system can be complicated. In a scenario when most of the users are storing weak passwords by current scoring algorithms, if a new password is created which might be relatively better but not strong on its own, the peer-feedback meter might show it as a much stronger password compared to others. Hence, this would lead to a general degradation of password strength quality over time. One way to avoid this is to use a combination of policies, such as enforcing complex character set and minimum password length to increase the entropy.

5.2 FUTURE WORK

Although user authentication mechanisms have improved over the years, more needs to be done. Our attempt was designed to investigate a direct relationship between peer influence and password strength through the use of peer-feedback meters. However, we did not conclusively establish if peer-feedback generated passwords led to more memorable passwords, which weighs heavily in the usability-security tradeoff. Future research should focus on investigating that relationship in detail by employing various other algorithms based on entropy to find a direct link between the peer-feedback and the increase in entropy.

We had only a single prototype design for the peer-feedback password meter. A variety of such peer-feedback meters should be tested to determine if they can improve the quality of generated passwords by delivering the feedback in a more user-friendly manner.

Additionally, our study had a small sample size. Conducting a large-scale study on systems which have a high volume of users would generate more reliable data. Moreover, conducting this study in such prototype peer-feedback meters on functioning systems would help evaluate our hypothesis as the data of users can be compared to those of others. Also, it is worth mentioning that due to constraints of this thesis work we did not check the password strength scores when no password meter is present at all. This would require a larger sample for the study and future research can design studies which compare all three scenarios.

This study was conducted in a lab setting whereas conducting it in realistic or field environment would produce results with higher external validity. Thus, future work should seek to collect data in a more naturalistic setting.

BIBLIOGRAPHY

[1] McAfee labs, “McAfee Threat Predictions 2016,” 2016.

[2] Experian Data Breach Resolution, “DATA BREACH INDUSTRY FORECAST,” no. 4, 2017.

[3] N. Symantec, “Norton Cyber Security Insights Report 2016,” *Nort. by Symantec*, pp. 3–9, 2016.

[4] Verizon, “2017 Data Breach Investigations Report,” no. 10, p. 74, 2017.

[5] W. Han, Z. Li, L. Yuan, and W. Xu, “Regional Patterns and Vulnerability Analysis of Chinese Web Passwords,” *IEEE Trans. Inf. Forensics Secur.*, vol. PP, no. 99, pp. 1–1, 2015.

[6] Trustwave, “2015 Trustwave Global Security Report,” p. 90, 2015.

[7] Verizon, “2016 Data Breach Investigations Report,” *Verizon Bus. J.*, no. 1, pp. 1–65, 2016.

[8] J. Godin, “Infosecurity Europe 2004: Office workers give away passwords for a chocolate bar,” *Univ. Wired*, no. April, p. 1, 2014.

[9] K. Security, “Crossing the password chasm,” no. January, 2017.

[10] E. H. Spafford, “OPUS: Preventing weak password choices,” *Comput. Secur.*, vol. 11, no. 3, pp. 273–278, 1992.

[11] R. Morris and K. Thompson, “Password security: a case history,” *Commun. ACM*, vol. 22, no. 11, pp. 594–597, Nov. 1979.

[12] D. V Klein, “Foiling the cracker: A survey of, and improvements to, password security,” *Proc. 2nd USENIX Secur. Work.*, pp. 5–14, 1990.

[13] X. D. C. De Carnavalet and M. Mannan, “A Large-Scale Evaluation of High-Impact Password Strength Meters,” *ACM Trans. Inf. Syst. Secur.*, vol. 18, no. 1, p. 1:1--1:32, 2015.

[14] A. Adams and M. A. Sasse, “Users are not the enemy,” *Commun. ACM*, vol. 42, no. 12, pp. 40–46, Dec. 1999.

[15] X. D. C. De Carnavalet and M. Mannan, “From Very Weak to Very Strong : Analyzing Password-Strength Meters,” *Ndss 2014*, no. February, pp. 23–26, 2014.

[16] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley, “Does My Password Go Up to Eleven? The Impact of Password Meters on Password Selection,” *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 2379–2388, 2013.

[17] B. Ur *et al.*, “How does your password measure up? The effect of strength meters on password creation,” 2012, pp. 65–80.

[18] L. F. Berkman, “Social support, social networks, social cohesion and health.,” *Soc. Work Health Care*, vol. 31, no. 2, pp. 3–14, 2000.

[19] P. DiGioia and P. Dourish, “Social navigation as a model for usable security,” in *Proceedings of the 2005 symposium on Usable privacy and security - SOUPS ’05*, 2005, pp. 101–108.

[20] L. O’Gorman, “Comparing passwords, tokens, and biometrics for user authentication,” *Proc. IEEE*, vol. 91, no. 12, pp. 2021–2040, Dec. 2003.

[21] M. Bishop and D. V. Klein, “Improving system security via proactive password checking,” *Comput. Secur.*, vol. 14, no. 3, pp. 233–249, Jan. 1995.

[22] S. Komanduri *et al.*, “Of Passwords and People: Measuring the Effect of Password-Composition Policies,” *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI ’11*, p.

2595, 2011.
[23] P. G. Inglesant and M. A. Sasse, "The true cost of unusable password policies," *Proc. 28th Int. Conf. Hum. factors Comput. Syst. - CHI '10*, p. 383, 2010.
[24] W. E. Burr *et al.*, "Electronic Authentication Guideline," Gaithersburg, MD, Nov. 2013.
[25] A. S. Brown, E. Bracken, S. Zoccoli, and K. Douglas, "Generating and remembering passwords," *Appl. Cogn. Psychol.*, vol. 18, no. 6, pp. 641–651, 2004.
[26] J. J. Yan, "A note on proactive password checking," in *Proceedings of the 2001 workshop on New security paradigms - NSPW '01*, 2001, p. 127.
[27] B. Ur *et al.*, "How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation Blase," in *Security'12 Proceedings of the 21st USENIX conference on Security symposium*, 2012, pp. 5–16.
[28] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. (Belin) Tai, J. Cook, and E. Eugene Schultz, "Improving password security and memorability to protect personal and organizational information," *Int. J. Hum. Comput. Stud.*, vol. 65, no. 8, pp. 744–757, Aug. 2007.
[29] M. Whitty, J. Doodson, S. Creese, and D. Hodges, "Individual differences in cyber security behaviors: An examination of who is sharing passwords," *Cyberpsychology, Behav. Soc. Netw.*, vol. 18, no. 1, pp. 3–7, Jan. 2015.
[30] B. Ur, J. Bees, S. M. Segreti, L. Bauer, N. Christin, and L. F. Cranor, "Do Users ' Perceptions of Password Security Match Reality ?," *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI '16*, pp. 3748–3760, 2016.
[31] A. Besmer, J. Watson, and H. R. Lipford, "The impact of social navigation on privacy policy configuration," in *Proceedings of the Sixth Symposium on Usable Privacy and Security - SOUPS '10*, 2010, p. 1.
[32] S. Egelman, D. Molnar, N. Christin, A. Acquisti, C. Herley, and S. Krishnamurthi, "Please Continue to Hold An empirical study on user tolerance of security delays," 2010.
[33] D. Kahneman and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econom. J. Econom. Soc.*, vol. 47, no. 3, pp. 263–291, 1979.
[34] I. Ajzen, "The theory of planned behavior," *Organ. Behav. Hum. Decis. Process.*, vol. 50, no. 2, pp. 179–211, 1991.
[35] W. Khern-am-nuai, W. Yang, and N. Li, "Using Context-Based Password Strength Meter to Nudge Users ' Password Generating Behavior : A Randomized Experiment," *SSRN*, p. 27, 2016.
[36] T. Piper and A. Blanco, "jQuery Password Strength Meter for Twitter Bootstrap." GitHub, 2016.
[37] W. C. Summers and E. Bosworth, "Password policy : The good , the bad , and the ugly," *Res. Gate*, no. February, 2015.
[38] D. Oblinger and J. Oblinger, "Is It Age or IT: First Steps Toward Understanding the Net Generation," *Educ. net Gener.*, vol. Chapter 2, no. 2, p. 2.1-2.20, 2005.
[39] J. Henrich, S. J. Heine, and A. Norenzayan, "Most people are not WEIRD," *Nature*, vol. 466, no. 7302, pp. 29–29, Jul. 2010.
[40] J. B. Kessler and L. Vesterlund, "The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects," in *Handbook of Experimental Economic Methodology*, no. 2006, Oxford University Press, 2015, pp. 391–406.

APPENDIX A

Each participant was presented with a survey on demographics and their internet security behavior during the study. Below is the complete questionnaire of the survey.

1. What year are you in college?

    - 1st year (freshman)
    - 2nd year (sophomore)
    - 3rd year (junior)
    - 4th+ year (senior)
    - 1st year of graduate school
    - 2nd year of graduate school
    - Post-Bac

2. What is/are your major(s) or intended major(s) (if not yet declared)?

    - American and Ethnic Studies (BA)
    - Applied Computing (BA)
    - Biology (BS)
    - Business Administration (BA)
    - Chemistry (BS)
    - Chemistry (BA)
    - Climate Science & Policy (BS)
    - Community Psychology (BA)
    - Computer Engineering (BS)
    - Computer Science & Software Engineering (BS)
    - Culture, Literature & the Arts (BA)
    - Educational Studies (BA)
    - Electrical Engineering (BS)
    - Environmental Science (BS)
    - Environmental Studies (BA)
    - Gender, Women, & Sexuality Studies (BA)
    - Global Studies (BA)
    - Health Studies (BA)
    - Individualized Study - Interdisciplinary Study (BA)
    - Interactive Media Design (BA)
    - Interdisciplinary Arts (BA)
    - Law, Economics & Public Policy (BA)
    - Mathematical Thinking and Visualization (BA)
    - Mathematics (BS)

- Mechanical Engineering (BS)
- Media & Communications Studies (BA)
- Nursing (BS), First Year Entry RN to BSN (freshman entry)
- Nursing (BS), RN to BSN (transfer entry)
- Physics (BA)
- Physics (BS)
- Science, Technology & Society (BA)
- Society, Ethics & Human Behavior (BA)
- Leadership MBA (LMBA) (Bellevue)
- Technology MBA (TMBA) (Bothell)
- Master of Arts in Cultural Studies
- Master of Arts in Policy Studies
- Master of Education
- Master of Education - Special Education Leadership (ECSEL)
- Master of Education Leadership Development for Educators (LEDE)
- Master of Education (Secondary/Middle-Level Endorsement)
- Master of Fine Arts in Creative Writing and Poetics
- Master of Nursing
- Master of Science in Accounting
- Master of Science in Computer Science & Software Engineering
- Master of Science in Cyber Security Engineering
- Master of Science in Electrical Engineering

3. Are you an international student?

    - Yes
    - No

4. What gender do you identify with?

    - Male
    - Female
    - Other
    - Do not wish to answer

5. What ethnicity do you primarily identify with?

    - Asian/Pacific Islander
    - Black/African-American
    - White/Caucasian
    - Hispanic
    - Native American/Alaskan Native
    - Other/Multi-Racial

6. What is your current age?

    - Under 18
    - 18-19
    - 20-24
    - 25-29
    - 30-34
    - 35-39
    - 40-44
    - 45-49
    - 50-54
    - 55-59
    - 60-64
    - 65-69
    - 70-74
    - 75-79
    - 80-84
    - 85-89
    - 90 & above

7. What title best describes the type of work you currently do?

    - Student
    - Accounting / Finance / Banking
    - Administration / Clerical / Reception
    - Advertisement / PR
    - Architecture / Design
    - Arts/Leisure / Entertainment
    - Beauty / Fashion
    - Buying / Purchasing
    - Construction
    - Consulting
    - Customer Service
    - Distribution
    - Education
    - Health Care (Physical & Mental)
    - Homemaker
    - Human resources management
    - Management (Senior / Corporate)
    - News / Information
    - Operations / Logistics
    - Planning (Meeting, Events, etc.)
    - Production
    - Real Estate

- Research
- Restaurant / Foodservice
- Retired
- Sales / Marketing
- Science / Technology / Programming
- Social service
- Unemployed
- Other

8. Please check the devices that you own AND have used within the past 3 months (select all that apply):

	iPhone/iPad/Apple/Mac	Android	Windows/Microsoft	Linux-based	Other
Smartphone(s)	☐	☐	☐	☐	☐
Tablet(s)	☐	☐	☐	☐	☐
Laptop(s)	☐	☐	☐	☐	☐
Desktop(s)	☐	☐	☐	☐	☐
Smartwatch(es)	☐	☐	☐	☐	☐

9. Please select the device that you use a majority of the time for computing purposes (NOT including the time you may use it for talking on the phone or texting):

○ Smartphone (iPhone)	○ Tablet (iPad)	○ Laptop (Mac)	○ Desktop (Mac)	○ Smartwatch (Apple)
○ Smartphone (Android)	○ Tablet (Android)	○ Laptop (Linux based)	○ Desktop (Linux based)	○ Smartwatch (Android)
○ Smartphone (Windows)	○ Tablet (Windows)	○ Laptop (WIndows)	○ Desktop (Windows)	○ Smartwatch (Microsoft)
○ Smartphone (other)	○ Tablet (other)	○ Laptop (other)	○ Desktop (other)	○ Smartwatch (other)

10. How often do you use the following social networking platforms?

	Never	Rarely (once every month or so)	Sometimes (once every week or two)	Often (once every day or two)	Very Often (several times a day)	All the time (checking it several times every hour I am awake)
Facebook	○	○	○	○	○	○
Twitter	○	○	○	○	○	○
LinkedIn	○	○	○	○	○	○
Pinterest	○	○	○	○	○	○
Google Plus	○	○	○	○	○	○
Tumblr	○	○	○	○	○	○
Instagram	○	○	○	○	○	○
VK	○	○	○	○	○	○
Flickr	○	○	○	○	○	○
Vine	○	○	○	○	○	○

11. How many friends do you have on the social networking platform Facebook?

- 0-9
- 10-24
- 25-49
- 50-99
- 100-199
- 200-299
- 300-499
- 500-749
- 750-999
- 1,000-1,500
- 1,500 or more

12. What best describes you as a computer user?

- Novice user (you just started using computers)
- Average user (you use word processors, spreadsheets, Email, surf the Web, etc.)
- Advanced user (you can install software, setup configurations, etc.)
- Expert user (you can setup operating systems, know some computer programming languages, etc.)

13. How many "passwords/pins/passphrases" do you have to remember at any one time in your everyday life? (for accessing anything)

- 1-2

- 3-4
- 5-6
- 7-8
- 9-10
- 11+

14. Do you use a Password Manager to remember passwords for you?

- Yes
- No

15. How often do you change your passwords for existing accounts?

- Weekly
- Monthly
- Every few months
- Yearly
- Once every few years or less frequently
- Only if I have been notified, there has been a problem, such as a data breach or compromise
- Never

16. I try to create secure passwords for (choose all that apply):

- all my accounts and websites
- my email account(s)
- accounts that hold my credit card information
- online banking sites
- online shopping sites
- online booking/reservation sites
- social network sites
- private life accounts
- professional life accounts
- rarely
- never

17. I have shared my passwords with others for the following types of accounts (choose all that apply):

- all my accounts and websites
- my email account(s)
- accounts that hold my credit card information
- online banking sites
- online shopping sites

- online booking/reservation sites
- social network sites
- private life accounts
- professional life accounts
- rarely
- never

18. In your opinion, what should the minimum length of a password be?

- at least 5 characters
- at least 6 characters
- at least 7 characters
- at least 8 characters
- at least 9 characters
- at least 10 characters
- more than 10 characters

19. Please indicate the extent to which you agree with the following statement:

I always check to make sure websites where I enter passwords use SSL (https://) before proceeding.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree
- Not sure what SSL (https://) is

20. Please indicate the extent to which you agree with the following statement:

I use unique passwords for each of my accounts.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

APPENDIX B

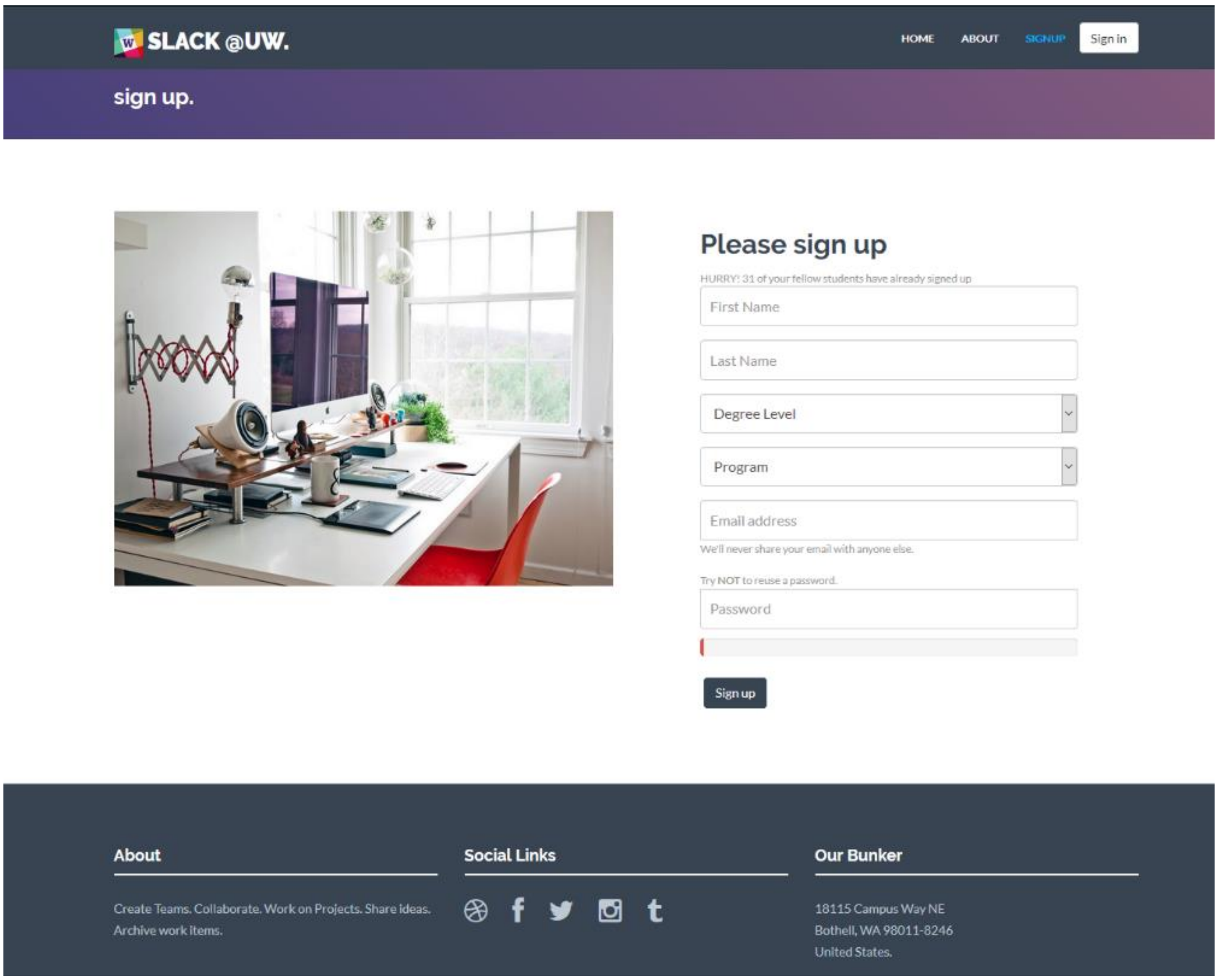

Figure 5-1. Web page with traditional password meter

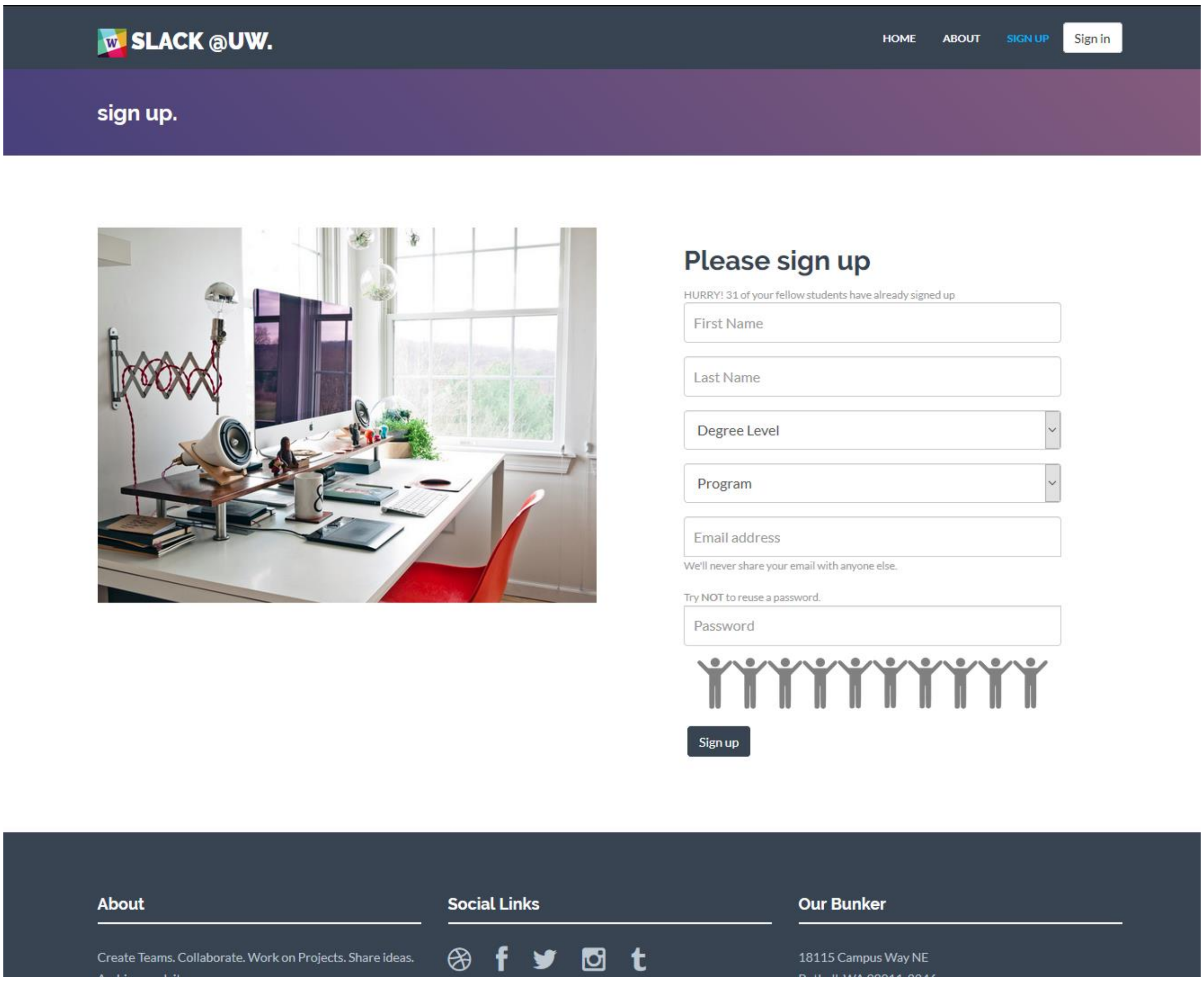

Figure 5-2. Web page with peer-feedback password meter

APPENDIX C

Complete survey results are shown below in the order they appear on the survey.

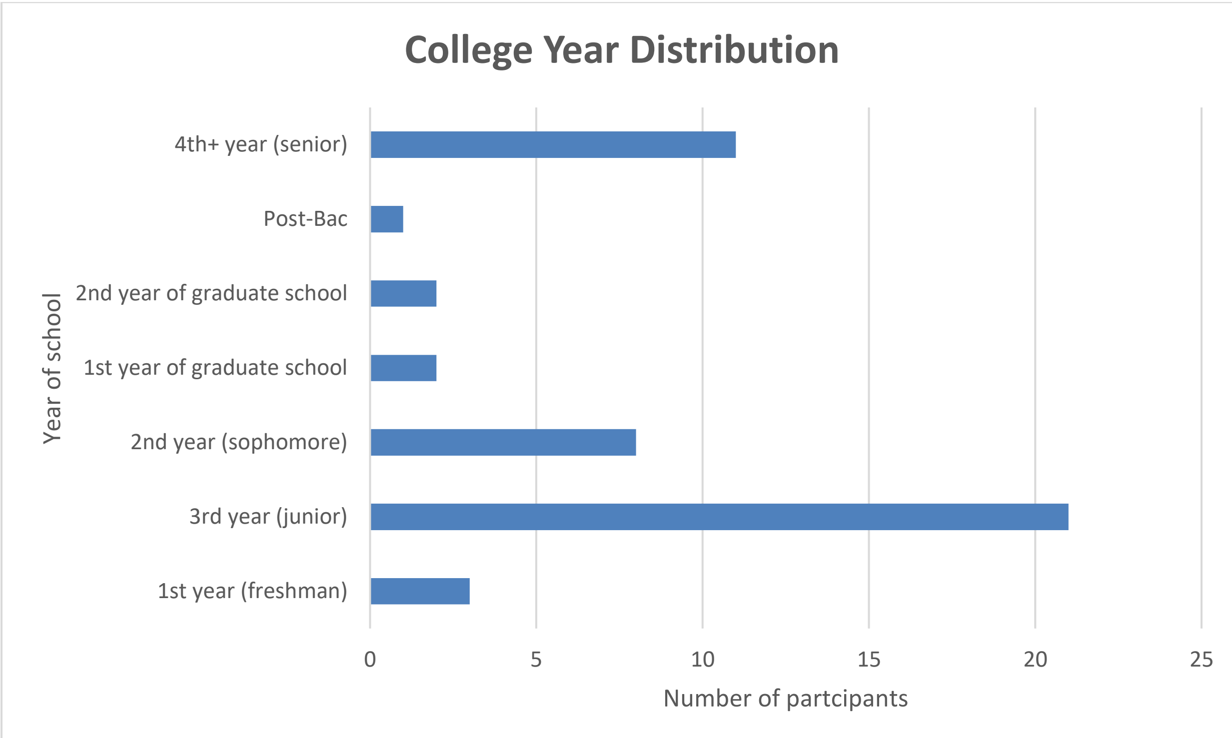

Figure 5-3. College year distribution

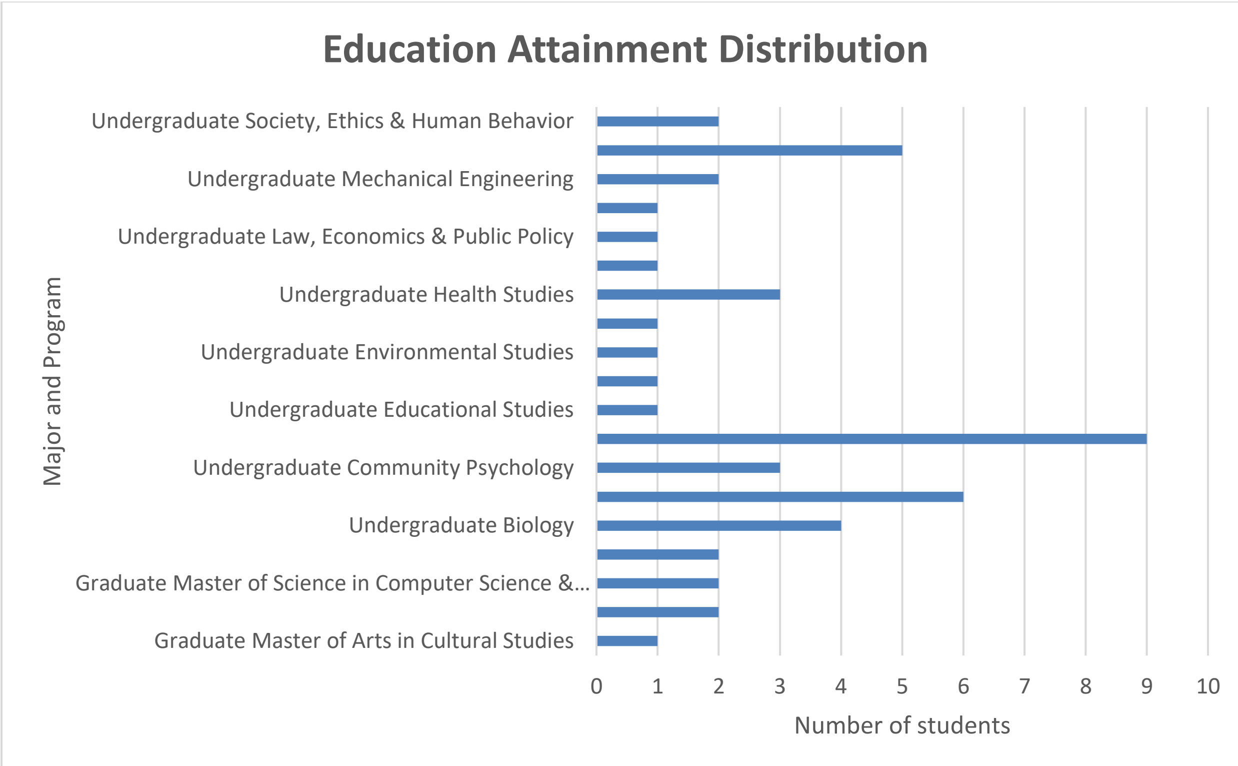

Figure 5-4. Major and program distribution among participants

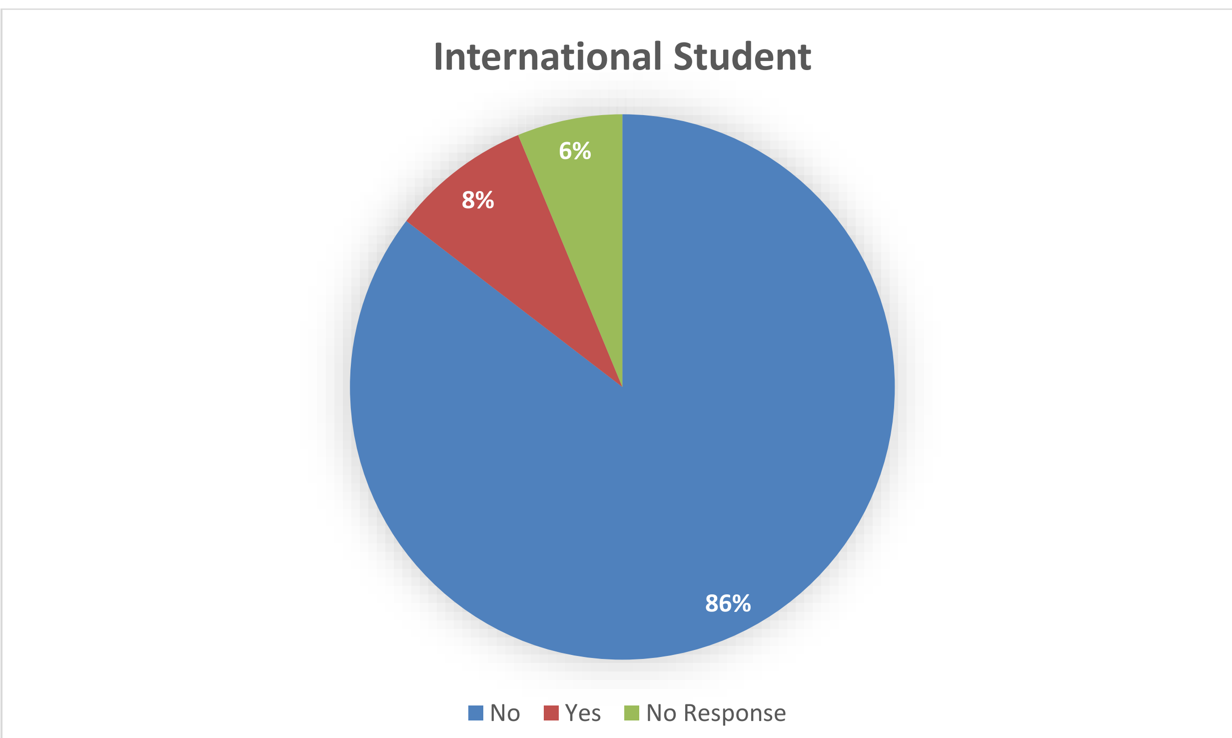

Figure 5-5. International students in sample population

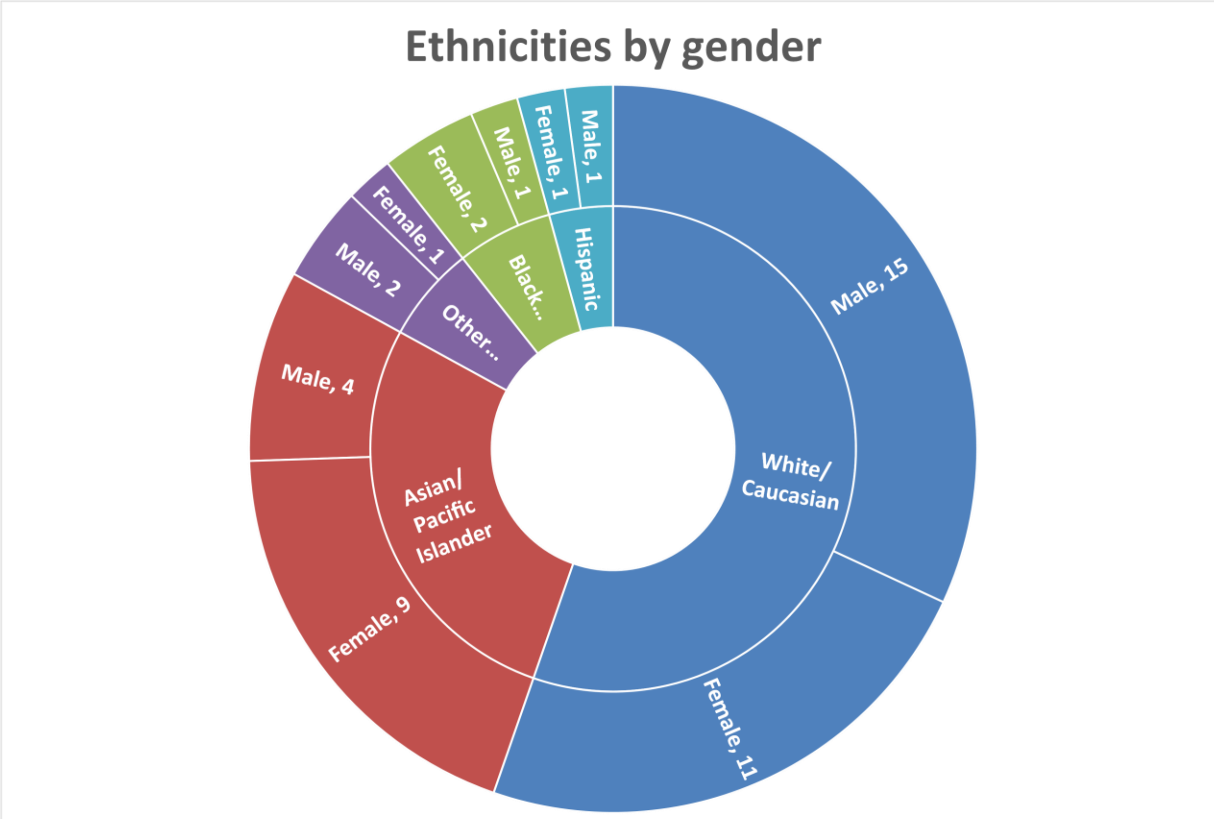

Figure 5-6. Ethnicities by gender distribution

Age Group	# of participants
18-19	9
20-24	28
25-29	6
30-34	2
35-39	2

Table 5-1. Participants' age group

Figure 5-7. Occupation of the participants

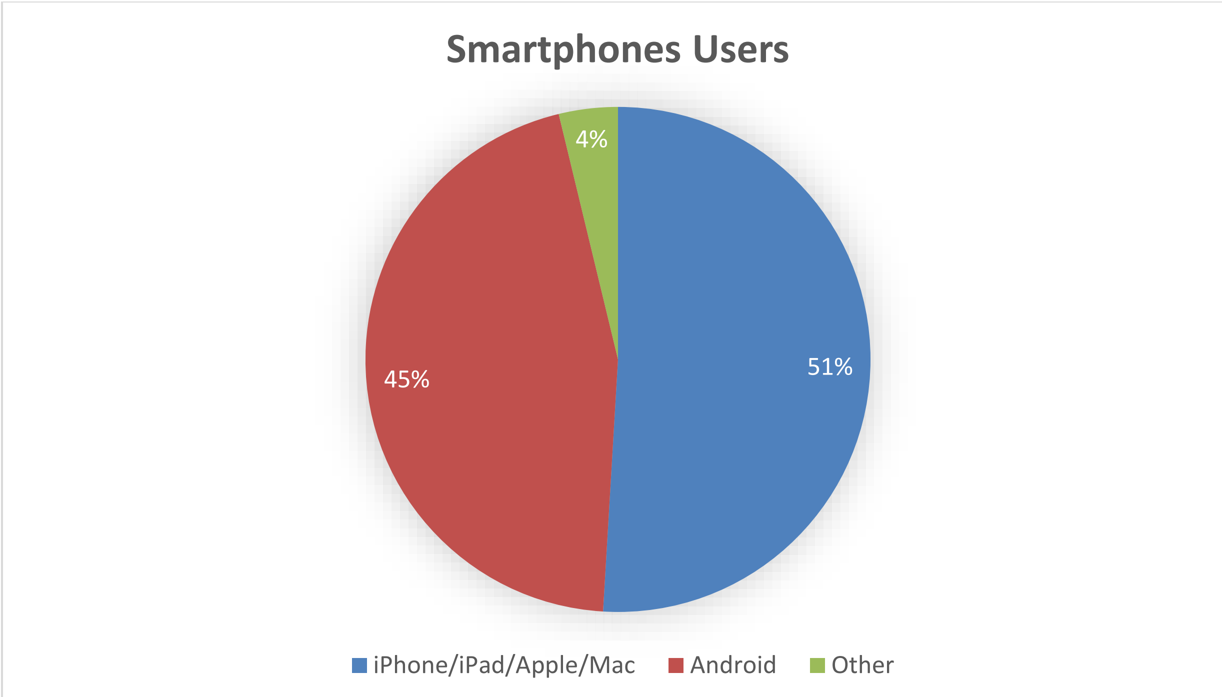

Figure 5-8. Type of smartphone users among the sample population

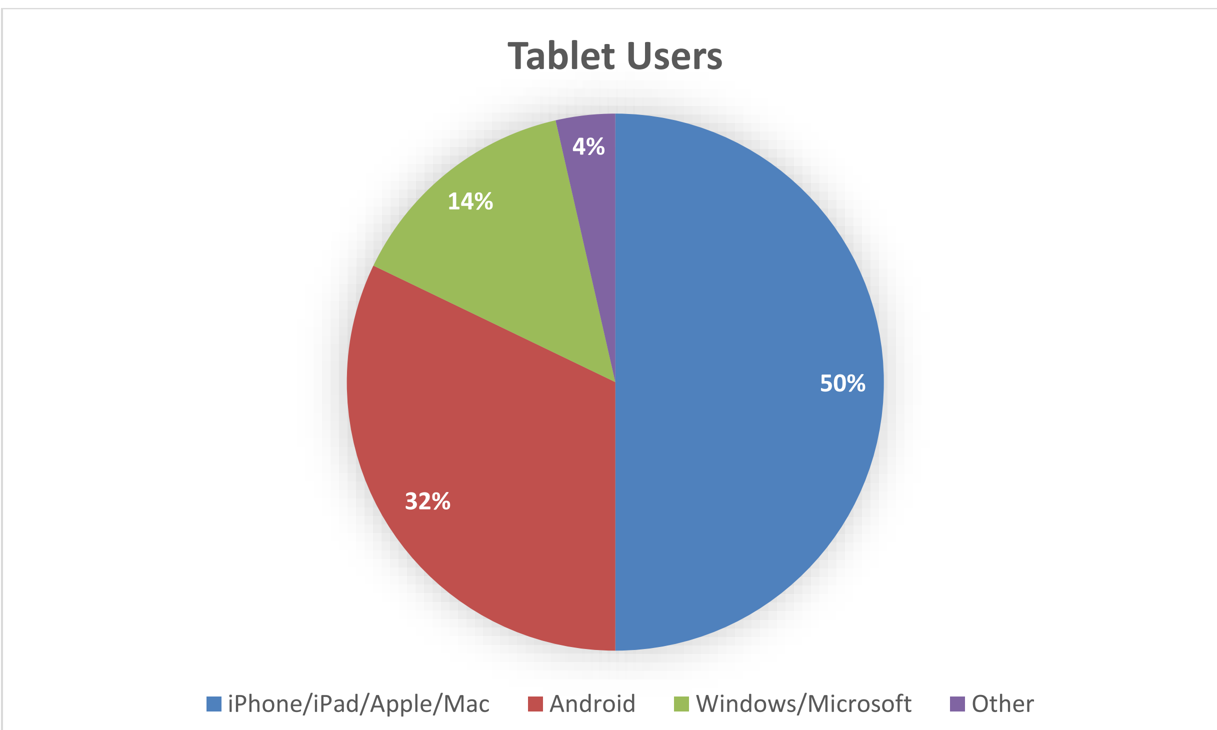

Figure 5-9. Type of tablet users among the sample population

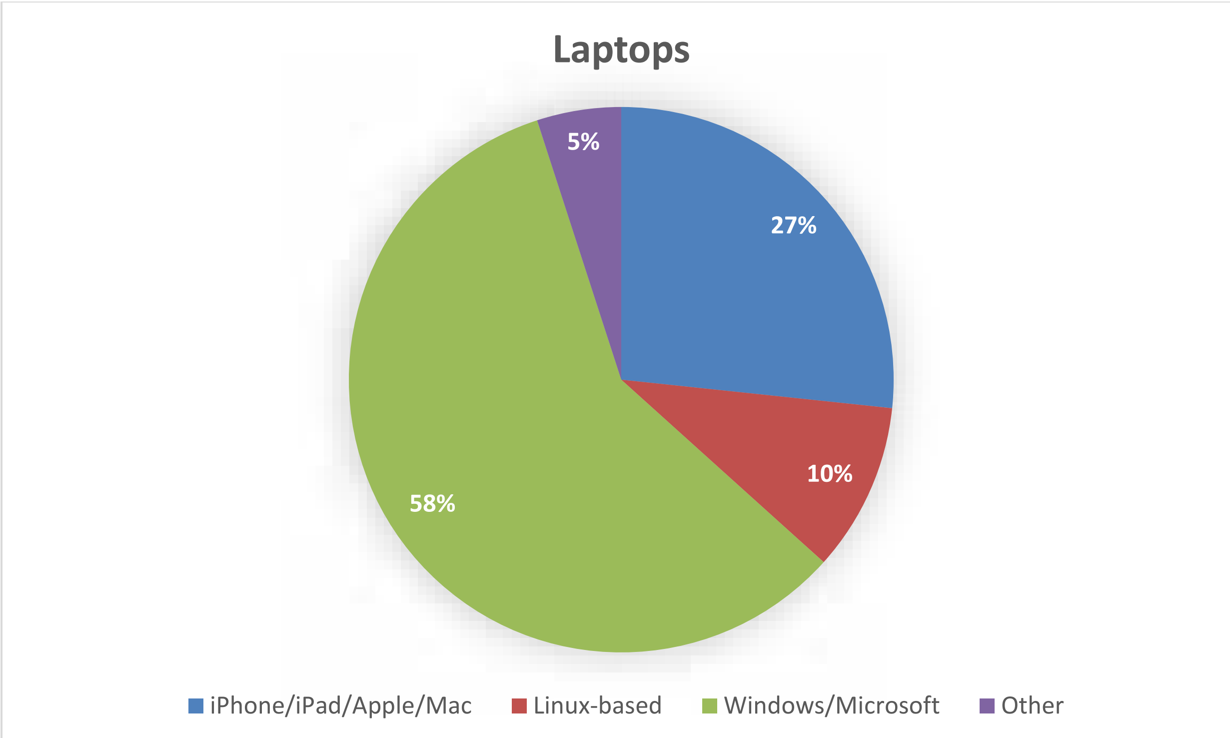

Figure 5-10. Type of laptop users among the sample population

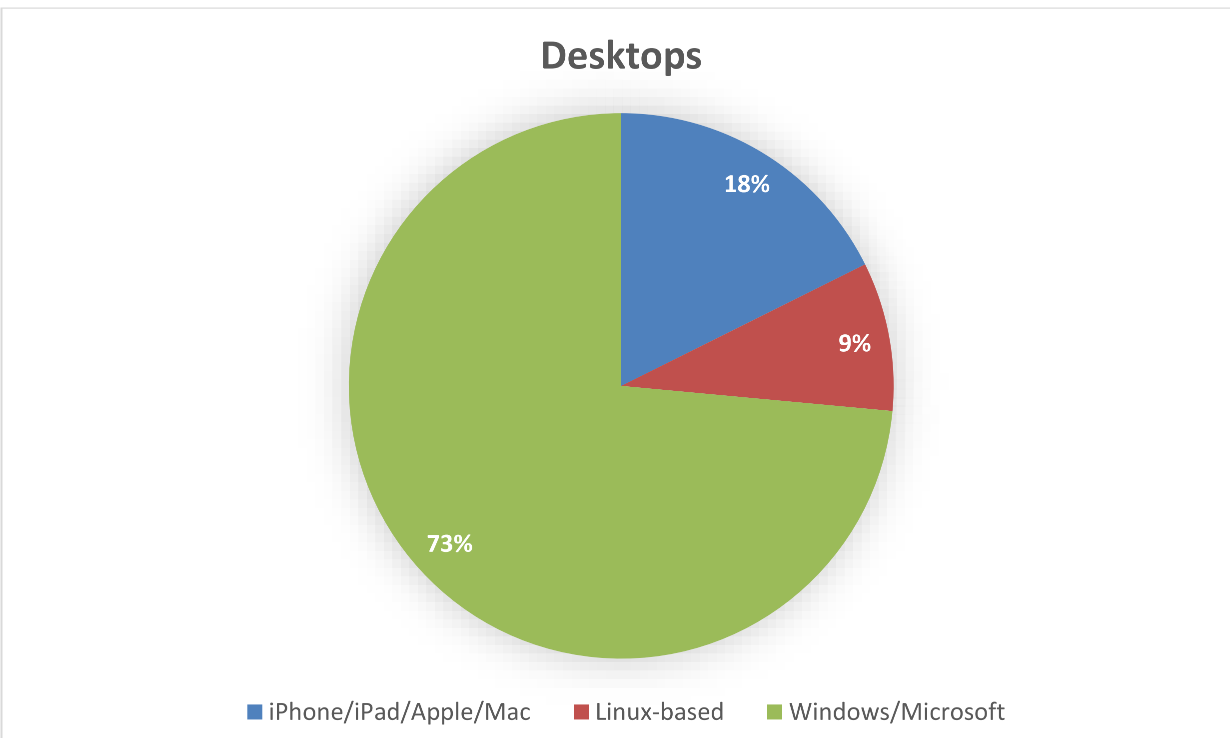

Figure 5-11. Type of desktop users among the sample population

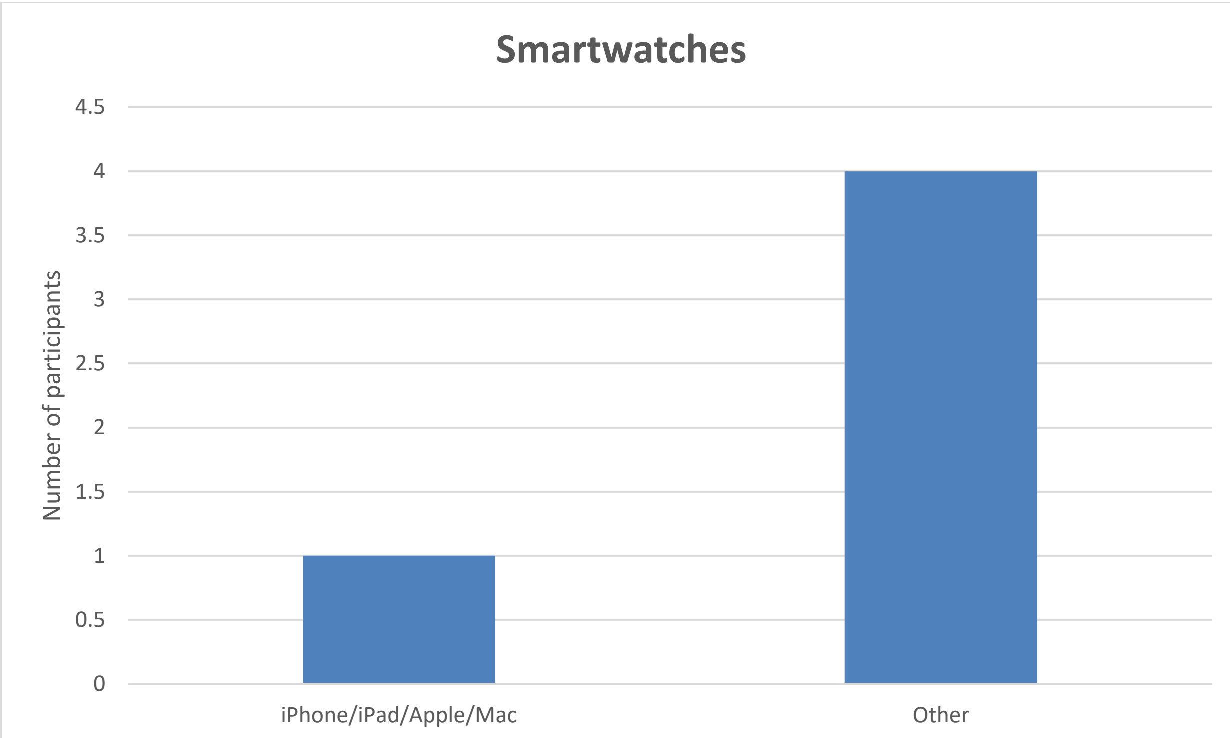

Figure 5-12. Smartwatches owner in sample population

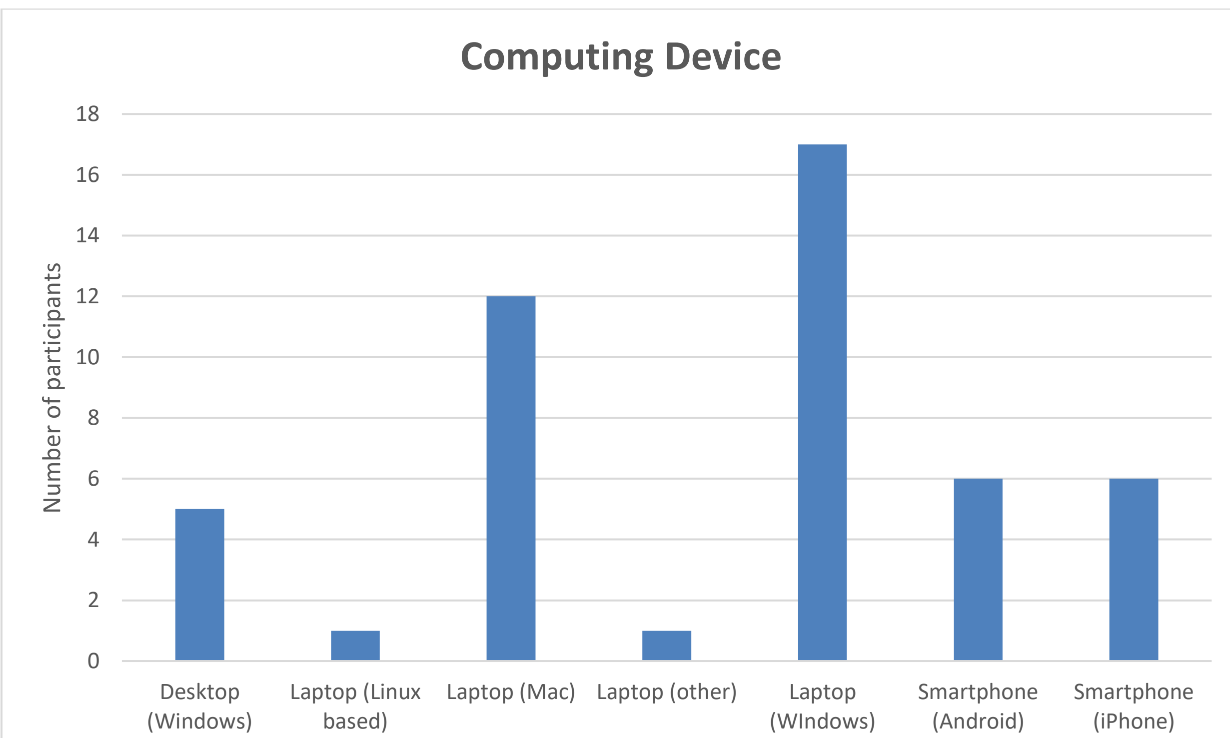

Figure 5-13. Primary computing device by users

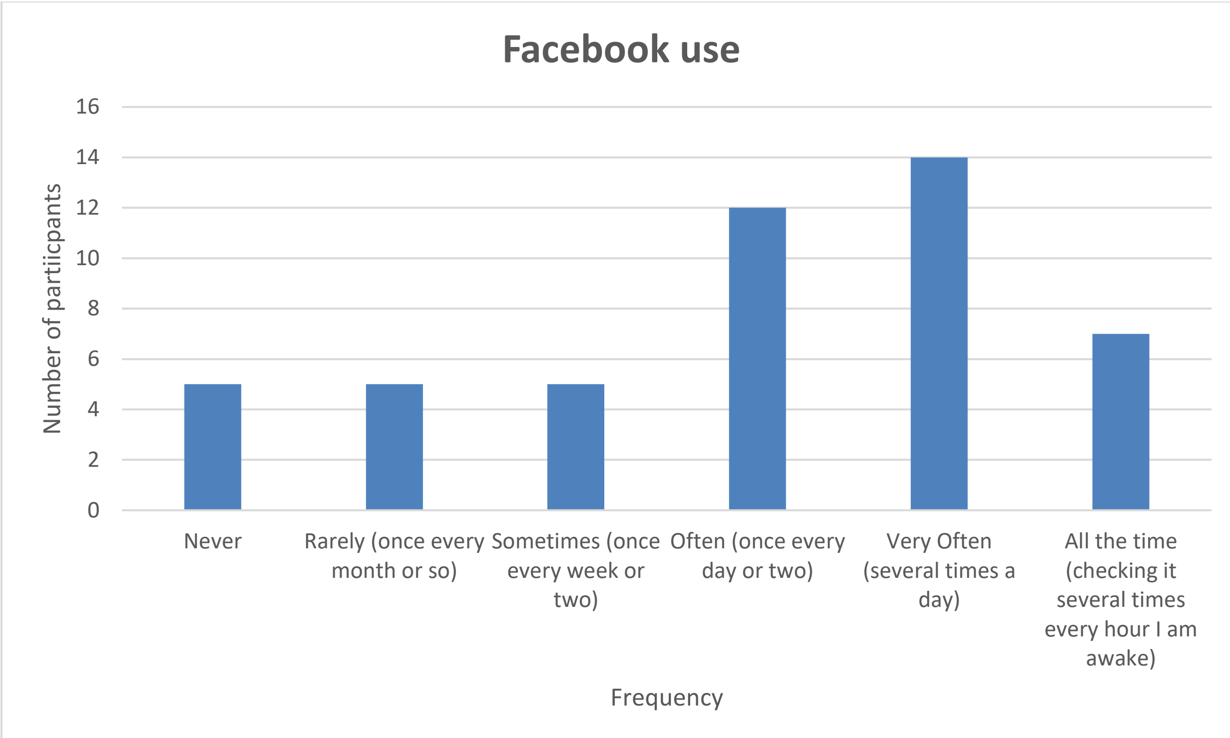

Figure 5-14. Frequency of Facebook use

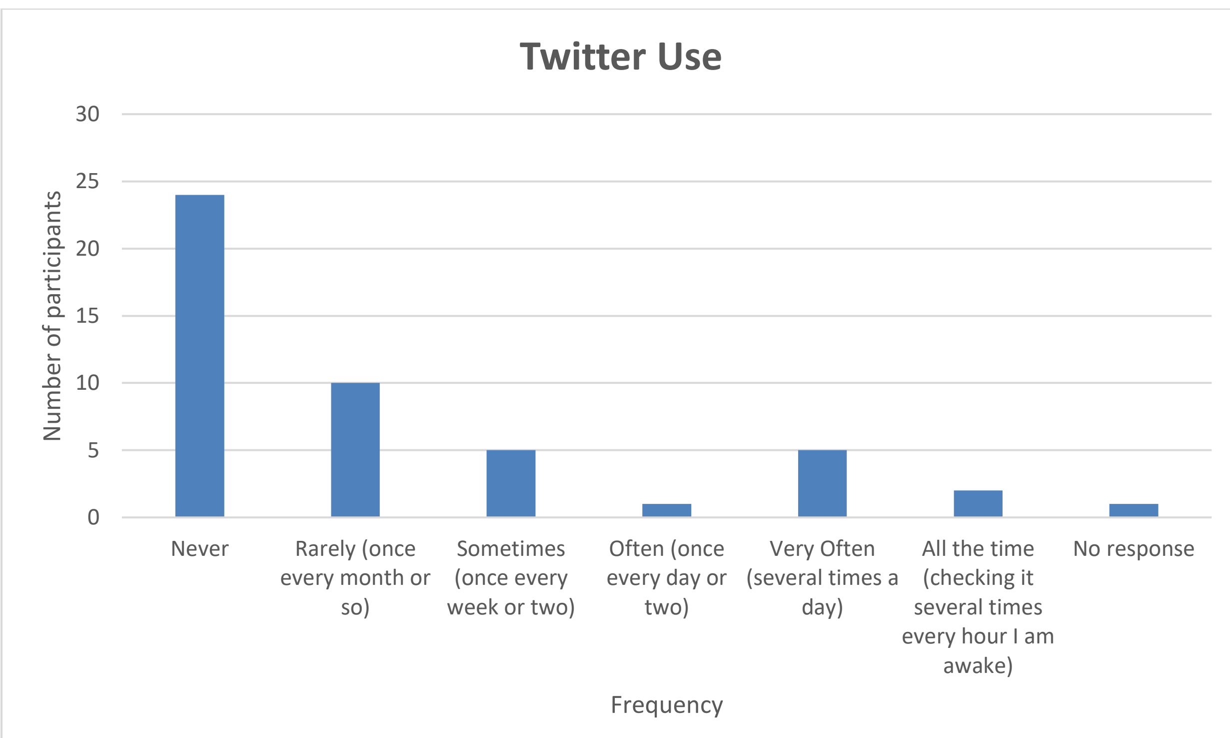

Figure 5-15. Frequency of Twitter use

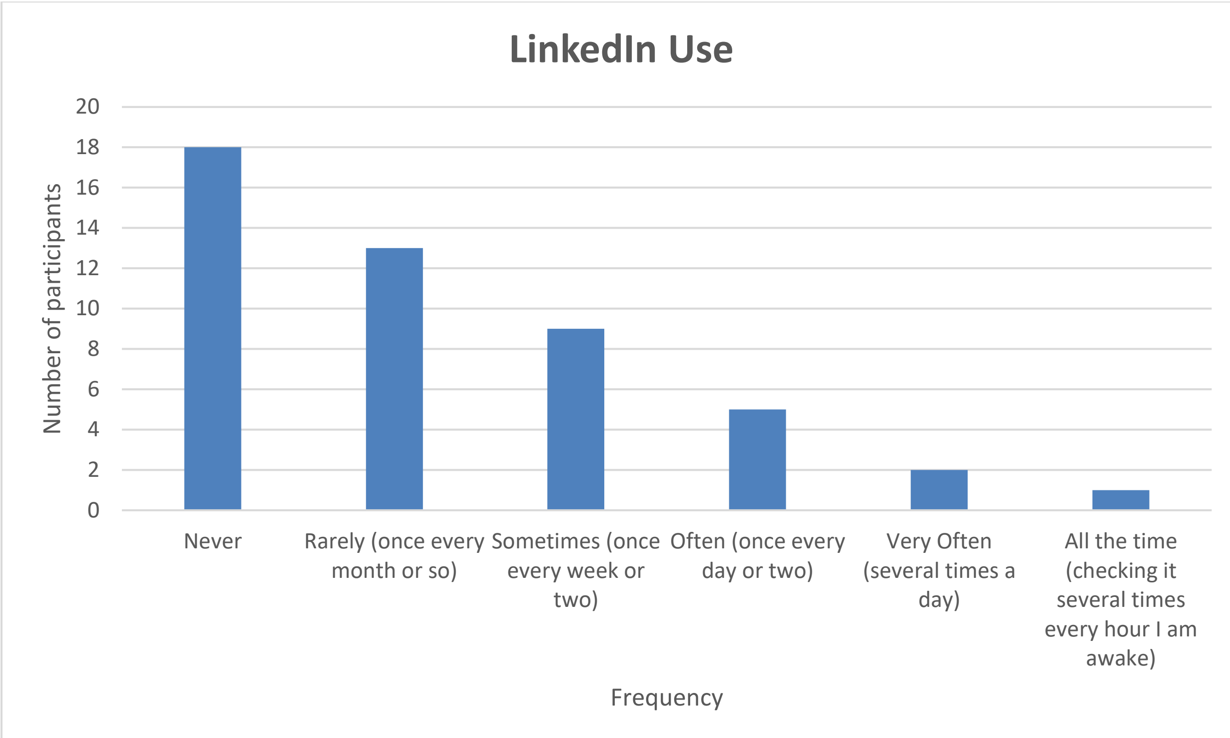

Figure 5-16. Frequency of LinkedIn use

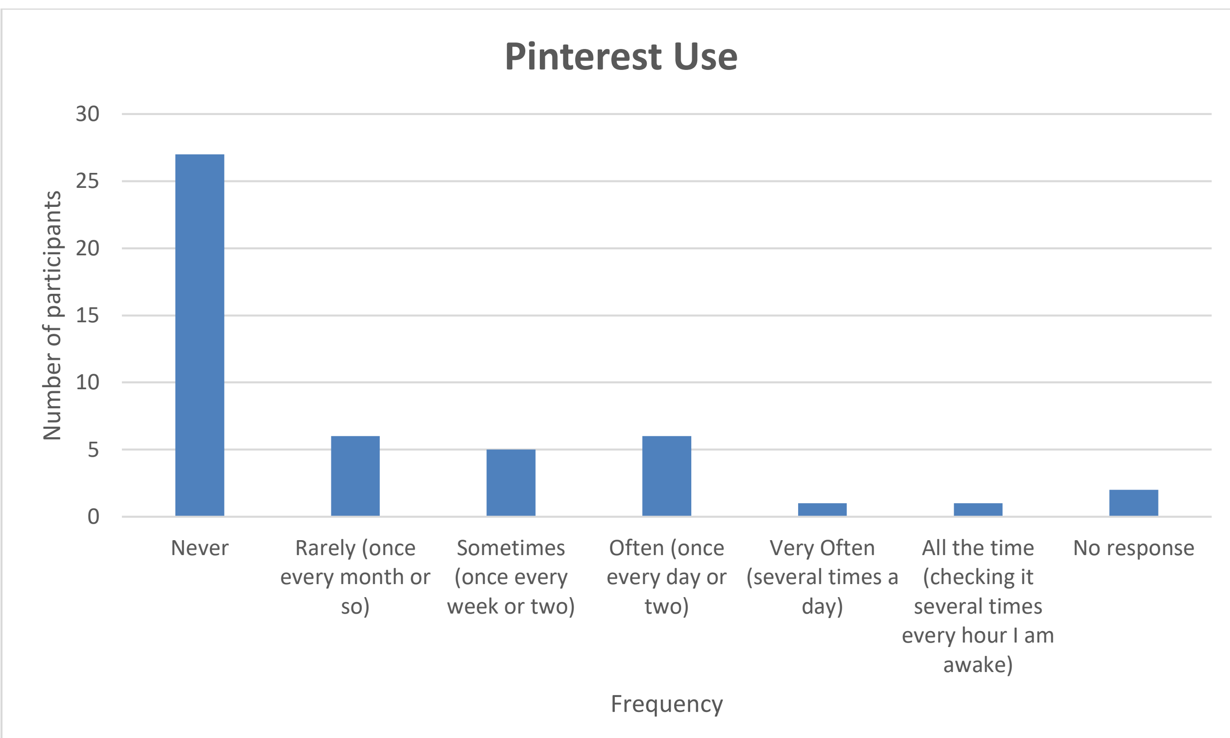

Figure 5-17. Frequency of Pinterest use

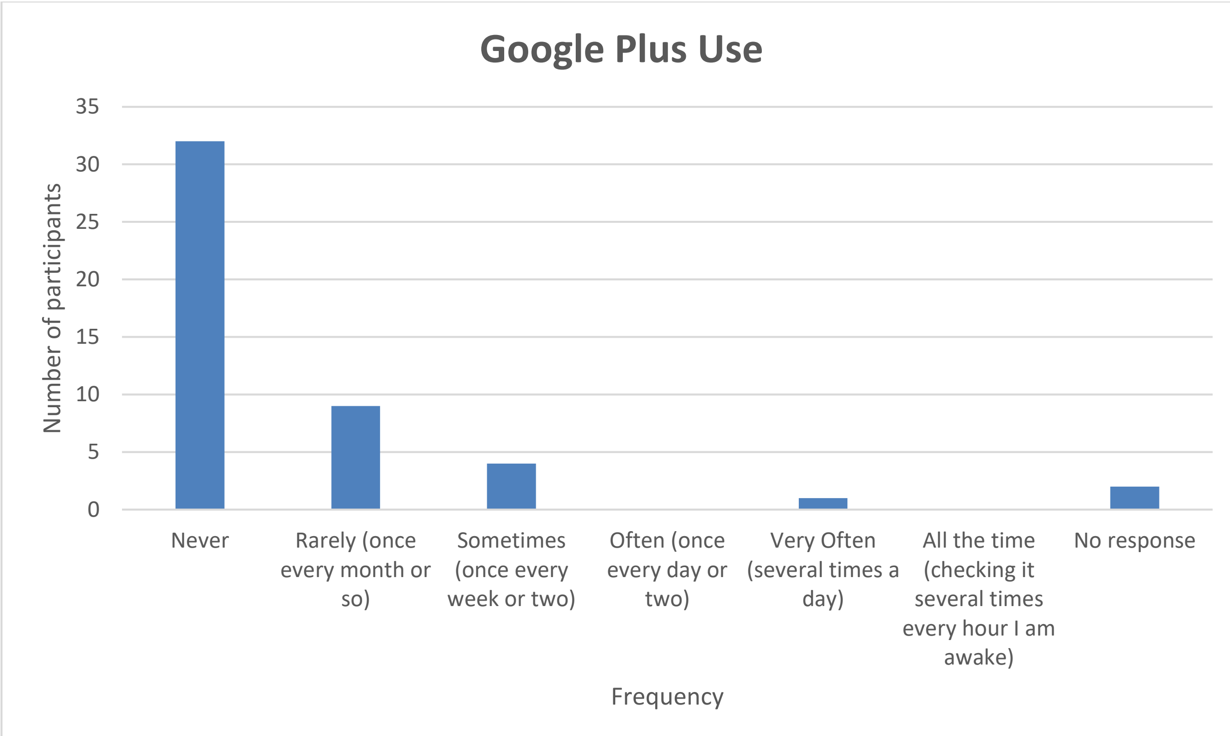

Figure 5-18. Frequency of Google Plus use

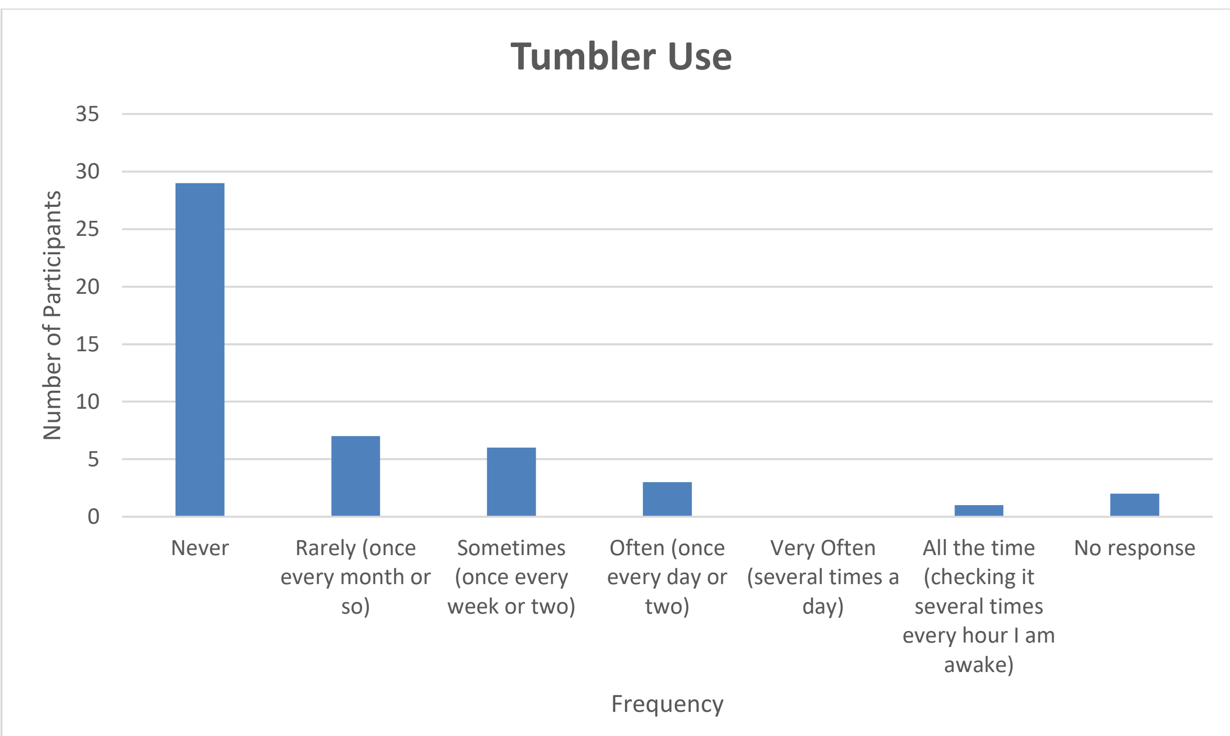

Figure 5-19. Frequency of Tumbler use

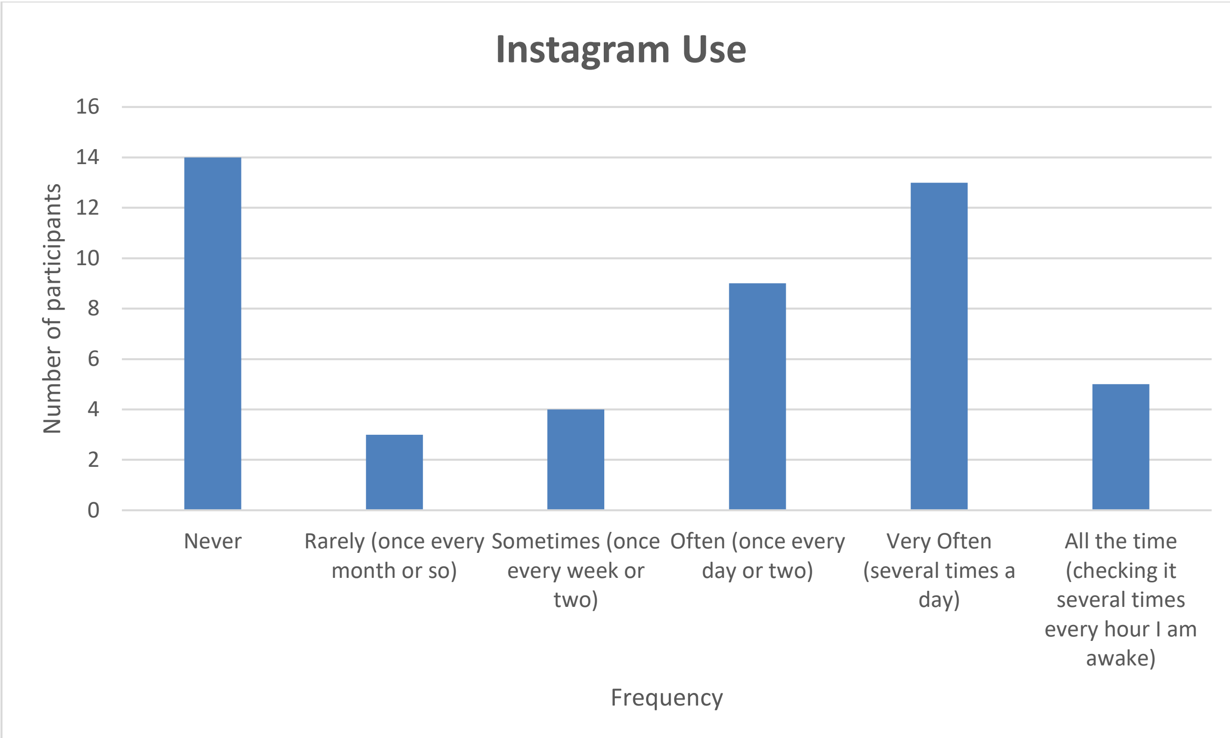

Figure 5-20. Frequency of Instagram use

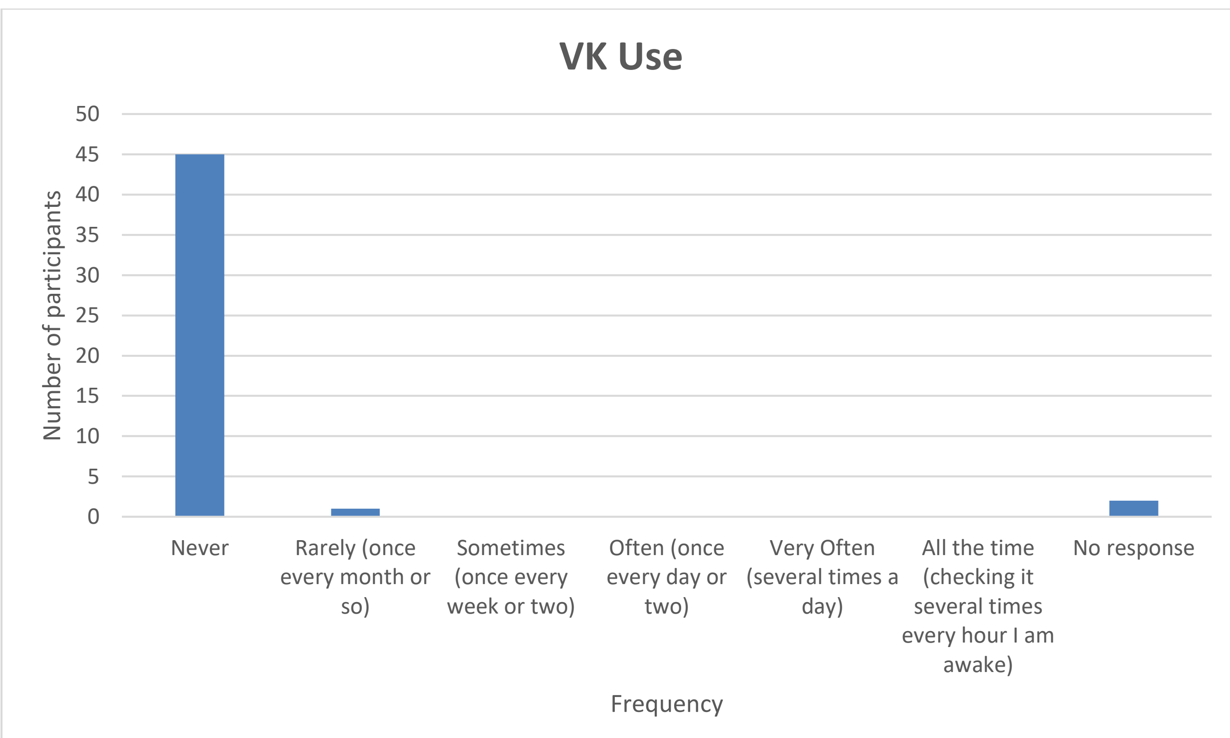

Figure 5-21. Frequency of VK use

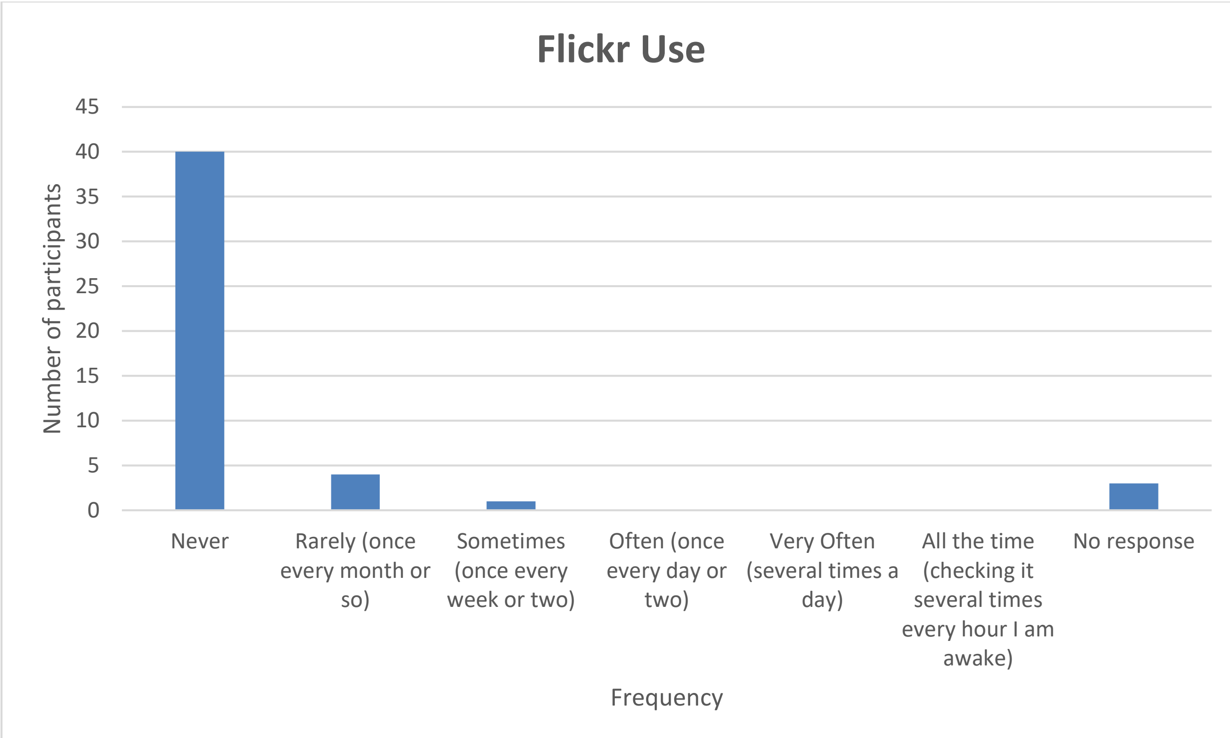

Figure 5-22. Frequency of Flickr use

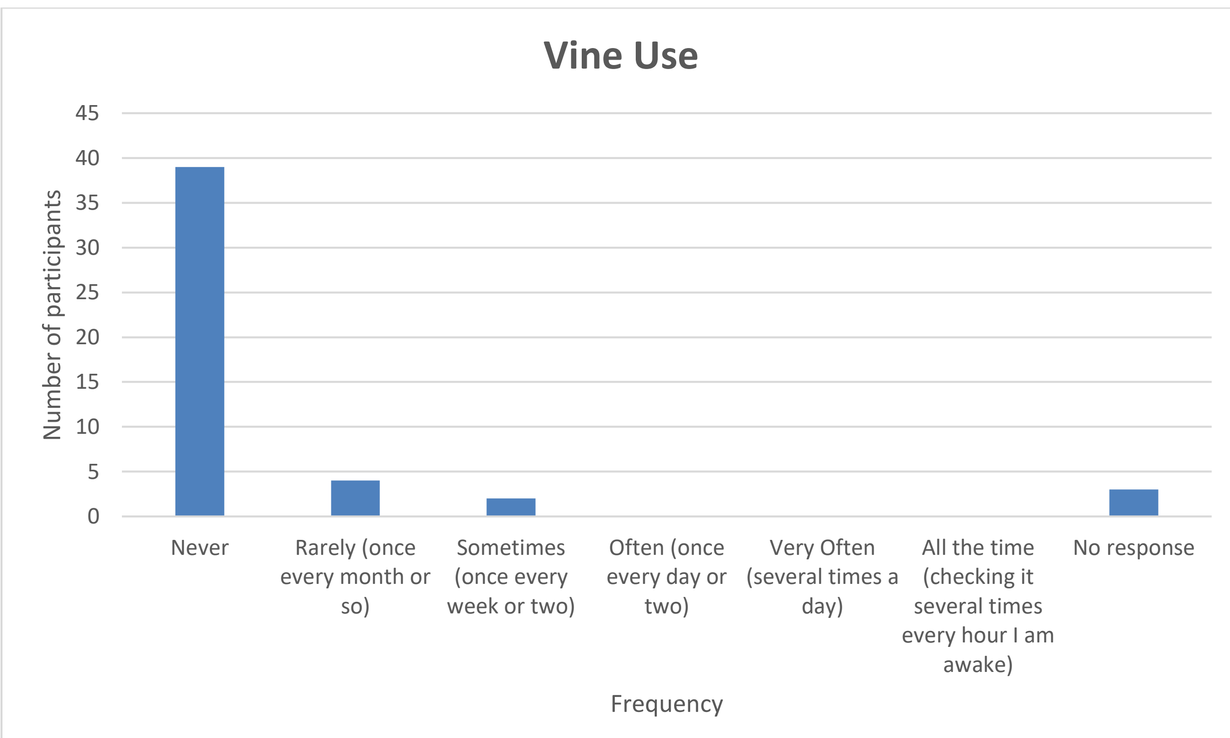

Figure 5-23. Frequency of Vine use

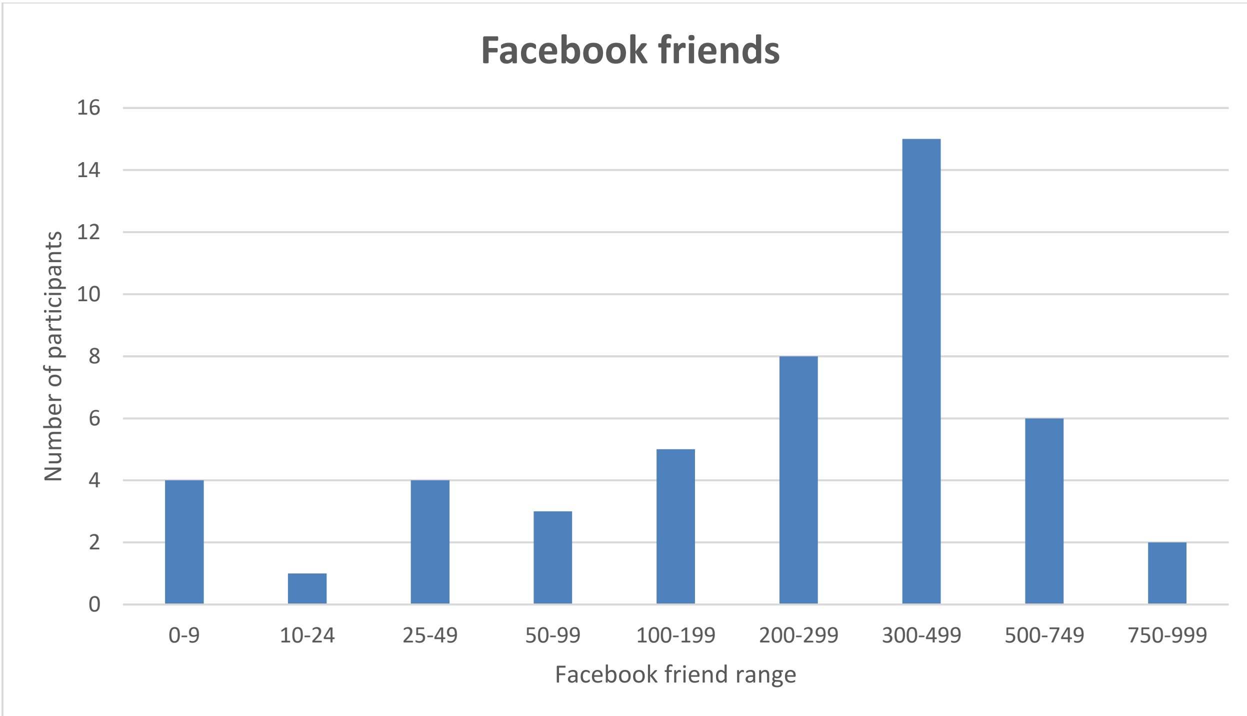

Figure 5-24. Facebook friends of users

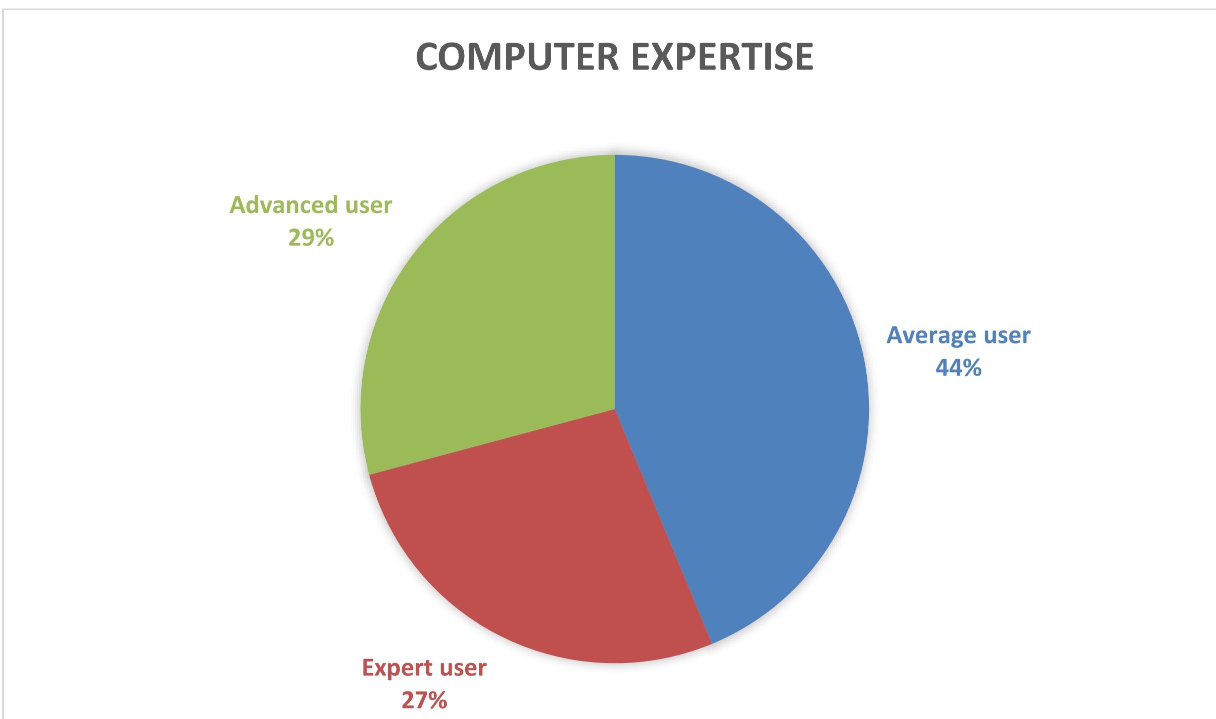

Figure 5-25. Computer expertise

Computer use expertise was defined as:

- Average user (you use word processors, spreadsheets, Email, surf the Web, etc.),
- Advanced user (you can install software, setup configurations, etc.),
- Expert user (you can setup operating systems, know some computer programming languages, etc.)

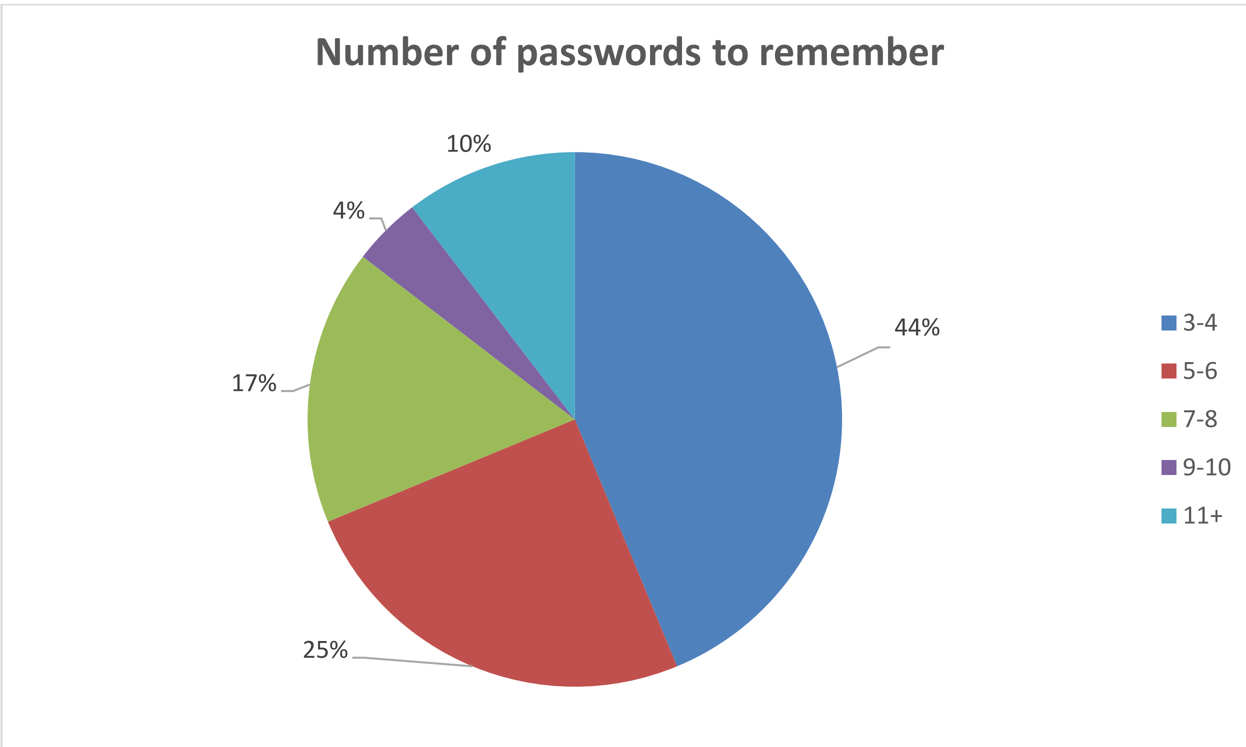

Figure 5-26. Number of passwords remembered by participants

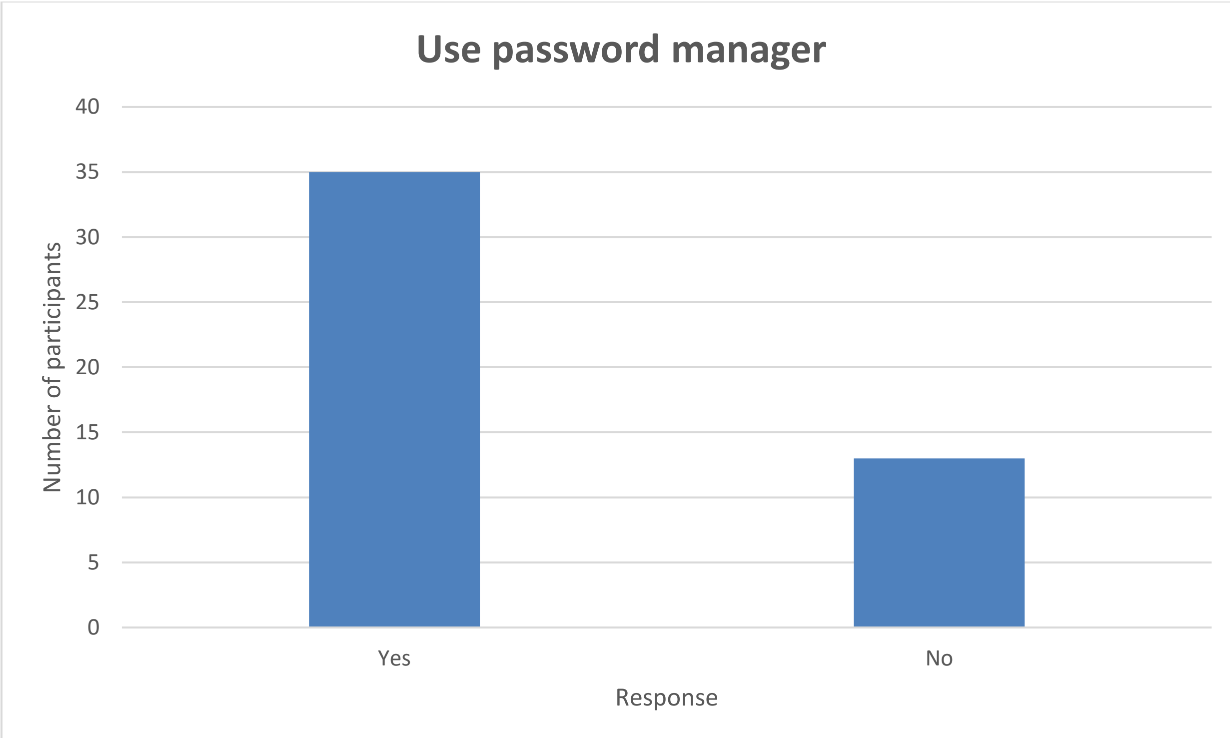

Figure 5-27. Use of a password manager by users

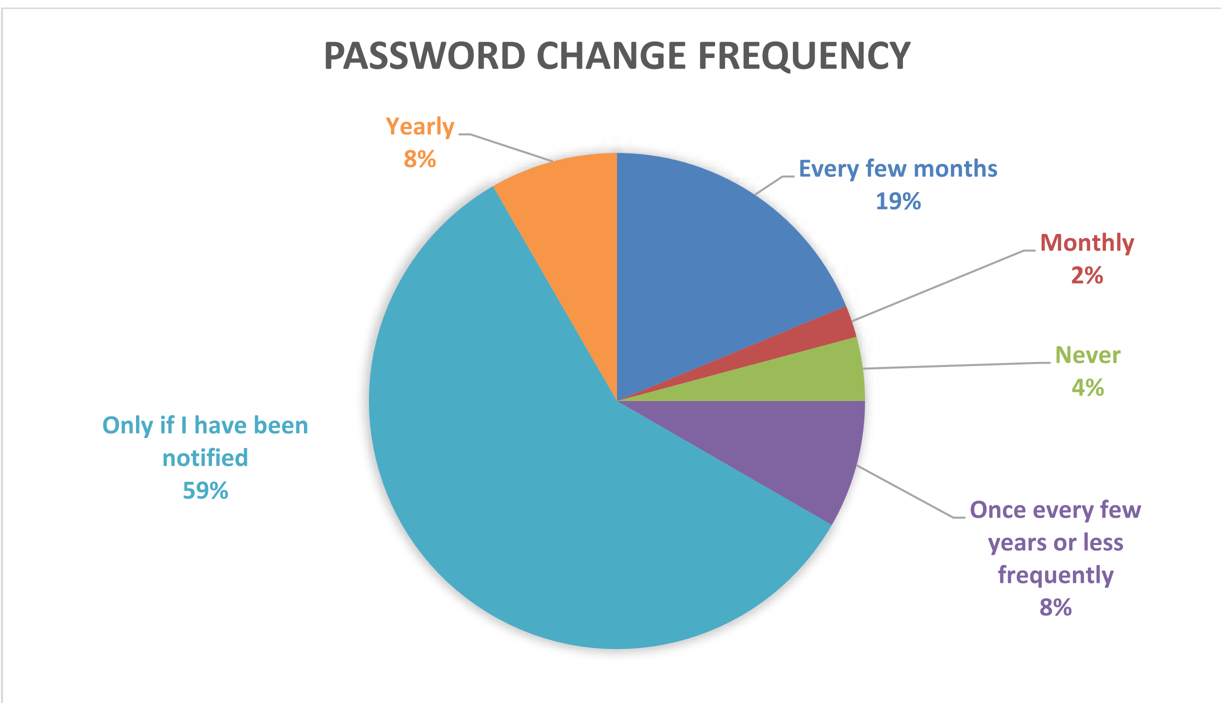

Figure 5-28. Frequency of password change amongst participants

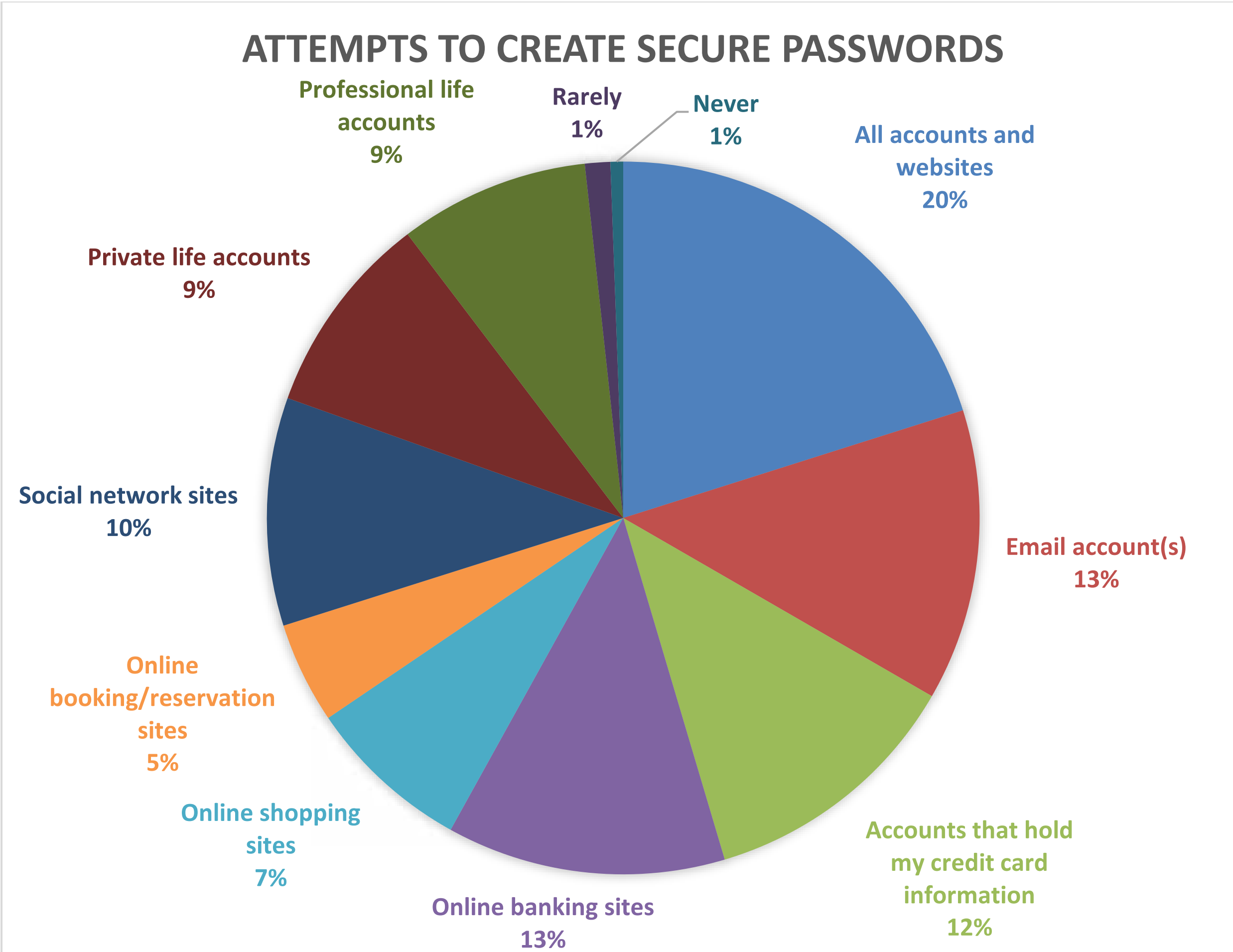

Figure 5-29. Secure password creation by account type

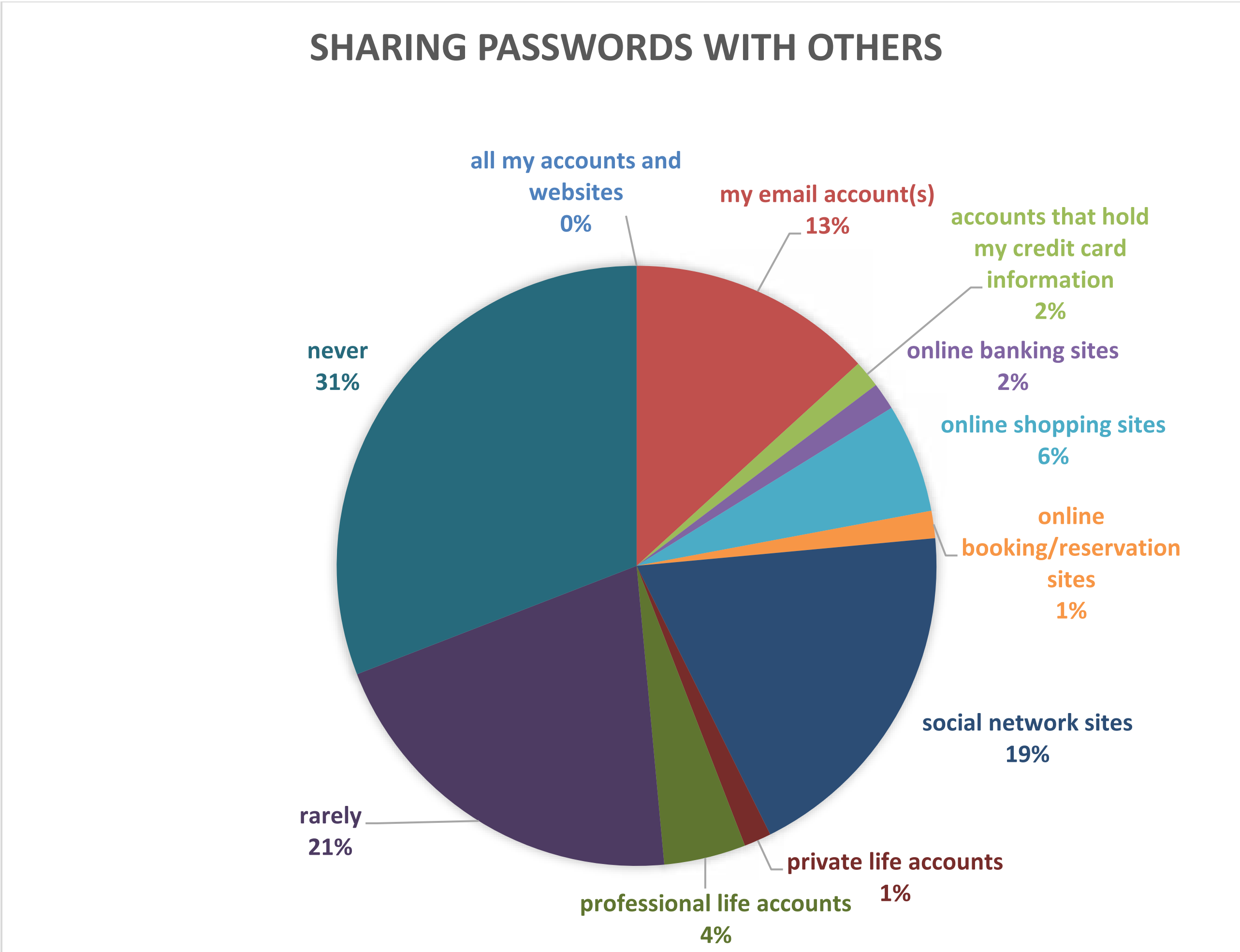

Figure 5-30. Sharing of passwords with others

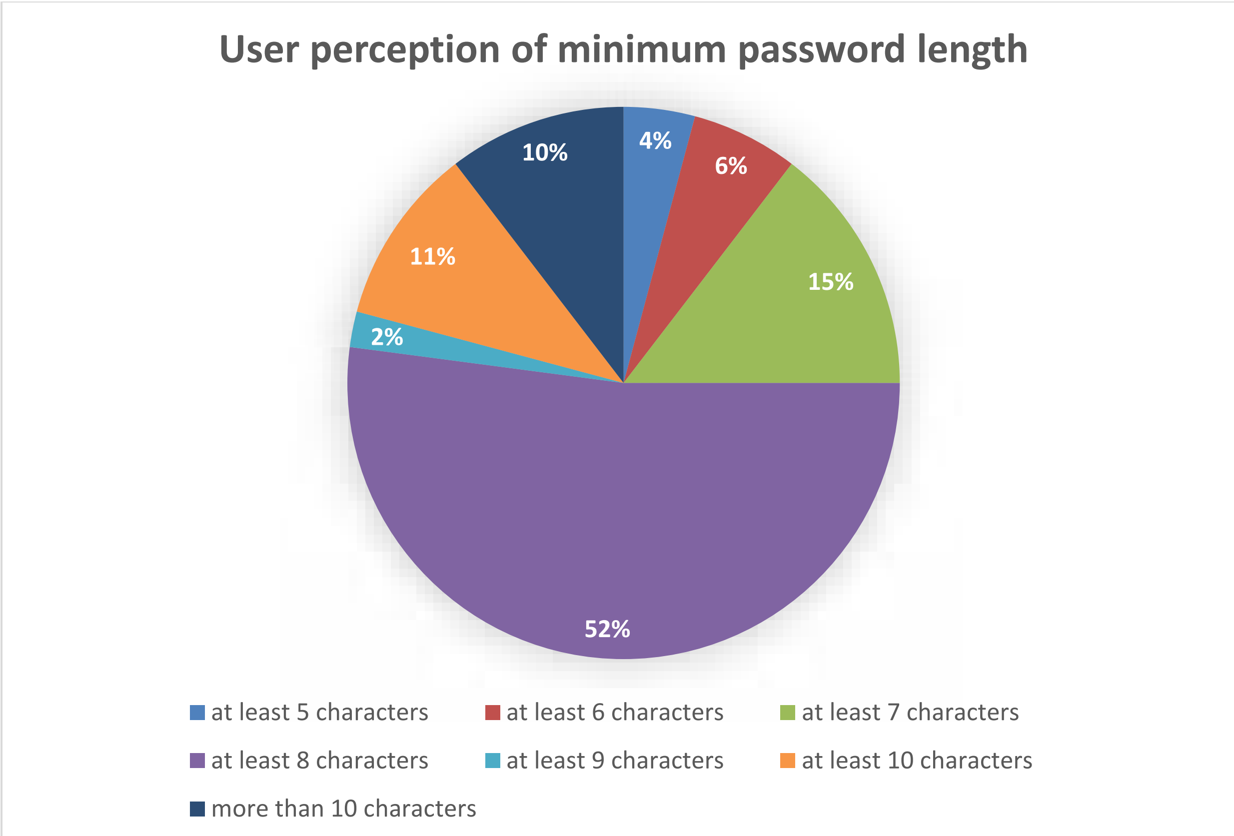

Figure 5-31. Perception of minimum password length by users

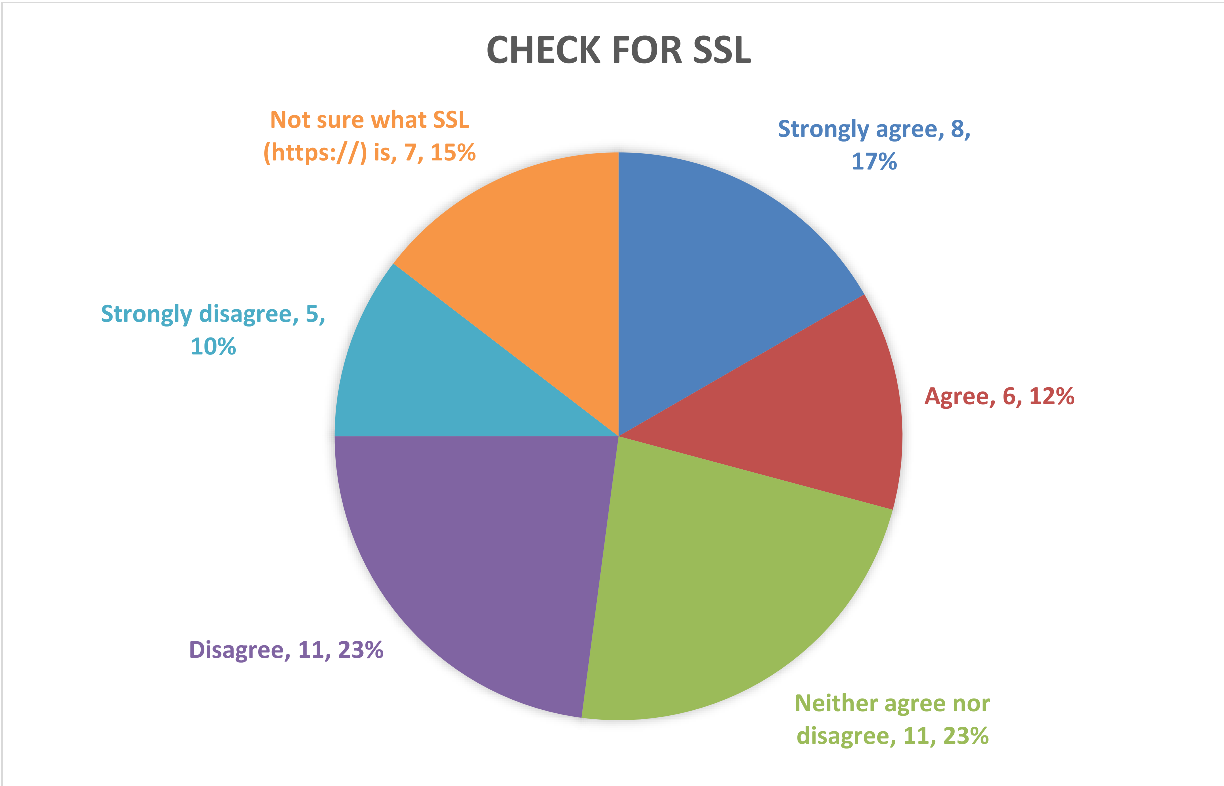

Figure 5-32. Users check for SSL (https://) in browser

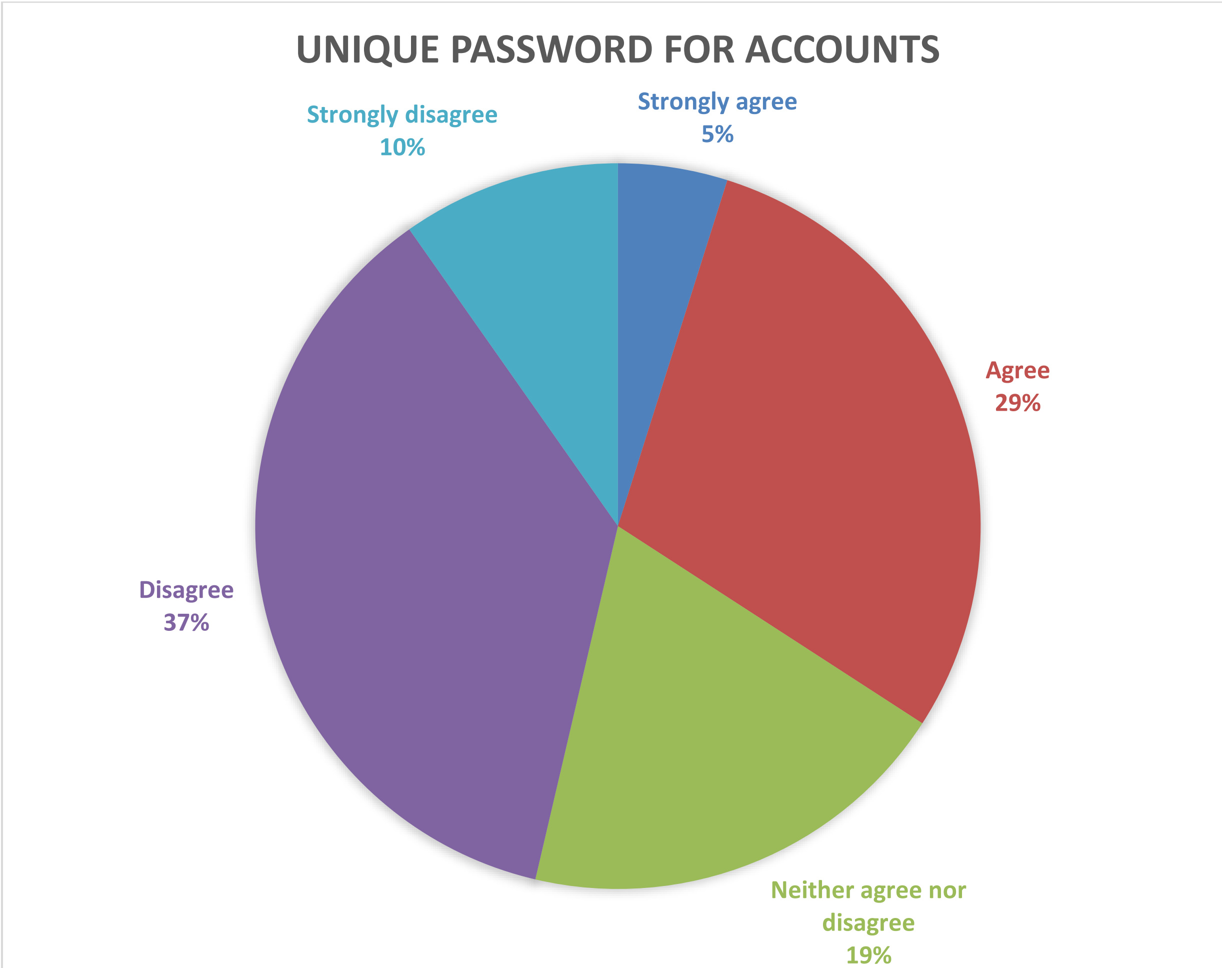

Figure 5-33. Do users create unique passwords