On Password Behaviours and Attitudes in Different Populations

Ruba Alomari, Julie Thorpe

Computer Science Department, University of Ontario Institute of Technology, Oshawa, Canada

Abstract

We explore the differences in password behaviours and attitudes of samples of university students, IT professionals, and the general population (non-student and non-IT professional). Currently, text-based password authentication is the most commonly deployed user authentication mechanism, despite numerous alternatives. Passwords are thus a popular research topic, where many password studies are done in universities where the majority of participants are students. Many studies also do not differentiate between IT professionals and others. Thus, we aim to gain insights about the effect of focusing on samples from university students and/or IT professionals. We conducted a 100-participant online user study involving three sessions over 8-10 days. The study tasks involved password creation and password recall. We also surveyed participant's password-related behaviours and preferences for different authentication methods. Our results provide insights about the effect of using a large proportion of university students and/or IT professionals in password studies. Our findings should be of interest to the research community and also provide useful data to authentication system designers regarding user preferences and attitudes.

Keywords: `Empirical studies, Passwords, Authentication, User Behaviour, User Studies`

Preprint submitted to Journal of Information Security and Applications *February 8, 2020*

1. Introduction

Authentication is the practice of identification in order to gain access to information or resources. It is the first line of defense for every system. Password authentication has been long used, and while other authentication methods such as tokens and biometrics have been developed, password authentication is still the most popular method. Thus, password research remains an important topic; such research often involves user studies, many of which are done with university student participants (e.g., [1, 2, 3, 4, 5]).

Recruiting participants for a research user study is a challenging task, whether the user study is conducted in a lab or online. Academic researchers often opt to conveniently advertise their user studies to the community within their institution, which is composed largely of students. The use of crowdsourcing sites facilitates recruitment for user studies, but this brings other challenges (technical, ethical, methodological, and dropout [6]). Given the prevalence of user studies that recruit from the student population, it is important to understand how this practice might affect user study results.

In the present work, we aim to gain insights into the ways that the target population might affect password behaviours and attitudes. We ran a password study with three groups: the Students group which contains university students as participants, the General group which contains participants from outside the university who are not students nor IT professionals, and the IT group which contains IT professionals who are practicing outside of a university environment. The research questions under investigation are: (1) In what ways might running password studies with students as participants lead to a biased result with regard to password questionnaire results, usability, and security? (2) Are IT professionals more likely to engage in secure password behaviours than non-IT professionals? (3) For the purpose of informing future research on alternatives to passwords, what new authentication scheme properties do these groups find the most desirable?

The study results showed mostly similar responses between the Students

group and the General group; however, we did find at least one notable difference—the Students group had significantly faster login times. This performance improvement is something that the research community should find of importance for the purpose of interpreting related results involving samples of university students.

Regarding the IT group, our results indicate that they have more confidence in their computer security knowledge and are more likely to employ some secure behaviours: they record fewer financial passwords, reuse fewer passwords, and are more likely to use a random password. Also, their passwords were stronger than the other groups against offline password guessing attacks.

We asked all groups about the properties of new authentication methods they would try, positioned in terms of an account type (financial or infrequently used web account) and in terms of security and usability trade-offs. For infrequently used web accounts, all groups appear to be most willing to try a method that is slower to input but easier to remember than passwords. Students also appear to be more open to methods (for infrequently used web accounts) that are easier to remember but less secure. For financial accounts, all groups were most willing to try a method that has slower input speed but stronger security than passwords. Our survey results provide useful data to researchers designing new authentication systems.

This paper contributes insights into some effects of using different groups (including university students) as participants in password user studies. It serves as a starting point in the study of password behaviours in different populations, finding some concrete differences between groups, and also detecting some areas that likely deserve further study with larger populations. It also provides insight regarding the most desirable features for users in new authentication methods. The remainder of the paper is organized as follows. Section 2 summarizes related work done in password studies; it focuses on related work pertaining to the ecological validity of password studies. Section 3 describes the user study we have conducted, and Section 4 presents the study results. Section 5 discusses limitations and ecological validity of our study. Finally, Section 6 provides some

discussion of our results and concluding remarks.

2. Related Work

Password studies can be divided into two main categories: studies on real-world passwords that are based on leaked password sets (such as RockYou and LinkedIn, e.g., [7, 8, 9]) and user studies that collect passwords in a controlled environment (e.g., [10, 11, 12]). The advantage of the real-world studies is that it is based on real passwords that people created for real systems. These studies provide accurate information about the passwords; however, it is limited by the system they were created for, and do not allow researchers to experiment with different settings. Controlled user studies give researchers the ability to study different security or usability aspects [1]. One major concern about user studies is the ecological validity of the study. Researchers often try to address this concern individually in their studies by trying to put controls in place to improve ecological validity such as: assigning passwords to participants when password usability is in question and not password creation [5], opting for online studies to increase sample size and diversity (MTurk is widely used for this purpose) [13], or hiding the study's interest in passwords in order not to influence the study participants [14].

We focus our discussion on related work in the ecological validity of password studies and those that focus on different populations. For surveys of the vast field of password systems research, please see [15, 16].

The Amazon Mechanical Turk (MTurk) is a crowdsourcing Internet marketplace that gives individuals or businesses (known as Requesters) access to a diverse, on-demand, scalable workforce. Amazon MTurk is becoming popular in user studies. Buhrmester et al. [17] stated that the MTurk population is significantly more diverse than samples used in typical lab-based studies that heavily favor college-student participants, and concluded that overall, MTurk can be used to obtain high-quality data inexpensively and rapidly. However, his work was based on psychology and social sciences which doesn't necessar-

ily apply to password studies, and it compared MTurk participants to college students and not to the general population. Thus, this study does not address whether either group (Mturk or students) provide a representative sample for study participants with regard to password studies. Other researchers have recognized the challenges with Mturk participants in password studies. Biddle et al. [18] found that when conducting studies using MTurk, the same challenges that Web-based studies face remain and may be magnified, such as dealing with skewed demographics (compared to actual target users) and ecological validity issues (i.e., is rapid task completion a primary motivation for MTurk workers?). Additional challenges include designing appropriate short, specific micro-tasks that are likely to be completed. For these reasons, the overall suitability of MTurk for authentication studies remains unknown. Komanduri et al. [13] described the ecological validity of studies as being difficult to demonstrate in any password study where participants know they are creating a password for a study, instead of creating a password for an account they value and expect to access repeatedly over time.

Fahl et al. [1] performed the first work that addressed the issue of ecological validity of password studies, with the type of study (real-world vs. user study) as the independent variable. Fahl et al. [1] compared user study passwords of 645 students to their real passwords created on the university's systems. They compared the passwords from an online user study setup and a laboratory study setup under priming and no-priming conditions. This study provided valuable information regarding password studies, finding that 29.9% of participants did not behave as they normally do, while 46.1% percent offered comparable data and 24.0% offered somewhat comparable data, concluding that password studies create useful data. Although there are some participants who do not behave realistically during password studies, on the whole, it recommends that more research is needed to be done to find out how to best interpret the results. Mazurek et al. [19] studied 25,000 real passwords of high-value university accounts for university faculty, staff, and students, and compared them to password sets previously collected in controlled experiments or leaked

from low-value accounts as well as passwords collected from MTurk. The study found more consistent similarities between the university passwords and passwords collected for research studies under similar composition policies than it did between the university passwords and subsets of passwords leaked from low-value accounts that happen to comply with the same policies [19]. Mazurek et al. also found that passwords created on MTurk are not a perfect substitute for high-value passwords either, as they were slightly weaker than the genuine ones. However, when used as training data for guessing real passwords, passwords from MTurk were just as effective as real passwords. The study concluded that passwords gathered from carefully controlled experimental studies may be an acceptable approximation of real world, high-value passwords, while being much easier to collect. While the study concluded that university students and faculty passwords are very comparable to passwords collected under carefully controlled studies with MTurk participants (which may or may not be representative of a student or general population), the study did not address the ecological validity of using students as study participants in comparison to the general population of non-MTurk users. To the best of our knowledge, there has been no study to explore the effect of using students as participants on the ecological validity of password user studies. As such, this is the first study concerning the issues of using university students as participants in password creation user studies.

In recent years, there has been increased interest in the password behaviours of IT and security experts. Stobert and Biddle [20] interviewed a small set of security experts to understand their password management strategies, finding that they tend to combine the use of password managers for important accounts with other practices (e.g., password reuse, writing down passwords, and infrequent password changes) for unimportant accounts. Loutfi and Jøsang [21] report on a survey of IT professionals in terms of their general password metrics, confidence, storage, uniqueness, and perception of account sensitivity. Their results suggest that the passwords of IT professionals may not be as secure as they perceive them to be, but do not compare these results to those of any other groups. Ion et al. [22] surveyed security experts and compared them to a

group of non-expert Mturk users. Their results relating to password behaviours indicate that security experts less frequently write down passwords, reuse passwords, and change their passwords than non-experts. Our work contributes further to our understanding of the password behaviours and attitudes of IT professionals.

3. User Study

We conducted an online user study to measure differences in password behaviours and attitudes between student participants, participants who are in the general (non-student) population, and participants who are IT professionals. We recruited three groups of participants: Students, General and IT. Members of the Students group were recruited through university-wide email advertisements. In our advertisement, we explicitly asked that volunteer students not be enrolled in an Information Technology Security program. IT was mostly recruited by posting advertisements in IT-related LinkedIn groups, but approx. 10% was sampled in the same way as General and moved to IT based on their self-reported profession. General was recruited by posting to a variety of Facebook groups. All interested participants then contacted the researchers. Students were compensated $10 each. Members of the IT and General groups were entered into a draw for a $50 pre-loaded credit card, where 4 winners were drawn per group. Students' compensation model was chosen as it is known to attract students in our university; other groups were compensated through more draws (with larger value) as this was considered more attractive to off-campus participants. The user study was approved by our university's Research Ethics Board. We ensured passwords were transmitted using SSL and to allow analysis of the unhashed passwords, they were stored with public key encryption, where the private key was stored separately and used only on an offline system for analysis. The user study was split into three sessions over approximately 8 days, with a time frame of up to 48 hours to complete Sessions 2 and Session 3. All sessions were held online so participants were able to participate from a place of their

choice and convenience. The three sessions were set up as follows:

- Session 1 (day one). Participants were directed to the online password creation system, where the text password creation policy was displayed (i.e., password must be ≥ 8 characters in length, with at least one special character, one number, and one uppercase character). This policy was used as it is a commonly recommended strong password policy. Participants were told "You will be asked to re-enter this password in 3 different sessions over the course of 8-9 days". Each participant was advised to create a text password that they can remember, but would be difficult for others to guess. They were also asked to not use a password they have used on any other system. After successfully creating and confirming their password they were redirected to a page to answer a questionnaire (approximately 10–15 minutes in length) to distract the user from their password and gather information on demographics and password behaviours. Finally, the user was asked to log in to the system using the password they created and confirmed earlier.

- Session 2 (day two). The user was sent an email containing a link to the online system, reminding them to log in for Session 2. The participant was asked to reproduce the text password they created in Session 1 and had the option to reset the password if forgotten.

- Session 3 (day eight). The user was sent an email containing a link to the online system, reminding them to log in for Session 3. The participant was asked to reproduce the password they last created on our system and had the option to reset the password if forgotten. Then they were redirected to a page to answer a questionnaire to gather information on their opinions on their study password and desirable authentication system properties (approximately 5 minutes in length).

4. Study Results

We recruited 100 participants for our study, in three different categories: 36 participants in the Students group, 31 participants in the IT group and 33 participants in the General group. We aimed for approximately 30 participants in each group, but accepted additional applicants in case of drop-outs. We describe the demographics of our groups in Section 4.1, the questionnaire data related to password behaviours in Section 4.2, the password data collected in Section 4.3, and the questionnaire data related to desirable properties of authentication methods in Section 4.4.

To gain insight regarding differences between our target populations, throughout the results we report p-values for each of the password behaviour questions, usability metrics, and security metrics collected. These p-values are computed through two-tailed tests to compare both Students vs. General groups, and Students vs. General vs. IT groups. We used different statistical tests based on the nature of the data, which are mentioned next to each question using the following acronyms: Mann-Whitney U test (MWU), Kruskal-Wallis test (K-W), Chi-squared test ($\mathcal{X}^2$), and Fisher's exact test (Fisher). We use Bonferroni correction to account for multiple tests, and report on both corrected and uncorrected results where applicable.

4.1. Demographics

The Students group contained 50 % females 50% males, the IT group contained 32% females and 68% males and the General group contained 73% females and 27% males. The majority of the students (89%) were aged less than 25 years old, the IT group majority (84%) were aged 30 years and older, the General population group majority (75%) 30 years and older as well. Our Students group had a variety of majors: Engineering (29%), Health Sciences and Nursing (25%), Criminology, Legal Studies, and Forensics (21%), Commerce (8%), Education (6%), Life Sciences (9%), CS (3%), and Communication (3%). Our General group had a variety of professional fields: Education (39%), Health (15%), Unemployed/NA (15%), Management (6%), Sales (6%), Business (3%), HR (3%),

Table 1: How often are you connected to the Internet?

Response	**Students**	**General**	**IT**
Connected online 24/7.	58%	55%	77%
Not always, but more than 5 times per day.	33%	24%	13%
2 to 5 times per day.	8%	15%	6%
Once a day.	0%	3%	3%
N/A	0%	3%	0%

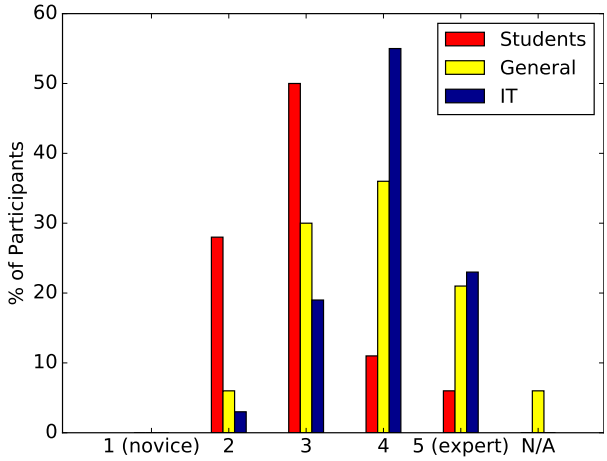

Figure 1: Computer skill level as reported by participants (note that values have been rounded, so bars may slightly vary in size).

Self-employed (3%), Telecom (3%), Banking (3%), and Sociology (3%). The users reported frequently being connected to the Internet, as described in Table 1. They were also asked to self-report their computer skill level, the results of which are given in Figure 1. The graph shows the majority of the IT group reported their computer skills as above average at 81%, versus approximately half of the General group, and 14% of the Students group. Their self-reported computer security knowledge is shown in Figure 2; the Students group apparently felt the least confident in their computer security knowledge, and the IT group the most confident.

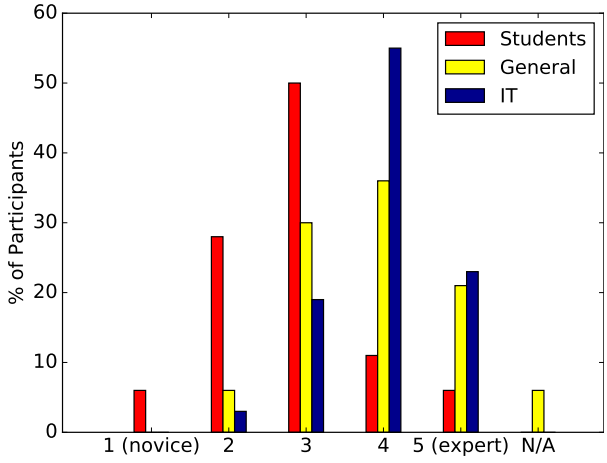

Figure 2: Self-reported Computer Security knowledge of the study participants.

4.2. Password Behaviour Questionnaire Results

We asked participants a number of questions to describe their password behaviours. In this section, we provide graphs and describe the data.

Password Recording. In response to the question: "Which of the following passwords do you record on paper or on a mobile, PC, laptop, or other devices?" Our results (shown in Fig. 3) indicate that a high percent of participants do record their passwords somewhere, especially in the Students group. Note that 'N/A' here means they do not save their passwords on any location. A $\mathcal{X}^2$ test showed no significant difference between any of the groups regarding not recording passwords for Students vs. General ($p = .422$) and Students vs. General vs. IT ($p = .110$). Our finding that the IT group was least likely to record passwords is consistent with findings in [22], which found that non-experts were more likely to report writing down passwords for some of their accounts. Also note that this is a multiple answer question, meaning that a participant can choose more than one password category (financial, email, social network, other web, and 'other' accounts) that they record. We found that financial account passwords are recorded by 44% of Students, 27% of the General group, and 13% of the IT group. A $\mathcal{X}^2$ test between the three groups showed $p = .017$, which is only significant prior to Bonferroni correction. No significant difference was

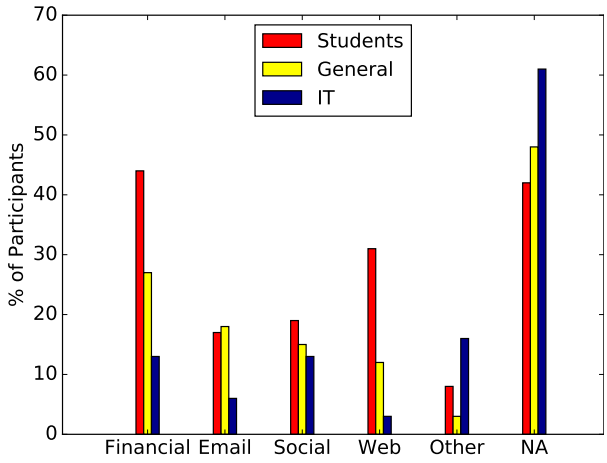

Figure 3: Which of the following passwords do you record on paper or on a mobile, PC, laptop, or other devices?

found between Students and General groups ($p = .138$). Email account passwords are recorded by 17% of the Students group, 18% of the General group, and 6% of the IT group. This might be attributed to a more frequent use of the email account in comparison to the financial account, or the popular practice of saving the email account password in the email client (e.g., on mobile devices), whereas no financial client will allow saving passwords in the client or the browser. A $\mathcal{X}^2$ test indicated no significant differences between any of the groups. The same conclusion might be applied to social network accounts; we ran a Fisher test which found no significant difference between any of the groups. A $\mathcal{X}^2$ test showed significant differences prior to Bonferroni correction in the recording of 'other web' passwords in Students vs. General groups ($p = .027$) and also between the three groups ($p = .003$); however, these results are not significant after correction. We note that it is possible that students simply use a wider variety of accounts. A higher percentage (16%) of the IT group responded that they record 'other' passwords; this might be due to the systems they need to administer as part of their job functions. However, a Fisher test indicates the differences between the three groups for 'other' accounts is not significant ($p = .203$).

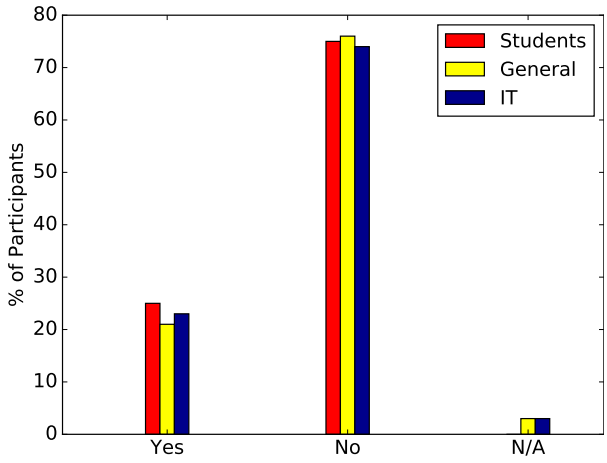

Figure 4: Did you, at any time during the study, write down or record your password in any way?

We asked participants if they recorded their study password and encouraged an honest answer. If they answered yes, we asked what they wrote down and verified it matched their study password. Figure 4 indicates that approx. 23% of the participants recorded the study password somewhere, and an average of 75% across the groups reported that they did not save the password anywhere. A $\mathcal{X}^2$ test showed no significant difference between the three groups ($p = .990$).

Password Reuse. We asked participants whether they sometimes reuse the same password in different applications, 100% of the Students group, 82% of the General group, and 77% of the IT group responded 'yes.' A Fisher test showed no significant difference between the Students and General groups ($p = .089$). When IT experts are included in the comparison, a Fisher test shows a significant difference ($p = .020$) prior to Bonferroni correction; however, this result is not significant after correction.

Figure 5 shows a graph of the number of passwords for each group. The rounded average number of distinct passwords for the Student group was 4, for the General group the average was 4, and for the IT group it was 7 (which, for this IT group only, is consistent with the results of Florêncio and Herley [10]). Figure 5 shows a peak of participants with 3 distinct passwords among all the

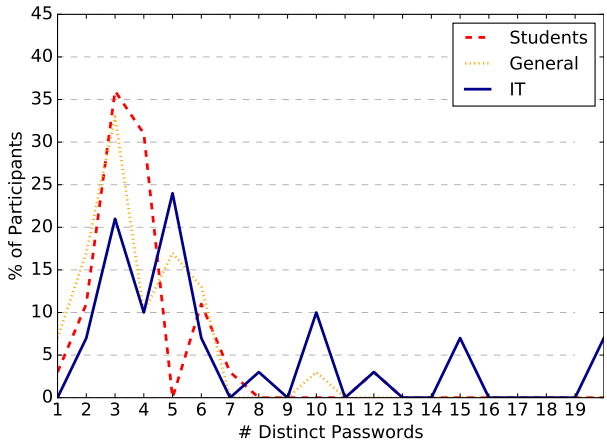

Figure 5: How many distinct passwords do you use?

groups, and another peak at 5 distinct passwords for the General and the IT groups. A MWU test indicates no significant difference between Students and General groups ($p = .619$); however, a K-W test between all three groups shows $p = .004$ (which is only significant before Bonferroni correction).

Password Selection Strategies. Figure 6 shows the strategies used by participants to select a password in the study. It shows that the most popular strategies among the groups were meaningful numbers and names. Random passwords are least popular with both the Students and General groups, but 26% of the IT group uses them. The Students group uses the most names at 67% (vs. 42% of the General and IT groups). We found no significant differences between the groups in the password creation strategies used after Bonferroni correction, but note that one was significant prior to correction: the use of names in passwords for Students versus General groups (MWU, $p = .044$) and all three groups (K-W, $p = .028$).

Password Sharing. We asked students if they share their e-learning password and asked the General and IT groups if they share their work password with colleagues to do a task when they are not available. Figure 7 shows the responses for this question. The groups had similar responses, although the question was not applicable to 15% of the General group (mostly because 9%

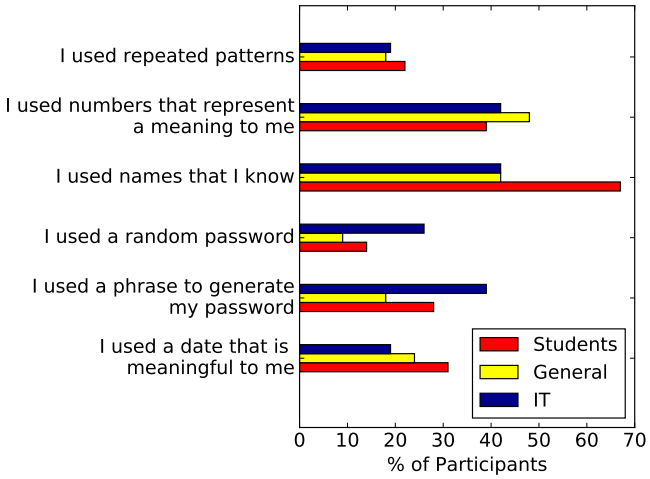

Figure 6: What was your strategy for selecting a text password?

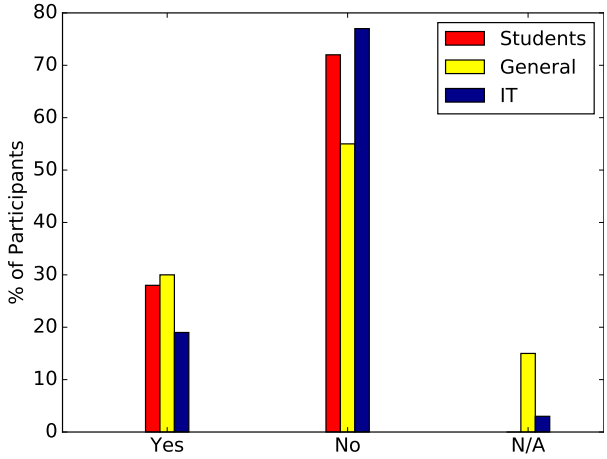

Figure 7: Do you sometimes share your work password or e-learning password with colleagues to do a task when you are not available to do it?

were unemployed and 3% were self-employed). A $\mathcal{X}^2$ test showed no significant difference between any of the three groups ($p = .409$).

Changing Expired Passwords. As the effectiveness of password ageing has been questioned [23], we wanted to know if the participants create new passwords that are a variation of their old ones. A high percentage of participants do use a variation of their old password to create their new password; the IT group uses this technique most (74%), the Students group comes next at 61%, and the General group at 58%. A $\mathcal{X}^2$ test showed no significant difference between any of the three groups ($p = .347$).

Password Change Triggers. To find out if participants change their password frequently and what triggers the password change, we asked the study participants what triggers their password changes. The results (see Figure 8) indicate the most common reason to change the password is 'only if forced by the system', followed by 'when I feel my account has been compromised.' Less popular reasons include when the participant receives an email indicating that someone tried to reset his/her password, and if they clicked a link and realized it was a scam. Fisher and $\mathcal{X}^2$ tests on the different reasons to change passwords showed no significant difference between any of the three groups.

Attitudes About Study Passwords. Figure 9 shows participants' agreement with a number of statements related to passwords created and the password creation process. We asked participants if the password they created is similar to a password they create for an account they don't care about. All the groups responded on average neutrally; however, when we asked the participants if the password they created was similar to one of the important account passwords they use, all group's responses are more agreeable. It is difficult to interpret this result, but possible reasons include not understanding that financial accounts might warrant a more complex password, pleasing the researcher [24], or not understanding the question. Based on Figure 9, it seems that most participants reported that they took the task of creating a password for the study seriously as they believed it was difficult to guess, and resembled a password they will create for an important account. They generally did not wish

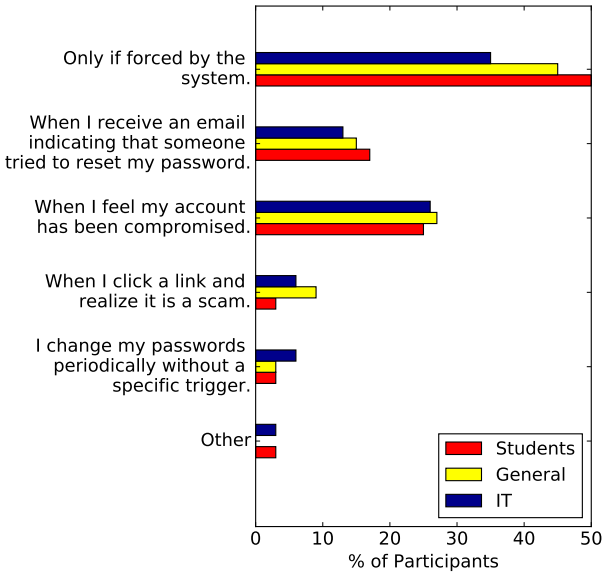

Figure 8: What triggers you to change your password?

to have guidance to create the password, but the IT group seems more open to this idea. Participants seem to agree it was generally easy to create a hard to guess password that they never used before. We found no significant differences after Bonferroni correction, but note that there were some of significance prior to correction: between the Students and General groups in answering if the study password created resembled an important web account password (MWU, $p = .036$), and between all three groups on whether they would like guidance in the password creation process (K-W, $p = .042$).

4.3. Password Analysis

Here we report on the passwords created and used as part of our study. Creation times, login times, password resets, and failed logins are reported as usability metrics in Sections 4.3.1-4.3.4. Password strength is reported in terms of guessability in Section 4.3.5.

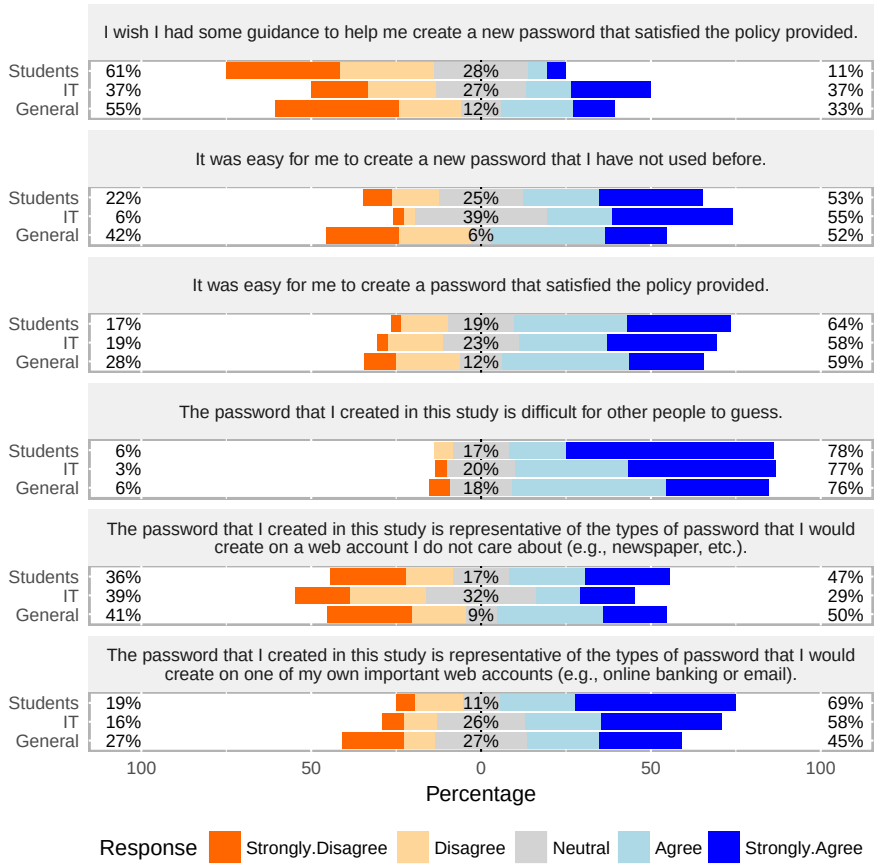

Figure 9: Ratings related to passwords created and the password creation process. The percentages on the left side are the % of participants who disagreed to some degree, the percentages on the right side are the % of participants who agreed to some degree, and the percentages in the middle presents neutral responses.

Table 2: Password creation time in seconds by group.

Group Name	AVG	STD	MAX	MIN	1QT	2QT	3QT
Students	83.1 s	142.9 s	606 s	6 s	19 s	34 s	73 s
General	90.9 s	209.1 s	885 s	9 s	24 s	34 s	53 s
IT	66.9 s	71.6 s	236 s	5 s	22 s	37 s	69 s

4.3.1. Creation Times

We calculated the creation time as the time taken from the time the 'Create password' screen is loaded to the time the 'Submit' button is clicked and the creation is successful. Table 2 shows the password creation time in seconds for each group. Password creation times showed no significant difference between Students and General groups (MWU, $p = .652$) nor any of the three groups (K-W, $p = .851$).

Table 2 shows that the Students and General groups took the longest time on average to create the password, followed by the IT group. It is clear from Table 2 that there is a minimal difference between the groups in the time they needed to create the password. That all three groups created passwords within similar times can be attributed to two reasons. First, the password policy enforced by our study is very common and familiar to most users. This likely translated to users not needing to take a long time trying to come up with a password meeting requirements they are not accustomed to. Second, we asked the participants about the strategy they used when creating a password (see Figure 6). The results indicated no significant differences in strategies used to create new passwords. If the groups generally use similar strategies, it seems sensible that their creation times are also similar.

4.3.2. Login Times

We calculated the login time as the time taken from the second the login pane is loaded to the time the Submit button is clicked. All login attempts are

Table 3: Login time in seconds by group.

Group	AVG	STD	MAX	MIN	1QT	2QT	3QT
Students	13.9 s	25.9 s	146 s	1 s	6 s	9 s	13 s
General	22.0 s	23.3 s	97 s	2 s	12 s	15 s	22.5 s
IT	14.5 s	13.9 s	100 s	2 s	8 s	12 s	17 s

included in this calculation (including failed logins). Table 3 shows the login time in seconds by group. Our data shows a significant difference (before and after Bonferroni correction) between the Students and General groups (MWU, $p = .000$) and for all three groups (K-W, $p = .000$).

Note that our analysis excludes 4 participants (3 from Students, 1 from General) from the login times statistics due to abnormal data that appears to have been caused by unusual browsers.

To avoid considering times when the participant has likely left the system on the login screen (e.g., switched to another task or left their computer), we remove all logins that are longer than 5 minutes; this resulted in 3 login attempts in the IT group being removed. If we focus on the medians, the Students group is the fastest at 9 seconds, the General group is the slowest at 15 seconds, and the IT group is between the two other groups at 12 seconds. For the first, second, and third quartiles, the General group is 6 or more seconds slower than the Students group, and 3-4 seconds slower than the IT group.

4.3.3. Password Resets

A password reset button appeared to the participants after the wrong password is entered three times in a row, but a reset was not required until after 10 incorrect login attempts. None of the participants during the study period had a forced password reset; all resets were voluntarily triggered by the participants. We opted to not show the reset password button until after the third login attempt to encourage the participants to try to remember their password rather than reset it immediately. Overall, 12% of the total number of participants (an

Table 4: Password resets by session by group.

Session	Students	General	IT
1	0.0%	3.0%	0.0%
2	2.8%	3.0%	3.2%
3	8.3%	6.1%	9.7%

actual count of 12 participants) reset their password during the study period. The number of password resets showed no significant difference between any of the three groups ($p = .304$).

As shown in Table 4, the three groups had very close percentages of password resets. Session 3 had the largest number of resets. One participant chose to reset their password at the end of Session 1 when they were required to re-enter the password they created before completing the Session 1 questionnaire. Three participants reset their password in Session 2, and 8 participants reset their password in Session 3. It is worth noting that a few of the participants took longer than 48 hours to log in to Sessions 1 and 2 (between 48 hours and 72 hours); however, none of those particular participants were disqualified (and none of them reset their passwords, so this choice did not affect these statistics). It is noted that the number of password resets increased as time passed. This is an expected behaviour, since text passwords are purely memory-based. However, human memory has limited temporal capacity, and failure to remember a password that has not been used for a period of time can be due to reasons such as decay and fading, interference, and retrieval failure [25].

4.3.4. Login Attempts

Here we examine the number of participants who were able to log in with a single attempt. Results show that Session 1 had 88% successful first login attempts, Session 2 had 78%, and Session 3 had 76%. When examining the percentage of first successful logins by group, the results indicate that IT was the least successful group. This could be due to the fact that IT staff manages so

Table 5: Number of login attempts before success.

Group	AVG	STD	MAX	MIN
Students	1.2294	0.6002	4	1
General	1.2000	0.5292	4	1
IT	1.3191	0.8149	5	1

many passwords in their work, so password fatigue is affecting them. However, the differences in the number of login attempts required before successful login showed no significant difference between the three groups (K-W, $p = .864$). The average number of attempts before successful login is shown in Table 5.

4.3.5. Password Strength

We evaluate password security by the guessability of the passwords using three well-known password guessing methods. For each guessing method we limited the number of guesses to 3 billion, and only use guesses that conform to the policy our passwords were created with; i.e., we assume the attacker would know the policy and make use of that information to optimize the attack. The guessing methods tested were:

1. John the Ripper (JtR). We used JtR 1.8 community enhanced version (bleeding jumbo) [26] with wordlist mode (default rules) and the passwords.txt wordlist (2,151,220 unique values) available at Dazzlepod [27], followed by Incremental mode.
2. Weir approach [28]. This method uses the guesses generated by the software available on Weir's personal website [29], trained on the RockYou dataset and the input dictionary (dic-0294).
3. Semantic approach [30, 31], trained on the RockYou dataset, using its own custom mangling rules.

Both JtR and Weir approaches guessed no passwords. We note that the Weir approach, when filtered according to the password creation policy used in

Table 6: Contents of passwords created for each group. Note that each password might contain words, names, and/or leet.

Contains	**General**	**IT**	**Students**
Leet	9.1%	32.3%	13.9%
Non-English words	3.0%	9.7%	0.0%
English words	33.3%	41.9%	63.9%
Words (any language)	36.3%	51.6%	63.9%
Non-English names	48.5%	22.6%	16.7%
English names	0.0%	3.2%	11.1%
Names (any language)	48.5%	25.8%	27.8%

our study, only generated approximately 300 million passwords. The Semantic approach was the only method to successfully guess any passwords within 3 billion guesses (see Figure 10). The number of passwords guessed in the Students and General groups is comparable (17% and 18% guessed respectively), and only one password (3%) was guessed in the IT group, but the differences between all three groups are not statistically significant (Fisher, $p = .166$).

To gain further insight into the passwords' security, we further manually analyzed the passwords in each group. The results are presented in Table 6. The IT group appears to use 'leet speak' in their passwords more often than other groups (e.g., replacing letters with numbers that look similar, such as the letter 'o' with the number '0'). Note that the semantic guesser does not incorporate leet [30], so a cracker that does incorporate leet might have better performance against the IT group. The Students group appears to use words the most often, and the General group uses names most often. We also analyzed the password lengths of each group, finding the averages were 11.9, 10.8, and 10.5 for the General, IT, and Students groups respectively.

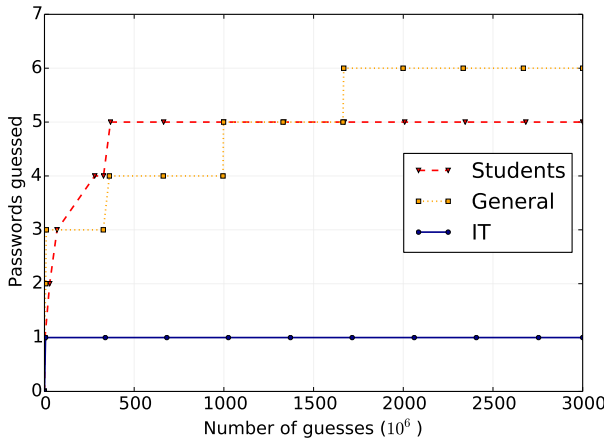

Figure 10: Passwords guessed using the Semantic approach for each group.

4.4. Desirable Authentication System Properties

We were interested in which authentication system properties are most desirable for users, for the purpose of informing future research in user authentication. We asked the participants to rate their interest in using new authentication methods that had the properties listed. Of course, users would like a system that has the best security and usability properties; however, the reality is often that there are trade-offs involved. Thus, we frame these questions in terms of trade-offs, to see which properties are valued above others. We also frame these questions for specific account types, and the results show that desirable properties differ for different accounts.

Figure 11 shows responses for online financial accounts. The responses indicate that methods providing less security were dismissed by the majority of each of the groups, regardless of the advantage offered in return. The most popular method for the financial accounts was one that is slower to input but provides stronger security. Other methods were on average neutrally received by the groups, except for slower to input, but is more entertaining to input than passwords which was rejected by the IT group.

Figure 12 shows responses to the same questions as in Figure 11, but for infrequently used web accounts. For such accounts, the least popular method

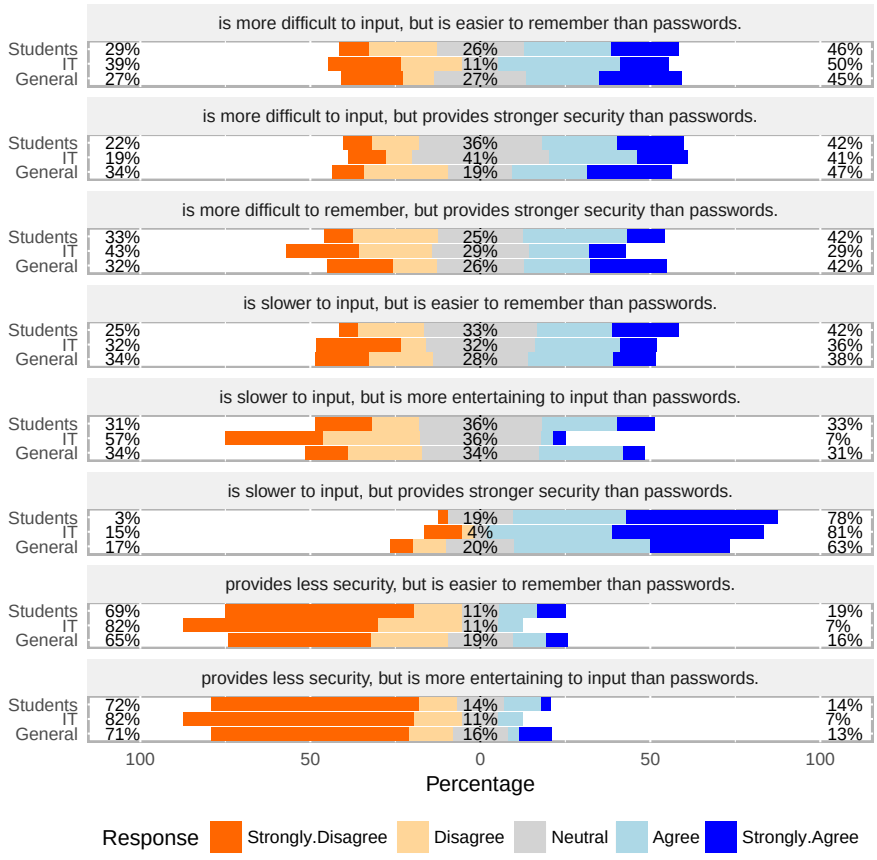

Figure 11: For my *online financial accounts*, I would try a new authentication method that... (Note that 100% of the Students group, 95% of the General population and 90% of the IT group answered this question). The percentages on the left side are the % of participants who disagreed to some degree, the percentages on the right side are the % of participants who agreed to some degree, and the percentages in the middle presents neutral responses.

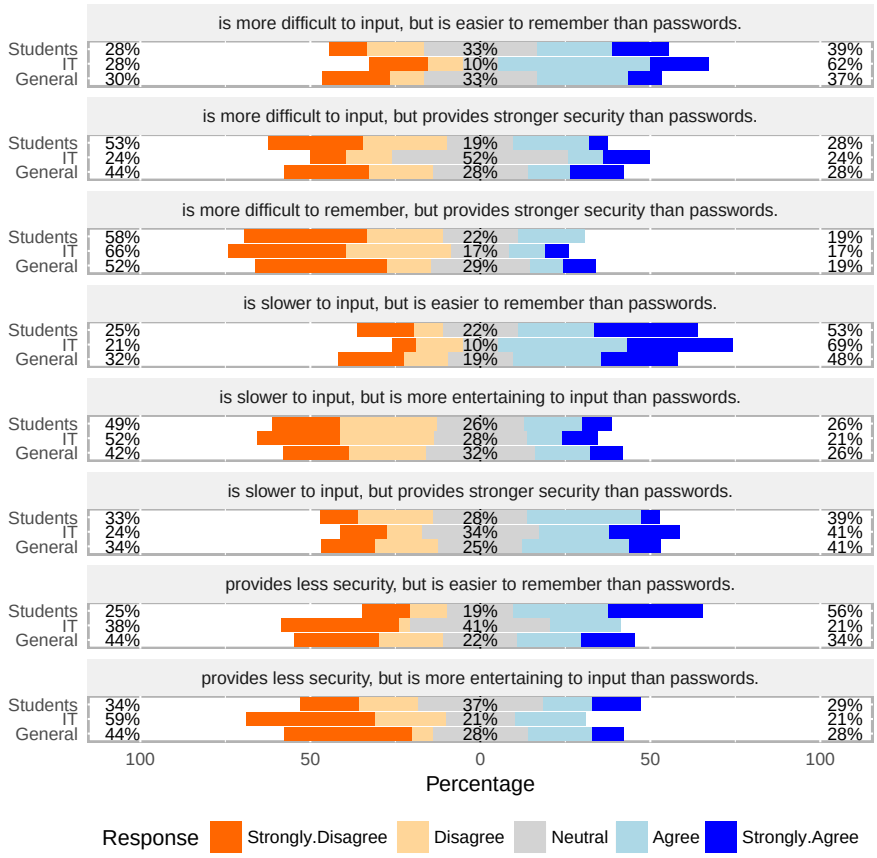

Figure 12: For *occasionally used web accounts*, I would try a new authentication method that... (Note that 100% of the Students group, 95% of the General population and 94% of the IT group answered this question). The percentages on the left side are the % of participants who disagreed to some degree, the percentages on the right side are the % of participants who agreed to some degree, and the percentages in the middle presents neutral responses.

is one that is more difficult to remember, but provides stronger security than passwords. A popular trade-off in this context for all groups is a method that is slower to input, but is easier to remember than passwords. These results show participants changed their preferences when it came to an account they rarely log in to; memorability became more of a concern than security. In this context, we notice the Students group is most open to sacrificing security for the sake of memorability. Less secure methods are not accepted by the IT group, and the groups seem to be neutral with regard to the other methods. The same question was asked for email accounts. Once again, less security was disregarded by all the groups, except for the General group which was neutral for methods that provide less security but are easier to remember. Popular methods were those more difficult to remember, but provide stronger security than passwords, and methods that are slower to input, but provide stronger security than passwords. Regarding email accounts, participants seem to assign similar importance to desirable features as for online financial accounts.

For the analysis in Figs. 9, 11, and 12, we exclude data from participants who provided clearly inconsistent answers (i.e., same response for all questions). This accounted for 8% of Students and 3% of General group responses.

5. Limitations And Ecological Validity

Our study was performed completely online in an attempt to enhance ecological validity; however, the participants were not necessarily assigning importance to the password they created, as it did not actually protect anything of value. We attempted to evaluate the effect of this with the last two questions presented in Figure 9. Responses showed that most, but not all, participants thought their password was representative of one for an important account (e.g., online banking). While our password guessing results also suggest the passwords were fairly strong, we caution that the methods used were trained on primarily English passwords, and our manual analysis found many of the passwords in this study contained non-English words and names (see Table 6).

Our sample sizes ranged from 31-36 participants per group. While it would have been easy to gather more university student participants, it was difficult to recruit for the IT group, and we wished to keep the groups of similar size. It is possible that gathering larger samples would detect statistical significance for additional measures. For example, 3% of the IT group's passwords were guessed (vs. approx. 17-18% for the other groups), but for these proportions and sample sizes, the IT group's passwords were not significantly stronger than the other groups. It would be interesting to collect more data to see if any further differences between groups become significant.

One goal of this study was to explore differences between the password behaviours of students and general (non-students). We caution that it is difficult to be sure that our sample is representative of the general (non-student) population; however, we did verify our sample was diverse in terms of employment, field of expertise, gender, age, computer skills, security knowledge, and time spent online. We also caution that our study only applies to the domain of passwords, and as such, the findings do not necessarily apply to other studies involving participants from these groups.

6. Discussion and Concluding Remarks

This paper explores the behaviour and password habits of three target populations: university students (a common sample for password user studies), general (non-student) participants, and IT professionals. Our results provide insight about the effects of sampling participants from specific target populations. Our results indicate that student participants have significantly faster login times than general (non-student) participants. The Students group was younger than the other groups, so this difference could be due to e.g., agility. Regardless of the reason, this result is something that the research community might wish to consider for the purpose of interpreting performance results when using student participants.

There are a number of data points we analyzed that were significant prior to

multiple test correction, but not after. It is possible that these particular data points may be worth further study. Such results include ones suggesting that IT professionals are more likely to employ some secure behaviours, e.g., recording fewer financial passwords, reusing fewer passwords, having more distinct passwords, and using more random passwords. Also of interest is that the Students group was more likely than the General group to believe their study password resembles a password they might use for an important web account. But again, this was only significant when uncorrected for multiple tests, thus it may be an issue worth further study.

The IT group also appeared to have the most accurate understanding of the guessability of their passwords; only 3% were guessed in our experiments vs. 17-18% from the General and Students groups. That all groups thought their passwords were similarly difficult to guess suggests that password security education for non-IT professionals could be improved.

In most of the study results, the Students group responses were very comparable to the General group responses while the IT group did have relatively different responses. The study results suggest that university students might represent a sensible sample of the general population despite the age difference between the groups. However, it is worth noting that the students had faster login times.

Our results also provide insight into what features for new authentication methods users find most desirable for certain accounts. Security and memorability appear to be the two most important features of authentication systems for users, and the most important feature depends on the type of account. For online financial accounts, participants preferred a system that provides stronger security regardless of other factors. The most preferred method was one that sacrifices input speed in favor of stronger security. For infrequently used web accounts, participants favored memorability, particularly over input speed. Students also appeared to be willing to sacrifice security in favor of memorability on infrequently used web accounts. Regarding email accounts, security was considered something that cannot be sacrificed, and the most popular trade-

offs were to sacrifice either memorability or login speed for improved security. These results can help inform future research into acceptable alternatives to passwords.

7. Acknowledgments

We thank the participants of our user study. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] S. Fahl, M. Harbach, Y. Acar, M. Smith, On the ecological validity of a password study, in: Proceedings of the Ninth Symposium on Usable Privacy and Security, SOUPS '13, ACM, 2013, pp. 13:1–13:13.

[2] A. Forget, S. Chiasson, P. C. van Oorschot, R. Biddle, Improving text passwords through persuasion, in: Proceedings of the 4th symposium on Usable privacy and security, ACM, 2008, pp. 1–12.

[3] J. Thorpe, B. MacRae, A. Salehi-Abari, Usability and security evaluation of geopass: A geographic location-password scheme, in: Proceedings of the Ninth Symposium on Usable Privacy and Security, SOUPS '13, ACM, 2013, pp. 14:1–14:14.

[4] A. E. Dirik, N. Memon, J.-C. Birget, Modeling user choice in the passpoints graphical password scheme, in: Proceedings of the 3rd symposium on Usable privacy and security, ACM, 2007, pp. 20–28.

[5] N. H. Zakaria, D. Griffiths, S. Brostoff, J. Yan, Shoulder surfing defence for recall-based graphical passwords, in: Proceedings of the Seventh Symposium on Usable Privacy and Security, SOUPS '11, ACM, 2011, pp. 6:1–6:12.

[6] M. Hoerger, Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary

participation in psychological research, Cyberpsychology, Behavior, and Social Networking 13 (6) (2010) 697–700.

[7] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, J. Lopez, Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms, in: 2012 IEEE Symposium on Security and Privacy, 2012, pp. 523–537. doi:10.1109/SP.2012.38.

[8] M. Weir, S. Aggarwal, M. Collins, H. Stern, Testing metrics for password creation policies by attacking large sets of revealed passwords, in: Proceedings of the 17th ACM Conference on Computer and Communications Security, CCS '10, ACM, New York, NY, USA, 2010, pp. 162–175. doi:10.1145/1866307.1866327.
URL http://doi.acm.org/10.1145/1866307.1866327

[9] A. Das, J. Bonneau, M. Caesar, N. Borisov, X. Wang, The tangled web of password reuse., in: Proceedings of Network and Distributed System Security Symposium (NDSS), Vol. 14, 2014, pp. 23–26.

[10] D. Florêncio, C. Herley, A large-scale study of web password habits, in: Proceedings of the 16th International Conference on World Wide Web, WWW '07, ACM, New York, NY, USA, 2007, pp. 657–666. doi:10.1145/1242572.1242661.

[11] M. Zviran, W. J. Haga, Password security: an empirical study, Journal of Management Information Systems 15 (4) (1999) 161–185.

[12] J. Zhang, X. Luo, S. Akkaladevi, J. Ziegelmayer, Improving multiple-password recall: an empirical study, European Journal of Information Systems 18 (2) (2009) 165–176.

[13] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, S. Egelman, Of passwords and people: measuring the effect of

password-composition policies, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2011, pp. 2595–2604.

[14] S. Haque, M. Wright, S. Scielzo, A study of user password strategy for multiple accounts, in: Proceedings of the third ACM conference on Data and application security and privacy, ACM, 2013, pp. 173–176.

[15] J. Bonneau, C. Herley, P. C. Van Oorschot, F. Stajano, The quest to replace passwords: A framework for comparative evaluation of web authentication schemes, in: 2012 IEEE Symposium on Security and Privacy, IEEE, 2012, pp. 553–567.

[16] D. Florêncio, C. Herley, P. C. Van Oorschot, An administrator's guide to internet password research., in: LISA, Vol. 14, 2014, pp. 35–52.

[17] M. Buhrmester, T. Kwang, S. D. Gosling, Amazon's mechanical turk a new source of inexpensive, yet high-quality, data?, Perspectives on psychological science 6 (1) (2011) 3–5.

[18] R. Biddle, S. Chiasson, P. Van Oorschot, Graphical passwords: Learning from the first twelve years, ACM Comput. Surv. 44 (4) (2012) 19:1–19:41.

[19] M. L. Mazurek, S. Komanduri, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, R. Shay, B. Ur, Measuring password guessability for an entire university, in: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, ACM, 2013, pp. 173–186.

[20] E. Stobert, R. Biddle, Expert password management, in: International Conference on Passwords, Springer, 2015, pp. 3–20.

[21] I. Loutfi, A. Jøsang, Passwords are not always stronger on the other side of the fence, in: Proceedings of the Usable Security Workshop (USEC), 2015.

[22] I. Ion, R. Reeder, S. Consolvo, "...no one can hack my mind": Comparing expert and non-expert security practices, in: Eleventh Symposium On Usable Privacy and Security (SOUPS 2015), USENIX Association, 2015, pp.

327–346.
URL https://www.usenix.org/conference/soups2015/proceedings/presentation/ion

[23] Y. Zhang, F. Monrose, M. K. Reiter, The security of modern password expiration: An algorithmic framework and empirical analysis, in: Proceedings of the 17th ACM conference on Computer and communications security, ACM, 2010, pp. 176–186.

[24] S. Schechter, Common pitfalls in writing about security and privacy human subjects experiments, and how to avoid them, http://research.microsoft.com/apps/pubs/default.aspx?id=179980, [Last Accessed: June, 2018].

[25] R. C. Atkinson, R. M. Shiffrin, Human memory: A proposed system and its control processes, Psychology of Learning and Motivation 2 (1968) 89–195.

[26] Magnumripper, Community enhanced version (bleeding jumbo) of john the ripper 1.8, https://github.com/magnumripper/JohnTheRipper/tree/c63b0187eab690ba92093a7d6182752527ecd26a, [Last Accessed: June, 2018].

[27] Dazzlepod, Dazzlepod disclosure project, http://dazzlepod.com/disclosure/, [Last Accessed: June, 2016].

[28] M. Weir, S. Aggarwal, B. De Medeiros, B. Glodek, Password cracking using probabilistic context-free grammars, in: 2009 30th IEEE Symposium on Security and Privacy, IEEE, 2009, pp. 391–405.

[29] M. Weir, Reusable security, https://sites.google.com/site/reusablesec/, [Last Accessed: June, 2018].

[30] R. Veras, C. Collins, J. Thorpe, On the semantic patterns of passwords and their security impact, in: Proceedings of NDSS, 2014.

[31] R. Veras, Parsing and semantic classification of passwords, https://github.com/vialab/semantic-guesser [Last Accessed: January, 2018] (2017).