

COMP 3008: Project 2

Winter 2018

Project Members

Jason Lai, Altin Rexhepaj, Randy Taylor, Devin W.

Table of Contents

| | |
|--|-----------|
| COMP 3008: Project 2 | 1 |
| Project Members | 1 |
| Table of Contents | 2 |
| Part 2: Design, Implementation, Statistical Inference | 3 |
| Part 2.1: Abstract | 3 |
| Rationale | 3 |
| Password Space Complexity | 4 |
| Part 2.2: Password Scheme Demonstration | 5 |
| Part 2.3: Quantitative Testing Framework Demonstration | 8 |
| Part 2.4,5,6: Survey & Results | 10 |
| Questions | 10 |
| Survey Results | 19 |
| Results in Comparison to Text21 | 19 |
| Number of Logins (Histogram) | 19 |
| Number of Successful Logins (Histogram) | 20 |
| Average Successful Login Time (Histogram) | 21 |
| Average Successful Login Time (Boxplot) | 21 |
| Number of Failed Logins (Histogram) | 22 |
| Average Failed Login Time (Histogram) | 23 |
| Average Failed Login Time (Boxplot) | 25 |
| Average Login Time (Histogram) | 25 |
| Average Login Time (Boxplot) | 26 |
| Graphs Unique to the Phonetic.pw Scheme Analysis | 28 |
| Number of Practices (Histogram) | 28 |
| Number of Successful Practices (Histogram) | 28 |
| Number of Failed Practices (Histogram) | 29 |
| Number of Hints (Histogram) | 29 |
| Average Hint Time (Histogram) | 30 |
| Average Hint Time (Boxplot) | 31 |
| Average Test Time (Histogram) | 31 |
| Average Test Time (Boxplot) | 31 |
| Conclusion | 32 |
| Appendix | 34 |
| Appendix A: Lookup table used in phonetic password scheme | 34 |
| Appendix B: Consent forms and participant signatures | 35 |

Part 2: Design, Implementation, Statistical Inference

Part 2.1: Abstract

Rationale

Based on our comparisons of Image21 and Text21, we decided to use a text-based scheme. However, we experienced some difficulty in remembering randomly generated passwords and wanted to provide some pattern to provide a memory aid. From this, our goal with our password scheme was to make more effective use of working memory using chunking. To do this, we decided to use a phonetic password scheme. The idea behind this is the password will be easier to remember if the user can pronounce it as a word. This encourages the user to process the data both textually and verbally. The user only needs to remember a few sounds instead of some number of individual letters. We initially chose to preserve the five-letter length of Text21.

We excluded graphemes, which are phonemes produced by multiple different combinations of letters, from the password scheme. This ensures that the user does not feel confused when attempting to sound out the password to remember it. For example, 'K' and 'CK' and 'F' and 'PH' are both graphemes, so we would only use what we felt was the most intuitive phoneme. In this case, 'K' and 'F.'

Since our five-letter word did not reach the 21 bit requirement, we chose to append a single digit to the end of our word since the math worked out almost perfectly. By adding a sixth letter to the end of the word we could have reached a 21 bit size, but there was some concern with the expected number of syllables per password. It was decided that a mono- or disyllabic word with a single number would be easier to remember than a trisyllabic word.

To create a word we randomly generate one letter, then generate a valid succeeding letter. The succeeding letter is selected from a lookup table to produce a phoneme, and this process repeats until the word has been fully generated. Each succeeding letter depends on its preceding letter. Our lookup table limits the number of graphemes, as mentioned previously. For example, the letter 'A' can be succeeded by 'K' but not by 'C,' 'CK,' or 'Q.' Please refer to Appendix A for the full lookup table.

Another issue we had to address was combinations of letters that are difficult to pronounce, like 'NW' or 'WB'. These have also been excluded from our lookup table. We have excluded the possibility of repeating letters since it would be impossible to recall duplicate letters based solely on the sound of the word they belong to.

By applying our rules, the scheme is able to consistently produce a readable word without creating ambiguity regarding its spelling. Thanks to this, the scheme should deliver good usability and memorability.

After comparing our proposed scheme to Text21, we chose to use the former over the latter because the sound of the word aids in the user's ability to memorize and recall the word, whereas a random password has no pattern to aid in either. Additionally, the presence of numbers in the middle of a word awkwardly breaks up the password. Our scheme does not have this problem, as the number is always at the end.

Password Space Complexity

1. The letters 'Q' and 'K' have the same phoneme (/k/), so we have removed 'Q' as a possibility in this scheme and kept 'K' (it is more natural), which reduces our phonetic alphabet to 25 letters.
2. The letter 'X' is very uncommon in the English language and is difficult to make phonetic words with as no letter can succeed it without causing phonetic overlap with the letter 'Z'. For example, "Xenon" is heard as "Zenon". Alternatively, no letter can precede 'X' without overlapping with the phonetic sound made from the series of the letters 'E', 'K', and 'S', which make 'Eks'. Excluding 'E', 'K', and 'S' entirely will reduce the password space significantly, so we have excluded 'X' instead.
 - a. Finally, our phonetic alphabet comprises of **24 letters**
3. Our word length for this scheme is **5 letters**, followed by **1 digit from 0 to 9**.
4. By the product rule, the total number of passwords is given by all the possibilities of phonetic 5 letter passwords multiplied by 10 possibilities of digits from the interval [0,9].
 - a. Using *calcPasswordComplexity()* in the directory *"/phonetic.pw/src/client/scheme.js"*, our password complexity comes out to exactly **2,036,790 $\approx 2^{21}$**

Total Password Complexity: 2,036,790 $\approx 2^{21}$

Part 2.2: Password Scheme Demonstration

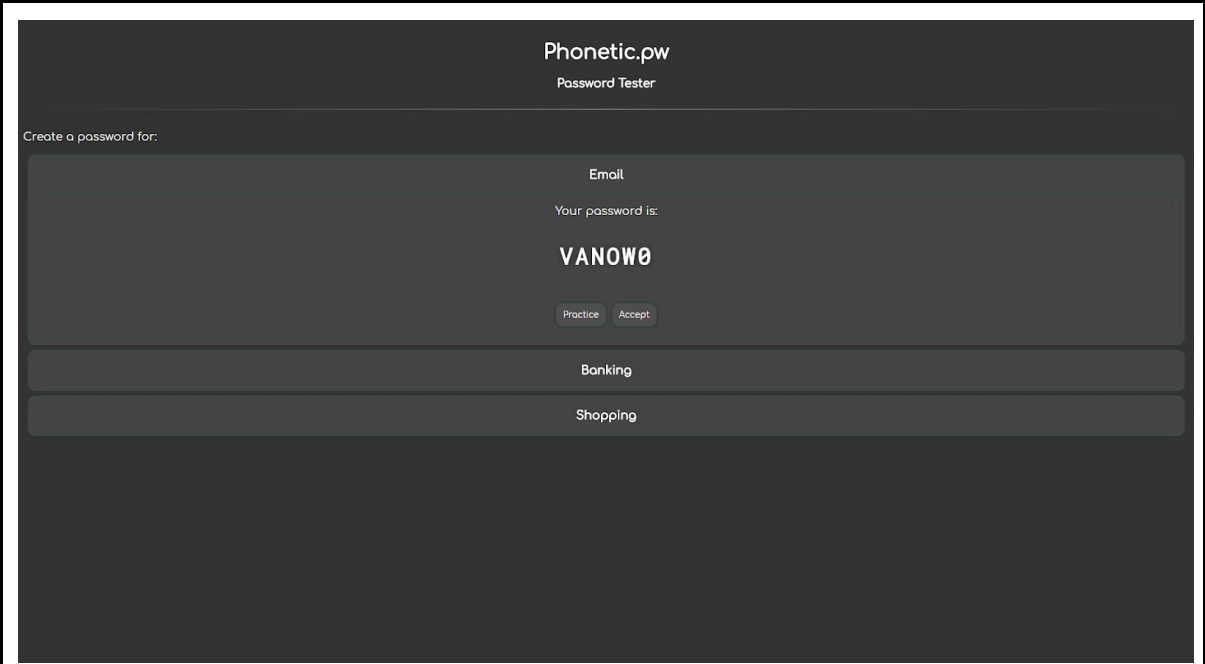


Figure 2.2.1: As the user opens the web-app, the password creation process for the email system automatically starts and displays a generated password. From here, the user has two options: Practice (which open the practice password entry) or Accept (which saves the generated password for the user).

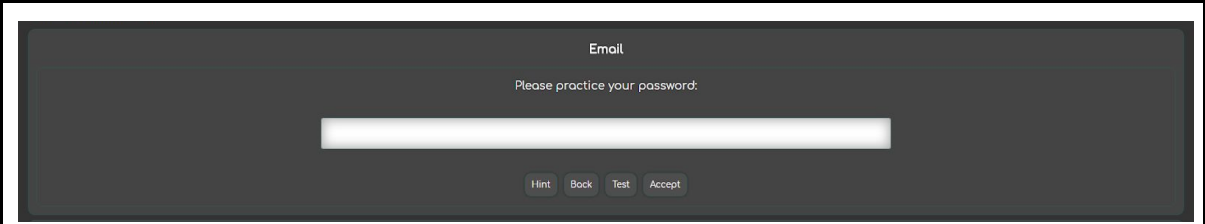


Figure 2.2.2: If the user decides to practice their password, this “password practicer” opens.

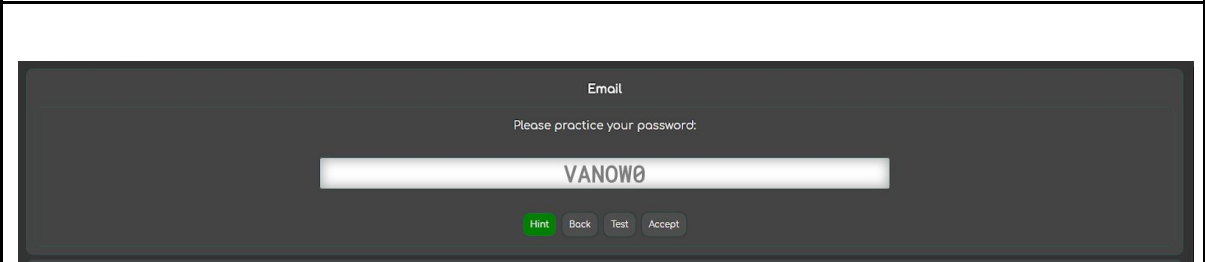


Figure 2.2.3: If the user needs a reminder of what the generated password is, they can click and hold the “hint” button that will show the password for as long as they hold it.

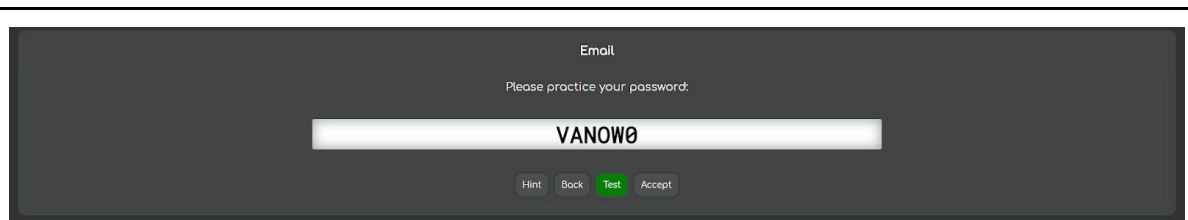


Figure 2.2.4: The user can also practice their password by typing their guess and pressing the “test” button.

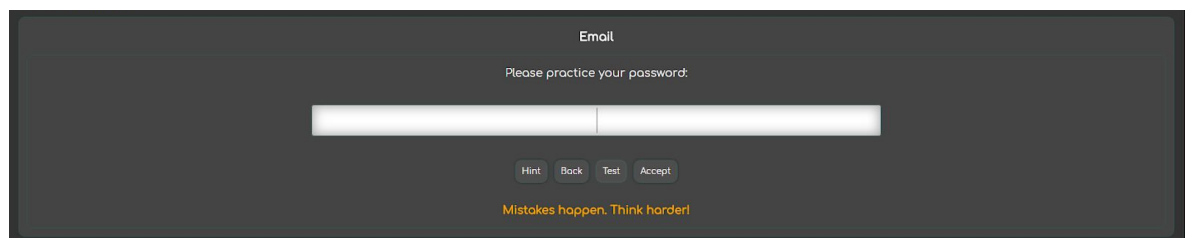


Figure 2.2.5: If they enter an incorrect password, there will be a brief, but encouraging message indicating that it was incorrect, and the input will be reset for them to try again, should they want to.

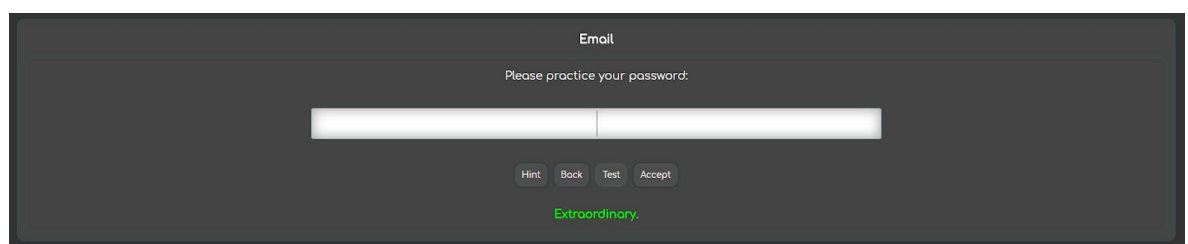


Figure 2.2.6: Likewise, if they enter the correct password, there will be a brief, but encouraging message indicating that it was correct, and the input will be reset for them to try again, should they want to.

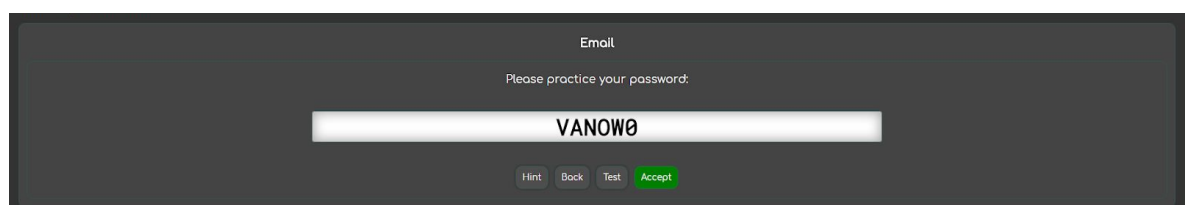


Figure 2.2.7: If the user feels they are ready, they accept either from the “password practicer” or they can click “back” and then “accept” in the initial “generator” page.

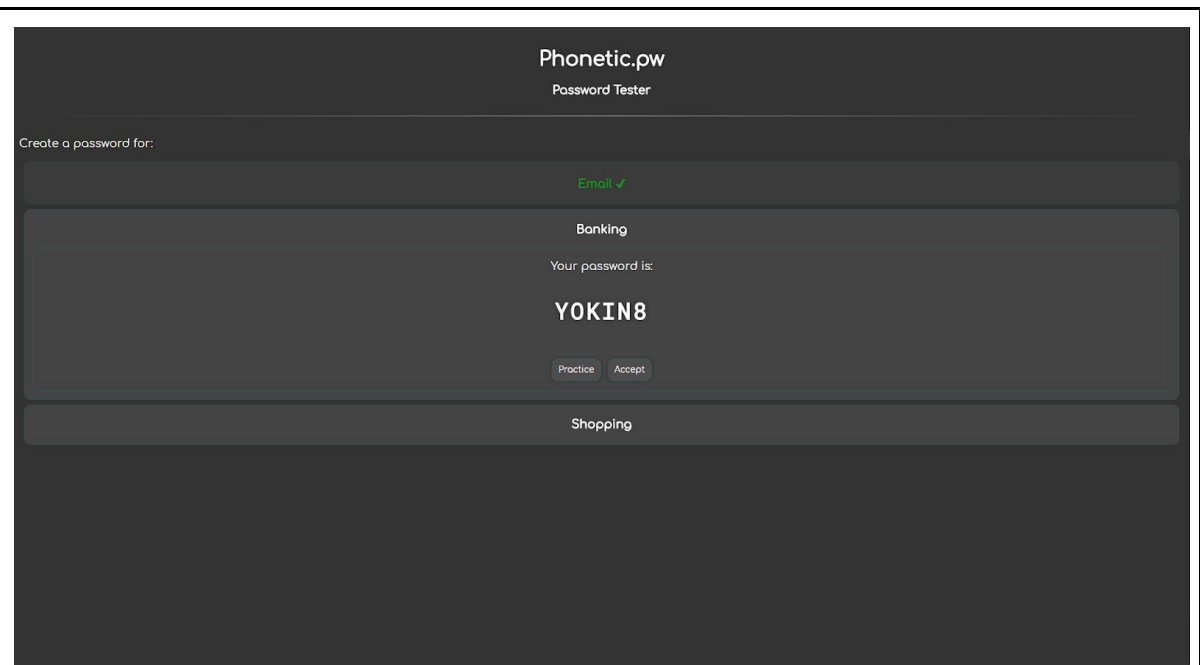


Figure 2.2.8: Once the user accepts, their password, they automatically move on to creating the next one.

Part 2.3: Quantitative Testing Framework Demonstration

The screenshot shows the 'Phonetic.pw Password Tester' interface. Under the heading 'Create a password for:', there are three rows, each with a green checkmark and a label: 'Email ✓', 'Banking ✓', and 'Shopping ✓'. Below this, under the heading 'Test your password for:', there is a section for the 'Banking' password. It shows 'Attempts Left: 3' and a text input field labeled 'Enter your password:'. The input field is currently empty, and a 'Submit' button is located below it.

Figure 2.3.1: Once all the passwords are generated, the user then enters the testing stage, where the order that the passwords are tested is randomized. Users are given 3 attempts.

This screenshot shows the same 'Phonetic.pw Password Tester' interface as Figure 2.3.1, but with the password 'YOKIN7' entered into the 'Enter your password:' field for the 'Banking' password. The 'Submit' button remains visible below the input field.

Figure 2.3.2: Once the user has typed in their password, they click the “submit” button to try logging in.

Phonetic.pw
Password Tester

Create a password for:

Email ✓

Banking ✓

Shopping ✓

Test your password for:

Banking
Attempts Left: 2

Enter your password:

Submit

It's on the tip of your tongue!

Figure 2.3.3: Similar to when practicing passwords in the creation stage, if the user enters an incorrect password there is a brief message indicating it. The “attempts left” counter is also decremented.

Phonetic.pw
Password Tester

Create a password for:

Email ✓

Banking ✗

Shopping ✓

Test your password for:

Banking ✗

Shopping ✓

Email ✓

The session is now over. Thank you for your participation.

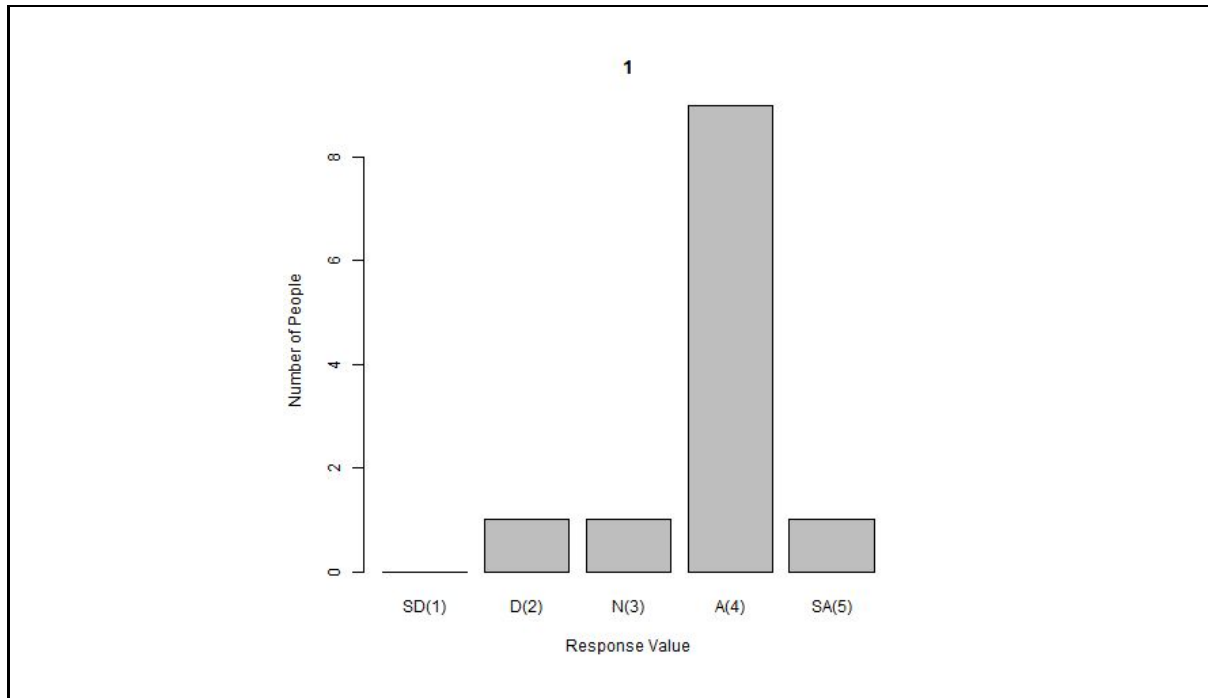
Figure 2.3.4: If the user fails to enter the correct password within 3 attempts, they fail the login test and the system title becomes red with an ‘✗’ beside it. Otherwise, if they succeed the login, the system title becomes green with a ‘✓’ beside it. Once the logins are completed for all the systems, the user has completed the testing stage.

Part 2.4,5,6: Survey & Results

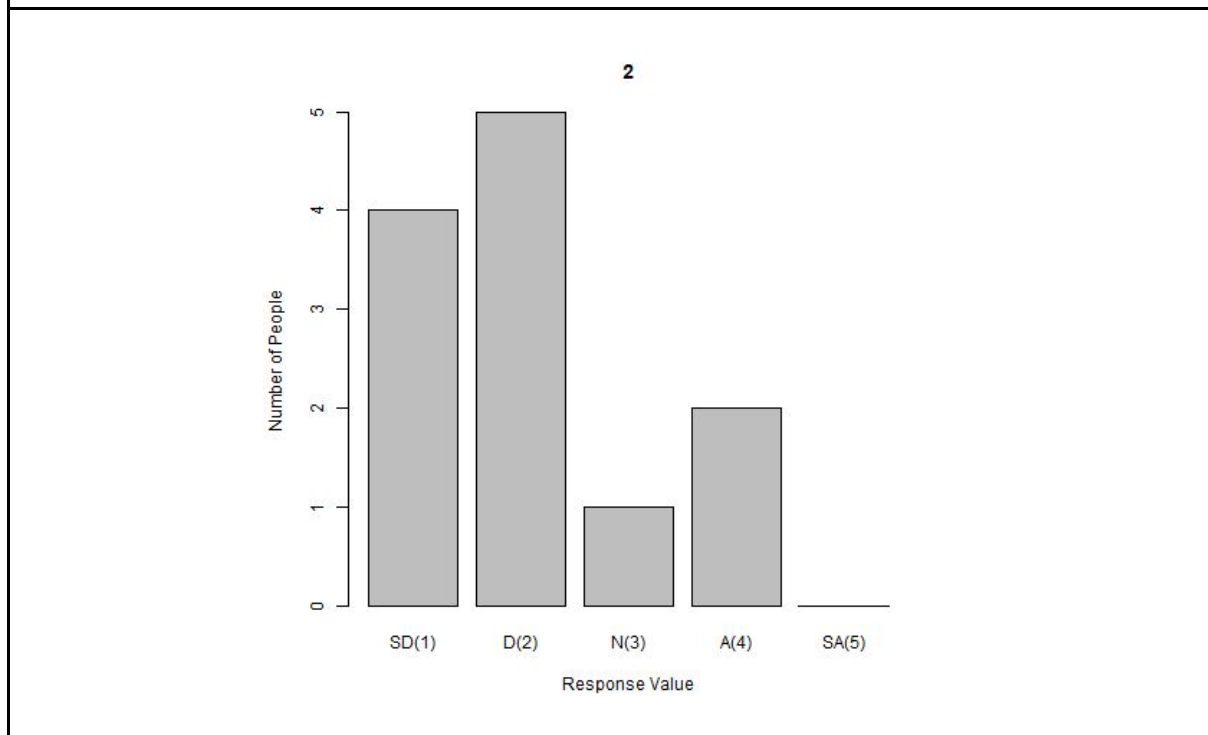
Survey: <https://hotsoft.carleton.ca/comp3008limesurvey/index.php/857752?lang=en>

Questions: See *survey_questions.pdf*

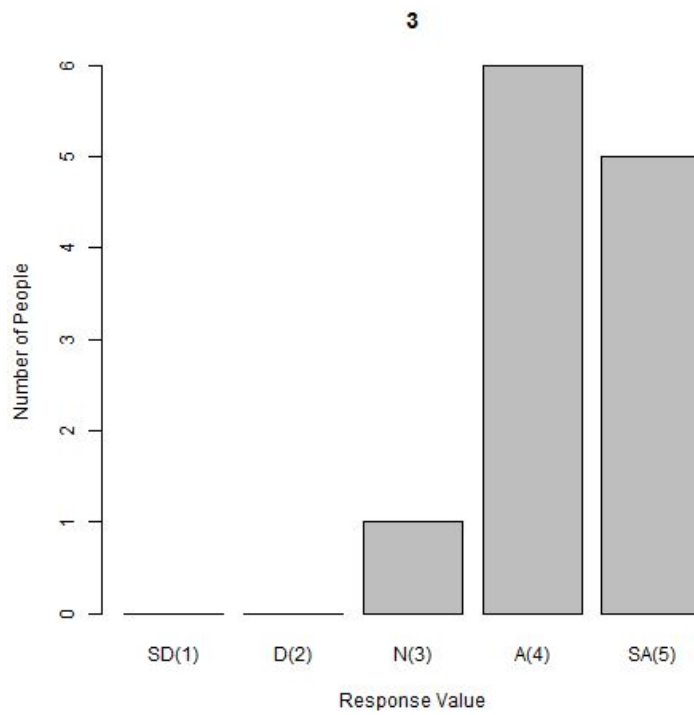
Questions



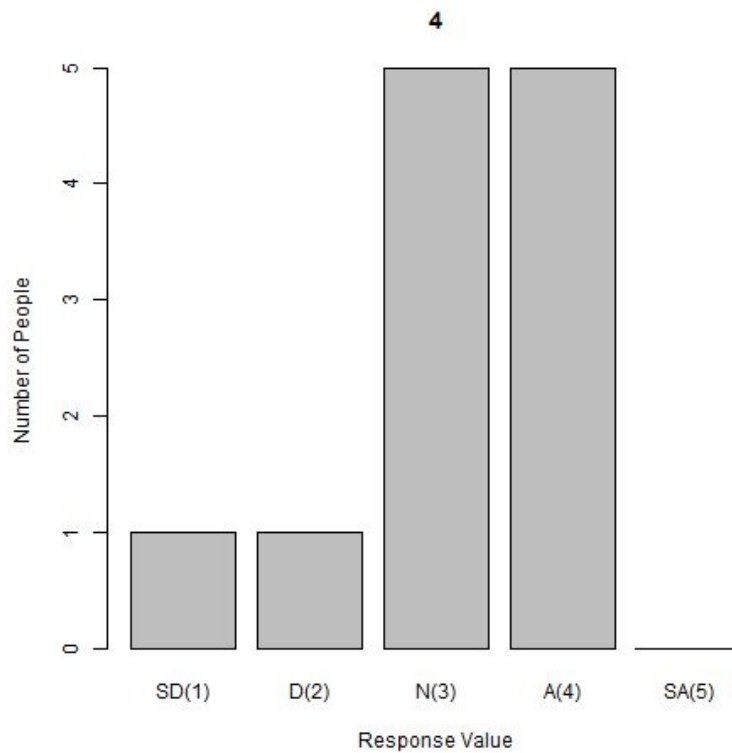
1. I thought the passwords generated by the scheme were easy to remember.



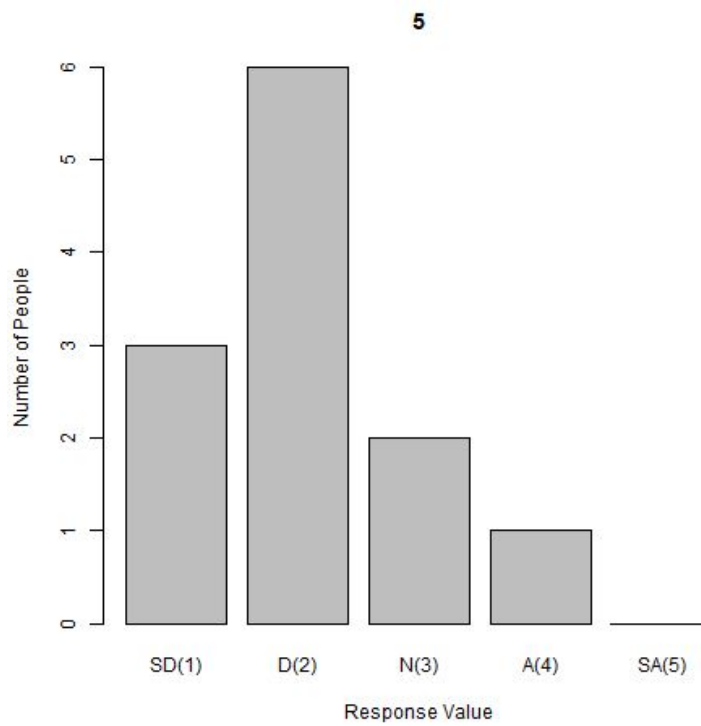
2. I find generating a random phonetic word impedes memorability.



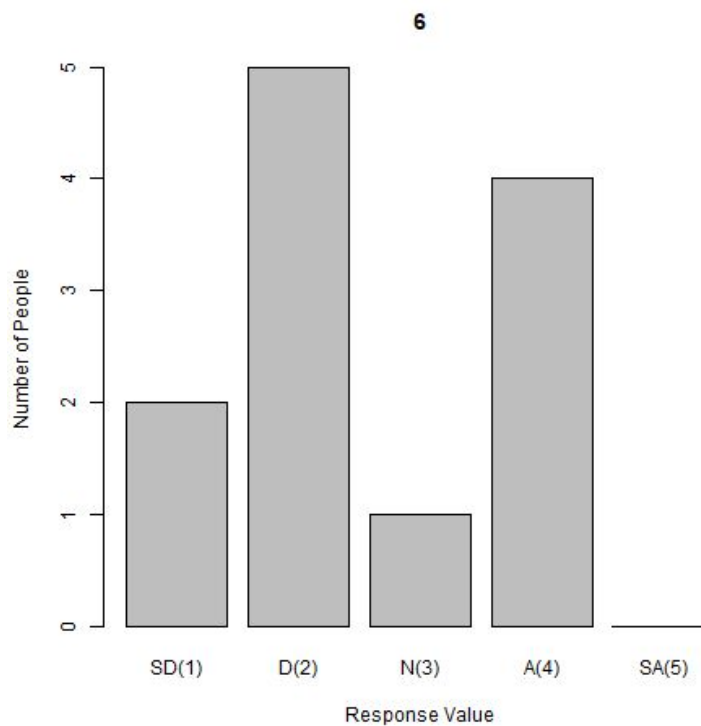
3. I felt the phonetic gimmick was helpful in creating a memorable password.



4. I found it easy to remember which passwords belong to which account.

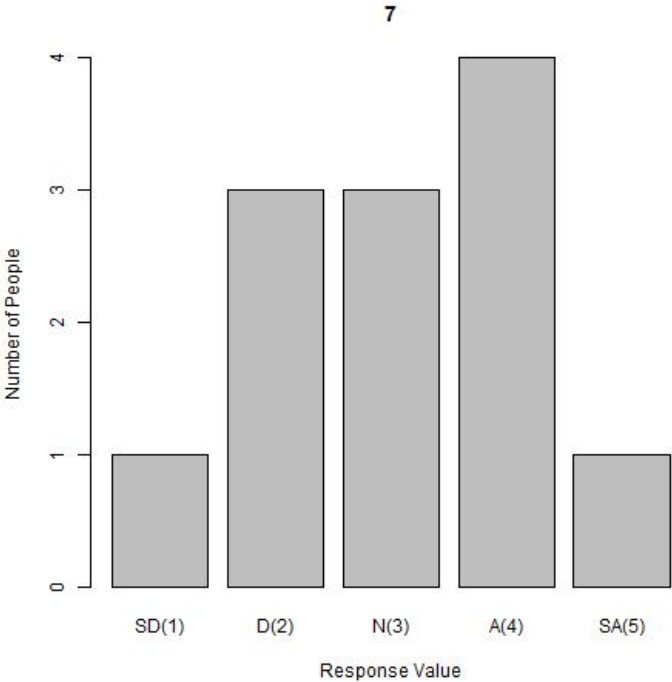


5. I found it difficult to remember which letters made up which sounds in my passwords.

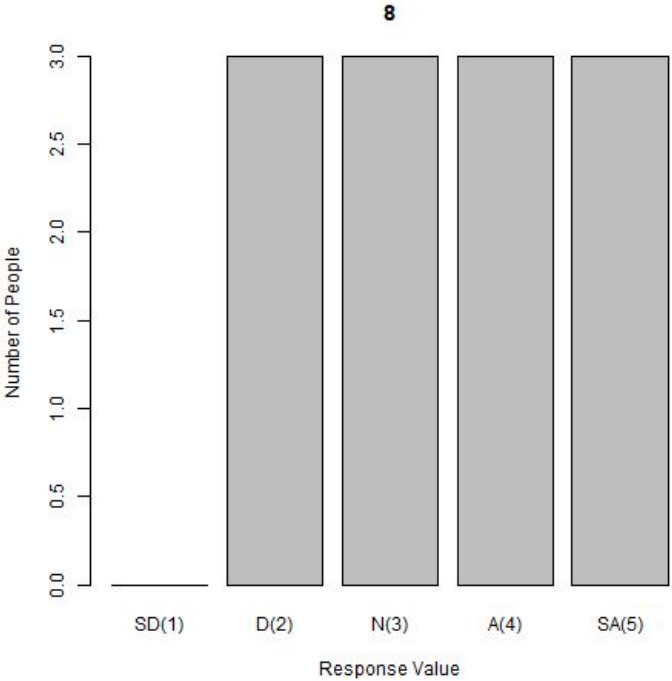


6. I think these passwords would be easy to remember, even if I had many of them

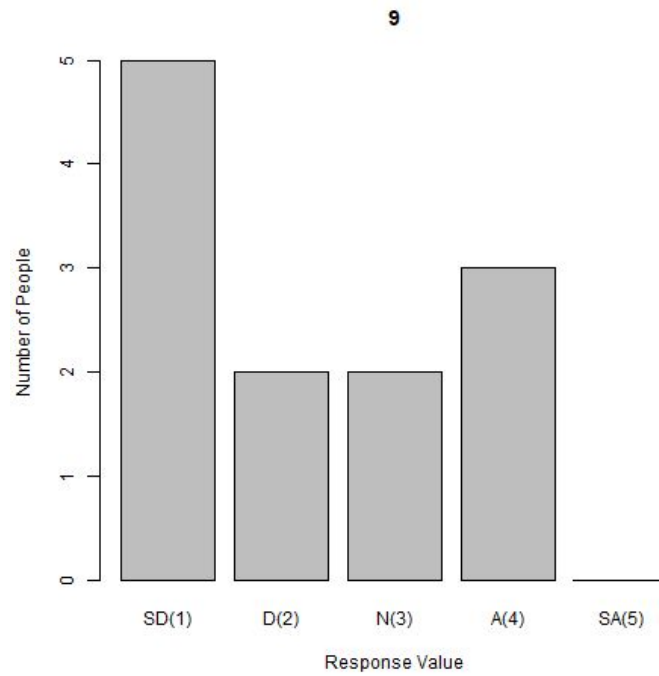
for many different accounts.



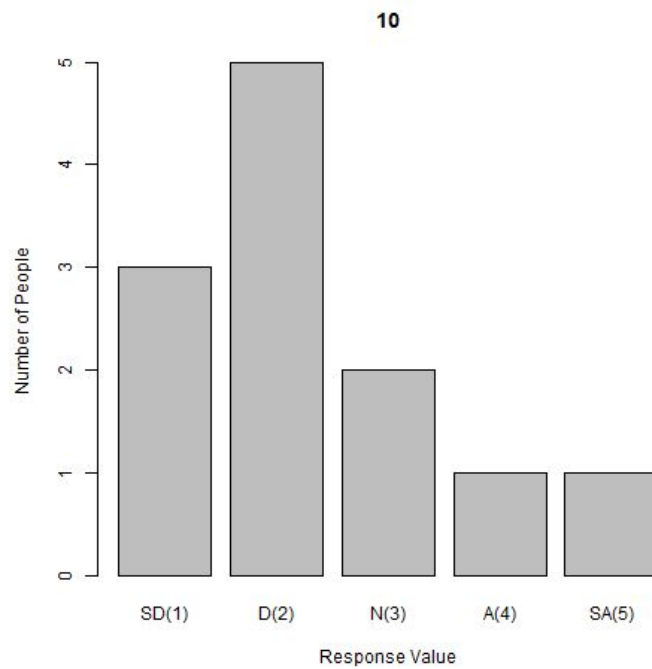
7. I feel I could remember these passwords long term.



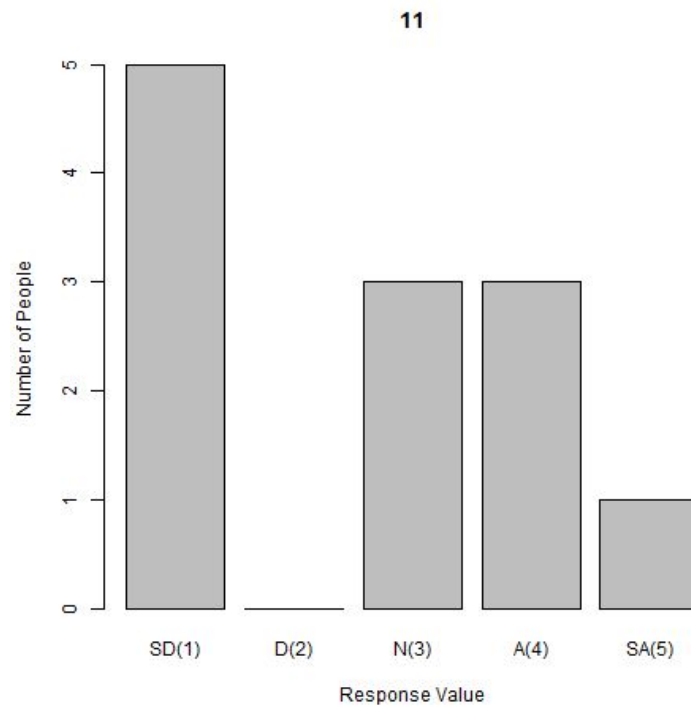
8. I needed to rely on an alternative memory trick or tricks to remember my passwords.



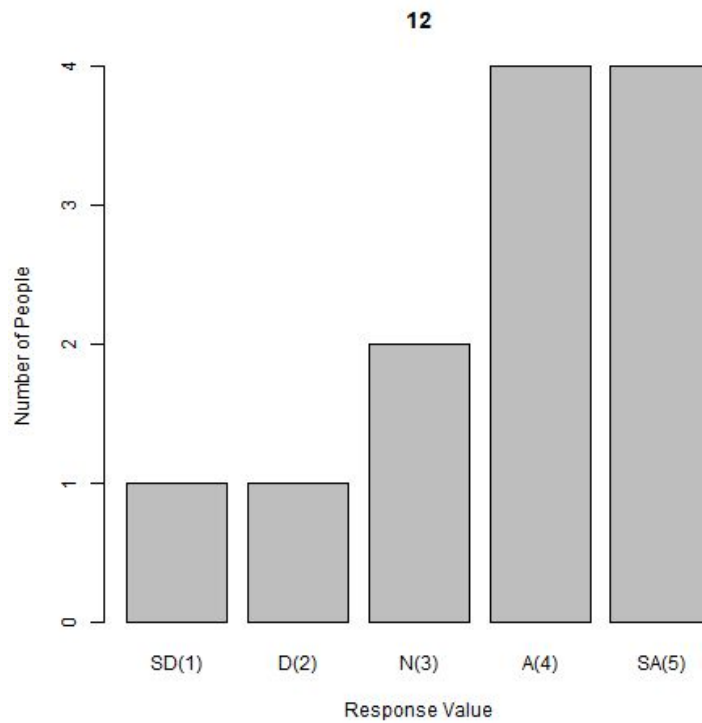
9. I think the numbers included at the end of the password helped me remember it.



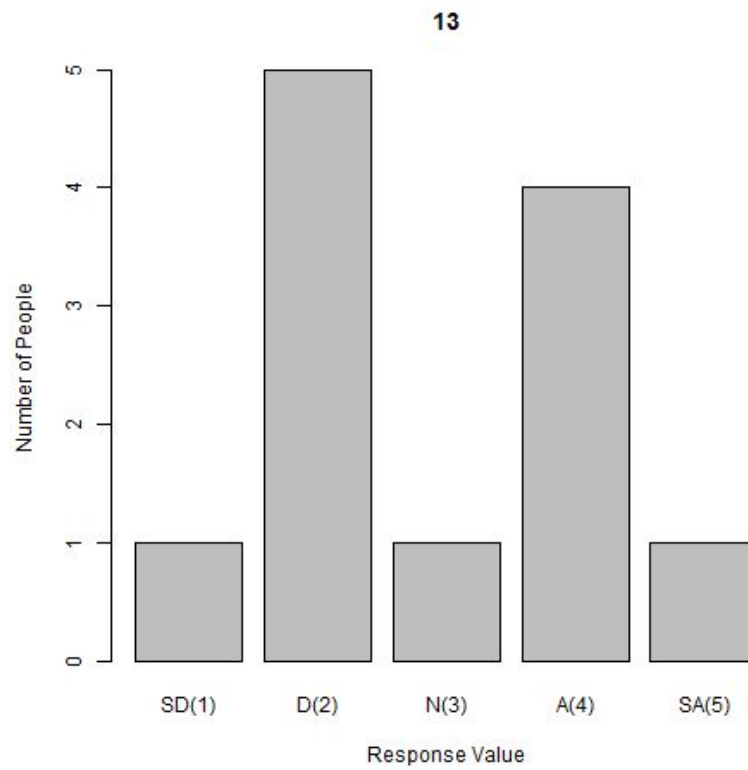
10. I found this scheme more difficult to use than user-chosen ones.



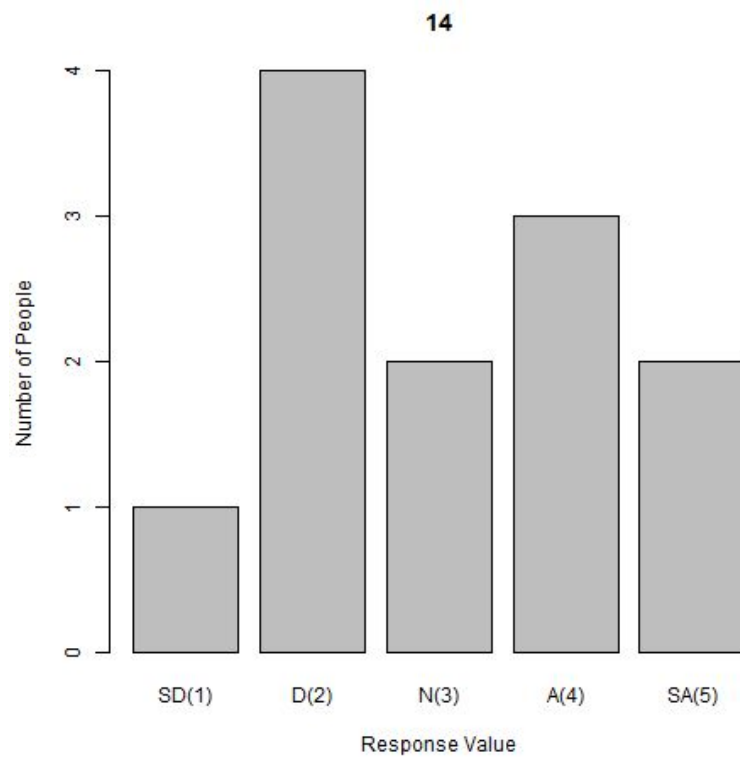
11. I would use this scheme over a user-chosen scheme.



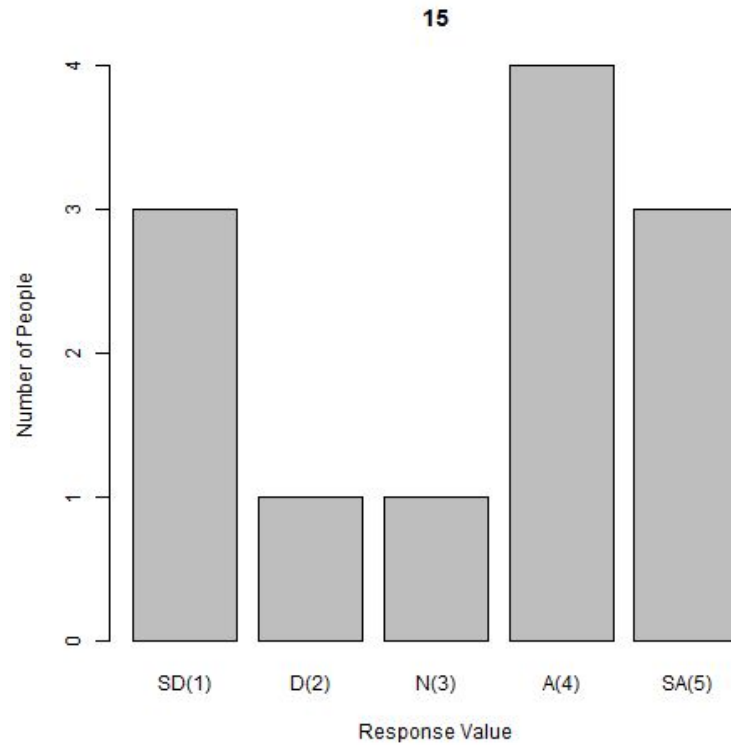
12. I would imagine that most people would adapt to this scheme very quickly.



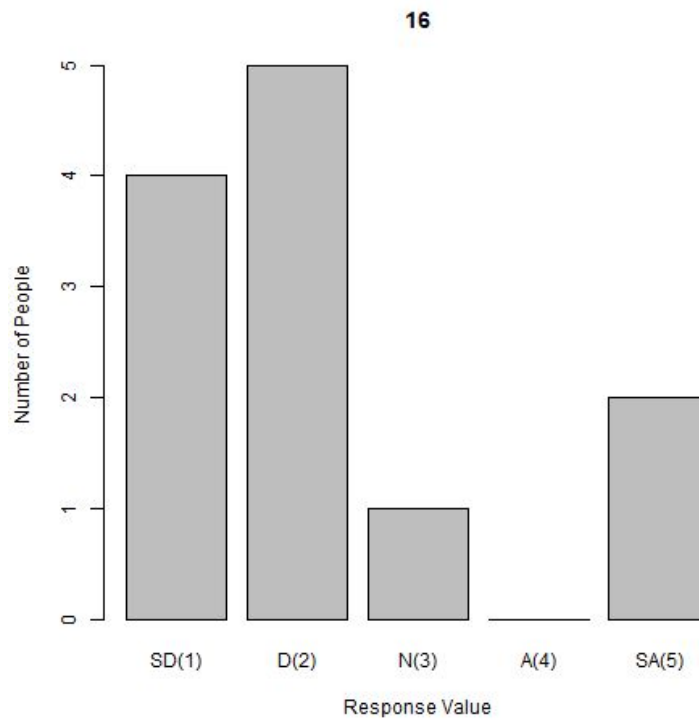
13. I would be worried about the security of my accounts using this scheme.



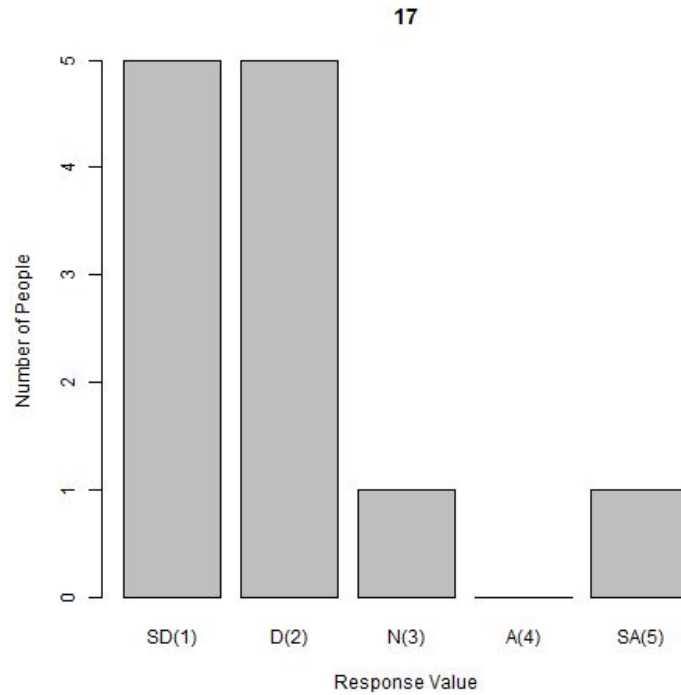
14. I feel safe with passwords generated using this scheme.



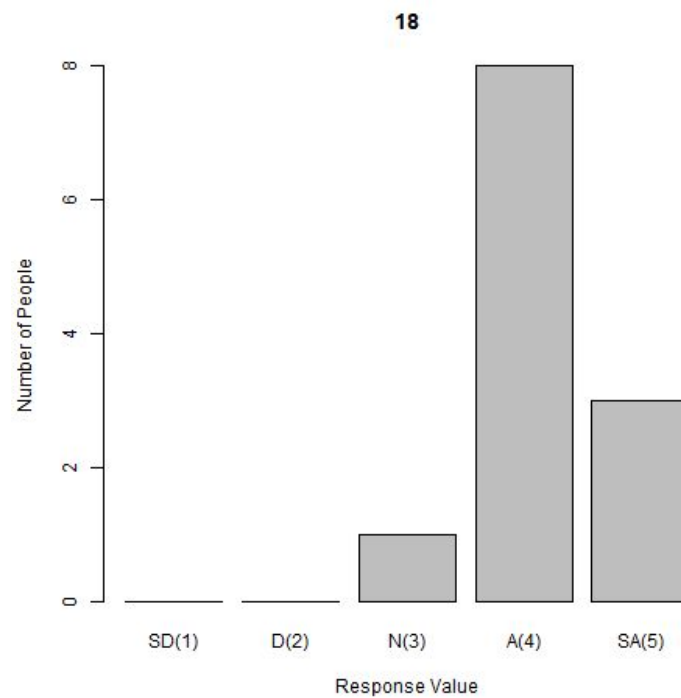
15. I found the password length to be too short.



16. I would prefer a scheme that relies more on visuals, like images.



17. I am more comfortable with other forms of input (e.g. clicking, drawing, etc.) than text input.



18. I find this password scheme works well overall.

Survey Results

From our survey it can be seen that people generally liked our scheme (1, 3, 5, 18). Most students found that the phonetic scheme helped them remember their passwords or were indifferent (2, 5). Unfortunately, users struggled to remember numbers in passwords more than we anticipated, meaning our attempt to add numbers to the end to improve the password space hindered memorability (9). However, people seemed to be uncomfortable with the idea of random passwords in general (10, 11). When asked, results were very mixed about users' opinions of the safety of our password scheme (13, 14). The important takeaway from this is randomly generated phonetic passwords can inspire confidence - assuming they are secure - but users doubt their usefulness and prefer to avoid random passwords if possible.

Results in Comparison to Text21

| Number of Logins (Histogram) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|--|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--------------|-------|-----|---|------|---|-------|---|-------|---|-------|---|-------|---|
| Phonetic.pw | Text21 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Number of Logins</p> <table border="1"> <caption>Data for Phonetic.pw Histogram</caption> <thead> <tr> <th>Logins</th> <th>Users</th> </tr> </thead> <tbody> <tr><td>3</td><td>3</td></tr> <tr><td>4</td><td>3</td></tr> <tr><td>5</td><td>1</td></tr> <tr><td>6</td><td>2</td></tr> <tr><td>7</td><td>0</td></tr> <tr><td>8</td><td>0</td></tr> <tr><td>9</td><td>3</td></tr> </tbody> </table> | Logins | Users | 3 | 3 | 4 | 3 | 5 | 1 | 6 | 2 | 7 | 0 | 8 | 0 | 9 | 3 | <p>Number of Logins</p> <table border="1"> <caption>Data for Text21 Histogram</caption> <thead> <tr> <th>Logins Range</th> <th>Users</th> </tr> </thead> <tbody> <tr><td>0-5</td><td>1</td></tr> <tr><td>5-10</td><td>1</td></tr> <tr><td>10-15</td><td>6</td></tr> <tr><td>15-20</td><td>6</td></tr> <tr><td>20-25</td><td>3</td></tr> <tr><td>25-30</td><td>1</td></tr> </tbody> </table> | Logins Range | Users | 0-5 | 1 | 5-10 | 1 | 10-15 | 6 | 15-20 | 6 | 20-25 | 3 | 25-30 | 1 |
| Logins | Users | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Logins Range | Users | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0-5 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5-10 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10-15 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 15-20 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20-25 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 25-30 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Mean: 5.916667 Median: 5.5 Standard Deviation: 2.314316</p> <p>There were 12 total users of this system who logged in a total of 71 times.</p> | <p>Mean: 16.61111 Median: 16 Standard Deviation: 4.900647</p> <p>There were 19 total users of this system who logged in a total of 299 times.</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>t = -8.0145 df = 25.812 p-value = 1.802e-08 alternative hypothesis: true difference in means is not equal to 0</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

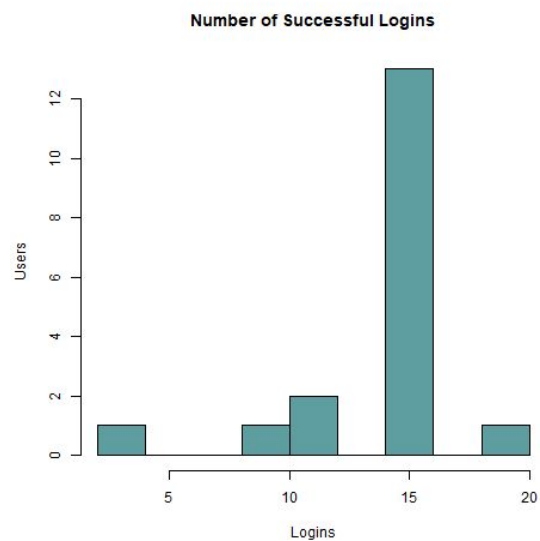
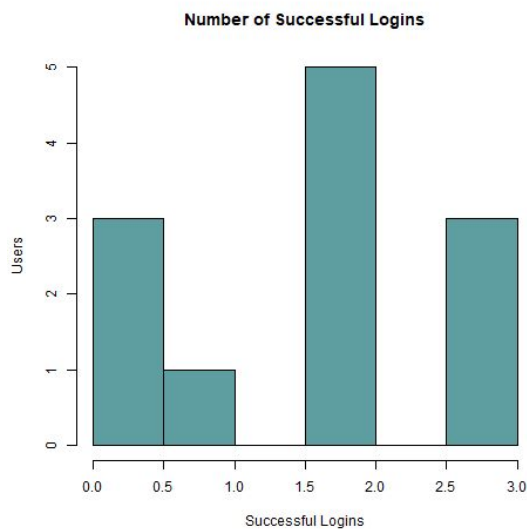
95 percent confidence interval:

-13.438281 -7.950608

decision: reject the null hypothesis

Though the averages of the total attempted logins per user between the two schemes differ greatly, this statistic is suspect as the context of the testing was different. Users for Phonetic.pw were only given the opportunity to log in through the testing framework, where they were only given 3 total attempts for each of 3 different accounts. Therefore, users could only have a maximum of 9 login attempts, while users of Text21 had no limit.

Number of Successful Logins (Histogram)



Mean: 1.666667

Median: 2

Standard Deviation: 1.154701

Users successfully logged in 20 times. This means they failed substantially more than they succeeded, which is not promising for this scheme.

Mean: 14.05556

Median: 15

Standard Deviation: 3.438061

Text21 users successfully logged in 253 times

$t = -14.139$

$df = 22.255$

$p\text{-value} = 1.335e-12$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

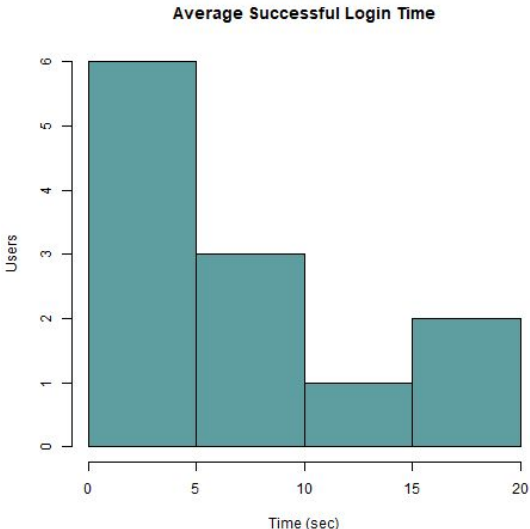
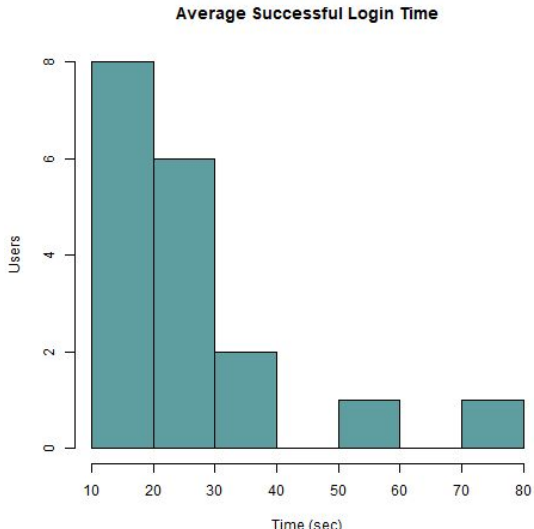
-14.20489 -10.57289

decision: reject the null hypothesis

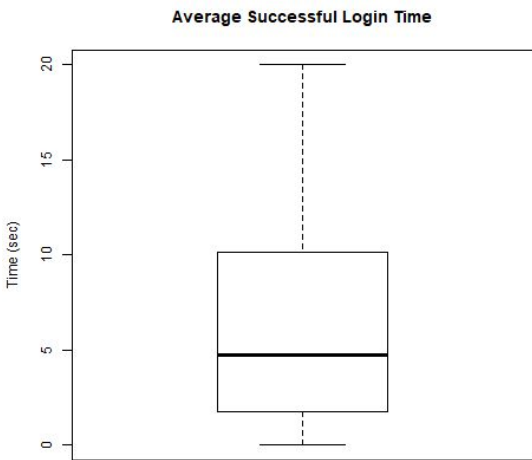
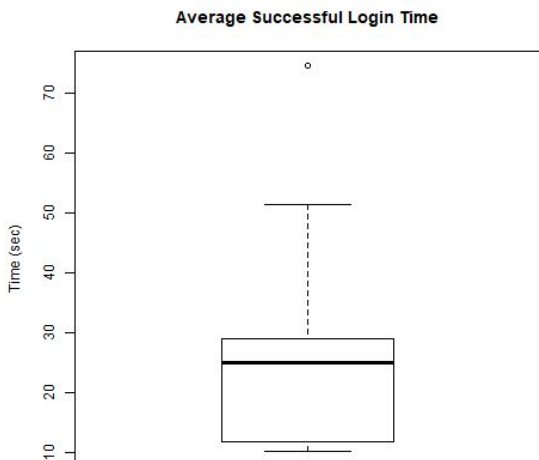
Phonetic.pw users had a 28.2% success rate, while Text21 users had an 84.6% success rate.

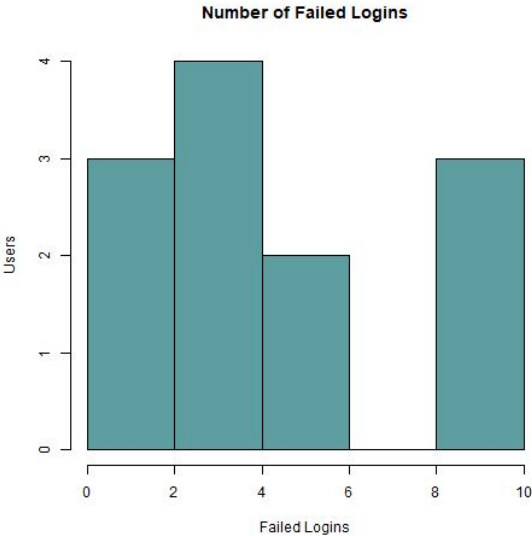
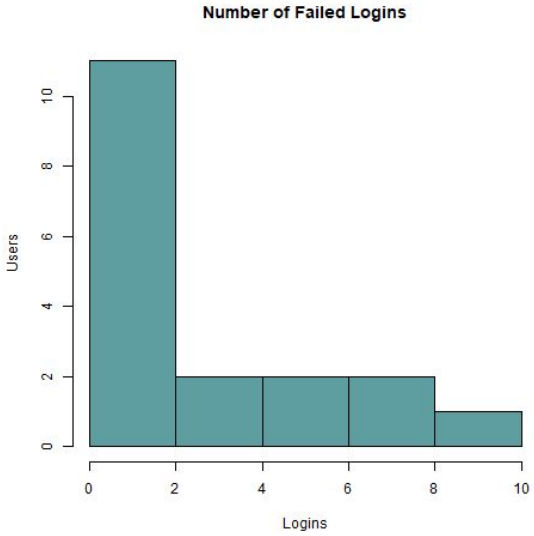
As was the case with the total number of logins, the number of successful logins is difficult to use in our comparison, as the Phonetic.pw users had limited attempts where Text21 users did not. The conditions of both studies were different as well, which could have affected success rates.

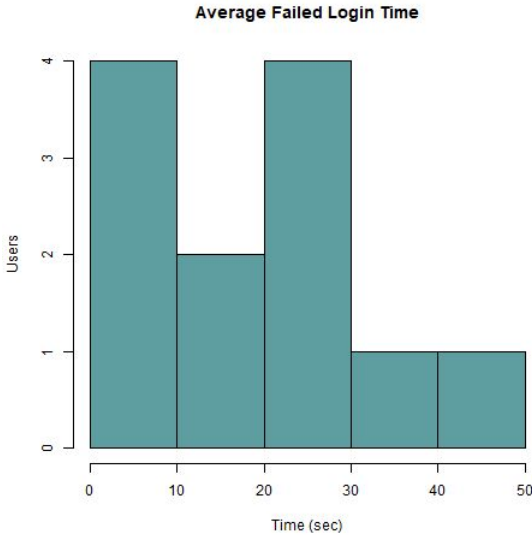
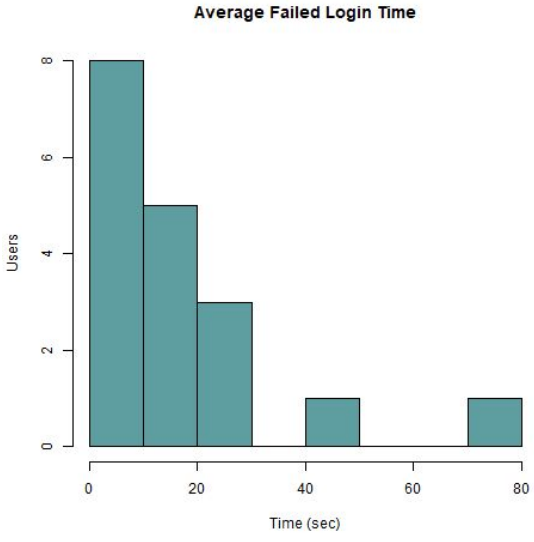
Average Successful Login Time (Histogram)

| <p style="text-align: center;">Average Successful Login Time</p>  <table border="1"> <caption>Data for Phonetic.pw Histogram</caption> <thead> <tr> <th>Time (sec)</th> <th>Users</th> </tr> </thead> <tbody> <tr><td>0-5</td><td>6</td></tr> <tr><td>5-10</td><td>3</td></tr> <tr><td>10-15</td><td>1</td></tr> <tr><td>15-20</td><td>2</td></tr> </tbody> </table> | Time (sec) | Users | 0-5 | 6 | 5-10 | 3 | 10-15 | 1 | 15-20 | 2 | <p style="text-align: center;">Average Successful Login Time</p>  <table border="1"> <caption>Data for Text21 Histogram</caption> <thead> <tr> <th>Time (sec)</th> <th>Users</th> </tr> </thead> <tbody> <tr><td>10-20</td><td>8</td></tr> <tr><td>20-30</td><td>6</td></tr> <tr><td>30-40</td><td>2</td></tr> <tr><td>40-50</td><td>0</td></tr> <tr><td>50-60</td><td>1</td></tr> <tr><td>60-70</td><td>0</td></tr> <tr><td>70-80</td><td>1</td></tr> </tbody> </table> | Time (sec) | Users | 10-20 | 8 | 20-30 | 6 | 30-40 | 2 | 40-50 | 0 | 50-60 | 1 | 60-70 | 0 | 70-80 | 1 |
|---|---|-------|-----|---|------|---|-------|---|-------|---|--|------------|-------|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|
| Time (sec) | Users | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0-5 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5-10 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10-15 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 15-20 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Time (sec) | Users | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10-20 | 8 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20-30 | 6 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 30-40 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 40-50 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 50-60 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 60-70 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 70-80 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Mean: 6.805556 Median: 4.75 Standard Deviation: 6.836043</p> <p>All users successfully logged in under 20 seconds. This is promising, as it shows that our scheme promotes speedy recall.</p> | <p>Mean: 25.57956 Median: 25.14167 Standard Deviation: 16.23751</p> <p>Most users took between 10 and 40 seconds to enter their password using Text21.</p> | | | | | | | | | | | | | | | | | | | | | | | | | | |

Average Successful Login Time (Boxplot)

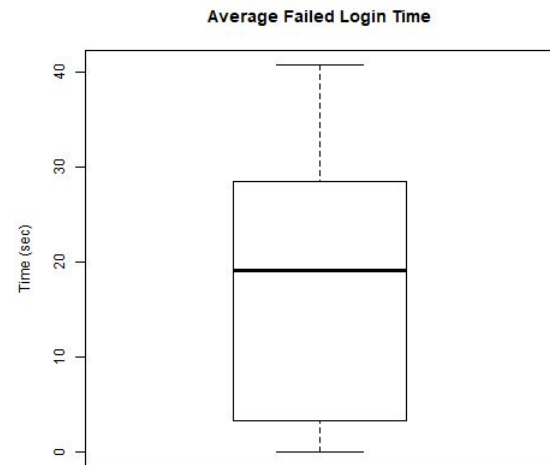
| | |
|---|--|
|  |  |
| <p>Most users fell in the third quartile, which is less than 10 seconds and more than 5 seconds. This is an acceptable login time.</p> | <p>For Text21, most users fell within the first quartile, which was less than 25 seconds. There was one outlier who took significantly more time</p> |
| <p>$t = -4.3599$ $df = 24.558$ $p\text{-value} = 0.0002025$ alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -27.650528 -9.897473 decision: reject the null hypothesis</p> <p>Looking at the average successful login times for both schemes, it can be seen that Phonetic.pw users were able to login much faster than those using Text21 on average. This indicates a faster recall time, possibly because a phonetic password makes it easier to remember than a random one. The difference between mean, median, and standard deviation for the Phonetic.pw data were also significantly lower. There are fewer outliers in the data, showing a higher level of consistency among the users, though this could be because of a limited sample space.</p> | |
| <p>Number of Failed Logins (Histogram)</p> | |

|  <table border="1"><caption>Data for Text21 Histogram</caption><thead><tr><th>Failed Logins Range</th><th>Number of Users</th></tr></thead><tbody><tr><td>0-1</td><td>3</td></tr><tr><td>1-2</td><td>4</td></tr><tr><td>2-3</td><td>2</td></tr><tr><td>3-4</td><td>0</td></tr><tr><td>4-5</td><td>0</td></tr><tr><td>5-6</td><td>0</td></tr><tr><td>6-7</td><td>0</td></tr><tr><td>7-8</td><td>3</td></tr><tr><td>8-9</td><td>0</td></tr><tr><td>9-10</td><td>0</td></tr></tbody></table> | Failed Logins Range | Number of Users | 0-1 | 3 | 1-2 | 4 | 2-3 | 2 | 3-4 | 0 | 4-5 | 0 | 5-6 | 0 | 6-7 | 0 | 7-8 | 3 | 8-9 | 0 | 9-10 | 0 |  <table border="1"><caption>Data for Text21 Histogram</caption><thead><tr><th>Logins Range</th><th>Number of Users</th></tr></thead><tbody><tr><td>0-1</td><td>10</td></tr><tr><td>1-2</td><td>2</td></tr><tr><td>2-3</td><td>2</td></tr><tr><td>3-4</td><td>2</td></tr><tr><td>4-5</td><td>2</td></tr><tr><td>5-6</td><td>2</td></tr><tr><td>6-7</td><td>2</td></tr><tr><td>7-8</td><td>2</td></tr><tr><td>8-9</td><td>1</td></tr><tr><td>9-10</td><td>0</td></tr></tbody></table> | Logins Range | Number of Users | 0-1 | 10 | 1-2 | 2 | 2-3 | 2 | 3-4 | 2 | 4-5 | 2 | 5-6 | 2 | 6-7 | 2 | 7-8 | 2 | 8-9 | 1 | 9-10 | 0 |
|--|--|-----------------|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|------|---|---|--------------|-----------------|-----|----|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|------|---|
| Failed Logins Range | Number of Users | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0-1 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1-2 | 4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2-3 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3-4 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4-5 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5-6 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6-7 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7-8 | 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8-9 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9-10 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Logins Range | Number of Users | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0-1 | 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1-2 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2-3 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3-4 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4-5 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5-6 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6-7 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7-8 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8-9 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9-10 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Mean: 4.25 Median: 3.5 Standard Deviation: 3.441062</p> <p>Users failed to login 51 times. It was observed during demos that the number at the end of the word gave users the most trouble. As with Text21, most users failed quickly, which is promising since they did not spend too much time trying to remember their password. Most of these failures were concentrated in three users, which suggests that the majority of users did not have many problems.</p> | <p>Mean: 2.555556 Median: 1 Standard Deviation: 3.329409</p> <p>Most users failed Text21 0 to 2 times. The rest were spread evenly between 2 to 10 times.</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>t = 1.3385 df = 23.174 p-value = 0.1937 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.9232384 4.3121273 decision: reject the null hypothesis</p> <p>We could not discuss much due to the difference in sample size.</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <p>Average Failed Login Time (Histogram)</p> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

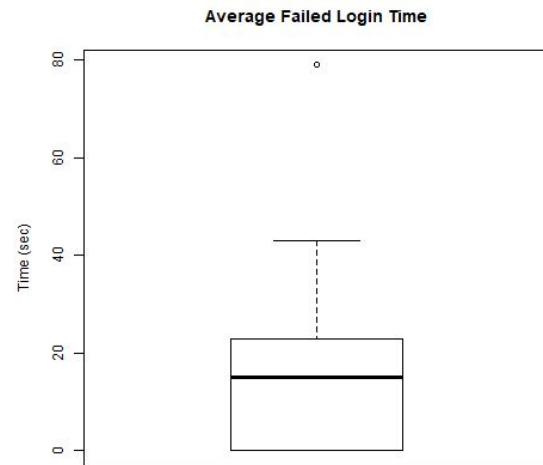
| | |
|--|--|
|  <p>Average Failed Login Time</p> <p>Users</p> <p>Time (sec)</p> |  <p>Average Failed Login Time</p> <p>Users</p> <p>Time (sec)</p> |
| <p>Mean: 17.99907 Median: 19.16667 Standard Deviation: 14.11323</p> <p>Most users tended to not fail quickly, taking longer than 10 seconds. This could indicate that they almost remembered their passwords but could not quite, or it could be that university students dislike failure and tried very hard to remember their passwords. This is not consistent with Text21 failure times, which suggests that the context of each study could have affected results.</p> <p>This could indicate that users were close to remembering their password, but either failed to differentiate which belongs to which account, or could not remember a portion of a password, like the number at the end.</p> | <p>Mean: 15.39544 Median: 15.02083 Standard Deviation: 20.20553</p> <p>When people failed to enter their password correctly with Text21 the failure generally occurred very quickly, the most common longest times being between 20 and 30 seconds.</p> |
| <p>t = 0.41543 df = 27.897 p-value = 0.681 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -10.23662 15.44390 decision: accept the null hypothesis. Mean failure time is similar.</p> | |

As was the case with the total number of logins, the number of failed logins is difficult to use in our comparison, as the Phonetic.pw users had limited attempts where Text21 users did not. The conditions of both studies were different as well, which could have affected failure rates.

Average Failed Login Time (Boxplot)



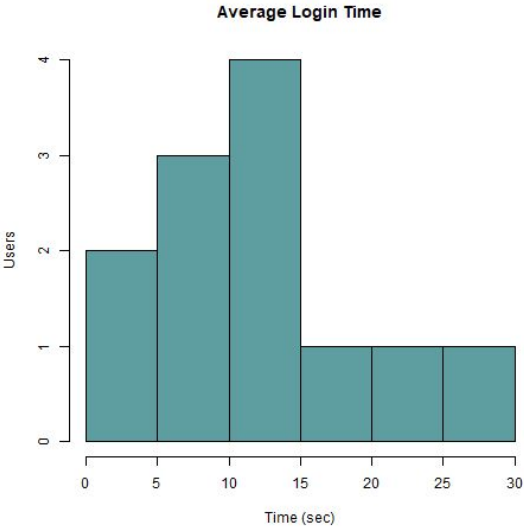
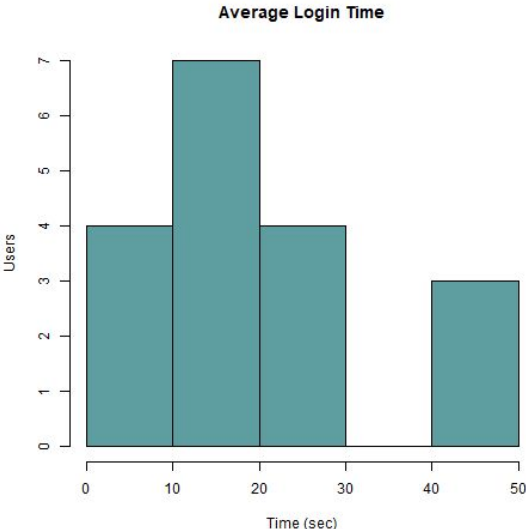
Most users fell within the first quartile, taking between 5 and 18 seconds to fail.

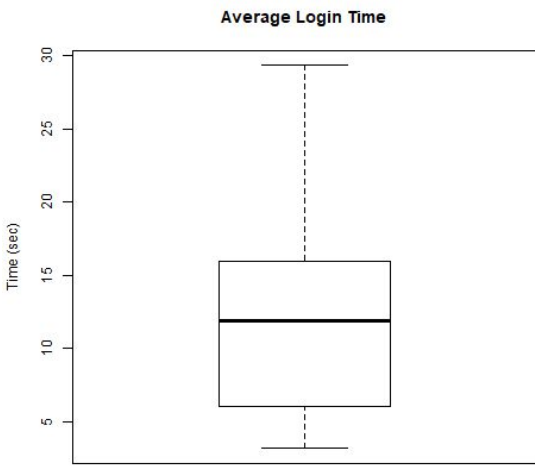
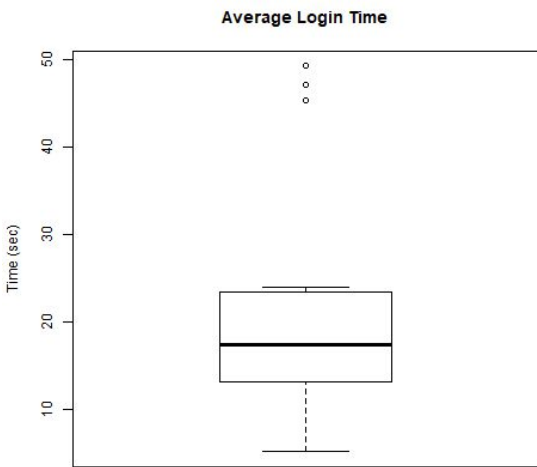


The average fail time for Text21 is concentrated in the first quartile, meaning users tend to enter their password quickly and then fail.

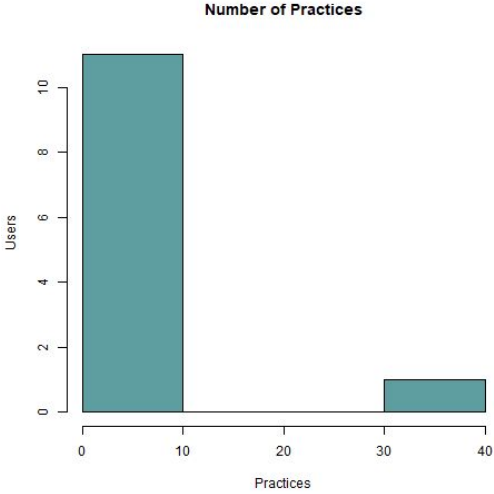
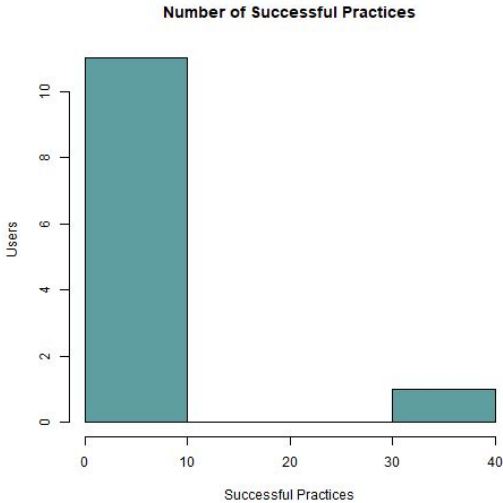
Looking at the average failed login times for both schemes, it can be seen that users experienced relatively similar failure times. However, failure times for Phonetic.pw were slightly higher on average. We can also see that the histogram is bimodal, whereas the histogram for Text21 is unimodal. This indicates that the data for Phonetic.pw is more widely distributed. Similar to the average successful login times, the standard deviation for Phonetic.pw was lower and there were fewer outliers, meaning more consistency among users.

Average Login Time (Histogram)

| | |
|---|---|
|  <p>Average Login Time</p> <p>Time (sec)</p> <p>Users</p> |  <p>Average Login Time</p> <p>Time (sec)</p> <p>Users</p> |
| <p>Mean: 12.40231 Median: 11.875 Standard Deviation: 8.035262</p> <p>This is the average login time of participants attempting the test portion of our program, where they had to remember the passwords with no hint option.</p> <p>Most users took less than 15 seconds to enter their passwords.</p> | <p>Mean: 20.4875 Median: 17.46875 Standard Deviation: 13.63761</p> <p>The majority of users took between 0 and 30 seconds to enter their passwords using Text21.</p> |
| <p>t = -2.0397 df = 27.705 p-value = 0.05102 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -16.20887425 0.03851084 decision: reject the null hypothesis. Recall speed for a phonetic password is initially faster than for random characters because remembering a pattern is easier than remembering randomness.</p> <p>On average, users tended to login faster using the Phonetic.pw scheme as opposed to the Text21 scheme. Most logins occurred within 15 seconds for the Phonetic.pw scheme, whereas with the Text21 scheme most logins took longer than 20 seconds.</p> | |
| <p>Average Login Time (Boxplot)</p> | |

| | |
|--|---|
|  |  |
| <p>Most users fell within the first quartile, taking less than 12 seconds to login. This is acceptable, but there were still a substantial number of users who took longer than 12 seconds, which is longer than desired when entering a password. These users struggled to either differentiate which password belongs to which account, or to remember what the password is.</p> | <p>The majority of users who used Text21 fell in the third quartile, with a median of 17.5. This is not a sizable majority compared to the first quartile. This means users were reasonably well distributed in terms of password entry times, aside from three outliers.</p> |
| <p>Looking at the average login times for both schemes, it can be seen that Phonetic.pw users successfully logged in or failed all of their attempts more quickly than Text21 users on average. We can also see that the distributions of both histograms are relatively similar, so comparing means is fairly reliable.</p> | |

Graphs Unique to the Phonetic.pw Scheme Analysis

| Number of Practices (Histogram) | | | | | | | |
|---|-------|----------------------------|-------|------|----|-------|---|
|  <table border="1"><caption>Data for Number of Practices Histogram</caption><thead><tr><th>Practices Range</th><th>Users</th></tr></thead><tbody><tr><td>0-10</td><td>83</td></tr><tr><td>30-40</td><td>1</td></tr></tbody></table> | | Practices Range | Users | 0-10 | 83 | 30-40 | 1 |
| Practices Range | Users | | | | | | |
| 0-10 | 83 | | | | | | |
| 30-40 | 1 | | | | | | |
| <p>Mean: 6.916667 Median: 4.5 Standard Deviation: 9.931203</p> <p>Most users practiced their passwords between 0 and 10 times, for a total of 83 practices, though 37 of these were performed by one user. It is likely that more practice would have yielded more success from participants, but the context of the demo may not have allowed this.</p> | | | | | | | |
| Number of Successful Practices (Histogram) | | | | | | | |
|  <table border="1"><caption>Data for Number of Successful Practices Histogram</caption><thead><tr><th>Successful Practices Range</th><th>Users</th></tr></thead><tbody><tr><td>0-10</td><td>83</td></tr><tr><td>30-40</td><td>1</td></tr></tbody></table> | | Successful Practices Range | Users | 0-10 | 83 | 30-40 | 1 |
| Successful Practices Range | Users | | | | | | |
| 0-10 | 83 | | | | | | |
| 30-40 | 1 | | | | | | |
| <p>Mean: 5.916667</p> | | | | | | | |

Median: 3.5

Standard Deviation: 9.680893

Most users were successful in their practices.

Most users completed between 0 and 10 practices for all 3 accounts. This shows that users felt confident rather quickly with their passwords.

Number of Failed Practices (Histogram)



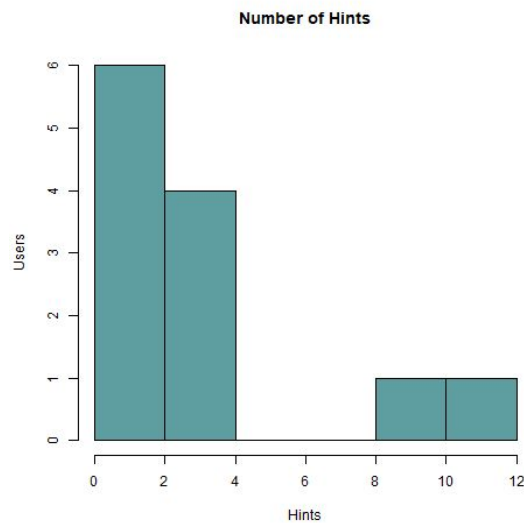
Mean: 1

Median: 0.5

Standard Deviation: 1.3484

The maximum number of practices failed was 4. This means most people were able to successfully remember their password in practice. Most users failed only zero to one practice attempt.

Number of Hints (Histogram)



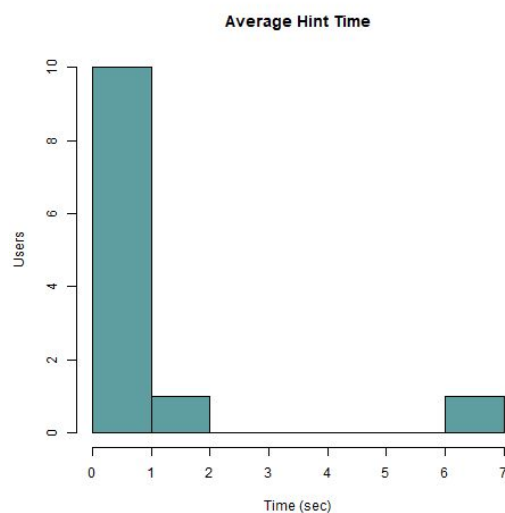
Mean: 3.083333

Median: 2.5

Standard Deviation: 3.872005

Most people used very few hints during practice, which, combined with the low number of fails, means most people were able to remember their passwords during practice or they felt confident in their ability to remember it. This shows that our passwords have good short-term memorability

Average Hint Time (Histogram)



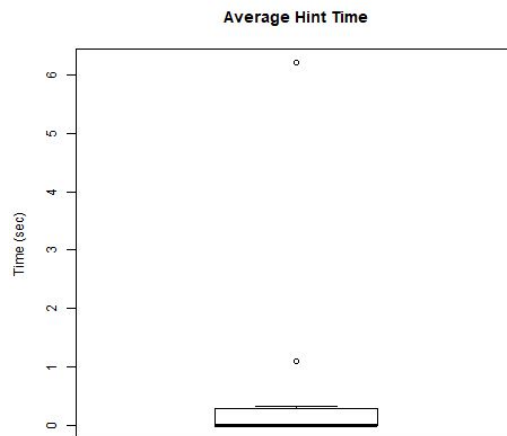
Mean: 0.6717593

Median: 0

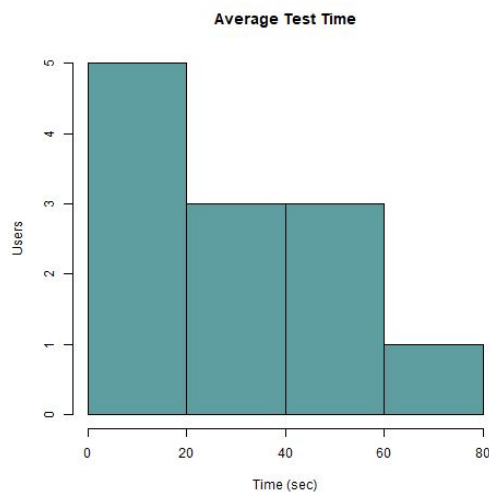
Standard Deviation: 1.769973

These results suggest that many users made use of our hint feature.

Average Hint Time (Boxplot)



Average Test Time (Histogram)



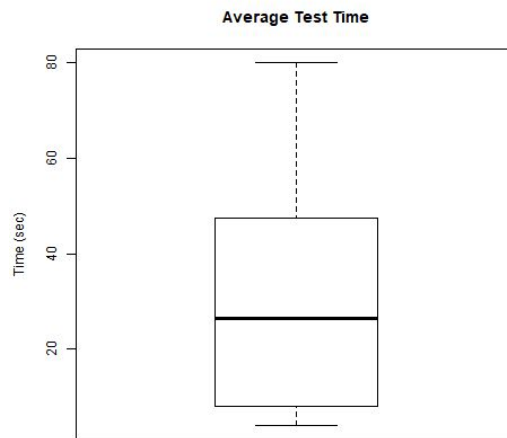
Mean: 29.66667

Median: 26.5

Standard Deviation: 24.03154

A perfect test would consist of three password entries. This means most users took about 10 seconds per password entry assuming they succeeded each time. These results could be skewed by users who gave up and quickly entered false values to exit the test. The bulk of the users completed the tests within 60 seconds.

Average Test Time (Boxplot)



Most students fell within the third quartile or beyond, which debunks the theory that many users could have given up quickly. This supports the theory that participants were trying hard to remember their passwords during the test, which reveals possible problems. Users may have had difficulty differentiating which passwords belonged to which account, and thus were trying to determine which one was correct, or simply struggled to remember a password for the account overall. Ideally, the scheme would allow for hasty recall with little to no mistakes.

Conclusion

While there are several comparisons from which we can find useful inferences, there are some that, while relevant to the overall comparison, can be considered suspect because of different testing contexts. In particular, all of the login count comparisons (total logins, successful logins, and failed logins) should be taken with a grain of salt as Text21 users were not confined to a scheme testing session and were able to attempt logging in as many times as they needed. Phonetic.pw users were only given a maximum of 9 attempts, spread over 3 accounts (3 attempts per account), so comparing the average number of logins is not valid in this case. Additionally, many of the histograms for Text21 contained outliers that should normally be of concern, however given that the sample of participants for Phonetic.pw was heavily limited (mostly Carleton computer science students), it is expected that there would be significantly fewer outliers. We have therefore dismissed outliers and focused on comparing means.

Through the analysis of average login times it became clear that our scheme performed better. Both successful and total login time averages for Phonetic.pw users were lower than those of Text21. This implies that the use of Phonetic.pw facilitated overall password recall speed. On the other hand, their average failed login times were higher, meaning they spent more time trying to recall it. This could mean that users felt they almost remembered the password, and wanted to dedicate the extra time to trying to remembering it. The higher fail

rate was likely caused by the protracted nature of our study. A format that promoted extended use would likely have yielded more promising results.

Considering user satisfaction, as was seen in our survey, most users dislike and distrust randomly chosen passwords. To solve this, we propose adding a “change” option if users do not like their randomly chosen password. A potential downside of this is users could consistently click change until they find a password they enjoy, which could potentially reduce the number of accepted passwords, thus weakening our scheme.

In comparing the results of our tests and the results of the survey, we have concluded that our scheme is better than Text21. However, we were not able to achieve a depth of analysis that we were completely satisfied with, given the difference in contexts of studies of the two schemes. This limited the data we could use for the comparison, as well as what inferential statistics were applicable. Part of the problem was due to the environment in which the study took place. The classroom was loud with other students testing their own software, and there was a lot of pressure on students to perform quickly. As a result, our scheme’s high fail rate is likely a product of the test setting. If we were to redo or continue this study, we would like to ensure that testing conditions are identical for both of the schemes to ensure consistency. We would also like to use more participants with the hope of broadening the sample scope would even the distribution of the data.

If we were to make changes to our scheme, we think one of its biggest flaws is its limited password space. With five letters being what we felt was the optimal length to make up a memorable “phonetic word”, we had to add something that raised the password complexity. However, we felt the addition of the one-digit number to the end, though it raised the complexity, hindered memorability. This is supported by the results of our survey. In future studies we would alter the scheme so that it reaches an acceptable password space without including numbers.

Appendix

Appendix A: Lookup table used in phonetic password scheme

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|----|----|----|----|----|---|----|----|
| b | a | a | a | -a | a | a | a | -a | a | a | a | a | a | -a | a | | a | a | a | -a | a | a | | a | a |
| -c | -c | -b | -b | b | -b | -b | -b | b | -b | -b | -b | -b | -b | b | -b | | -b | -b | -b | b | -b | -b | | -b | -b |
| d | -d | -d | -c | -c | -c | -c | -c | -c | -c | -c | -c | c | -c | -c | -c | | -c | -c | -c | -c | -c | -c | | -c | -c |
| -e | e | e | e | d | -d | -d | -d | d | -d | -d | -d | -c | -d | d | -d | | -d | -d | -d | d | -d | -d | | -d | -d |
| f | -f | -f | -f | f | e | e | e | -e | e | e | e | e | e | -e | e | | e | e | e | -e | e | e | | -e | e |
| g | -g | -g | -g | g | -g | -f | -f | f | -f | -f | -f | -f | -f | f | -f | | -f | -f | -f | f | -f | -f | | -f | -f |
| h | -h | h | -h | h | -h | -h | -g | g | -g | -g | -g | -g | -g | g | -g | | -g | -g | -g | g | -g | -g | | -g | -g |
| i | i | i | i | -i | i | i | i | -h | -h | -h | -h | -h | -h | -h | h | | -h | h | h | h | -h | -h | | -h | -h |
| j | -j | -j | -j | j | -j | -j | -j | j | i | i | i | i | i | -i | i | | i | i | i | -i | i | i | | i | i |
| k | -k | -k | -k | k | -k | k | -k | k | -k | -j | -j | -j | -j | j | -j | | -j | -j | -j | j | -j | -j | | -j | -j |
| l | -l | l | -l | l | l | l | -l | l | -l | l | -k | -k | -k | k | -k | | -k | k | k | k | -k | -k | | -k | -k |
| m | -m | -m | -m | m | -m | -m | -m | m | -m | -m | -m | -l | -l | l | l | | -l | l | l | l | -l | -l | | -l | -l |
| n | -n | -n | -n | n | -n | -n | -n | n | -n | -n | -m | -n | -m | m | -m | | -m | -m | -m | m | -m | -m | | -m | -m |
| -o | o | o | o | -o | o | o | o | -o | o | o | o | o | o | n | -n | | -n | n | -n | n | -n | -n | | -n | -n |
| p | -p | -p | -p | p | -p | -p | -p | p | -p | -p | -p | -p | -p | p | o | | o | o | o | -o | o | o | | o | o |
| -q | -q | -q | -q | -q | -q | -q | -q | -q | -q | -q | -q | -q | -q | -q | -q | | -p | p | -p | p | -p | -p | | -p | -p |
| r | r | r | r | r | r | r | -r | r | -r | r | -r | r | r | r | r | | -q | -q | -q | -q | -q | -q | | -q | -q |
| s | -s | -s | -s | s | -s | -s | -s | s | -s | -s | -s | -s | -s | s | -s | | -s | r | r | r | -r | -r | | -r | -r |
| t | -t | -t | -t | t | -t | -t | -t | t | -t | -t | -t | -t | -t | t | -t | | -t | t | s | s | -s | -s | | -s | -s |
| -u | u | u | u | -u | u | u | u | -u | u | u | u | u | u | -u | u | | u | u | u | t | -t | -t | | -t | -t |
| v | -v | -v | -v | v | -v | -v | -v | v | -v | -v | -v | -v | -v | v | -v | | -v | -v | -v | v | u | u | | -u | u |
| w | -w | -w | -w | w | -w | -w | -w | -w | -w | -w | -w | -w | -w | w | -w | | -w | w | -w | -w | -w | -v | | -v | -v |
| -x | -x | -x | -x | -x | -x | -x | -x | -x | -x | -x | -x | -x | -x | -x | -x | | -x | -x | -x | -x | -x | -x | | -w | -w |
| y | -y | -y | -y | y | -y | -y | -y | -y | -y | -y | -y | -y | -y | -y | -y | | -y | -y | -y | -y | -y | -y | | -x | -x |
| z | -z | -z | -z | z | -z | -z | -z | -z | -z | -z | -z | -z | -z | z | -z | | -z | -z | -z | z | -z | -z | | -z | -y |

Phonetic.pw lookup table

| | |
|--------------------------------------|--|
| Available, preceding letter (parent) | Each letter is generated based on the letter that precedes it in order to make a phoneme. The process repeats until the desired word length is reached. Consecutive letters are not allowed. |
| Available, succeeding letter (child) | |
| Unavailable letter | |

Appendix B: Consent forms and participant signatures

Omitted.