# Quantitative Usability Evaluation

Carleton University, Ottawa, Canada School of Computer Science Winter Term, 2018 COMP3008 Project 2

#### **Authors**

Name	Student Number	Email
Tamara Alhajj	100948027	tamara.alhajj@carleton.ca
Mohamad Yassine	100966528	mohamad.yassine@carleton.ca
Mohamed Gahelrasoul	101007118	mohamed.gahelrasoul@carleton.ca
Roman Kishinevsky	101009733	roman.kishinevsky@carleton.ca
Evan Daniel	100887258	evan.daniel@carleton.ca

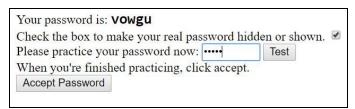
### Table Of Contents

Part 1: Sample Data and Descriptive Statistics	3
Section 1.1 Comparison of TEXT21 and IMAGEPT21	3
Section 1.2 Documentation for Log Data Processing Software	5
Section 1.3 Comparing the Usability of the Password Schemes	6
Descriptive statistics calculated for Both Password Schemes	6
Graphs for Both Password Schemes	7
Part 2: Design, Implementation, Statistical Inference	11
Section 2.1 New Authentication Scheme	11
Section 2.2 Implementation of Password Scheme	13
Development	13
Section 2.3 Documentation of Framework	13
Section 2.4 Questionnaire	16
Section 2.5 Consent Forms	17
Section 2.6	17
Descriptive statistics	17
Graphs for our Password Scheme	17
Inferential Statistics	18
Survey Results	19
Citations	22
Appendix	23

### Part 1: Sample Data and Descriptive Statistics

#### Section 1.1 Comparison of TEXT21 and IMAGEPT21

TEXT21 is a password scheme of 5 characters consisting of lowercase letters, from a to z, and numbers, from 0 to 9. So, each of the 5 characters can be chosen 36 ways, with repetition. This implies password space of this scheme is 36<sup>5</sup>, which about 60 million.



**Figure 1.1.1** Screenshot from the training page<sup>[1]</sup> of the TEXT21 password scheme

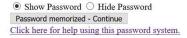
Some advantages of the TEXT21 scheme include:

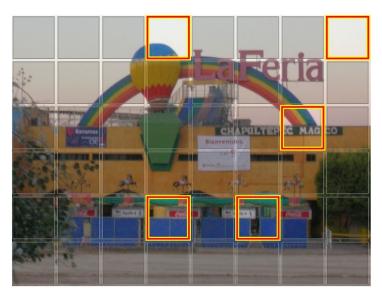
- The password space of the text scheme is larger than that of IMAGEPT21, and thus is more secure.
- Many authentication systems are text based, which makes this password scheme more familiar to a user
- Login in time is very fast as most users are accustomed to typing
- Users cannot choose their own password; random text with high variation will result in a more secure password
- This scheme could be made to be more secure by allowing use of uppercase letters, symbols, or more characters.
- The best password would be both easy to recall, yet difficult to crack. To achieve this, we can take an easy password like "Dog" and morph it into something more secure, whilst maintaining the same memorability, such as "...D0g...".

Some disadvantages of the TEXT21 scheme include:

- Although more secure than the IMAGEPT21 scheme, the time required to brute force search a TEXT21 password is only a fraction of a second<sup>[2]</sup>.
- Random text with high variation is not easy for a user to recall. Moreover, as characters count increases so would the cognitive load to recall said password.
- User may be tempted to write down or save password, to alleviate the need for recall; however, this is an insecure method

The IMAGEPT21 authentication scheme is a password of 5 random tiles, chosen from a image made of 48 tiles. Thus, the password word space of this scheme is 48 choose 5, which is about 1.7 million.





**Figure 1.1.2** Screenshot from the training page<sup>[1]</sup> of the IMAGEPT21 password scheme Some advantages of the IMAGEPT21 scheme include:

- Graphical passwords rely on recognition of visual cues, making for a more memorable password.
- Conveniently, the cued highlighted practice simplifies password memorisation.
   The textual password can also be shown to the user during practice but does not have as strong an effect.
- Visual markers to help the user remember.
- Usability is easy on mobile devices.

Some disadvantages of the *IMAGEPT21* scheme include:

- The password space of the image scheme is smaller, and thus is less secure.
- Longer input time than text based passwords.
- Some pictures may not have visual markers, making it harder to remember the specific quadrants.
- Shoulder-surfing makes it easy to obtain the password of a user.
- User cannot store password, unless they store it semantically, or by possibly taking a screen capture of the password.

#### Section 1.2 Documentation for Log Data Processing Software

The source code for the log data processing software is the file "Part1Section2.R". For this file to run you will need to include the provided csv files:

- "Text21.csv"
- "Imagept21.csv"

The resulting csv data is in a file called "Part1DescriptiveStats.csv".

### Approach

First take each data set and create 3 subsets for each one. One contains the logs for the login attempts, one for the successful logins, and one for the failed logins. Then loop through each data set, row by row, until I find the event that marks when each user enters the login page. Keep track that timestamp and compare it to the timestamp when the user attempts to login. Successful login times will go in one dataframe while failing login times will go in the other. Next, use the 3 subsets from earlier to create 3 frequency tables for each user. These will store each user's total number of login attempts, the number of successful logins, and the number of failed logins.

Calculate the average of the successful login time and the average failed login time for each user using the data frames created earlier.

Afterwhich, create a table for each password scheme that contains all the summarized data for each user (total logins, total successful, total failed, average successful login time, average failed login time).

Finally, combine these two tables into a single table and discriminate between them by adding a column that stores the password scheme used.

Also we have excluded the one major outlier of this data set, as this user took 17217.73333 seconds to log in successfully. This is about 4 hours taken to simply log in; for obvious reasons, this is not representative of the data as a whole. Thus user 28 has been omitted.

#### Pseudocode for Log Data Processing Software

01| Import CSV data 02| Combine text and image data frames Compute average login time differences 03| 04| repeated procedure for 05| Successful Text Login 06| Failed Text Login 07| Successful Image Login |80 Failed Text Login 09| Create a table for each password scheme include all summarized data 10| 11| Combine two tables into a single table 12| Add column for scheme type 13| Remove row of outlier Write filtered data to CSV file 14|

### Section 1.3 Comparing the Usability of the Password Schemes

Descriptive statistics calculated for Both Password Schemes

Password Scheme	Number of Logins	Successful Logins	Failed Logins	Average Successful Login Time	Average Failed Login Time
testpasstiles	19.64285714	14.92857143	4.714285714	18.7467437	24.30327381
testtextrandom	16.61111111	14.0555556	2.55555556	9.95436566	10.83154762

Figure 1.3.1: Table of calculated means

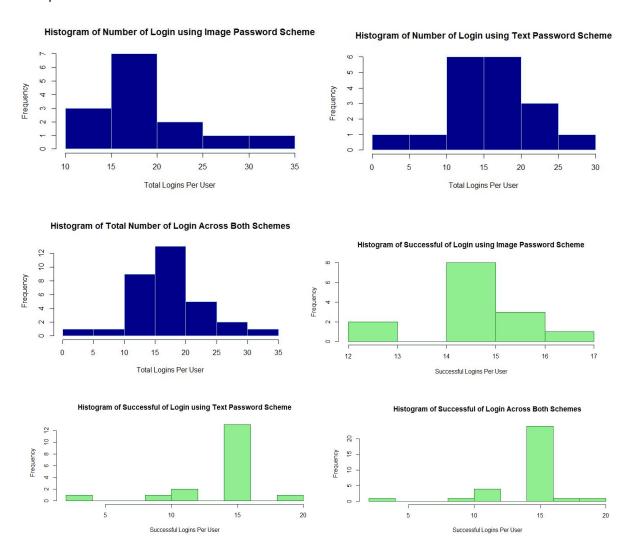
Password Scheme	Number of Logins	Successful Logins	Failed Logins	Average Successful Login Time	Average Failed Login Time
testpasstiles	5.343682724	1.384768001	4.496640993	7.748479856	14.30666573
testtextrandom	4.900646883	3.438060709	3.329409455	4.24446424	5.648491406

Figure 1.3.2: Table of calculated standard deviations

Password Scheme	Number of Logins	Successful Logins	Failed Logins	Average Successful Login Time	Average Failed Login Time
testpasstiles	18.5	15	3.5	20.41458333	19.6666667
testtextrandom	16	15	1	9.066666667	9.0625

Figure 1.3.1: Table of calculated medians

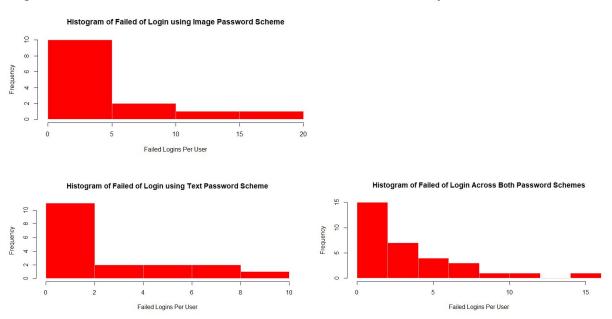
### Graphs for Both Password Schemes



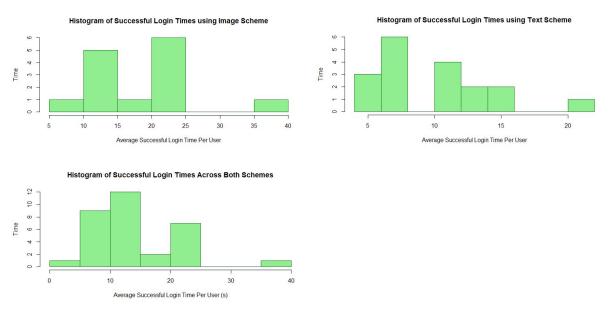
The blue graph displayed for the Image Scheme is skewed to the right, showing a low number of logins, whereas the Text Scheme is skewed to the left, indicating a higher

number of logins. However, when combined, we have a more normally distributed number of logins.

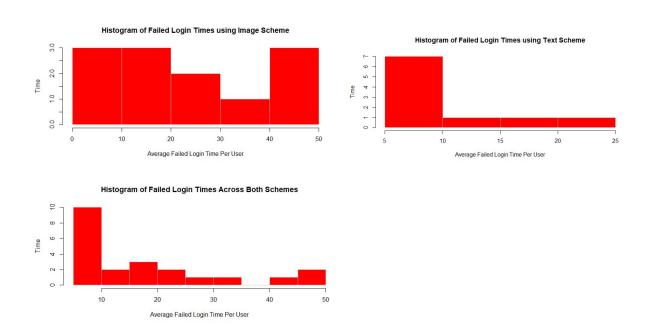
In regards to the successful logins, the displayed graphs are both skewed left, showing that there are about the same number of success. Still, the Text Scheme has a slightly larger success rate. When combined, there is a more normally distributed success rate.



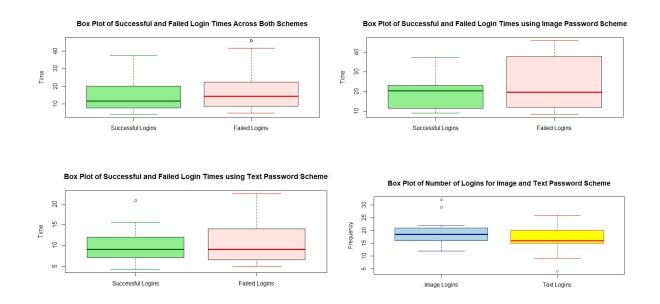
The red graphs are skewed to the right. All the graphs show that it didn't matter whether is was text based or image based, most users didn't have trouble logging in. There were a couple outliers who had issues logging in but overall most of the users had no issue.



The depicted graphs show a bimodal pattern in each one. Displaying the average time it takes for successful logins shows that users generally take a bit more time when entering a Image based pattern, which took about 10-15 seconds or 20-25. In regards to Text based, it shows that users only took about 5-8 seconds, sometimes up to 10-12 seconds. In a combined graph it shows that it takes about 7-15 seconds to login.



The bimodal shape for the first graph, *Histogram of Failed Login Times using Image Scheme*, is an indicator of users who either spent way too much time trying to remember or too little time, hence the bimodal shape. The Text based scheme seemed to still take a shorter time hence the skewed to the right shape, even though it was still failed attempts. As the for the third graph, the shape of the graph is skewed to the right. This indicates that even though the Image Scheme did on average take more time, both schemes usually took about the same time when compared to each other.



Overall there is not a major difference between the two schemes. However, the time taken to login in is clearly shorter when entering a TEXT21 password, rather than an IMAGEPT21 password. This is an indicator that the image scheme takes a longer time to recognize and enter, where as the textbase is purely based on memorization. Which in this case can be slightly faster and more efficient as demonstrated by the data.

# Part 2: Design, Implementation, Statistical Inference

#### Section 2.1 New Authentication Scheme

The analysis of the prior section leads us to discuss a new knowledge-based authentication scheme. Which will assign passwords randomly, just as TEXT21 and IMAGEPT21 do, rather than let users choose them. A reasonably sized password space should be at least 2<sup>21</sup> possible passwords, however our new password scheme turns out to be much more secure. The new scheme does not use characters, words, images, or any combination of those; rather we have implemented something novel: rhythmic authentication. The development of a touch-based scheme provides simplicity, as well as efficiency when entering a password. Moreover, the new scheme is best suited for use on mobile devices, by using vibrations, and optionally sounds, to generate a unique rhythmic pattern. The user would then memorize based on the sequence of vibrations or sounds in later input as a password.

The generated passwords range from 4 to 6 vibrations. The possible time intervals between vibrations is 200 milliseconds to 1000 milliseconds. So, at most, the longest password would be 6000 milliseconds, and the shortest would be 800 milliseconds. Such a range allows for a large number of combinations, to ensure variability and security. The user is given an interface that displays a box in which they can enter their generated password. The reason for including only one box is to keep a simplistic design, which helps to minimize the required cognitive load on the user. Keeping a simple and obvious layout helps the user have a swift and easy login, avoiding any unnecessary inputs that will cause confusion. The box is initially grey, however, once a password has been generated, the box changes colours to indicate that the user may then enter their password. Such a scheme would allow the user to recall passwords due to the rhythm of the vibrations. As an added measure, having an audible sequence that plays with each vibration would aid in further solidifying the password pattern within the user's memory.

In consideration of the widespread use of mobile devices, having such a password scheme allows for high user efficiency and ease of use. In essence, because people have become so dependent on their phones, it makes it an obvious choice to have a quick and discreet password entry for personal accounts on their devices. When a user generates a new password, the only person capable of acquiring that password is the holder of the device, making it unique and only known to the user. In circumstances where the user may be entering their password in public, having a

password that does not require key entering will save the user from having to hide their password entry from possible onlookers.

#### Calculations of Password Space:

Minimum Rest Time (between beats): 200ms
Maximum Rest Time (between beats): 1000ms
Thus, the range is 1000ms - 200ms = 800ms

Margin of Error (Ratio): 0.35

Hence, the total number of rest possibilities:  $800 \times 0.35 = 280$ 

Number of Possible Rests: 4 or 5 or 6

→ by Rule of Sum & Rule of Product, the Total Password Space is

 $280^4 + 280^5 + 280^6$ 

 $\approx 4.84 \text{ x } 10^{14} \text{ possible passwords}$ 

Which is much larger than 2<sup>21</sup> required password space, and thus quite secure.

### Section 2.2 Implementation of Password Scheme

#### Development

The application was written primarily in Android software development kit using Java Development Kit 8.0. All code has been thoroughly commented and documented in the provided compressed file (ZIP). The compressed file also contains a "readme.txt" file which explains how to go about compiling the code and testing it.

#### Section 2.3 Documentation of Framework



Figure 2.3.1: Landing page screenshots of new scheme

The user is first presented with the landing page that is simple and easy to navigate. Upon startup the application automatically generates a username and prepares the log file. First, the user picks out the category for which they need to store a password for. They are given multiple options such as Bank, Email, Shopping, Social, and Other (*Figure 2.3.1*). Once the user has selected their category of choice, they are now able to tap "Generate". Generate simply generates a sequence of vibrations and/or tones (if sound is enabled) in a certain rhythm.

Once the playback of the rhythm is done, the main button changes color (*Figure 2.3.2*), indicating that the user is now able to repeat the pattern they just experienced. In the case the user has forgotten the sequence, they are able to repeat it by simply tapping "Repeat". The repeat button will repeat the *last* generated sequence. Once the user has successfully entered the sequence in the correct rhythm. The repeat button will not repeat any sequence.

Furthermore, as the user starts to enter the sequence a timer is initiated and starts counting down from seven seconds. When the timer stops, the algorithm compares the

user's input versus the generated sequence.



Figure 2.3.2 : User input

Using a mathematical model the program then either proceeds with the input as successful or informs the user their input is incorrect, as seen in *Figure 2.3.3*. The program gives the user a total of three attempts at entering the sequence. It displays the number of failures the user has had so far. At this stage the user is only able to recreate the sequence until they either succeed or fail three consecutive times.

Once the user has failed all three attempts, they are no longer able to enter a sequence as the button has been disabled, in addition, the button changes to a grey colour which would make it easier for the user to acknowledge the button has been disabled. (*Figure 2.3.5*)

In the testing stages the testers have the ability of changing the username on the go without having to relaunch the application. Simply tapping on the profile icon will bring up a dialog (Figure 2.3.4) informing the tester that the previous user's information and data will be stored and a new blank user will be made. In the backend, the user data such as the user ID, sequences with respect to their categories. The backend also deals with the storage of the user sequences with respect to their categories. This is used when the user is being asked to log back in to their Email, Bank ... etc



Figure 2.3.3: User failed to enter correct sequence.

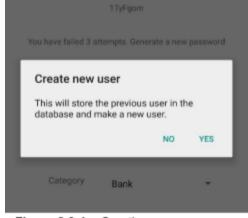


Figure 2.3.4: Creating new users

In addition, the application has the ability to provide the sequence in vibrate and sound. As shown in "Figure 2.3.2", the user has the option of enabling a checkbox titled "Sound". This will allow the user to receive auditory feedback from the application instead of just haptic. From our technical beta testing stages we recognized that some phones had different vibrators of various strengths, which would have rendered our authentication method fairly unusable for users with weak vibrating phones. As a result, we added the auditory feature

to also provide a sound feedback that will aid those users to be able to learn their

sequences better. Moreover, our auditory (ears) system has great pattern recognition. Adding this additional feature of sound will aid in the recognition and memorisation of various patterns compared to only haptic.



Once the user has been asked to generate a sequence for all the possible categories, they are asked to tap on "Login". This will randomize the categories (*Figure 2.3.6*). The user is now prompted to recreate the sequence, in the same rhythm, associated with the provided category. At this point

Figure 2.3.5 : User failure after three attempts

the button has been enabled once again and the colour has changed. Indicating to the user that they are now able to interact the button in order to re-enter their sequence.

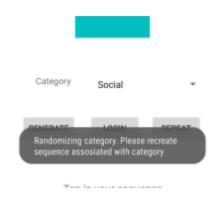


Figure 2.3.6: Login Button

In the case the user has failed to repeat a certain sequence, the system will provide a total of three attempts. If all failed, the system will provide the user with feedback, see *Figure 2.3.5*, and proceed to move forward to the next category. The program will mark the previous category as failed. Once the user has entered a correct sequence that matched the associated category, it will prompt the user and inform them they have successfully entered the authentication sequence, see Figure 2.3.7.

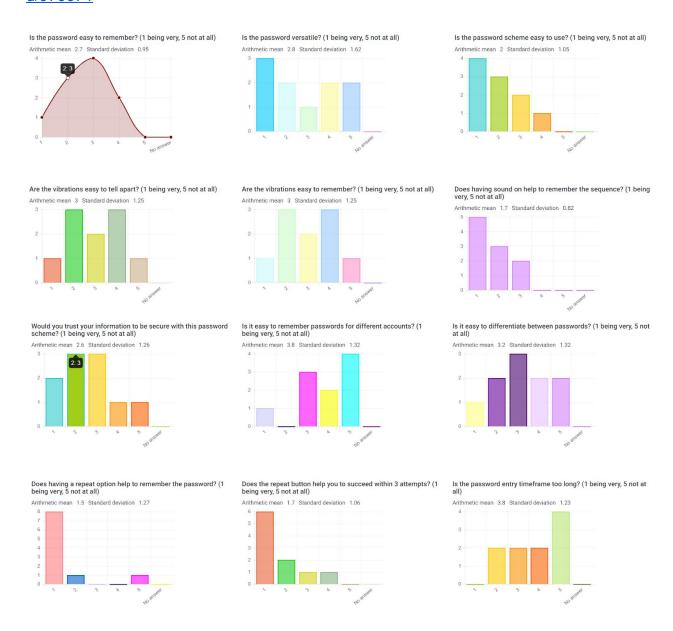
Once all categories are done the program logs the information into a file to be analysed. Every attempt that was made whether successful or not is saved in a log file that contains various other information such as time, user information and preferences, user and algorithm sequence information, and successes or failures for each user.

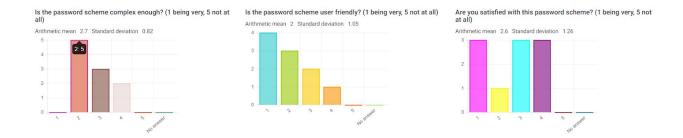


Figure 2.3.7 : Successful Login Attempt

#### Section 2.4 Questionnaire

Our questionnaire to investigate the user's perception of the new password scheme, in comparison to normal user-chosen text passwords can be found at the following site: <a href="https://hotsoft.carleton.ca/comp3008limesurvey/index.php/admin/survey/sa/view/surveyid/976674">https://hotsoft.carleton.ca/comp3008limesurvey/index.php/admin/survey/sa/view/surveyid/976674</a>





#### Section 2.5 Consent Forms

The completed consent forms from user tests have been added to the end of the report. Please refer to the Appendix for more information.

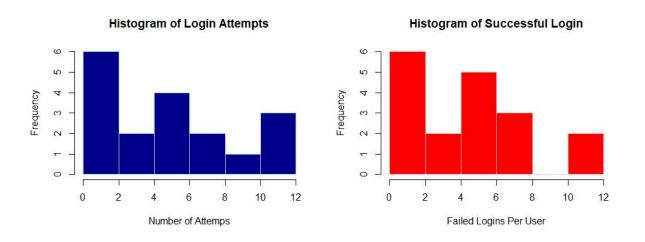
#### Section 2.6

#### Descriptive statistics

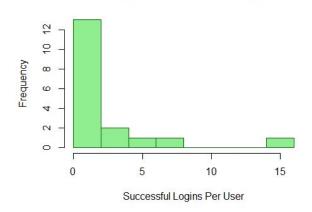
Stat	Total Attempts	Success Count	Failure Count
Mean	5.277778	2.555556	4.833333
SD	4.055578	3.791976	3.63399
Median	6	1	5.5

Figure 2.6.1 Statistical data regarding the rhythmic scheme per user

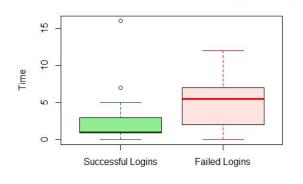
### Graphs for our Password Scheme



#### Histogram of Successful Login



#### Box Plot of Successful and Failed Login Attemps



#### Inferential Statistics

The following chart is a t-Test for successful and failed login counts, between the text based and rhythmic password scheme. Login times are not included since the time for the rhythmic password is set constant and unchanged for any one user.

Data Category	t	Degrees of freedom	p_value
success	9.532106025	33.67874078	4.32E-11
failure	-1.960768183	33.7427581	0.058203171

Figure 2.6.2 t-Test analysis of variance

From the histogram of successful logins for the rhythmic password scheme we can clearly note that few users were consistently successful in their login attempts. Moreover, we can see from the box pot of the successful and failed attempts, regarding the rhythmic scheme, that there is a wider spread of failures. In addition, being heavy-tailed implies that the distribution in question is not a normal one. The outliers seen in above the successful box are the users who were musically inclined and thus able to keep a beat.

In the analysis of variance test, figure 2.6.2, regarding successful password entry, the null hypothesis is that text based passwords have a higher success rate then the rhythmic one. Since the p-value is much smaller than 0.5, this indicates strong evidence against the null hypothesis. As a result, we can reject the null in favor of the alternative hypothesis, that the rhythmic password scheme has a higher successful login rate than

that of TEXT21. However, the p-value regarding failed login count between TEXT21 and the rhythmic scheme is marginal. Thus, we fail to reject the null; although, with a p-value so close to 0.05, we cannot interpret this part of the analysis with any strong conclusions. Most significantly, we must consider for all conclusions we have made, that the sample size of this study is around 100. This directly affects our statistical margin of error, resulting in about 10%[4]

#### Survey Results

Questionnaire overview:

Answers are 1 (very) to 5 (not at all)

- 1. Is the password easy to remember?
- 2. Is the password versatile?
- 3. Is the password scheme easy to use?
- 4. Are the vibrations easy to tell apart?
- 5. Are the vibrations easy to remember?
- 6. Does having sound on help to remember the sequence?
- 7. Would you trust your information to be secure with this password scheme?
- 8. Is it easy to remember passwords for different accounts?
- 9. Is it easy to differentiate between passwords?
- 10. Does having a repeat option help to remember the password?
- 11. Does the repeat button help you to succeed within 3 attempts?
- 12. Is the password entry time frame too long?
- 13. is the password scheme complex enough?
- 14. Is the password scheme user friendly?
- 15. Are you satisfied with this password scheme?

### Frequency of Rating for Each Question

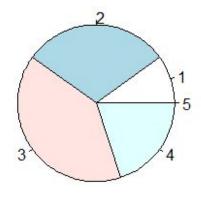
Question	rate 1	rate 2	rate 3	rate 4	rate 5
1	1	3	4	2	0
2	3	2	1	2	2
3	4	3	2	1	0
4	1	3	2	3	1
5	1	3	2	3	1
6	5	3	2	0	0
7	2	3	3	1	1
8	1	0	3	2	4
9	1	2	3	2	2
10	8	1	0	0	1
11	6	2	1	1	0
12	0	2	2	2	4
13	0	5	3	2	0
14	4	3	2	1	0
15	3	1	3	3	0

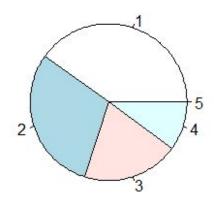
#### Feedback and Improvements

The user surveys displayed as pie charts below, clearly depict that users considered the scheme to be relatively easy to use, though the generated password patterns were challenging for some. The interface was ideal for users to input and repeat their passwords, however, keeping track of multiple passwords was convulating for users. Overall, the participants were divided on their opinions of the rhythmic password scheme. Since the password space is significantly large, the margin of error when validating a password could be increased. This would shrink the space but allow for ease of use for users whom are not rhythmically inclined. Moreover, shortening the login in time would also prove advantageous, as many users felt it too be too long.

# Pie Charts on Questionnaire Rating Frequencies

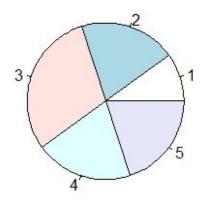
# Is the password easy to remember? Is the password scheme easy to use?

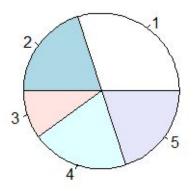




### Is it easy to differentiate between passwords?

Is the password versatile?





### Citations

[1] - SVP3008 Password Scheme Demonstration <a href="https://mvp.soft.carleton.ca/svp3008/">https://mvp.soft.carleton.ca/svp3008/</a>

[2] - Interactive Brute Force Password "Search Space" Calculator <a href="https://www.grc.com/haystack.htm">https://www.grc.com/haystack.htm</a>

[3] - Created questionnaire using the COMP3008 survey system at: https://hotsoft.carleton.ca/comp3008limesurvey

[4] - Sample size in relation to margin of error <a href="https://www.sciencebuddies.org/science-fair-projects/references/sample-size-surveys">https://www.sciencebuddies.org/science-fair-projects/references/sample-size-surveys</a>



Carleton University, School of Computer Science COMP3008: Human-Computer Interaction, Winter, 2018

Interaction Design Project User Consent Form

Instructor: Prof. Robert Biddle

Office: HP5169, Tel. Ext. 6317, Email: robert.biddle@carleton.ca

COMP3008 Student Names:

Mohamad Vassine, Tamara Alhajj, Roman Kishinevsky, Evan Daniel, Mohamed Gishelrass

Note: This project was reviewed by the Carleton University Research Ethics Board (CUREB-B), which provided clearance to carry out the research: (Clearance #105985).

If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca

Study Purpose: This project is to enable students in COMP3008 to experience working with potential users to better understand testing of software using quantitative measures and analysis. Working with potential users is considered best practice in the design of interactive computer systems, and is identified in international standards such as ISO 9241-210:2010 — Human-centred design for interactive systems. Part of the process is first obtaining informed consent from prospective participants, and that is the reason for this form.

Study Procedure: In the project, the proposed system is a kind of password system, where users will be asked to use the system, remember some random passwords, and then try to use them to simulate login to a system. At no time will you be asked for any of your real passwords on any real systems. You will also be asked to fill out a questionnaire about your perceptions of the software and related issues.

Risks, Benefits, Compensation: We are not aware of any risks associated with this study. The benefits are that COMP3008 students gain experience in this aspect of Human Computer Interaction Design, and that you may gain some insight about the processes involved. There will be no financial compensation for your participation in the study.

Consent and Withdrawal: We require your consent before you can participate in the student, which you may indicate by signing your initials in the space provided below. You may choose to withdraw from the study at any time and without explanation, in which case any collected data will be discarded.

Anonymity and Confidentiality: The study involves gaining better general understanding of how students plan their degree programs, but does not involve asking you any specific personal information, nor any specific details of your degree or courses. We do not record or even ask your full name. We will not record or divulge any personal information about you.

By signing with your initials below, you consent to participate in the study.

tials:	Date:		
	03 - 04 - 2019		

Agreement	Initials	Date
I consent to	101	2018 0 / 2 /
particpate.	-Unle.	April 3rd.
I consent to	1	2018
particpate.	1//~	09/03
I consent to	11.	2018 0 1 0
particpate.		03.11 0 )
I consent to	0.1	2018 1
particpate.	m	Apr 4
I consent to	De	2018 100
particpate.	5	2018 1/ / /
I consent to	RS	2018 14 114
particpate.	1)/	2018
I consent to	$\propto$	2018 2/11 - 2
particpate.		2010
I consent to	Dew	201864-03
particpate.	U CW	2010
I consent to	10-1	2018-04-03
particpate.	10	2018 3 //
I consent to	MG	2018 3/4
particpate.	70(0)	2018
I consent to		2016
particpate.  I consent to		2018
particpate.		2016
I consent to		2018
particpate.		2010
I consent to		2018
particpate.		2010
I consent to		2018
particpate.		
I consent to		2018
particpate.		
I consent to		2018
particpate.		
I consent to		2018
particpate.		
I consent to		2018
particpate.		
I consent to		2018
particpate.		
I consent to		2018
particpate.		
I consent to		2018
particpate.		
I consent to		2018
particpate.		,
I consent to		2018
particpate.		