

Clemson University TigerPrints

All Theses

Theses

12-2011

Evaluating the Usability of System-Generated and User-Generated Passwords of Approximately Minimum Equal Security

Sourav Bhuyan

Clemson University, sbhuyan@g.clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

 Part of the [Engineering Commons](#)

Recommended Citation

Bhuyan, Sourav, "Evaluating the Usability of System-Generated and User-Generated Passwords of Approximately Minimum Equal Security" (2011). *All Theses*. 1267.

https://tigerprints.clemson.edu/all_theses/1267

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

EVALUATING THE USABILITY OF SYSTEM-GENERATED AND USER-
GENERATED PASSWORDS OF APPROXIMATELY MINIMUM EQUAL SECURITY

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Industrial Engineering

by
Sourav Bhuyan
December 2011

Accepted by:
Dr. Joel S. Greenstein, Committee Chair
Dr. Anand K. Gramopadhye
Dr. Byung Rae Cho

ABSTRACT

System-generated or user-generated text-based passwords are commonly used by the users to authenticate access to their electronic assets. These passwords may vary in usability and memorability depending on the type of password generation, composition and length. However, little past research has compared usability and memorability of passwords, satisfying minimum entropy for a secure password. This study compared three password policy conditions, assigning/generating passwords of approximately equal minimum security, i.e. 6-character alphanumeric system-generated passwords, minimum 8-character restricted user-generated passwords and minimum 16-character unrestricted user-generated passwords.

The study involved 54 participants, equally divided into three groups, 18 in each password policy condition. The study took place over two sessions, with a period of 5-7 days in between them. In the first session, depending on the password policy condition, the participants were either assigned or asked to create a password. The participants were then asked to recall their passwords in the same session and after 5-7 days in the second session. The three password policy conditions were compared with respect to the dependent variables-- the time taken to create the password account, the password creation error rates, the time taken to recall and recall error rates for both sessions, the number of unrecoverable passwords in the second session, the proximity of the recalled password to the stored password measured by Damerau-Levenshtein and Jaro-Winkler edit distances, and the subjective ratings for the NASA task load indices and the System Usability Scale questionnaire.

There was significant difference between the password policy condition for the time taken to create a password account, password creation error rates, time taken to recall the passwords and temporal demand index of the NASA-TLX questionnaire. Across the task sessions, there were statistically significant differences for time taken to recall system-generated passwords, recall error rates, performance index of the NASA-TLX questionnaire and the SUS score. There was no significant difference for recall error rates and unrecoverable passwords among password policy conditions.

The results of this study suggest that the overall performance of the 8-character password was weaker compared to system-generated and 16-character passwords. The qualitative analysis of the comments made by the participants and the additional analysis of the user-generated passwords suggests that the participants showed bias towards the commonly used 8-character password policy condition. However, this bias did not translate into better memorability of the 8-character password. The performance and the positive trends exhibited by 16-character passwords indicate a potential area for the password application designers to explore.

DEDICATION

This thesis is dedicated to my loving parents, my father Pranab Kumar Bhuyan and my mother Anita Bhuyan, and to my sister Shraddha Bhuyan and my good friends.

ACKNOWLEDGMENTS

I am immensely thankful and express my gratitude to my advisor, Dr. Joel S. Greenstein. His valuable advice, insight and continuous efforts to make this process simple and easy helped me tremendously during my research. I would also like to thank my committee members, Dr. Anand K. Gramopadhye and Dr. Byung Rae Cho, for their invaluable recommendations and suggestions during the course of my research. I am thankful to Dr. DeWayne Moore for his valuable guidance on statistical data analysis. I am extremely grateful to Ms. Barbara Ramirez for her continuous guidance in documenting my research with the correct emphasis and structure.

I would like to thank my colleagues who are or were part of the Human Computer System Lab, Kapil Chalil Madathil, Kevin Juang, Sanjaykumar Ranganayakulu, Meera Ramachandran, Vikas Vadlapatla and Rachana Ranade, for their help and support during various stages of my research. Last but not the least, I would like to thank my family and my good friends who helped me in their own ways and constantly supported me during my study at Clemson University.

TABLE OF CONTENTS

| | Page |
|---|------|
| TITLE PAGE | i |
| ABSTRACT..... | ii |
| DEDICATION | iv |
| ACKNOWLEDGMENTS | v |
| LIST OF TABLES | viii |
| LIST OF FIGURES | xi |
| CHAPTER | |
| I. INTRODUCTION | 1 |
| II. LITERATURE REVIEW | 4 |
| III. RESEARCH HYPOTHESES | 18 |
| IV. METHOD | 21 |
| Participants..... | 21 |
| Testing environment | 22 |
| Experimental design..... | 23 |
| Tasks | 27 |
| Procedure | 36 |
| V. RESULTS | 40 |
| Objective measures | 41 |
| Time taken to create password account | 41 |
| Password creation error rates | 44 |
| Time taken to recall passwords..... | 47 |
| Recall error rates | 53 |
| Unrecoverable passwords | 57 |
| Edit distances | 58 |

| | |
|---|-----|
| Damerau-Levenshtein edit distances | 58 |
| Jaro-Winkler proximities | 59 |
| Subjective measures..... | 59 |
| NASA task load indices | 59 |
| Mental demand..... | 61 |
| Physical demand | 64 |
| Temporal demand | 67 |
| Performance | 71 |
| Effort | 74 |
| Frustration | 77 |
| System Usability Scale questionnaire | 80 |
| VI. DISCUSSION | 87 |
| Password account creation | 89 |
| Recall task in first session..... | 91 |
| Recall task in second session | 92 |
| Across task sessions | 93 |
| Qualitative analysis of participants comments | 94 |
| Analysis of user-generated password..... | 97 |
| VII. CONCLUSION..... | 99 |
| APPENDICES | 102 |
| A. Consent form for study participants..... | 102 |
| B. Pre-test questionnaire for study participants..... | 104 |
| C. Methodologies for remembering passwords | 105 |
| D. System Usability Scale questionnaire | 106 |
| E. NASA-TLX questionnaire | 108 |
| F. Consent form for pilot study participants | 109 |
| G. Preference ranking questionnaire..... | 111 |
| REFERENCES | 112 |

LIST OF TABLES

| Table | Page |
|---|------|
| 4.1 Three levels of password composition..... | 24 |
| 4.2 3x3 factorial design..... | 25 |
| 4.3 3x2 factorial design..... | 25 |
| 5.1 Descriptive statistics for the transposed value of password account creation time..... | 42 |
| 5.2 One-way ANOVA data for the transposed value of account creation time..... | 43 |
| 5.3 Descriptive statistics for error rates during password account creation | 45 |
| 5.4 One-way ANOVA data for error rates..... | 46 |
| 5.5 Descriptive statistics for the transposed value for the recall times for first session..... | 47 |
| 5.6 Descriptive statistics for the transposed value for the recall times for second session..... | 48 |
| 5.7 Two-way mixed ANOVA data for recall times | 49 |
| 5.8 Descriptive statistics for recall error rates for first session..... | 54 |
| 5.9 Descriptive statistics for recall error rates for second session | 54 |
| 5.10 Two-way mixed ANOVA data for recall error rates | 55 |
| 5.11 NASA-TLX rating scale definitions | 60 |
| 5.12 Descriptive statistics for mental demand during password creation..... | 61 |
| 5.13 Descriptive statistics for mental demand during recall in first session..... | 62 |
| 5.14 Descriptive statistics for mental demand during recall in second session | 62 |
| 5.15 Two-way mixed ANOVA data for mental demand..... | 63 |
| 5.16 Descriptive statistics for physical demand during password creation | 65 |

| | | |
|------|--|----|
| 5.17 | Descriptive statistics for physical demand during recall in first session | 65 |
| 5.18 | Descriptive statistics for physical demand during recall in second session..... | 66 |
| 5.19 | Two-way mixed ANOVA table for physical demand | 66 |
| 5.20 | Descriptive statistics for temporal demand during password creation | 68 |
| 5.21 | Descriptive statistics for temporal demand during recall in first session | 69 |
| 5.22 | Descriptive statistics for temporal demand during recall in second session..... | 69 |
| 5.23 | Two-way mixed ANOVA data for temporal demand..... | 70 |
| 5.24 | Descriptive statistics for performance during password creation | 71 |
| 5.25 | Descriptive statistics for performance during recall in first session | 72 |
| 5.26 | Descriptive statistics for performance during recall in second session | 72 |
| 5.27 | Two-way mixed ANOVA data for performance | 73 |
| 5.28 | Descriptive statistics for effort during password creation | 75 |
| 5.29 | Descriptive statistics for effort during recall in first session | 75 |
| 5.30 | Descriptive statistics for effort during recall in second session..... | 76 |
| 5.31 | Two-way mixed ANOVA data for effort..... | 76 |
| 5.32 | Descriptive statistics for frustration during password creation..... | 78 |
| 5.33 | Descriptive statistics for frustration during recall in first session..... | 78 |
| 5.34 | Descriptive statistics for frustration during recall in second session..... | 79 |
| 5.35 | Two-way mixed ANOVA data for frustration..... | 79 |
| 5.36 | Descriptive statistics of the SUS scores for password account creation..... | 81 |
| 5.37 | Descriptive statistics of the SUS scores for password recall in first session | 81 |
| 5.38 | Descriptive statistics of the SUS scores for password recall in second session | 82 |
| 5.39 | Two-way mixed ANOVA data for SUS score..... | 83 |
| 6.1 | Relative performance of policy conditions during password account creation | 87 |
| 6.2 | Relative performance of policy conditions during first session recall..... | 88 |

| | | |
|-----|--|----|
| 6.3 | Relative performance of policy conditions during second session recall | 89 |
| 6.4 | Categories of participant comments..... | 95 |
| 6.5 | 8-character passwords..... | 98 |
| 6.6 | 16-character passwords..... | 98 |

LIST OF FIGURES

| Figure | Page |
|---|------|
| 4.1 6-character alphanumeric password creation..... | 29 |
| 4.2 8-character password creation..... | 30 |
| 4.3 16-character password creation..... | 31 |
| 4.4 Response popup window for failed 6-character alphanumeric password creation..... | 32 |
| 4.5 Response popup window for failed 8-character password creation..... | 33 |
| 4.6 Response popup window for failed 16-character password creation..... | 34 |
| 4.7 Password recall pop-up window | 35 |
| 4.8 Failed password recall attempt..... | 36 |
| 4.9 Procedure flow for first and second session | 39 |
| 5.1 Mean password account creation time (seconds)..... | 43 |
| 5.2 Reflected transposed values for password account creation time..... | 44 |
| 5.3 Mean error rates during creation of password accounts | 46 |
| 5.4 Interaction effect plots of the time taken to recall password (seconds)..... | 51 |
| 5.5 Interaction effect plots of the reflected transformed values of time taken to recall | 53 |
| 5.6 Interaction effect plots for recall error rates | 57 |
| 5.7 Mean rating for mental demand | 64 |
| 5.8 Mean rating for physical demand | 67 |
| 5.9 Mean rating for temporal demand..... | 70 |
| 5.10 Mean rating for performance | 74 |

| | | |
|------|-------------------------------------|----|
| 5.11 | Mean rating for effort..... | 77 |
| 5.12 | Mean rating for frustration..... | 80 |
| 5.13 | Mean SUS-Creation task..... | 84 |
| 5.14 | Mean SUS-First session recall | 85 |
| 5.15 | Mean SUS-Second session recall..... | 86 |
| 6.1 | Pareto chart analysis | 97 |

1. INTRODUCTION

Computer authentication systems in the 1970s and early 1980s were primarily used by defense facilities, organizations and universities to control access to their sensitive assets. These authentication systems employed a two-part procedure, user identification and user authentication. The users identified themselves by logging in using the alphanumeric id they had created. While in the authentication procedure, they “shared a secret” in the form of a password with a computer to establish their credentials (Brostoff & Sasse, 2000). This password was either assigned or created, and subsequently memorized by the users. This concept of user ids and passwords was found to be a cost-effective and efficient method of maintaining security (Conklin, Dietrich, & Walz, 2004). One of the key elements in these systems was the reliance on human cognitive ability to remember both, the most important being the password (Conklin et al., 2004). Since the users were expected to remember their passwords when prompted without writing them down, these authentication systems were also called knowledge-based.

The earliest passwords were generated by the system and assigned to the user employees to ensure overall security (Adams, Sasse, & Lunt, 1997) (Adams & Sasse, 1999). However, as they were composed of apparently random characters having no meaning for the users, they were more difficult to remember than user-generated ones (Zviran & Haga, 1993). This high degree of complexity caused users to externalize them by writing them down, leading to potential breaches in security (Zviran & Haga, 1993). It led to user-generated passwords becoming widely used (Adams et al., 1997) even though

system-generated ones are more difficult to guess (Zviran & Haga, 1993). To enhance the security of user-generated passwords, they can be selected from a large domain of character sets, giving them the appearance of being randomly generated (Zviran & Haga, 1993). However, password guidelines that encourage users to do this, though they may help to create passwords that are difficult to crack, become difficult to use (Conklin et al., 2004). The limitations associated with restrictions on user-generated passwords include the time needed to generate an acceptable one, the guidelines that result in less memorable ones than those generated without them, and the additional restrictions that may cause more entry errors and lengthen the login procedure (Proctor, Mei-ching Lien, Vu, Schultz, & Salvendy, 2002). This issue concerning password generation is made more complex because users also tend to form their own mental models of good passwords regardless of the instructions provided, favoring memorability over security (Forget, Chiasson, & Biddle, 2007). As a result, users circumvent password guidelines when given a chance, meaning that their passwords are still subject to being breached by brute force attacks. In such attacks, the intruder creates and matches with the target password all possible combinations using a standard US keyboard of 94 characters (Allendoerfer, K., & Pai, S., 2005). In order to protect against such attacks, password guidelines recommend the use of all character sets and longer passwords (Allendoerfer et al., 2005).

With the advent of PCs in offices, school and homes, the user base has grown both in number and in its demographics (Conklin et al., 2004). In addition, the increased use of the internet has led to an increase in the number of password applications (Conklin et al.,

2004). Users now have multiple web accounts ranging from banking to retail, each with a different password (Conklin et al., 2004), creating a significant usability problem (Brostoff et al., 2000). To address these memorability issues, alternate authentication systems, such as biometric systems and image-based passwords systems, were introduced in the 1990s. Biometrics utilizes physical attributes such as finger prints, the retina, or characteristic behavior such as the signature and voice of the user for authentication (Clarke & Furnell, 2005). However, these authentication systems are expensive, obtrusive, difficult to implement on a large scale and have low user acceptance (Proctor et al., 2002). Similarly, image-based passwords, relying on the heuristic of recognition being more memorable than recall, are not as prevalent due to such reasons as user resistance to change from text-based passwords and the cost of modifying existing systems (Jeyaraman & Topkara, 2005). Text-based passwords remain the most common form of authentication (Forget et al., 2007), with user-generated passwords being preferred because of their meaningfulness to the user and greater memorability. Recall of material usually is better if users generate it rather than merely having it provided for them (Proctor et al., 2002).

To improve security and usability of user-generated passwords, proactive user-generated password checking, developed to ensure that user-generated passwords satisfy the composition guidelines, is frequently implemented (Proctor et al., 2002). These composition guidelines generally constrain user-generated passwords with respect to length, composition of character sets and inclusion in a dictionary (Herley, 2009). In early research, Zviran et al. (1993) compared the memorability of system-generated and

user-generated passwords. More recently, researchers have compared the usability of different user-generated password composition schemes. However, the passwords created using different composition schemes in these studies achieved different levels of minimum security, making comparisons across them difficult. To expand on this research, this study compared passwords satisfying NIST Level 2 security requirements that were either assigned by the system or created by the user using two different composition schemes.

2. LITERATURE REVIEW

Random system-generated passwords were one of the first types to be implemented in organizations to protect electronic information. Organizations assigned system-generated passwords to the employees who either memorized or kept a record of them and logged-in to access systems which were password protected. By assigning passwords to employees, organizations ensured that the combinations of characters were secure.

However, the responsibility for the security of electronic information has shifted from the organizations to the users. The users create their own passwords for applications on their personal computers, password systems or operating systems. These user-generated passwords are easier to remember than the random system-generated passwords that users were assigned. Towards the end of the 1990s and at the beginning of the new millennium, the increased usage of internet-based technologies saw a higher incorporation of user-generated password authentication systems for web sites, online

applications and offline services, meaning the number of passwords per user has increased.

To compare the usability and preferences of user-generated passwords and randomly assigned passwords, Zviran et al. (1993) had 103 participants create two user-generated passwords in addition to being assigned an eight-character random one as part of their study. One of the user-generated passwords was a maximum of 8 characters long and the other was an alphanumeric of up to 80 characters (passphrase). After three months, the participants' recall success rate was the highest for the 8-character user-generated passwords, followed by assigned random passwords and then the 80-character passwords. These results were supported by the data obtained with a subjective questionnaire in which the participants ranked the 8-character user-generated passwords highest for appeal and ease-of-recall. These passwords were further analyzed to determine the characteristics affecting their recall. The results revealed that 92% were composed of only lower case letters, suggesting better memorability of passwords of this composition.

Extending the focus of this study, Adams et al. (1997) investigated the memorability and cognitive demands of user-generated passwords. Their analysis of the responses of 139 participants revealed that fifty percent of them externalized their passwords and/or created similar passwords to cope with the cognitive demands associated with recalling multiple ones. These results were confirmed by the in-depth interviews of 30 of these 139 participants. In addition, these interviews revealed that the use of common words and personal data compromised the security of the passwords. As an extension of this study,

Adams et al. (1999) proposed user-centered design of password systems and educating users on password guidelines to cope with the cognitive demands associated with multiple passwords. One of the objectives of this study was to investigate the effectiveness of the early generic composition guidelines recommended by the Federal Information Processing Standards (FIPS, 1985). Its criteria for the creation of user-generated passwords of varying levels of security include length and the character sets used, with the recommendation of a length of at least 4 characters and a composition including numbers. However, cognitive demand and insufficient feedback on the strength of the multiple passwords caused users to focus on memorability rather than security. One of the recommendations of Adams et al. (1999) was to provide adequate online feedback on the strength of the password entered as well as password composition guidelines during creation to mitigate the need for having to change passwords at regular intervals.

In a quantitative study of the memorability and composition of user-generated passwords, Sasse, Brostoff, & Weirich, (2001) conducted a study in which 144 British Telecom employees were asked to describe the reason for the need to reset their passwords and to report the number of passwords they used at work. It was found that the employees had an average of 16 passwords. The passwords which were infrequently used were forgotten the most easily, followed by the moderately and then frequently used passwords. However, the 6-digit passwords for accessing voicemails yielded different results; irrespective of the frequency of use, their recall rate was low. Unlike passwords, they were forgotten even after short durations of non-use. These results indicated the

correlation between the composition of passwords and their memorability and recall, further supporting the conclusion that the compositions of passwords affect their memorability and recall irrespective of the frequency of their use.

To understand the effect of various composition schemes and additional guidelines/restrictions on password usability, Proctor et al. (2002) conducted an experiment involving 24 participants. For the first condition, called “minimal,” the participants created a password of at least 5 characters. The second, called “additional,” incorporated the additional guidelines of having at least one member from all the character sets on a keyboard, at most one character from the username and no consecutive similar characters. The participants were asked to rate each of the two passwords on difficulty of generation and recall using a 7-point Likert scale, with 7 being the most difficult. In the results, statistically significant difference was found between the time taken to generate and recall minimal condition passwords and additional condition passwords. Passwords with the additional composition restrictions were significantly harder to generate and remember than those based on the minimal requirements. All of the passwords created were subsequently subjected to a password cracking software. The results further revealed that 18 of 24 minimal condition passwords were cracked compared to eight of 24 in the additional condition, indicating the level of lower strength of passwords in the former.

Using a similar procedure, Proctor et al. (2002) conducted a second experiment which required a minimum length of at least 8 characters for the passwords. Similarly, the

results of the second experiment found a statistically significant difference between the time taken to generate and recall minimal condition passwords and those requiring additional guidelines. Also similar to the first experiment, the qualitative data found that passwords with additional guidelines were significantly harder to generate and remember compared to passwords with only a length restriction. In addition, passwords in this experiment were also subjected to cracking software, the results finding that four of twenty-four minimal conditions passwords and three of twenty-four additional condition passwords could be cracked. The results concerning the breached passwords from both experiments suggested that the increase in the minimum length of minimal condition passwords from 5 to 8 characters that led to an increase in their recall time were as resistant to password cracking software as the minimum 8-character password incorporating additional guidelines.

To understand the effect of user-generated passwords versus randomly selected passwords on memorability, Yan, Blackwell, Anderson, & Grant (2004) compared the ease of memorizing two user-generated passwords constructed based on different composition guidelines and one randomly selected password. A total of 288 participants were divided into three groups. The participants in the first group created a password of a minimum of seven characters including at least one number. The participants in the second created a password by randomly selecting eight characters from a list of printed letters and numbers with their eyes closed. In the last group, the participants create a mnemonic-based password by choosing any character from each word of a phrase and representing it as a lower or upper case letter, a number or a special character. The

participants were asked to create and keep a written record of their passwords until they memorized them. After four months, the participants received a two-question email asking them to subjectively rate the ease of memorizing the passwords from 1 to 5, with 5 being impossible, and to specify the duration participants referred to a written record until they memorized their passwords. The first group considered memorizing their passwords the easiest, rating it a 1.52, which was the lowest among the three groups. They also indicated that it took 0.7 weeks to memorize the passwords compared to 0.6 weeks for the mnemonics-based passwords and 4.8 weeks for the random passwords. These results suggest that random passwords are far less memorable than user-generated ones. However, the passwords of the three groups were of varying levels of strength due to their length and composition, thus affecting the generalizability of the results. In addition, keeping a written record of the passwords by the participants could also be considered a limitation of a study associated with a knowledge-based authentication system.

To investigate the effect of password construction guidelines on user behavior, Kuo et al. (2006) surveyed 290 participants. In the survey, a scheme of seven guidelines was given to all the participants, including recommendations to include numbers, lower and upper case letters, and special characters. The guidelines also recommended that passwords be long enough, not include dictionary words, not be related to the web site they were created for, and not be in a non-English language. The responses from the participants indicated that the number of guidelines they considered depended on whether they had received training on them earlier. There was a statistically significant difference between the number of guidelines considered by the participants who were aware of

password composition guidelines and the participants who were not. These results suggest that educating users on password composition guidelines affects user behavior, helping them to compose passwords that are both memorable and secure.

The increase in the variety and number of internet-based technologies and their password authentication systems has led to multiple password composition schemes. To study the issues of inconsistent password composition guidelines on an organizational level, Allendoerfer et al. (2006) interviewed 52 Federal Aviation Administration (FAA) employees, documenting their user experience with FAA password guidelines and systems. The employees experienced increased cognitive demands due to inconsistency in these guidelines, especially if they were similar but not exactly the same. Based on these results, Allendoerfer et al. (2006) recommended consistent password guidelines for all organizational password systems.

To address this issue, the National Institute of Standards and Technology (NIST) recommended the following user-generated password composition guidelines for all electronic authentication purposes (Burr, National Institute of Standards, & Technology, 2006):

- 1) A minimum of 8 characters selected from the keyboard of 94 printable characters
- 2) At least one upper case letter, one lower case letter, one number and one special character
- 3) No common words or permutations of usernames

User-generated passwords created using these guidelines have an estimated guessing entropy of 30 bits, satisfying the Level 2 security recommended by NIST for password authentication. For system-generated random passwords composed from the 94 characters on the keyboard, NIST estimates that five characters will satisfy its Level 2 security recommendation (Burr, Dodson, & Polk, 2006). However, since these guidelines are only a recommendation, they are not widely implemented.

A more recent study investigating various user-generated password construction schemes was conducted by Vu et al. (2007). They investigated the number of attempts and the time required to generate passwords. They also evaluated the number of login errors and the time required to recall these passwords after a short and long duration of time. In the first of three experiments involving a total of 32 participants, 16 participants created passwords for three accounts and the remaining 16 created them for five. These user-generated passwords were restricted to at least six characters including an upper and a lower case letter, a digit and a special character. These passwords were also required to be unique for each of the three or five accounts and could not contain the participant's username or any variations of it. The participants were asked to recall and login to their accounts four times in a random order 5 minutes after creation and after a week during a second recall session. The group with five accounts made significantly more recall errors than the group with three accounts. Most importantly, the experiment indicated that creating unique passwords for increasingly more accounts increased memory load. This conclusion was supported by the finding that 69 percent of the participants having five

accounts were unable recall their passwords after a week compared to 19 percent of the participants with only three accounts.

In the second experiment, also including two recall sessions, 20 participants created unique passwords for three accounts using the first letters of at least six words of a meaningful sentence constructed by them. The password for the remaining 20 participants also used the first letters of at least six words of a meaningful self-constructed sentence, but incorporated a digit and a special character. The results indicated that the difference in the password generation time between the two groups was statistically significant, with the first group taking 50.9 seconds and the latter 84.9 seconds. The results also indicated that the login errors and login times for the second group were twice that for the first for both short-term and long-term recall. Sixty-two percent of the user-generated passwords containing only letters were breached by cracking software compared to 2 percent of the passwords including a digit and special character. These results indicated that including a number and a special character increased password security.

The first part of the third experiment compared generation times, login times, and login errors for passwords created based on four conditions. Of 60 participants, 30 created passwords from the first letters of the words of a meaningful sentence they constructed including a digit and a special character, similar to the second condition of the previous experiment. The remaining participants created mnemonic-based passwords by replacing entire words of the sentences constructed with similar sounding words and

special characters as well as visually similar numbers and special characters. Fifteen of the first group of 30 participants created passwords in the first session and were then asked to enter them a week later in a second session. The remaining participants of the first group created a password, entered it after a short break of 5 minutes and were then asked to enter it again a week later in a second session. The second group of 30 participants who created mnemonic-based passwords were similarly split into two sub-groups, 15 involved in long-term recall and the remaining 15 subjected to both short-term and long-term recall. The results found that login times for passwords for long-term recall following short-term recall were 25 seconds faster than login times for only long-term recall for both first letter and mnemonic condition passwords. These results suggest that short-term recall improved the ability of the participants to remember their passwords.

In the second part of the third experiment, 15 new participants created passwords composed of the first letters of at least six words of a sentence constructed by them which also included a digit and a special character. These passwords were immediately entered with no short-term delay and then were recalled after a week. The results revealed that these participants took 45 seconds to login immediately and 47 after a week, significantly longer than mean login time of 21 seconds for the passwords created using the same guidelines in the first part that were recalled after a short-term and long-term period. These results suggested that the five-minute delay between password creation and short-term recall helped participants to remember these passwords better.

Although consistency of password composition schemes across systems can be achieved in organizations, it is difficult to implement such consistency across web sites. Florencio & Herley (2010) analyzed the password composition guidelines for 75 web sites with medium to heavy internet traffic, including bank, government, university, brokerage and defense web sites, to name a few. The password composition guidelines of these web sites ranged from 1-character unrestricted passwords to 12-character passwords including all character sets. The researchers found no correlation between the strength of the passwords resulting from the password policies and the value of the assets, the number of users, the size of the web site and the number of attacks on the web site. Some of the commercial and social networking web sites that earned revenue with each login imposed fewer restrictions on password composition, accommodating passwords that were easy to create, recall and use for multiple login attempts. Complex password composition guidelines, if implemented on such web sites, might cause revenue losses due to a decrease in user traffic. This discussion suggests that password composition guidelines should create usable passwords compatible with the nature of a web site and its users.

Komanduri, Shay, Kelley, Mazurek, Bauer and Christin (2011) compared passwords created by 5,000 participants, each assigned to one of five conditions across two sessions. In the first condition, participants were asked to create at least eight-character passwords for the purpose of completing a survey, referred to collectively as the Basic8Survey, with no restrictions. In the second condition, participants created passwords for the purpose of creating an e-mail account, called the Basic8, based on the same guidelines. The third

condition, the Basic16, asked participants to create passwords of at least 16 characters with no restriction on the character sets used. The fourth, named Dictionary8, created passwords of at least 8 characters with no restrictions and with a dictionary check to prevent the use of commonly used strings. The fifth condition, the Comprehensive8, asked participants to create 8 character minimum passwords of at least a number, a special character and both cases of letters, with a dictionary check.

The participants were asked to enter their respective passwords twice, the second time to confirm the first entry. On successfully logging-in, the participants were asked to complete a survey asking for their demographics, their rating of the password creation process, and the strategies employed. The participants were again asked to re-enter their passwords with a maximum of five attempts permitted. After two days, in the second session the participants were asked to enter these passwords via an email with a maximum of five attempts permitted to recall their passwords. The participants also answered survey questions on password creation, storage and usage. The results from the first session found that among the participants with at least one password creation failure, Dictionary8 passwords were easier to create with fewer attempts required than the Comprehensive8 passwords, but took significantly more attempts to create than the Basic8, Basic8Survey and Basic16 passwords. Considering the cumulative password creation failed attempts, Comprehensive8 passwords resulted in the highest number of attempts with a mean of 3.35 followed by Dictionary8 passwords, the mean of which was significantly higher than for the Basic16 passwords. Basic8 passwords exhibited significantly fewer numbers of attempts at 1.13 compared to the other four conditions,

with the Basic8Survey mean being marginally higher at 1.17. Approximately 25 percent of the participants completely failed to create acceptable passwords in the Comprehensive8. However, the completion failure rates for the participants in the other conditions were significantly lower, all under 19 percent. The participants rated Comprehensive8 passwords as significantly more difficult to recollect followed by the Basic16 passwords.

In addition, the participants who did not externalize their passwords required an average of 1.22 attempts to recall their passwords, with the difference in number of recall attempts for each condition being significantly different from one another. Most of the participants in the Comprehensive8 condition agreed that the creation of these passwords was “annoying.” However, 67 participants in this condition believed that these passwords would make their main email accounts more secure, suggesting that the perceived strength of passwords may affect participant willingness to use those that are “annoying” to create. Other findings suggest that Basic16 passwords were as secure as Comprehensive8 passwords, but were relatively more usable. However, the study did not consider the factor of password creation times in either session.

Alternate authentication technologies like biometric and image-based password systems have been introduced to improve the memorability and reduce the cognitive demands of user-generated passwords. Image-based password systems depend on user recognition rather than recall. As a result, they are expected to be more memorable and less mentally demanding than traditional text-based passwords. Dhamija & Perrig, (2000)

compared image-based passwords to user-generated passwords with respect to task completion time and error rates. Twenty participants created passwords and immediately logged in and then re-logged in after a week. The user-generated passwords took less time to create and login than the image-based passwords. After a week, the login times for user-generated passwords were shorter than the login times for image-based passwords, even though the login error rate for user-generated passwords increased. One participant failed to login during the initial session using his/her user-generated password. There were no failed logins in the initial session for the image-based password. In the second session, user-generated passwords resulted in four more failed or unrecoverable logins than image-based passwords. Dhamija et al. (2000) did not analyze the data on task completion time and error rates for statistical significance. The results indicated that user-generated passwords had shorter task completion times but were not as memorable as the image-based passwords.

In a concurrent study, Brostoff et al. (2000) compared user-generated passwords to an image-based password system called Passfaces, recruiting 34 under-graduate students each of whom created both types of passwords. Logins, login failure rates, time before first use and the number of login attempts were recorded after the participants were asked to login again 3 months after creation. The results revealed that user-generated passwords had a higher login failure rate of 15.1% compared to 4.9% for Passfaces, a statistically significant difference. The time before first use of the user-generated passwords and Passfaces was also significantly different, the time taken for the former being less than for the latter. The number of login attempts for the user-generated passwords was found

to be three times higher than the number of login attempts for Passfaces, also statistically significant. These analyses show that participants using user-generated passwords took less time to learn the system but had higher login errors, similar to the previous study. The studies conducted by Dhamija et al. (2000) and Brostoff et al. (2000) found that the usability of image-based passwords was marginally better than that of user-generated passwords. User-generated passwords allow users to complete logins in a shorter amount of time and reduce learning time, but these advantages are offset by a higher number of login errors.

Previous research has compared the memorability and usability of system-generated passwords and user-generated passwords. Passwords created under various composition schemes have been compared in terms of login errors, task completion times, and recall rates after a short- or long-term period or both. However, there is limited research comparing user-generated passwords created under various conditions having approximately minimum equal entropy other than that of Komanduri et al. (2011). This study extended the research by Komanduri et al. (2011) by comparing the usability of assigned system-generated passwords with user-generated passwords created under two composition schemes, with all passwords satisfying the NIST Level 2 requirements of 30 bits of entropy.

3. RESEARCH HYPOTHESES

This study compared the usability of three types of text-based passwords of approximately equal minimum security:

- 1) An assigned 6-character system-generated password selected randomly from any of the 36 alphanumeric characters available on the standard QWERTY keyboard.
- 2) A user-generated password of at least eight characters, with at least one lower case letter, one upper case letter, a number and one special character. This password must also pass a dictionary check.
- 3) A user-generated password of at least 16 characters with no restrictions; this password must also pass a dictionary check.

To compare the usability of the passwords created under these three conditions, the following hypotheses were proposed:

Hypothesis 1: The time taken to create a password account for the system-generated password will be less than that required for the 8-character user-generated or 16-character unrestricted user-generated password.

This result is expected due to the shorter length of the system-generated password.

Hypothesis 2a: The number of attempts required to create a valid password in the first session will be lower for the 16-character unrestricted user-generated password than for the 8-character restricted user-generated password.

These results are expected due to the lower complexity of the 16-character passwords than 8-character passwords.

Hypothesis 2b: The number of unrecoverable passwords and Damerau-Levenshtein edit distance in the second session will be lower for the 16-character unrestricted passwords than for the other passwords.

These results are expected due to the lower complexity of the 16-character unrestricted passwords than the 8-character restricted passwords and the 6-character alphanumeric system-generated passwords.

Hypothesis 2c: The Jaro-Winkler proximities in the second session will be higher for the 16-character unrestricted password than for the other passwords.

These results are expected due to the lower complexity of the 16-character unrestricted passwords than the 8-character restricted passwords and the 6-character alphanumeric system-generated passwords.

Hypothesis 2d: The NASA-TLX indices, time taken to recall and recall error rates will be lower for the 16-character unrestricted password than for either of the other passwords in both sessions.

These results are expected due to the lower complexity of the 16-character unrestricted passwords than the other passwords.

Hypothesis 2e: The ease-of-use will be higher for the 16-character unrestricted password than for either of the other passwords in both sessions.

These results are expected due to the lower complexity of the 16-character unrestricted passwords than the other passwords.

4. METHOD

Participants

Fifty-four undergraduate, graduate students and staff members participated in this study. They were recruited through an email or a verbal invitation describing the experimental study. Students interested in participating were pre-screened via questionnaire to determine their eligibility. To be eligible, participants should have had prior experience using the Internet for a minimum of one year and in constructing passwords for user accounts on the Web. The 54 participants were randomly divided into three groups, 18 in Condition 1 who were assigned alphanumeric system-generated passwords of 6 characters, each randomly selected from any of the 36 alphanumeric characters available on the standard QWERTY keyboard; 18 in Condition 2 creating passwords of a minimum of 8 characters composed of at least one lower case and one upper case letter, a number and a special character, subject to a dictionary check; and 18 participants in Condition 3 creating passwords with a minimum of 16 characters of any type characters, subject to a dictionary check. For Condition 1, prior to the experimental study, 40 Clemson University students were surveyed. This questionnaire asked respondents to rank three assigned alphanumeric system-generated passwords that varied in terms of character sets used and length but shared similar entropies (see Appendix G). They were asked to rank their most preferred form as one and their least preferred as

three. Statistical analysis of the responses indicated that 6-character alphanumeric system-generated passwords were preferred compared to 5 character completely random system-generated passwords and 7-character lower-case letter passwords. Based on these results, the 6-character alphanumeric password was selected as the system-generated password for the experimental study.

Testing Environment

This study was conducted in the Human Computer Systems Laboratory in Freeman Hall. The experimental set-up consisted of a desktop computer, table, chair, paper and pencil. The participants were assigned a 6-character alphanumeric password or asked to create a password by entering it into a popup window generated by the application using the guidelines provided. Those participants who created passwords were also given instructions on memory aids such as mnemonics and passphrases. The computer presented a password login application into which all 54 participants in the experimental study entered their passwords during the first session. This application provided immediate feedback on whether the passwords created conformed to the stipulated password policies before storing them. The participants were then asked to take a five-minute break, engaging themselves in a distraction task. The objective of the task was to reach the highest score in a game of Angry Birds (Rovio Entertainment Ltd., 2009). If the participants reached the highest score before five minutes, they were asked to continue to the next level of the game. At the end of the break, the participants were given five opportunities to enter their password. During the second session of the study, the

participants used the password login application to enter the passwords they were assigned or had created in the first session with a maximum of five attempts permitted. The researcher was present in the laboratory with the participants to provide guidance during both sessions.

Experimental Design

This experiment is considered to be both a one-factor design with three levels and a two-factor design with two or three levels. The independent variable of the former investigates the password composition scheme at the three levels defined in Table 4.1 below:

Table 4.1: Three levels of password composition

| Condition 1 | Condition 2 | Condition 3 ¹ |
|--|---|---|
| 6 characters | Minimum of 8 characters | Minimum of 16 characters |
| Alphanumeric characters selected from any of the 36 characters available on the standard QWERTY keyboard | Characters selected from any of the 94 characters available on the standard QWERTY keyboard | Characters selected from any of the 94 characters available on the standard QWERTY keyboard |
| System-generated and assigned | At least one lower and one upper case letter, one number and one special character | User-generated |
| No common words or repeated character sequences of length three or greater or permutations of usernames | No common words or repeated character sequences of length three or greater or permutations of usernames | No common words or repeated character sequences of length three or greater or permutations of usernames |
| | User-generated | |

¹(Burr et al., 2006)

However, for dependent variables recorded two or three times over task sessions, the experiment was a two-factor design with two or three levels. The second independent variable of the study were the task sessions defined in Tables 4.2 and 4.3 below:

Table 4.2: 3x3 factorial design

| IVs | Task Session 1: Creation | Task Session 1: Recall | Task Session 2: Recall |
|-------------|-----------------------------|---------------------------|---------------------------|
| Condition 1 | | | |
| Condition 2 | | | |
| Condition 3 | | | |

Table 4.3: 3x2 factorial design

| IVs | Session 1: Recall | Session 2: Recall |
|-------------|-------------------|-------------------|
| Condition 1 | | |
| Condition 2 | | |
| Condition 3 | | |

All three password schemes resulted in passwords that possess a minimum guessing entropy of 30 ± 2 bits (Burr et al., 2006). According to NIST guidelines (Burr et al., 2006), passwords of 16 characters or more do not need dictionary checks. However, we included a dictionary check for Condition 3 for consistency purposes. The dependent variables included objective and subjective measures of performance for each participant.

The objective measures for the first session consisted of the time taken to create an account and a password, the error rate during creation, the recall login error rates and the recall time. These error rates were calculated by dividing the total number of incorrect attempts by the total attempts made to complete the task. Additionally, the Damerau-

Levenshtein and Jaro-Winkler proximities of the incorrectly recalled passwords for each participant were recorded per recall session. The Damerau-Levenshtein distance between the recalled password and the stored password is the minimum number of operations, i.e. addition, subtraction, substitution or transposition, needed to transform the recalled password to the stored password. The Jaro-Winkler distance is a measure of similarity between the recalled password and the stored password. The objective measures in the second session included the recall login error rates, the recall time and the Damerau-Levenshtein and Jaro-Winkler proximities of incorrectly recalled passwords for each participant.

Subjective data were obtained through the System Usability Scale (SUS) (see Appendix D) administered to the participants at the end of each session of the experimental study. The questionnaire at the end of the first session addressed the ease of creating the password and the ease-of-use and memorability of the passwords for this session; the questionnaire administered to the participants at the end of the second session addressed the memorability of the passwords created. In addition, at the end of each session, the NASA-TLX questionnaire (see Appendix E) was administered to the participants to measure mental, physical and temporal workload as well as the performance, effort and frustration component of the workload.

The data collected for each dependent measure were statistically analyzed using a one-way analysis of variance or two-way mixed analysis of variance. The locus of any statistically significant difference was determined using an LSD post-hoc test.

Tasks

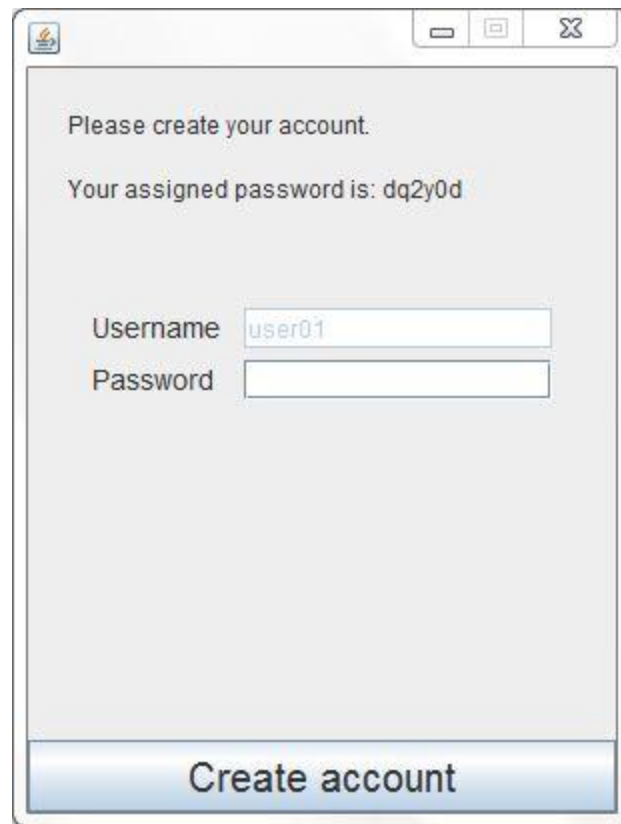
The experimental study was conducted in two sessions. In the first, the participants performed the following set of tasks:

- 1) Condition 1 participants received a 6-character alphanumeric password generated by the application on a pop-up window. See Figure 4.1.
- 2) Participants in Conditions 2 and 3 were provided a set of guidelines to create a password according to the password composition scheme assigned to them. See Figures 4.2 and 4.3.
- 3) The usernames for the participants were automatically generated and pre-populated in the popup window above the space for the password entry.
- 4) These assigned or created passwords were entered into the same popup window in the space provided.
- 5) After the entry of the passwords, all participants checked the feedback provided by the password login application.
- 6) If the feedback indicated that the password did not conform to the one that was assigned or to its requirements, Step 4 was repeated. See Figures 4.4, 4.5 and 4.6.
- 7) If the password entered was correct, the participants were asked to take a five-minute break in which they played the computer game Angry Birds (Rovio Entertainment Ltd., 2009).
- 8) After the break, the participants logged in using their assigned or created passwords. A total of five attempts were permitted for entering the password correctly for the first time. See Figures 4.7 and 4.8.

- 9) On entering the password successfully or exhausting all five recall attempts, the participants completed the System Usability Scale (SUS) and the NASA-TLX questionnaire

The participants were asked to return 5 to 7 days later, depending on their availability, to determine the memorability of their passwords by performing the following tasks:

- 1) All participants entered the password they were assigned or that they created in the first session into the login application.
- 2) A total of five attempts were permissible for entering the password correctly.
- 3) On entering the password successfully for the first time, the participants completed the System Usability Scale (SUS) and the NASA-TLX questionnaire.



A screenshot of a software window titled "Please create your account." The window has a standard OS-style title bar with minimize, maximize, and close buttons. The main content area is light gray and contains the text "Your assigned password is: dq2y0d". Below this, there are two input fields: "Username" with the value "user01" and "Password" which is empty. At the bottom of the window is a large blue button with the text "Create account".

Please create your account.

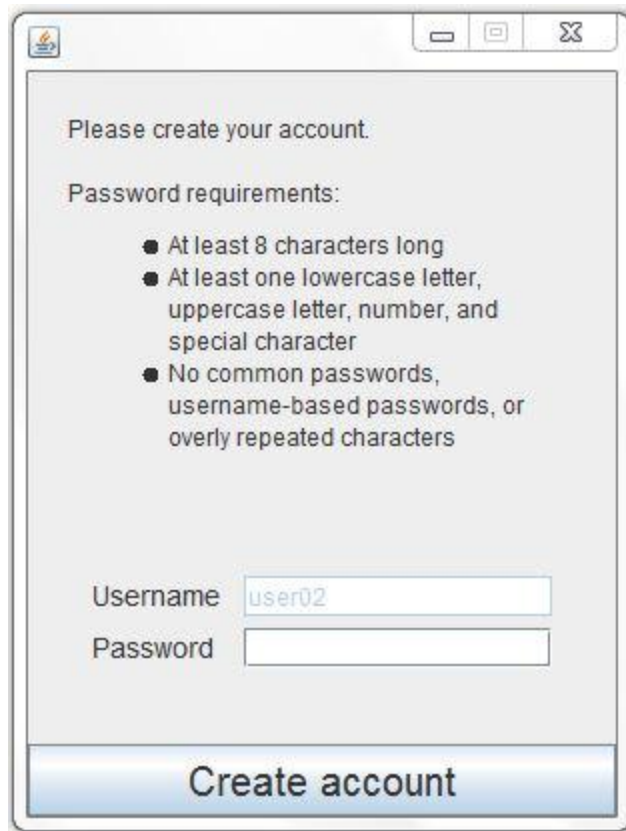
Your assigned password is: dq2y0d

Username

Password

Create account

Figure 4.1: 6-character alphanumeric password creation



Please create your account.

Password requirements:

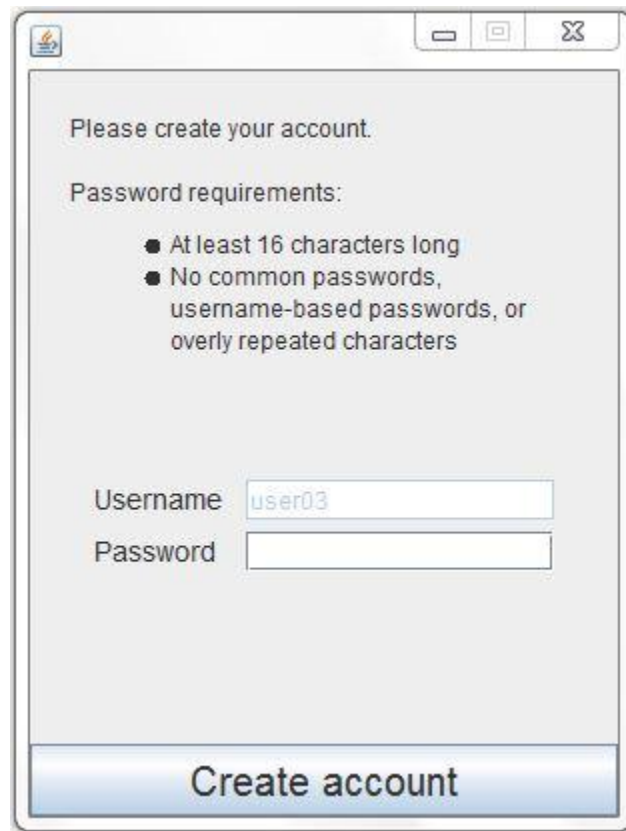
- At least 8 characters long
- At least one lowercase letter, uppercase letter, number, and special character
- No common passwords, username-based passwords, or overly repeated characters

Username

Password

Create account

Figure 4.2: 8-character password creation



A screenshot of a software dialog box for account creation. The dialog has a title bar with a small icon on the left and standard window controls (minimize, maximize, close) on the right. The main content area is light gray and contains the text "Please create your account." followed by "Password requirements:". Below this, there are two bullet points: "● At least 16 characters long" and "● No common passwords, username-based passwords, or overly repeated characters". Further down, there are two input fields: "Username" with the text "user03" and "Password" which is empty. At the bottom of the dialog is a blue button with the text "Create account".

Please create your account.

Password requirements:

- At least 16 characters long
- No common passwords, username-based passwords, or overly repeated characters

Username

Password

Create account

Figure 4.3: 16-character password creation

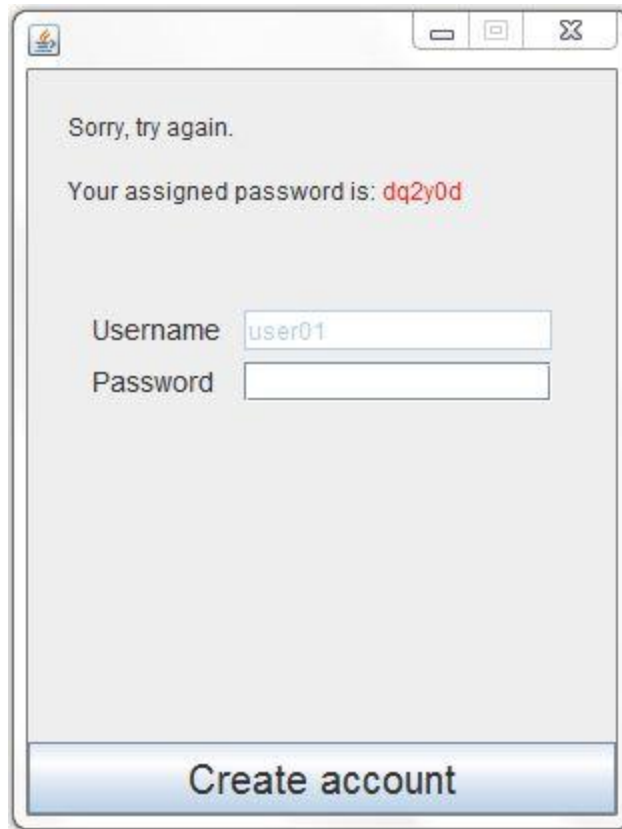


Figure 4.4: Response popup window for failed 6-character alphanumeric password creation

Sorry, try again.

Password requirements:

- At least 8 characters long
- At least one lowercase letter, uppercase letter, number, and special character
- No common passwords, username-based passwords, or overly repeated characters

Username

Password

Create account

Figure 4.5: Response popup window for failed 8-character password creation

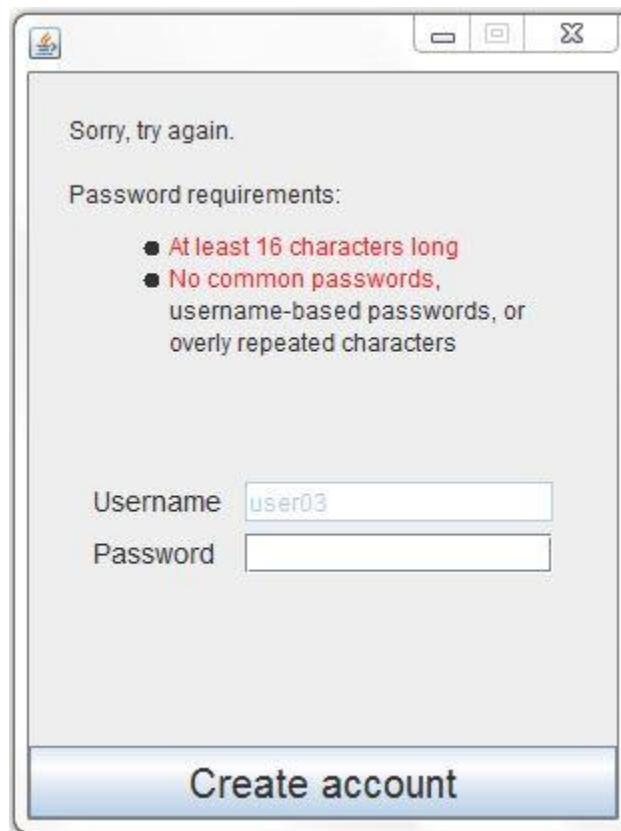


Figure 4.6: Response popup window for failed 16-character password creation

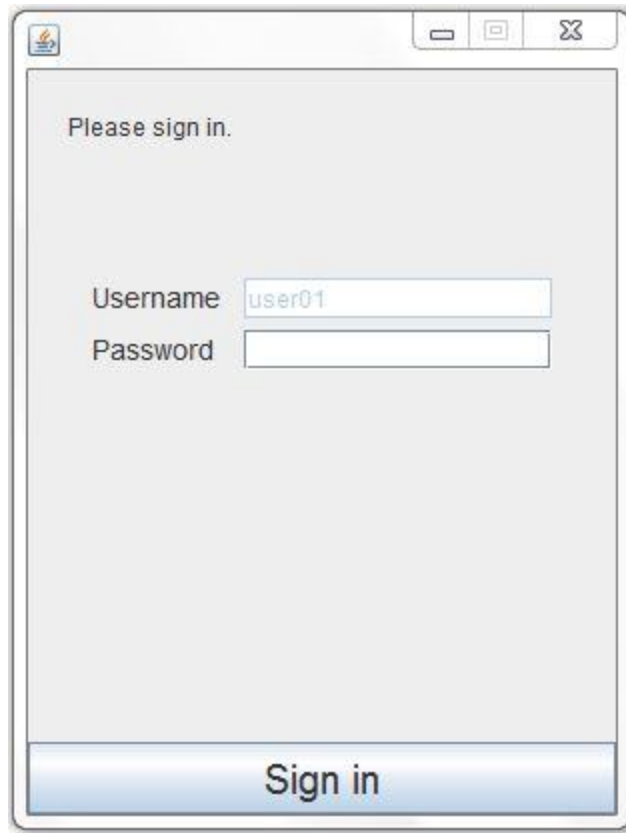


Figure 4.7: Password recall pop-up window

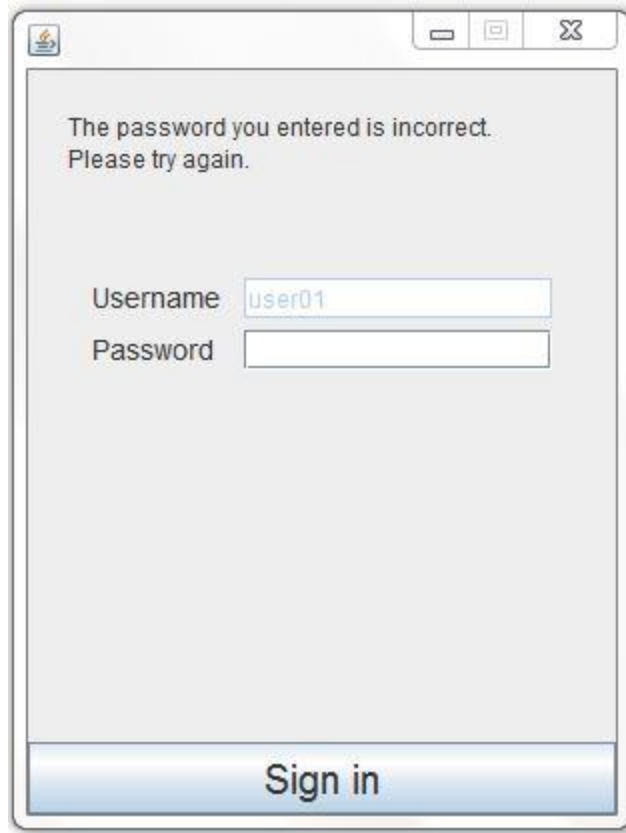


Figure 4.8: Failed password recall attempt

Procedure

At the beginning of the first session, the researcher greeted the participants, who were then seated in front of a desktop computer on a table in the Human Computer Systems Laboratory. The researcher provided a brief overview of the experiment to the participants. After the participants read and signed the informed consent form (see Appendix A), they completed a pre-study questionnaire (see Appendix B) asking for demographics, information on their Internet experience and their previous experience in creating user accounts on the Internet. On completion of the pre-study questionnaire, the

researcher provided training on memory tools such as mnemonics and passphrases for Conditions 2 and 3, and the types of passwords that would not be accepted by a dictionary check for Conditions 2 and 3 (see Appendix C). These techniques could be used to assist in the creation and memorization of passwords. The duration of this training was approximately 5 minutes. The participants were then asked to memorize the passwords that they would be assigned or that they would create to avoid externalizing them.

After the completion of training, the participants were either assigned or they created passwords conforming to the password guidelines they were provided. They then subsequently entered them into the password login application on the desktop computer. The application provided immediate feedback as to correctness in the assigned password condition. In the user-generated password conditions, the application provided feedback as to whether the passwords created conformed to the required guidelines. Participants who failed were asked to re-create the passwords. The time taken and the number of errors committed during the creation of correct passwords in the first session were recorded. After a five-minute break, the participants entered the passwords assigned or created into the application, with five attempts allowed. The time taken to recall the password and the login error rates were recorded. On completion of this task, the participants were asked to complete the System Usability Scale (SUS) questionnaire (see Appendix D). These questions use a 5-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Then, each participant was administered the NASA Task Load Index questionnaire to assess the performance, effort, frustration, mental, physical

and temporal demands experienced during the creation of the password (see Appendix E). At the end of the session, the researcher asked the participant to schedule a date and time for the second session of the experimental study. The duration of the first phase of the study was approximately thirty minutes.

At the beginning of the second session, the researcher briefed the participants on the task to be completed. The researcher asked them to recall their passwords from the first session and to enter them into the password login application on the desktop computer. The time taken to recall the password was recorded. A maximum of five attempts was allowed for the participants to recall their passwords correctly; if the participants failed to be able to do so in five attempts, the password was specified as unrecoverable. The Damerau-Levenshtein and Jaro-Winkler proximities for the unsuccessful attempts were recorded. The participants were asked to complete the System Usability Scale questionnaire (see Appendix D). The researcher then administered the NASA Task Load Index questionnaire to the participants to assess the performance, effort, frustration, mental, physical and temporal demands experienced during the login task (see Appendix E). The duration of the second phase was approximately 20 minutes. See procedure flow for first and second session in Figure 4.9.

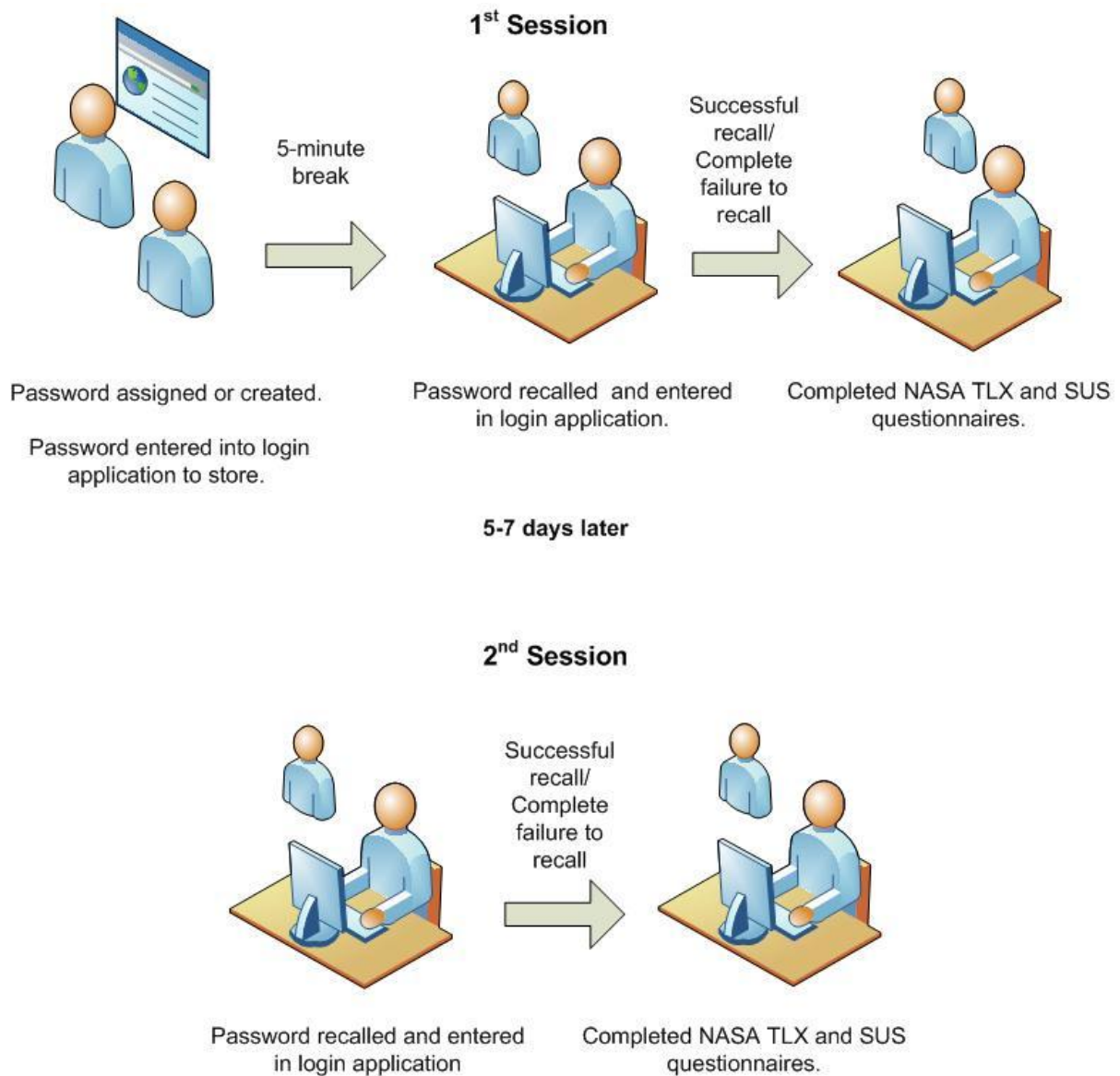


Figure 4.9: Procedure flow for first and second session

5. RESULTS

The data collected across all the task sessions for all the participants were screened and checked for the normality. These results showed that the times taken to create password accounts in the first session, the recall times in the first and second sessions, and the edit distances of Damerau-Levenshtein and Jaro-Winkler in the second session were non-normal with high skewness values. The data from these dependent variables were then transformed using the reciprocal function to normalize them.

The results were subsequently subjected to one-way or two-way mixed ANOVA, depending on whether the dependent variable was measured once or more than once the over task sessions. One-way ANOVA was applied to the three password policy conditions for the time taken to create passwords accounts (including memorizing and entering the password into the login application), the password creation error rates, and the Damerau-Levenshtein and Jaro-Winkler edit distances. The Damerau-Levenshtein and Jaro-Winkler edit distances for the first recall session were recorded for the fifty-six participants who successfully recalled their passwords within five attempts. Consequently, one-way ANOVA of the Damerau-Levenshtein and Jaro-Winkler edit distances for the first recall session or a two-way mixed ANOVA could not be conducted due to the unequal sample size of participants in the first session.

Two-way mixed ANOVA was applied for the three password policy conditions across the two task sessions for the time taken to recall the passwords and for the recall error rates as well as for the subjective ratings from the NASA-TLX metrics and the System

Usability Scale (SUS) scores. The second independent variable, task session, involved two levels (the first session recall and the second session recall task) or three levels (the first session password creation task and the first and the second session recall tasks), depending on whether the measures of the dependent variables were repeated. More specifically, the time taken to recall the password and the recall error rates had two levels for each task session, i.e. recall task in first and second session. The subjective measures, NASA-TLX indices and the SUS scores had three levels for each task session, i.e. password creation task in first session and recall task in first and second session.

Objective Measures

The objective measures recorded for the first session were the time taken to create a password account and the error rates for password creation. For both sessions, the time taken to recall the password, the recall error rate and the edit distances were recorded.

Time taken to create a password account. The time taken to create a password account in the first session includes the time taken to receive an assigned password or to create a password, to memorize the assigned/created passwords and to enter the passwords into the application. These steps were measured from the time a password was assigned or requested to be created to the time the account was created. The descriptive statistics for this metric are provided in Table 5.1. Mean, standard deviation and error in the table are transposed (reciprocal) values of the original time recorded in seconds:

Table 5.1: Descriptive statistics for the transposed value of password account creation
time

| Creation Time | N | Mean | Standard Deviation | Std. Error |
|------------------|----|--------|-----------------------|---------------|
| System-generated | 18 | 0.0278 | 0.01612 | 0.00380 |
| 8-character | 18 | 0.0191 | 0.00979 | 0.00231 |
| 16-character | 18 | 0.0131 | 0.00634 | 0.00150 |
| Total | 54 | 0.0200 | 0.01280 | 0.00174 |

A one-way between subjects ANOVA was conducted to test the effect of the password policy conditions on the time taken to create a password account. The results indicated this effect was significant, $F(2, 51)=7.395$, $p=0.002$. Subsequent post-hoc analysis revealed that the time to create a password account was less for the system-generated passwords than for either the 8-character ($p=0.028$) or the 16-character passwords ($p<0.001$). There was no statistically significant difference between the 8-character and the 16-character password. The one-way ANOVA table for the transposed value of the time taken to create password accounts is provided in Table 5.2, and the original and reflected transposed values of creation time are plotted in Figures 5.1 and 5.2. These reflected transposed values were obtained by subtracting the transposed values from a constant so that bar graphs are in the same direction as the original values:

Table 5.2: One-way ANOVA data for the transposed value of account creation time

| Creation Time | SS | df | Mean Squares | F | Sig. |
|----------------|------|----|-----------------|-------|-------|
| Between Groups | .002 | 2 | .001 | 7.395 | 0.002 |
| Within Groups | .007 | 51 | .000 | | |
| Total | .009 | 53 | | | |

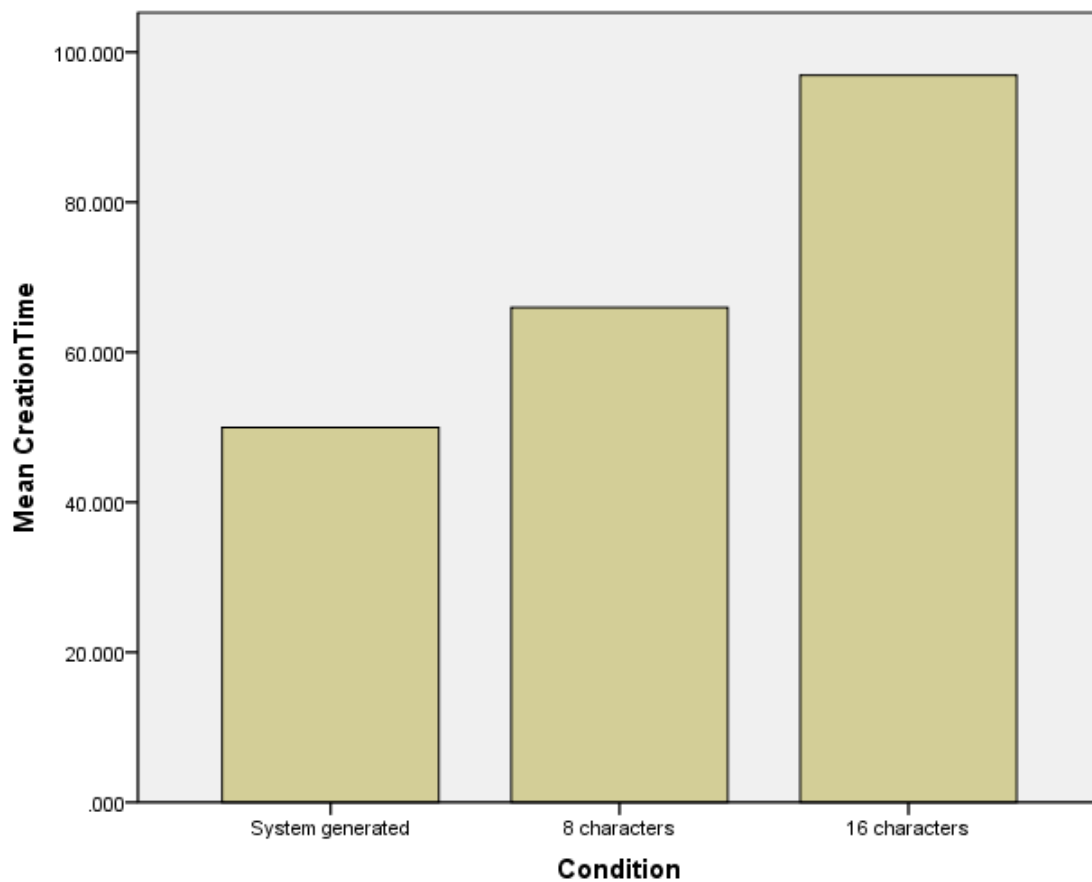


Figure 5.1: Mean password account creation time (seconds)

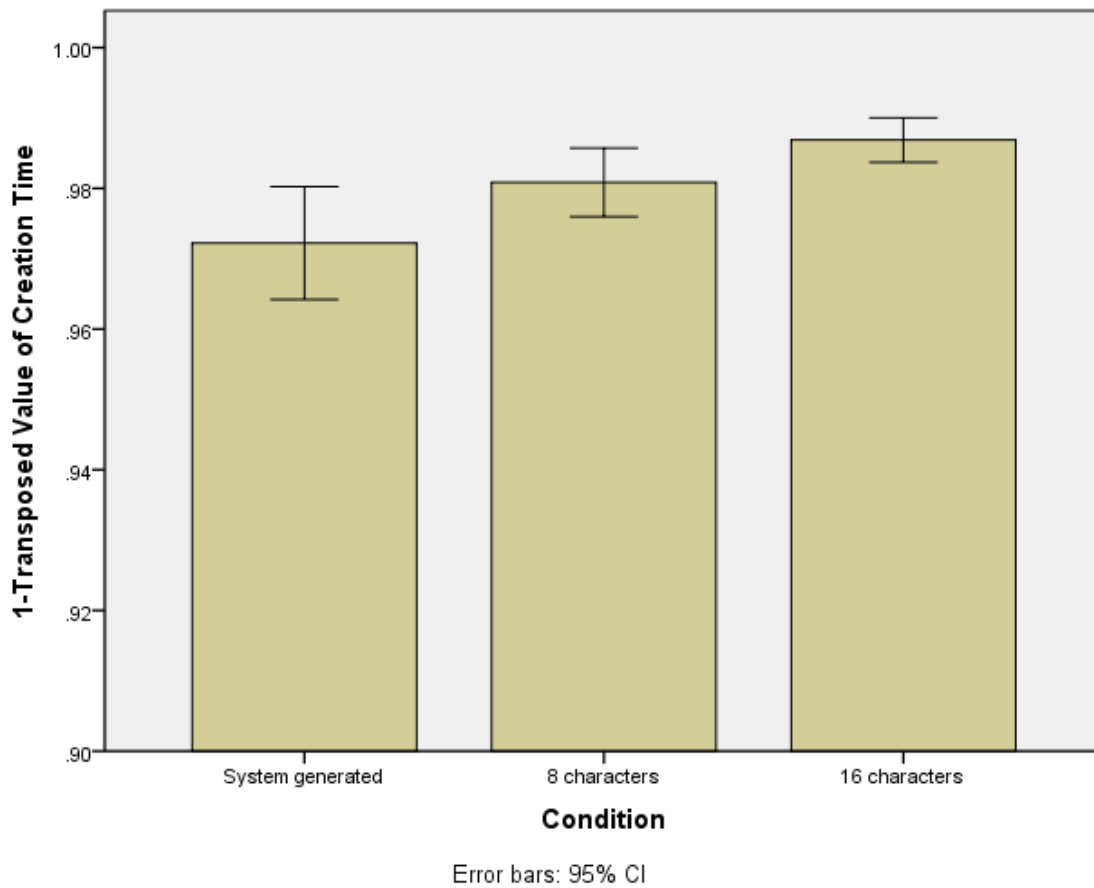


Figure 5.2²: Reflected transposed values for password account creation time

²(The transposed values are reflected in this graph so that the higher values correspond to the longer creation time)

Password creation error rates. Password creation error rates, which measure the number of attempts taken to create an account conforming to the password policy condition, were measured by dividing the number of errors by the total number of attempts taken to create the password account. The descriptive statistics for this metric are provided in Table 5.3:

Table 5.3: Descriptive statistics for error rates during password account creation

| Error Rate | N | Mean | Standard Deviation | Std. Error |
|------------------|----|--------|-----------------------|---------------|
| System-generated | 18 | 0.0000 | 0.000000 | 0.0000 |
| 8-character | 18 | 0.2130 | 0.278983 | 0.06575 |
| 16-character | 18 | 0.1111 | 0.213896 | 0.05041 |
| Total | 54 | 0.1080 | 0.217598 | 0.02961 |

A one-way between subjects ANOVA was conducted to test the effect of the password policy condition on password creation error rates. The results indicated that this effect was significant, $F(2, 51)=4.959$, $p=0.011$. Subsequent post-hoc analysis revealed that the error rate for creating a password account was lower for the system-generated passwords than for the 8-character ($p=0.003$). However, there was no significant difference between the system-generated and 16-character passwords ($p > 0.05$) or the 8-character and 16-character passwords ($p > 0.05$). The one-way ANOVA error rates are shown in Table 5.4, and the mean error rates are plotted in Figure 5.3.

Table 5.4: One-way ANOVA data for error rates

| Error Rate | SS | df | Mean Squares | F | Sig. |
|----------------|-------|----|-----------------|-------|-------|
| Between Groups | .409 | 2 | .204 | 4.959 | 0.011 |
| Within Groups | 2.101 | 51 | .041 | | |
| Total | 2.509 | 53 | | | |

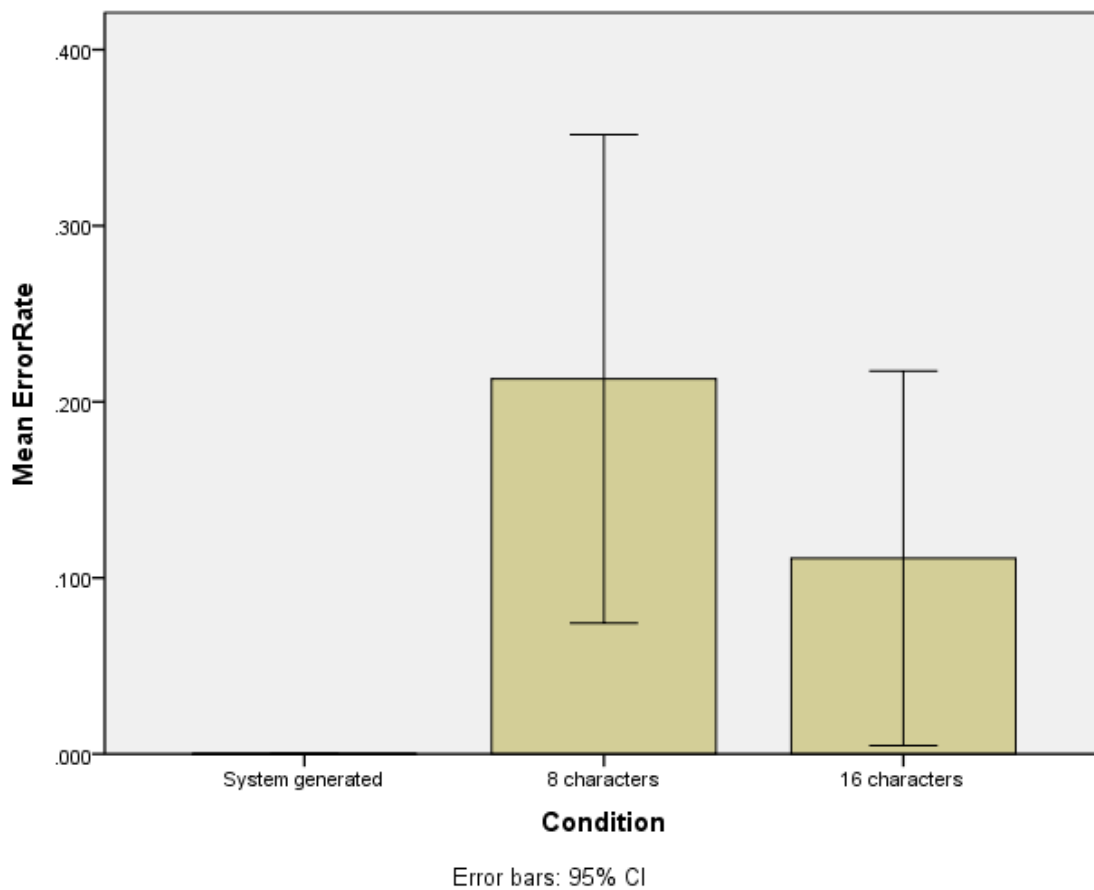


Figure 5.3: Mean error rates during creation of password accounts

Time taken to recall passwords. The time taken to recall the password, which includes the time taken to enter the passwords into the login application, was measured from the time the login application appeared to the time the participants completed the task. The descriptive statistics for this metric for both sessions are provided in Tables 5.5 and 5.6. Mean, standard deviation and error in the table are the transposed (reciprocal) values of the original time recorded in seconds:

Table 5.5: Descriptive statistics for the transposed value for the recall times for first session

| 1 st Session Recall Times | N | Mean | Standard Deviation | Std. Error |
|---|----|--------|-----------------------|---------------|
| System-generated | 18 | 0.1542 | .05008 | 0.01181 |
| 8-character | 18 | 0.0802 | .05369 | 0.01266 |
| 16-character | 18 | 0.0627 | .03346 | 0.00789 |
| Total | 54 | 0.0990 | .06074 | 0.00827 |

Table 5.6: Descriptive statistics for the transposed value for the recall times for second session

| 2 nd Session Recall Times | N | Mean | Standard Deviation | Std. Error |
|---|----|--------|-----------------------|---------------|
| System-generated | 18 | 0.0859 | .06584 | 0.01552 |
| 8-character | 18 | 0.0552 | .03805 | 0.00897 |
| 16-character | 18 | 0.0555 | .03549 | 0.00837 |
| Total | 54 | 0.0655 | .04970 | 0.00676 |

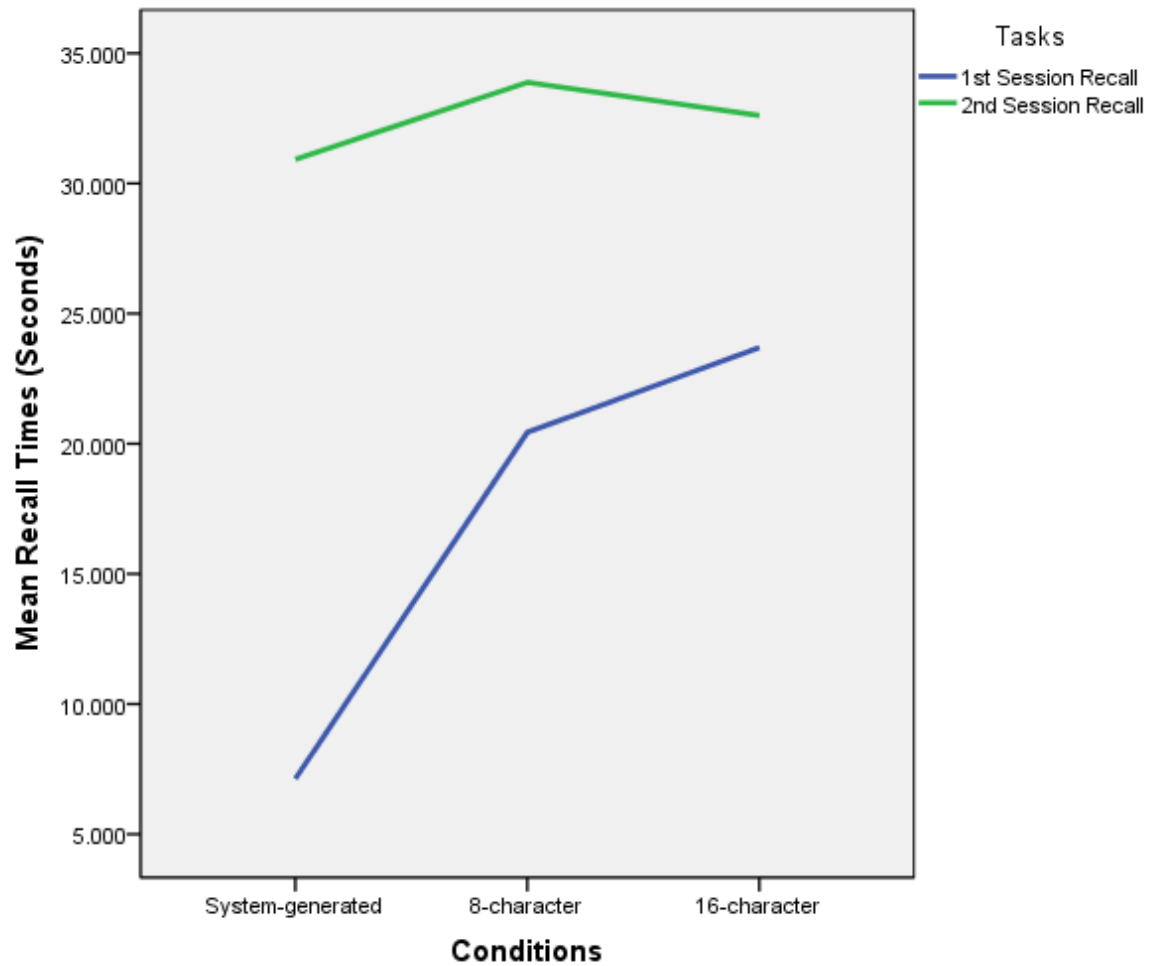
A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and the task sessions on the time taken to recall the passwords. The result indicated that the main effect was significant for both tasks, $F(1, 51)=15.634$, $p<0.001$ and password creation condition $F(2, 51)=15.170$, $p<0.001$. Subsequent post-hoc analysis of the task session main effect revealed that the time taken to recall a password was less for the first session than for the second ($p<0.001$). Analysis of the password creation condition main effect revealed that the time taken to recall a password was less for the system-generated passwords than for either the 8-character ($p<0.001$) or the 16-character passwords ($p<0.001$). The difference between the 8-character and 16-character password condition was not significant. The two-way mixed ANOVA data for the transposed value of the recall times are provided in Table 5.7:

Table 5.7: Two-way mixed ANOVA data for recall times

| Recall Times | SS | df | Mean Squares | F | Sig. |
|----------------------------|-------|----|-----------------|--------|-------|
| Task Sessions | .030 | 1 | .030 | 15.634 | 0.000 |
| Conditions | 0.078 | 2 | .039 | 15.170 | 0.000 |
| Task Sessions x Conditions | 0.018 | 2 | .009 | 4.584 | 0.015 |
| Error (Within-subjects) | .099 | 51 | .002 | | |
| Error (Between-subjects) | .132 | 51 | .003 | | |

The interaction effect of password policy conditions and task sessions on the time taken to recall passwords was significant, $F(2, 51)=4.584$, $p=0.015$. The interaction effects are plotted in Figure 5.4 and 5.5. Subsequent simple effects analysis of this interaction revealed that the time taken to recall a password in the first recall session was less for the system-generated passwords than for both the 8-character passwords ($p<0.001$) and the 16-character passwords ($p<0.001$). The results showed no significant difference between the 8-character and 16-character password conditions ($p >0.05$). There was no statistical significance across password conditions in the second session. A repeated measure ANOVA was conducted for the simple effect analysis to test the effect of task session on the time taken to recall system-generated passwords, 8-character passwords and 16-character passwords. The results indicated the effect of the task session on the system-generated passwords was significant, $F(1, 17)=17.527$, ($p=0.001$). Post-hoc analysis revealed that the time taken to recall system-generated passwords for the first

session was lower than for the second session ($p=0.001$). The effects of task sessions on the time taken to recall 8-character and 16-character passwords were not significant.



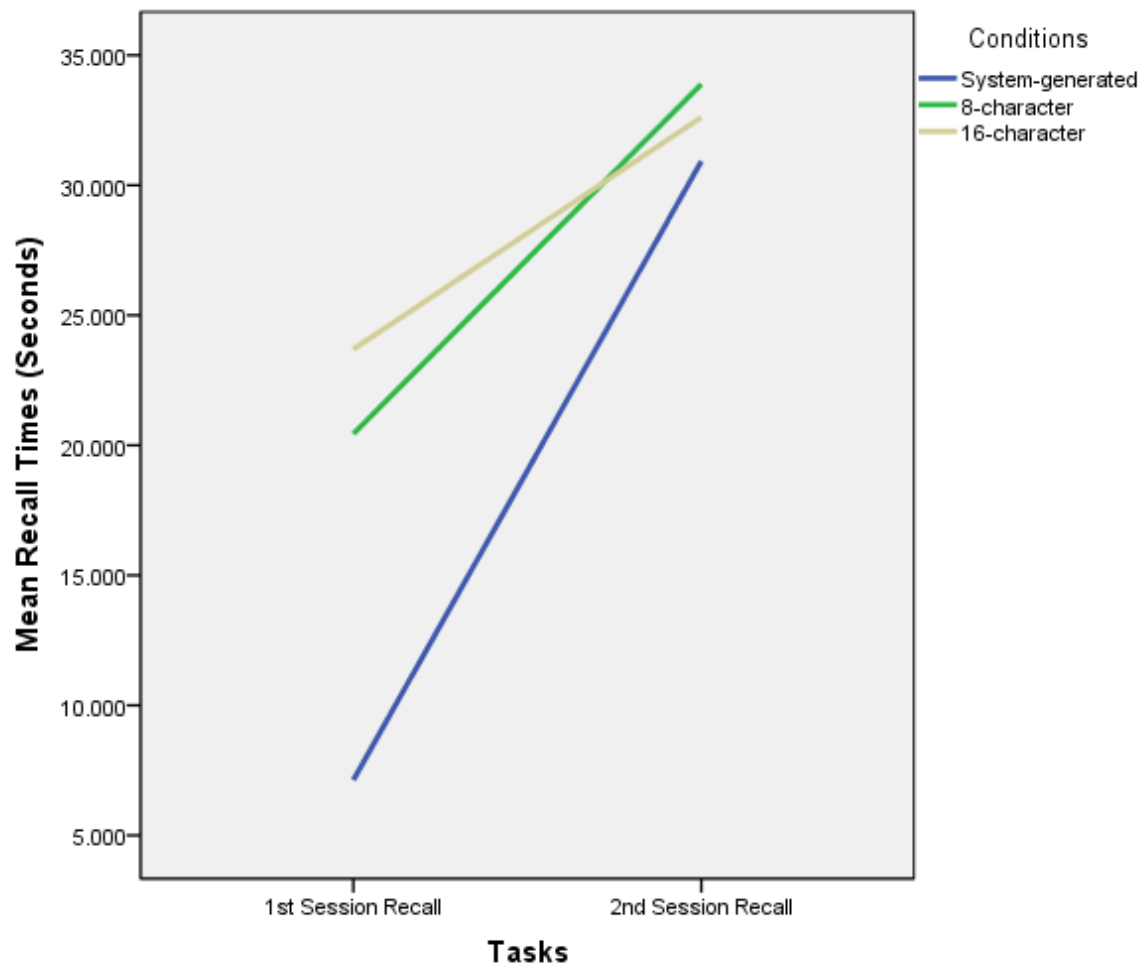
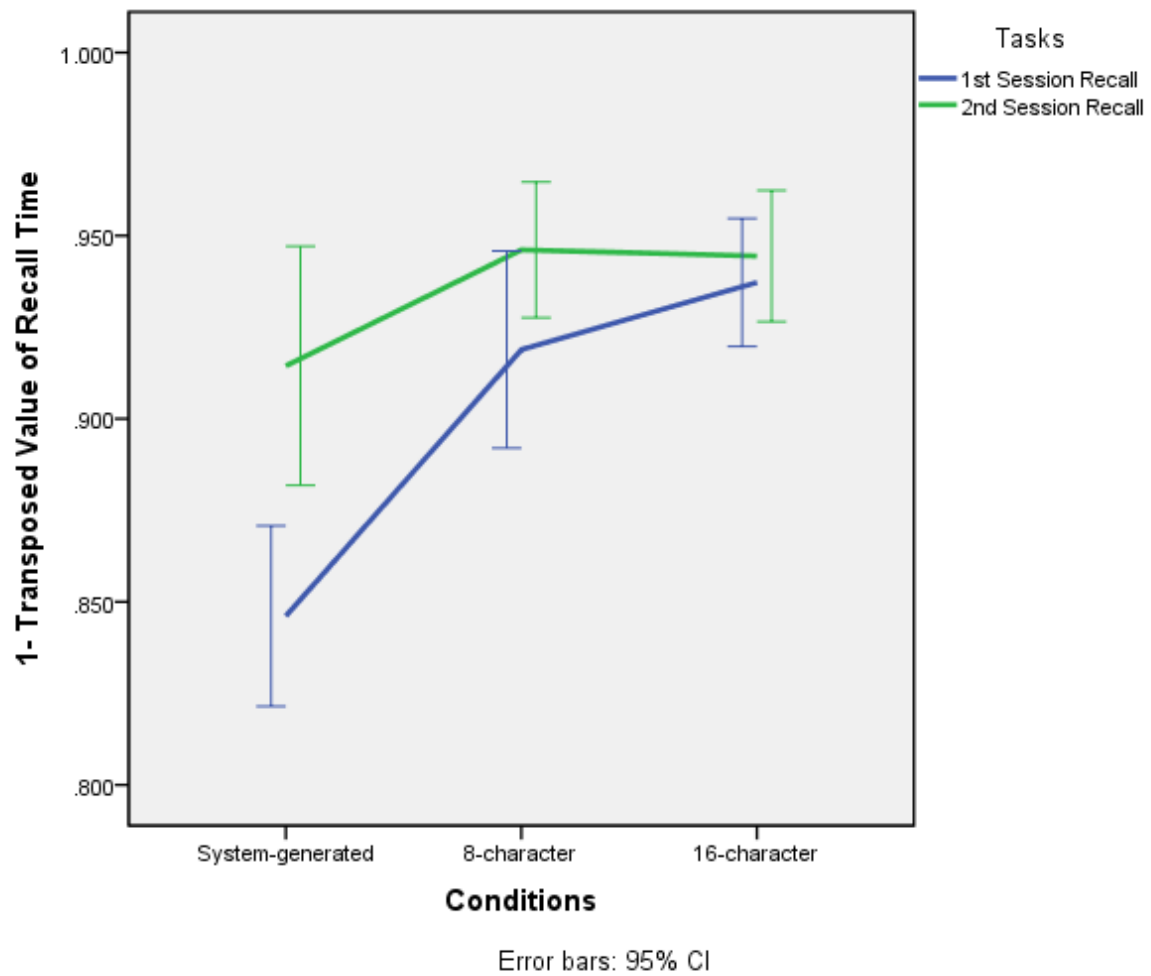


Figure 5.4: Interaction effect plots of the time taken to recall password (seconds)



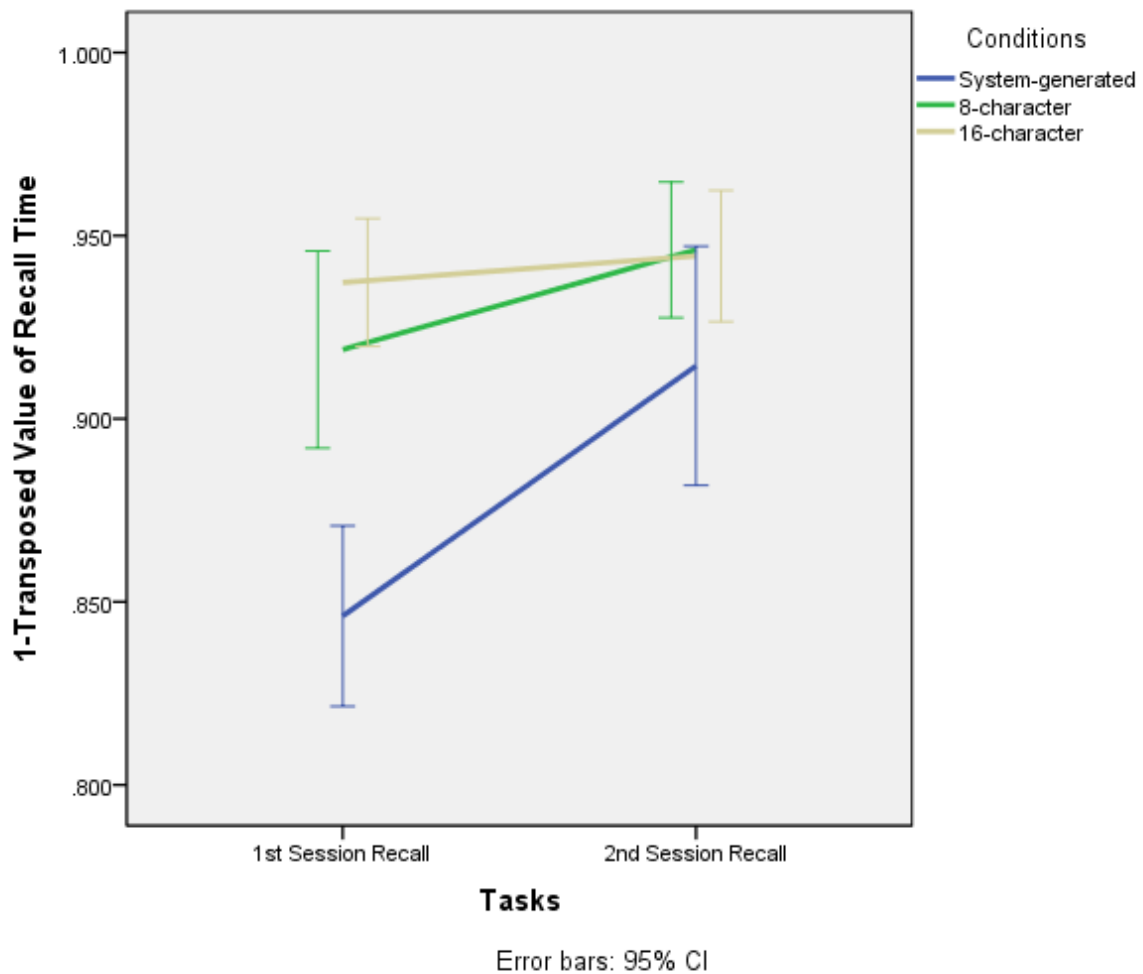


Figure 5.5³: Interaction effect plots of the reflected transformed values of time taken to recall

³(The transposed values are reflected in these graphs so that the higher values correspond to the longer recall time)

Recall error rates. The error rates, which specify the number of attempts taken to recall passwords in both sessions, were measured by dividing the number of errors by the

total number of attempts taken to recall the password. The descriptive statistics for this metric are provided in Tables 5.8 and 5.9:

Table 5.8: Descriptive statistics for recall error rates for first session

| 1 st Session Error Rates | N | Mean | Standard Deviation | Std. Error |
|--|----|---------|-----------------------|---------------|
| System-generated | 18 | 0.000 | .00000 | 0.00000 |
| 8-character | 18 | 0.21761 | .28748 | 0.06776 |
| 16-character | 18 | 0.12967 | .25288 | 0.05960 |
| Total | 54 | 0.11567 | .23486 | 0.31961 |

Table 5.9: Descriptive statistics for recall error rates for second session

| 2 nd Session Error Rates | N | Mean | Standard Deviation | Std. Error |
|--|----|---------|-----------------------|---------------|
| System-generated | 18 | 0.2777 | .42779 | 0.10083 |
| 8-character | 18 | 0.30094 | .41139 | 0.09696 |
| 16-character | 18 | 0.20372 | .35956 | 0.08475 |
| Total | 54 | 0.26081 | .39523 | 0.05378 |

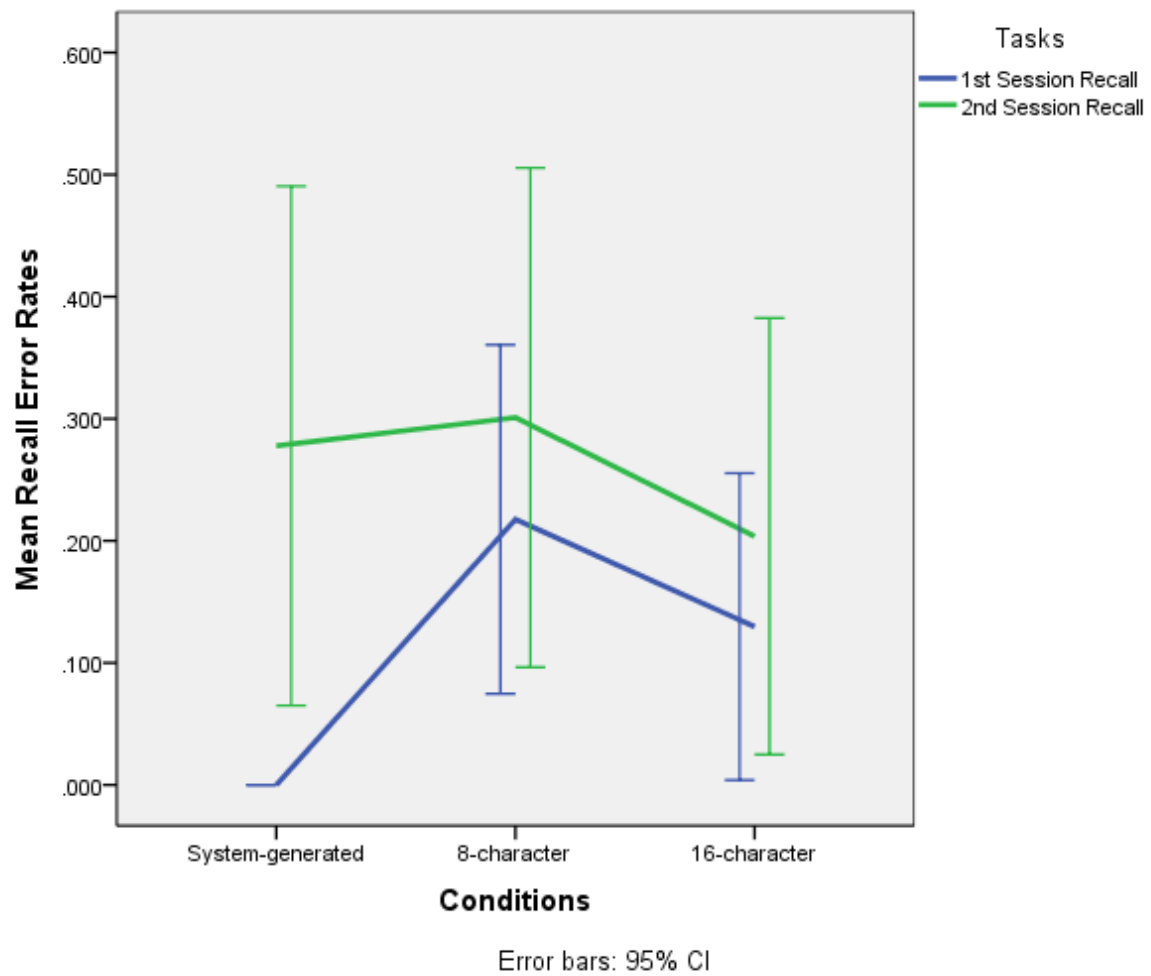
A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and the recall task sessions on recall error rates. The results of this analysis were significant for task sessions, $F(1, 51)=5.274$, $p=0.026$ but not significant for password policy conditions. Post-hoc analysis of the main effect of the

task session revealed that the recall error rate was lower for the first recall session than for the second ($p=0.026$). The interaction effect of password policy condition and recall task session on recall error rates was not significant ($p>0.05$). The two-way mixed ANOVA for error rates is provided in Table 5.10:

Table 5.10: Two-way mixed ANOVA data for recall error rates

| Error Rates | SS | df | Mean Squares | F | Sig. |
|----------------------------|-------|----|-----------------|-------|-------|
| Task Sessions | .568 | 1 | .568 | 5.274 | 0.026 |
| Conditions | 0.286 | 2 | .143 | 1.407 | 0.254 |
| Task Sessions x Conditions | 0.238 | 2 | .119 | 1.106 | 0.339 |
| Error (Within-subjects) | 5.493 | 51 | .108 | | |
| Error (Between-subjects) | 5.185 | 51 | .102 | | |

The descriptive statistics for recall error rates in the first session show that there were no errors in recalling system-generated passwords. This suggests that there must be a significant difference across password policy conditions in the first session. A one-way between subjects ANOVA was conducted to test the effect of the password policy conditions on recall error rates in the first session. The results indicated this effect was significant, $F(2, 51)=4.414$, $p=0.012$. This result should, however, be interpreted cautiously because the variance of recall error rates for system-generated passwords in the first session was zero. Zero variability violates the assumptions of the analytical techniques employed. The interaction effects are plotted in Figure 5.6.



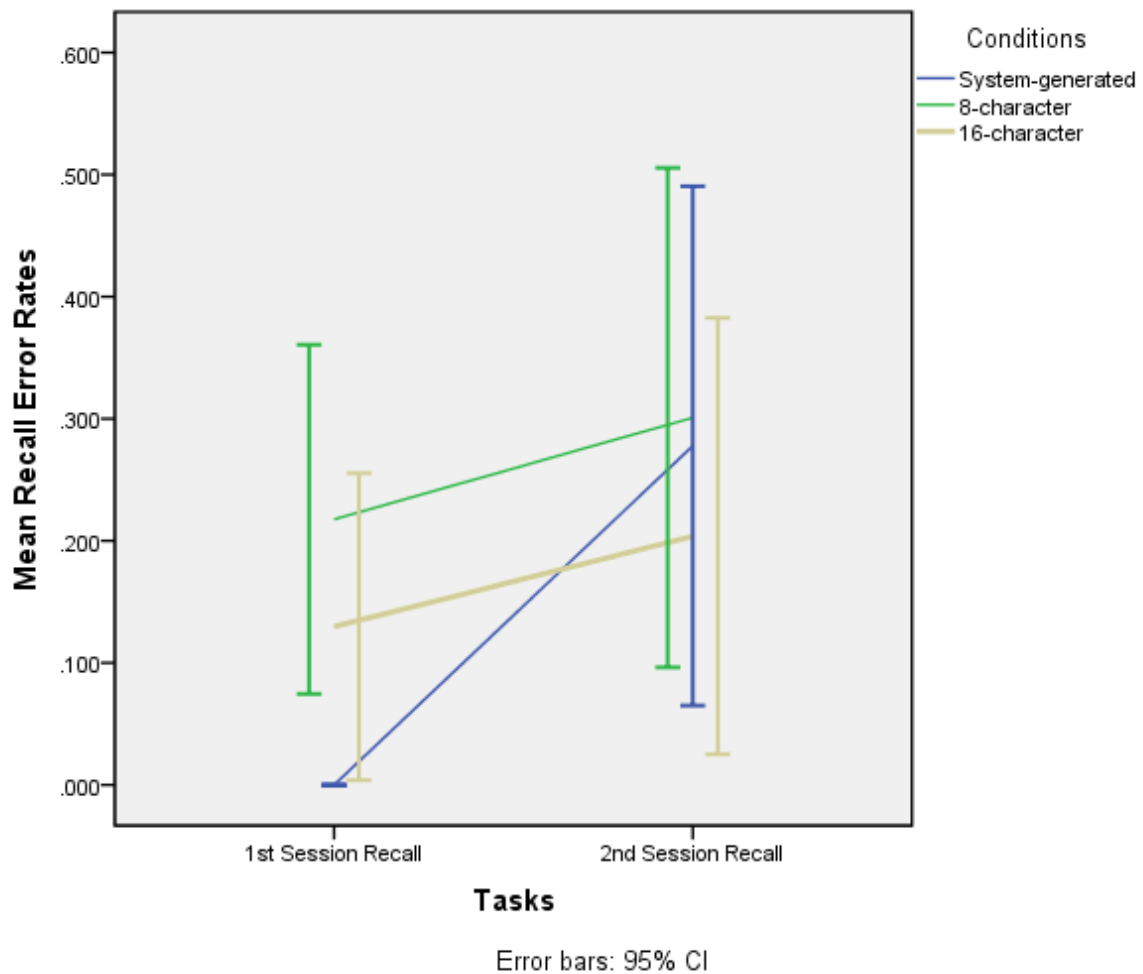


Figure 5.6: Interaction effect plots for recall error rates

Unrecoverable passwords. Three system-generated, three 8-character and two 16-character passwords could not be recalled in the first session. In the second session, four system-generated, three 8-character and two 16-character passwords could not be recalled.

The Fisher Exact Probability test was conducted to explore possible significant difference in the number of unrecoverable password across password policy conditions in the second session. The analysis used a conservative p-value of 0.897, revealing no statistically significant difference.

Edit distances. The edit distances were recorded in both sessions of the recall tasks when participants failed to recall their passwords. However, since participants were required to recall their passwords in the first session, participants who failed to do so were replaced by new ones. Consequently, only edit distances for the second session were statistically analyzed.

Damerau-Levenshtein edit distances. The Damerau-Levenshtein distance between the recalled and the stored passwords is the minimum number of operations needed to transform recalled passwords into those stored. From the incorrectly recalled passwords in the second session, the four system-generated recorded values were 4, 6, 1 and 1; the three 8-character passwords recorded of values were 1, 5 and 3; and the two 16-character passwords recorded values were 3 and 2. The remaining passwords that were correctly recalled recorded a value of zero.

Data for this dependent variable were non-normal. After reciprocal transformation, the skewness value remained lower than -2 with a high kurtosis value. These data suggest that this dependent variable was zero inflated with eighty-three percent of the data being zero.

Jaro-Winkler proximities. The Jaro-Winkler distance is a measure of difference between the stored and the recalled passwords. From the incorrectly recalled passwords in the second session, the four system-generated recorded values were 0.000, 0.944, 0.944 and 0.889; the three 8-character passwords recorded of values were 0.917, 0.778 and 0.963; and two 16-character passwords recorded values were 0.946 and 0.931. The remaining passwords that were correctly recalled recorded a value of one.

Data for this dependent variable were also non-normal. After reciprocal transformation, the skewness value remained higher than +2 along with a high kurtosis value. These data suggest that this dependent variable was one inflated with eighty-three percent of the data being one.

Subjective Measures

NASA Task Load Indices. The NASA-TLX assesses workload on six 7-point scales of mental, physical and temporal loads, performance, effort, and frustration with low and high end points. The NASA-TLX questionnaires were administered at the end of each task session, i.e., after first session--creation, first session--recall and second session--recall. The description of each subscale is provided in Table 5.11.

| Title | Endpoints | Descriptions |
|-----------------|-----------|--|
| Mental Demand | Low/High | How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? |
| Physical Demand | Low/High | How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious? |
| Temporal Demand | Low/High | How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic? |
| Performance | Good/Poor | How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals? |
| Effort | Low/High | How hard did you have to work (mentally and physically) to accomplish your level of performance? |
| Frustration | Low/High | How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task? |

Table 5.11: NASA-TLX rating scale definitions (Hart, 2006)

Mental Demand. A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and task sessions on the mental demand experienced by participants while creating and recalling passwords. The results indicated that main effects were not significant, sphericity assumed, $F(2, 102)=2.059$, $p>0.05$ for task sessions and $F(2, 51)=2.268$, $p>0.05$ for password policy. The interaction effect was also not significant, sphericity assumed, $F(4, 102)=1.155$, $p>0.05$. The descriptive statistics and two-way mixed ANOVA data for mental demand are provided in Tables 5.12, 5.13, 5.14 and 5.15:

Table 5.12: Descriptive statistics for mental demand during password creation

| 1 st Session Password creation | N | Mean | Standard Deviation | Std. Error |
|--|----|------|-----------------------|---------------|
| System-generated | 18 | 3.00 | 1.372 | 0.323 |
| 8-character | 18 | 4.00 | 1.749 | 0.412 |
| 16-character | 18 | 3.89 | 1.779 | 0.419 |
| Total | 54 | 3.63 | 1.674 | 0.228 |

Table 5.13: Descriptive statistics for mental demand during recall in first session

| 1 st Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 2.56 | 1.464 | 0.345 |
| 8-character | 18 | 3.83 | 1.948 | 0.459 |
| 16-character | 18 | 2.72 | 2.109 | 0.497 |
| Total | 54 | 2.04 | 1.913 | 0.260 |

Table 5.14: Descriptive statistics for mental demand during recall in second session

| 2 nd Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 3.33 | 2.000 | 0.471 |
| 8-character | 18 | 3.83 | 1.823 | 0.430 |
| 16-character | 18 | 2.89 | 2.220 | 0.523 |
| Total | 54 | 3.35 | 2.020 | 0.275 |

Table 5.15: Two-way mixed ANOVA data for mental demand

| Mental Demand | SS | df | Mean Squares | F | Sig. |
|----------------------------|---------|-----|-----------------|-------|-------|
| Task Sessions | 9.494 | 2 | 4.747 | 2.059 | 0.133 |
| Conditions | 25.568 | 2 | 12.784 | 2.268 | 0.114 |
| Task Sessions x Conditions | 10.654 | 4 | 2.664 | 1.155 | 0.335 |
| Error (Within-subject) | 235.185 | 102 | 2.306 | | |
| Error (Between-subject) | 5.185 | 51 | 5.636 | | |

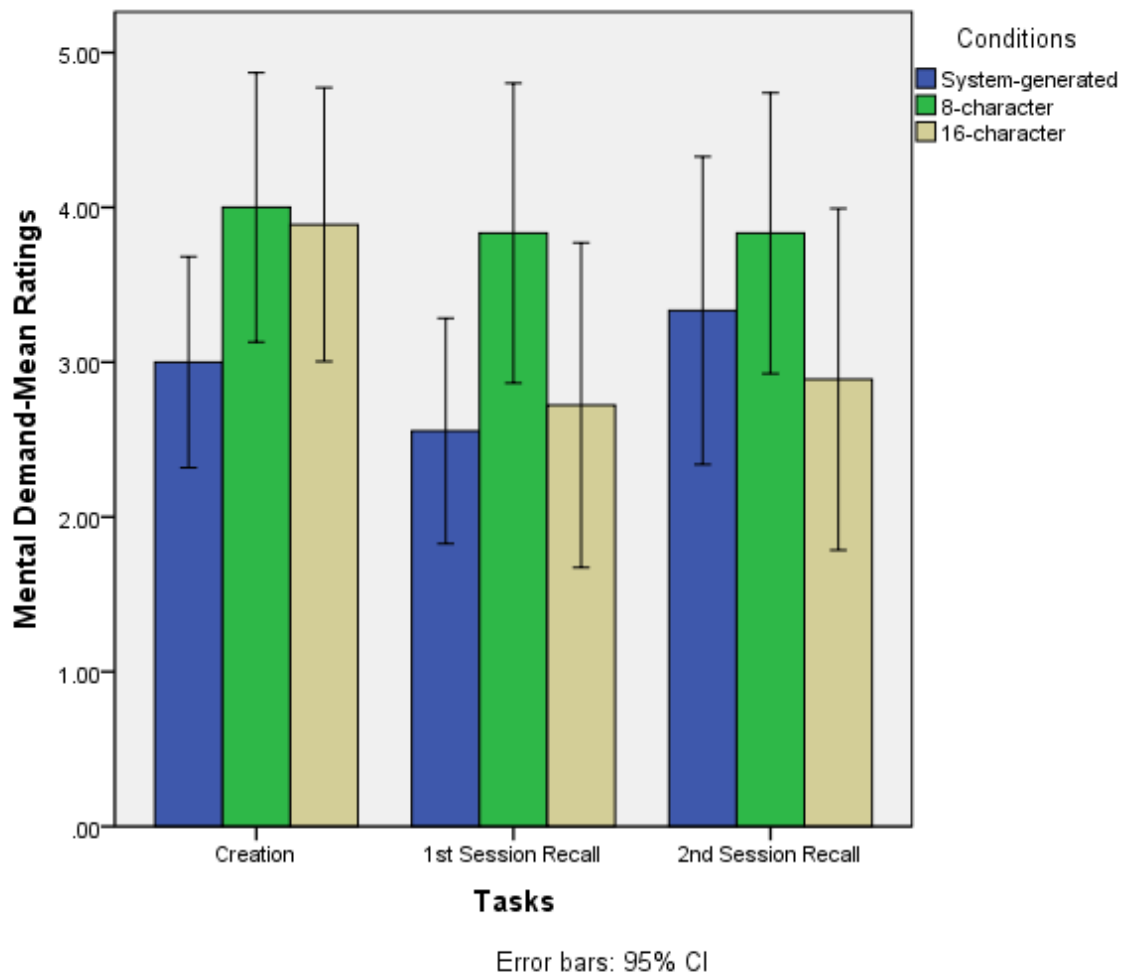


Figure 5.7: Mean rating for mental demand

Physical Demand. A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and task sessions on the physical demand experienced by participants while creating and recalling passwords. The results indicated the main effects were not significant, sphericity assumed, $F(2, 102)=0.567$, $p>0.05$ for task sessions and $F(2, 51)=1.126$, $p>0.05$ for password conditions. The

interaction effect was not significant, sphericity assumed, $F(4, 102)=0.693$, $p>0.05$. The descriptive statistics and two-way mixed ANOVA data for physical demand are provided in Tables 5.16, 5.17, 5.18 and 5.19:

Table 5.16: Descriptive statistics for physical demand during password creation

| 1 st Session Password creation | N | Mean | Standard Deviation | Std. Error |
|--|----|------|-----------------------|---------------|
| System-generated | 18 | 1.89 | 1.183 | 0.279 |
| 8-character | 18 | 2.06 | 1.162 | 0.274 |
| 16-character | 18 | 1.61 | 1.037 | 0.244 |
| Total | 54 | 1.85 | 1.123 | 0.153 |

Table 5.17: Descriptive statistics for physical demand during recall in first session

| 1 st Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 1.33 | .767 | 0.181 |
| 8-character | 18 | 2.06 | 1.392 | 0.328 |
| 16-character | 18 | 1.67 | 1.138 | 0.268 |
| Total | 54 | 1.69 | 1.146 | 0.156 |

Table 5.18: Descriptive statistics for physical demand during recall in second session

| 2 nd Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 1.72 | 1.179 | 0.278 |
| 8-character | 18 | 2.00 | 1.237 | 0.291 |
| 16-character | 18 | 1.83 | 1.043 | 0.246 |
| Total | 54 | 1.85 | 1.139 | 0.155 |

Table 5.19: Two-way mixed ANOVA table for physical demand

| Physical Demand | SS | df | Mean Squares | F | Sig. |
|----------------------------|---------|-----|-----------------|-------|-------|
| Task Sessions | 1.000 | 2 | 0.500 | 0.567 | 0.569 |
| Conditions | 4.778 | 2 | 2.389 | 1.126 | 0.332 |
| Task Sessions x Conditions | 2.444 | 4 | 0.661 | 0.693 | 0.598 |
| Error (Within-subjects) | 89.889 | 102 | 0.881 | | |
| Error (Between-subjects) | 108.167 | 51 | 2.121 | | |

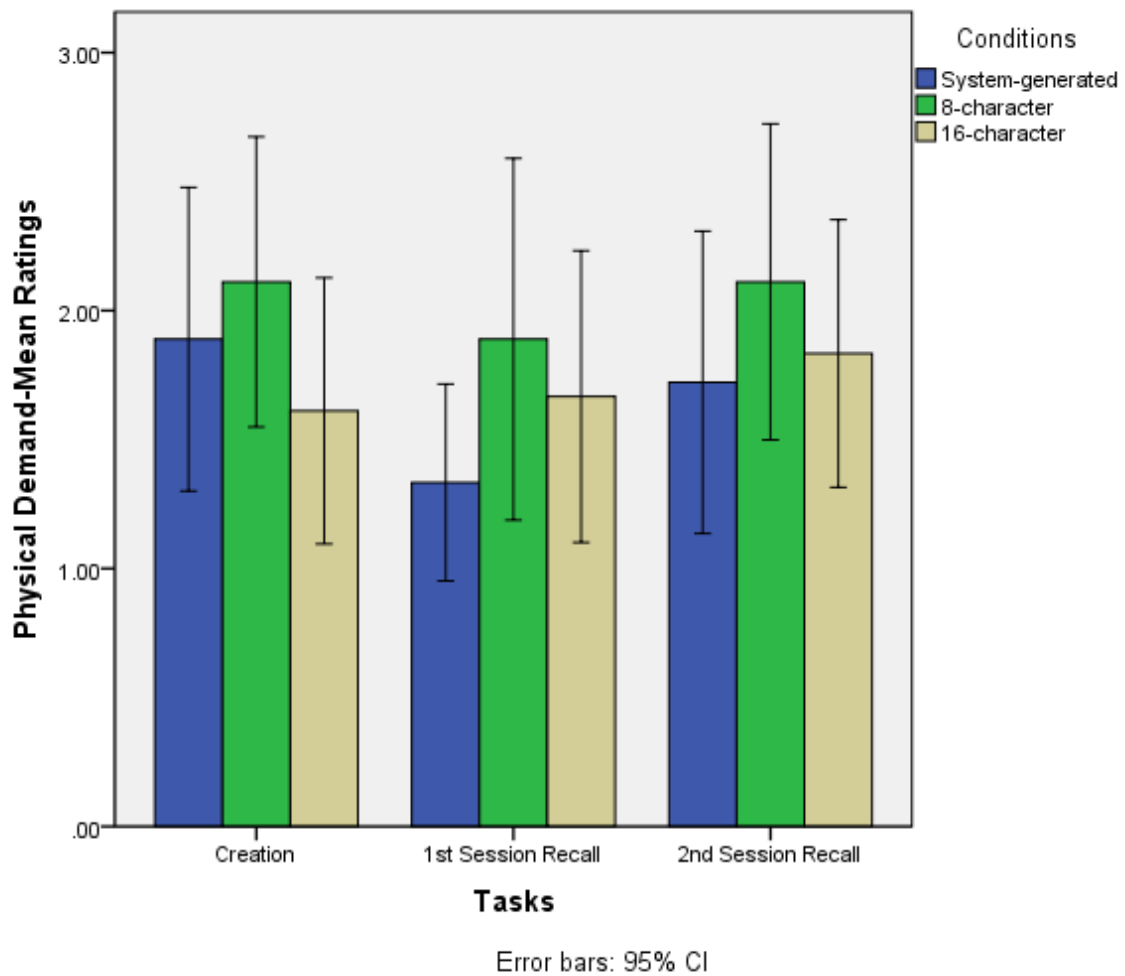


Figure 5.8: Mean rating for physical demand

Temporal Demand. A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and task sessions on the temporal demand experienced by participants while creating and recalling passwords. The results indicated that the main effect of task session approached significance, Wilks' Lambda, $F(2, 50)=2.723$, $p=0.075$. The main effect for password policy was significant, $F(2,$

51)=4.860, $p=0.012$. Subsequent post-hoc analysis of the task session main effect revealed that the temporal demand was higher during the creation of the password account than for the recall of the same password in the same session ($p=0.039$). There was no significant difference between the recall of the password in the first and second sessions ($p>0.05$). Post-hoc analysis of the main effects of password policy revealed that the temporal demand was higher for the 8-character user-generated passwords than for system-generated passwords ($p=0.003$). The interaction effect was not significant, Wilks' Lambda, $F(4, 100)=0.072$, $p>0.05$. The descriptive statistics and two-way mixed ANOVA data for temporal demand are provided in Table 5.20, 5.21, 5.22 and 5.23.

Table 5.20: Descriptive statistics for temporal demand during password creation

| 1 st Session Password creation | N | Mean | Standard Deviation | Std. Error |
|--|----|------|-----------------------|---------------|
| System-generated | 18 | 2.17 | 1.249 | 0.294 |
| 8-character | 18 | 3.28 | 1.742 | 0.4113 |
| 16-character | 18 | 2.67 | 1.715 | 0.404 |
| Total | 54 | 2.70 | 1.621 | 0.221 |

Table 5.21: Descriptive statistics for temporal demand during recall in first session

| 1 st Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 1.83 | 1.425 | 0.336 |
| 8-character | 18 | 2.89 | 1.779 | 0.419 |
| 16-character | 18 | 2.11 | 1.278 | 0.301 |
| Total | 54 | 2.28 | 1.547 | 0.211 |

Table 5.22: Descriptive statistics for temporal demand during recall in second session

| 2 nd Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 1.72 | 0.826 | 0.195 |
| 8-character | 18 | 2.72 | 1.674 | 0.394 |
| 16-character | 18 | 2.22 | 1.865 | 0.440 |
| Total | 54 | 2.22 | 1.550 | 0.211 |

Table 5.23: Two-way mixed ANOVA data for temporal demand

| Temporal Demand | df | F | Sig. |
|----------------------------|-----|-------|-------|
| Task Sessions | 2 | 2.723 | 0.075 |
| Conditions | 2 | 4.860 | 0.012 |
| Task Sessions x Conditions | 4 | 0.072 | 0.990 |
| Error (Within-subjects) | 100 | | |
| Error (Between-subjects) | 51 | | |
| Error (Task Sessions) | 50 | | |

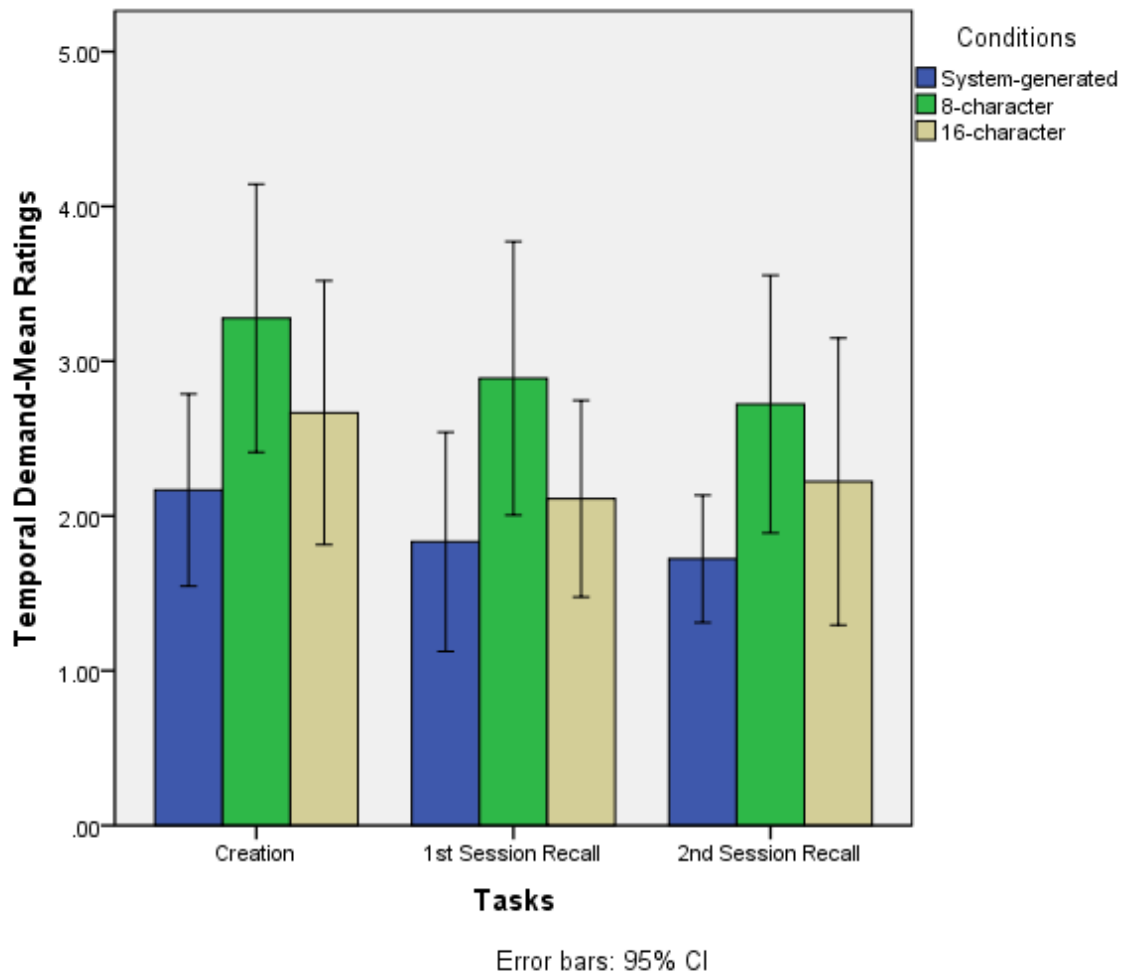


Figure 5.9: Mean rating for temporal demand

Performance. A two-way mixed ANOVA was conducted to test the main and interaction effects of password policy and task session on the performance component of the NASA-TLX while creating and recalling passwords. The results indicated the main effect of the task session was significant, Wilks' Lambda, $F(2, 50)=7.058$, $p=0.002$ and main effect of the password policy was not significant, $F(2, 51)=2.405$, $p>0.05$. Subsequent post-hoc analysis of the within-subject main effects revealed that the performance component was higher for the creation of the password account than for the recall of the same password in the same session ($p=0.002$) and higher for recall in the second session than for recall in the first session ($p=0.019$). The interaction effect was not significant, Wilks' Lambda, $F(4, 100)=0.582$, $p>0.05$. The descriptive statistics and two-way mixed ANOVA data for performance are provided in Tables 5.24, 5.25, 5.26 and 5.27:

Table 5.24: Descriptive statistics for performance during password creation

| 1 st Session Password creation | N | Mean | Standard Deviation | Std. Error |
|--|----|------|-----------------------|---------------|
| System-generated | 18 | 2.28 | 1.406 | 0.331 |
| 8-character | 18 | 2.83 | 1.581 | 0.373 |
| 16-character | 18 | 2.11 | 1.605 | 0.378 |
| Total | 54 | 2.41 | 1.536 | 0.209 |

Table 5.25: Descriptive statistics for performance during recall in first session

| 1 st Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 1.50 | 1.339 | 0.316 |
| 8-character | 18 | 2.39 | 1.614 | 0.380 |
| 16-character | 18 | 1.56 | 1.294 | 0.305 |
| Total | 54 | 1.81 | 1.455 | 0.198 |

Table 5.26: Descriptive statistics for performance during recall in second session

| 2 nd Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 2.72 | 2.469 | 0.582 |
| 8-character | 18 | 3.06 | 2.235 | 0.527 |
| 16-character | 18 | 1.78 | 1.734 | 0.409 |
| Total | 54 | 2.52 | 2.196 | 0.299 |

Table 5.27: Two-way mixed ANOVA data for performance

| Performance | df | F | Sig. |
|----------------------------|-----|-------|-------|
| Task Sessions | 2 | 7.058 | 0.002 |
| Conditions | 2 | 2.405 | 0.100 |
| Task Sessions x Conditions | 4 | 0.582 | 0.676 |
| Error (Within-subjects) | 100 | | |
| Error (Between-subjects) | 51 | | |
| Error (Task Sessions) | 50 | | |

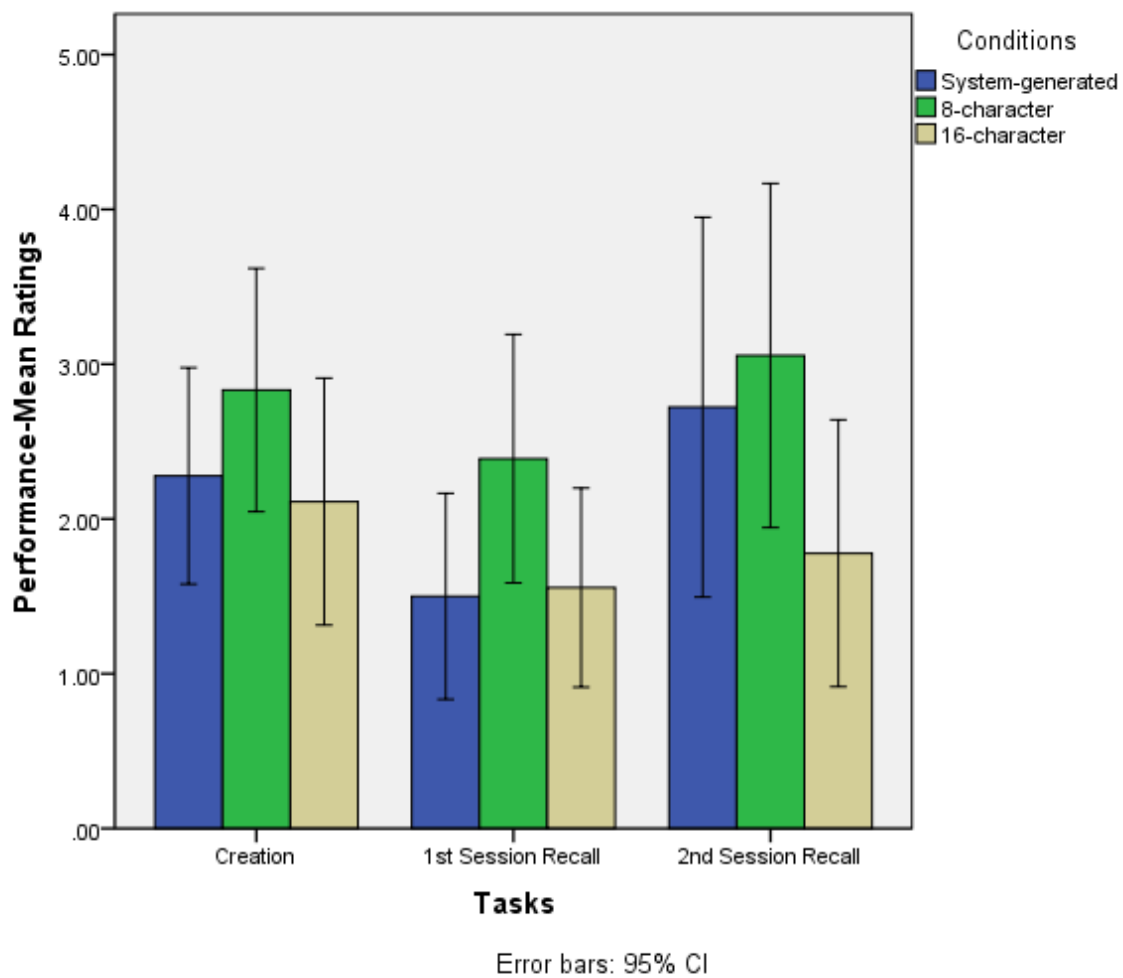


Figure 5.10: Mean rating for performance

Effort. A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and task sessions on the effort required by the participants while creating and recalling passwords. The results indicated that the main effects were not significant, Wilks' Lambda, $F(2, 50)=1.661$, $p>0.05$ for the task session and $F(2, 51)=1.817$, $p>0.05$ for the password policy. The interaction effect was also not

significant, Wilks' Lambda, $F(4, 100)=0.827$, $p>0.05$. The descriptive statistics and two-way mixed ANOVA data for effort are provided in Tables 5.28, 5.29, 5.30 and 5.31:

Table 5.28: Descriptive statistics for effort during password creation

| 1 st Session Password creation | N | Mean | Standard Deviation | Std. Error |
|--|----|------|-----------------------|---------------|
| System-generated | 18 | 2.83 | 1.425 | 0.336 |
| 8-character | 18 | 3.22 | 1.734 | 0.409 |
| 16-character | 18 | 3.17 | 1.505 | 0.355 |
| Total | 54 | 3.07 | 1.540 | 0.210 |

Table 5.29: Descriptive statistics for effort during recall in first session

| 1 st Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 2.06 | 1.259 | 0.297 |
| 8-character | 18 | 3.33 | 1.879 | 0.443 |
| 16-character | 18 | 2.72 | 1.742 | 0.411 |
| Total | 54 | 2.70 | 1.700 | 0.231 |

Table 5.30: Descriptive statistics for effort during recall in second session

| 1 st Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 2.78 | 1.734 | 0.409 |
| 8-character | 18 | 3.50 | 1.757 | 0.414 |
| 16-character | 18 | 2.89 | 2.026 | 0.478 |
| Total | 54 | 3.06 | 1.837 | 0.250 |

Table 5.31: Two-way mixed ANOVA data for effort

| Effort | df | F | Sig. |
|----------------------------|-----|-------|-------|
| Task Sessions | 2 | 1.661 | 0.200 |
| Conditions | 2 | 1.817 | 0.173 |
| Task Sessions x Conditions | 4 | 0.827 | 0.511 |
| Error (Within-subjects) | 100 | | |
| Error (Between-subjects) | 51 | | |
| Error (Task Sessions) | 50 | | |

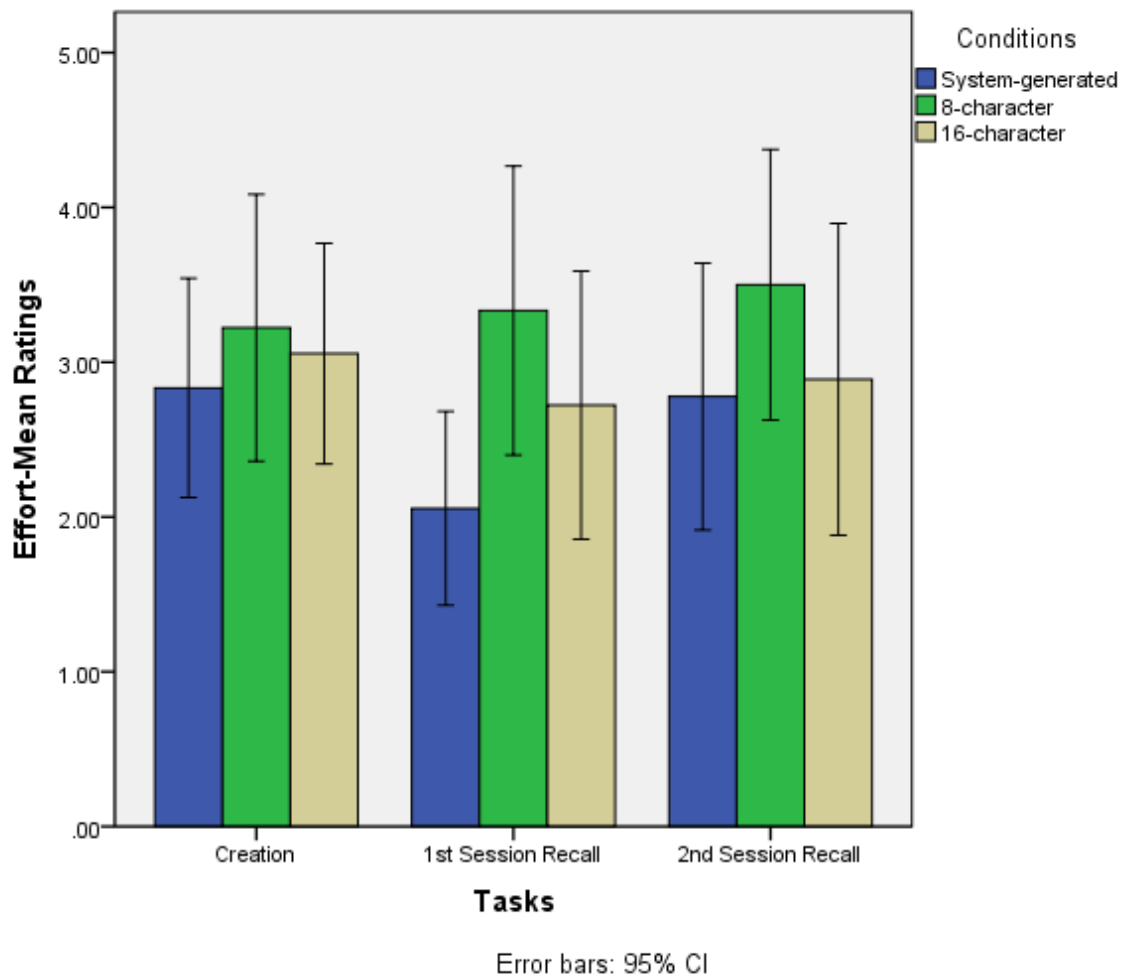


Figure 5.11: Mean rating for effort

Frustration. A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and task sessions on the frustration experienced by the participants while creating and recalling passwords. The results indicated that main effects were not significant, Wilks' Lambda, $F(2, 50)=0.235$, $p>0.05$ for the task session and $F(2, 51)=2.037$, $p>0.05$ for the password policy. The interaction

effect was also not significant, Wilks' Lambda, $F(4, 100)=1.147, p>0.05$. The descriptive statistics and two-way mixed ANOVA data for frustration are provided in Tables 5.32, 5.33, 5.34 and 5.35:

Table 5.32: Descriptive statistics for frustration during password creation

| 1 st Session Password creation | N | Mean | Standard Deviation | Std. Error |
|--|----|------|-----------------------|---------------|
| System-generated | 18 | 2.17 | 1.425 | 0.336 |
| 8-character | 18 | 2.72 | 1.708 | 0.403 |
| 16-character | 18 | 2.39 | 1.819 | 0.429 |
| Total | 54 | 2.43 | 1.644 | 0.224 |

Table 5.33: Descriptive statistics for frustration during recall in first session

| 1 st Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 1.61 | 1.092 | 0.257 |
| 8-character | 18 | 2.89 | 1.641 | 0.387 |
| 16-character | 18 | 2.50 | 1.543 | 0.364 |
| Total | 54 | 2.33 | 1.517 | 0.206 |

Table 5.34: Descriptive statistics for frustration during recall in second session

| 2 nd Session Recall | N | Mean | Standard Deviation | Std. Error |
|-----------------------------------|----|------|-----------------------|---------------|
| System-generated | 18 | 2.06 | 1.798 | 0.424 |
| 8-character | 18 | 2.56 | 1.790 | 0.422 |
| 16-character | 18 | 2.06 | 1.474 | 0.347 |
| Total | 54 | 2.22 | 1.679 | 0.228 |

Table 5.35: Two-way mixed ANOVA data for frustration

| Frustration | df | F | Sig. |
|----------------------------|-----|-------|-------|
| Task Sessions | 2 | 0.235 | 0.791 |
| Conditions | 2 | 2.037 | 0.141 |
| Task Sessions x Conditions | 4 | 1.147 | 0.339 |
| Error (Within-subjects) | 100 | | |
| Error (Between-subjects) | 51 | | |
| Error (Task Sessions) | 50 | | |

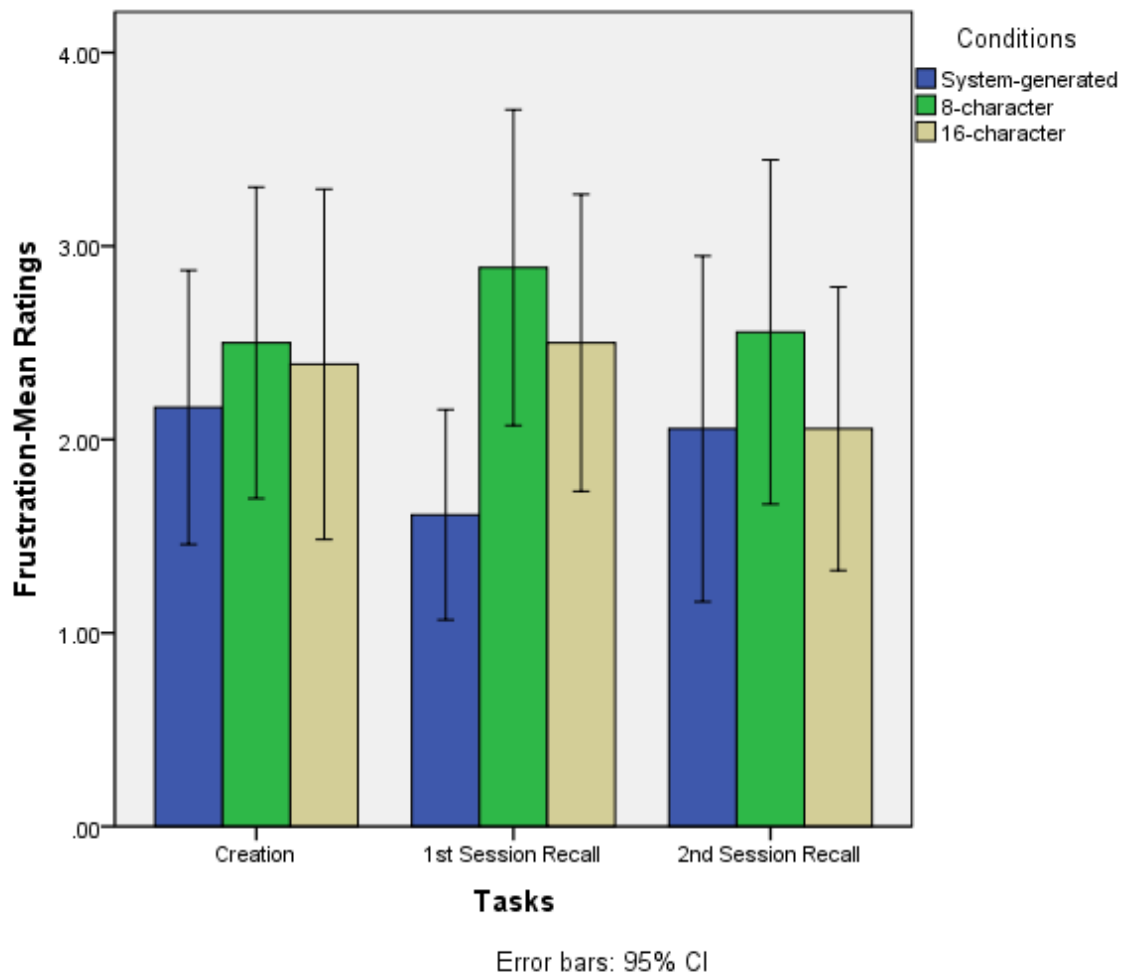


Figure 5.12: Mean rating for frustration

System Usability Scale (SUS) questionnaire. The System Usability Scale, a ten-item Likert-scale questionnaire that records global subjective assessment of the usability of a system, has a range of scores from 0-100. The SUS questionnaires were administered at the end of each task, i.e., 1st session--creation, 1st session--recall and 2nd session--recall. Refer to tables 5.36, 5.37 and 5.38:

Table 5.36: Descriptive statistics of the SUS scores for password account creation

| 1 st Session-Creation SUS score | N | Mean | Standard Deviation |
|---|----|---------|-----------------------|
| System-generated | 18 | 63.4722 | 19.65023 |
| 8-character | 18 | 68.0556 | 16.19176 |
| 16-character | 18 | 66.1111 | 17.55710 |
| Total | 54 | 65.8796 | 17.61858 |

Table 5.37: Descriptive statistics of the SUS scores for password recall in first session

| 2 nd Session-Recall SUS score | N | Mean | Standard Deviation |
|---|----|---------|-----------------------|
| System-generated | 18 | 56.6667 | 23.68606 |
| 8-character | 18 | 61.9444 | 16.63968 |
| 16-character | 18 | 61.8056 | 21.31358 |
| Total | 54 | 60.1389 | 20.50895 |

Table 5.38: Descriptive statistics of the SUS scores for password recall in second session

| 1 st Session Error Rates | N | Mean | Standard Deviation |
|--|----|---------|-----------------------|
| System-generated | 18 | 56.5278 | 19.38653 |
| 8-character | 18 | 62.0833 | 18.03285 |
| 16-character | 18 | 66.5278 | 21.43970 |
| Total | 54 | 61.7130 | 19.73183 |

A two-way mixed ANOVA was conducted to test the main and interaction effects of the password policy conditions and the task sessions on system usability while creating and recalling passwords. The results indicated that the main effect of task session was significant, sphericity assumed, $F(2, 102) = 3.766, p = 0.026$. The main effect of password policy was not significant, sphericity assumed, $F(2, 51) = 0.633, p > 0.05$. Post-hoc analysis of the task session main effect revealed that the SUS score was higher during the creation of the password account than during the recall of the same password in the same session ($p = 0.007$). There were no significant differences between other task sessions. The interaction effect was not significant, sphericity assumed, $F(4, 102) = 0.597, p > 0.05$. The two-way mixed ANOVA data for the SUS scores are provided in Table 5.39.

Table 5.39: Two-way mixed ANOVA data for SUS score

| SUS score | SS | df | Mean Squares | F | Sig. |
|----------------------------|-----------|-----|-----------------|-------|-------|
| Task Sessions | 950.309 | 2 | 475.154 | 3.766 | 0.026 |
| Conditions | 1118.596 | 2 | 559.298 | 0.633 | 0.535 |
| Task Sessions x Conditions | 301.312 | 4 | 75.328 | 0.597 | 0.666 |
| Error (Within-subjects) | 89.889 | 102 | 126.169 | | |
| Error (Between-subjects) | 45090.856 | 51 | 884.134 | | |

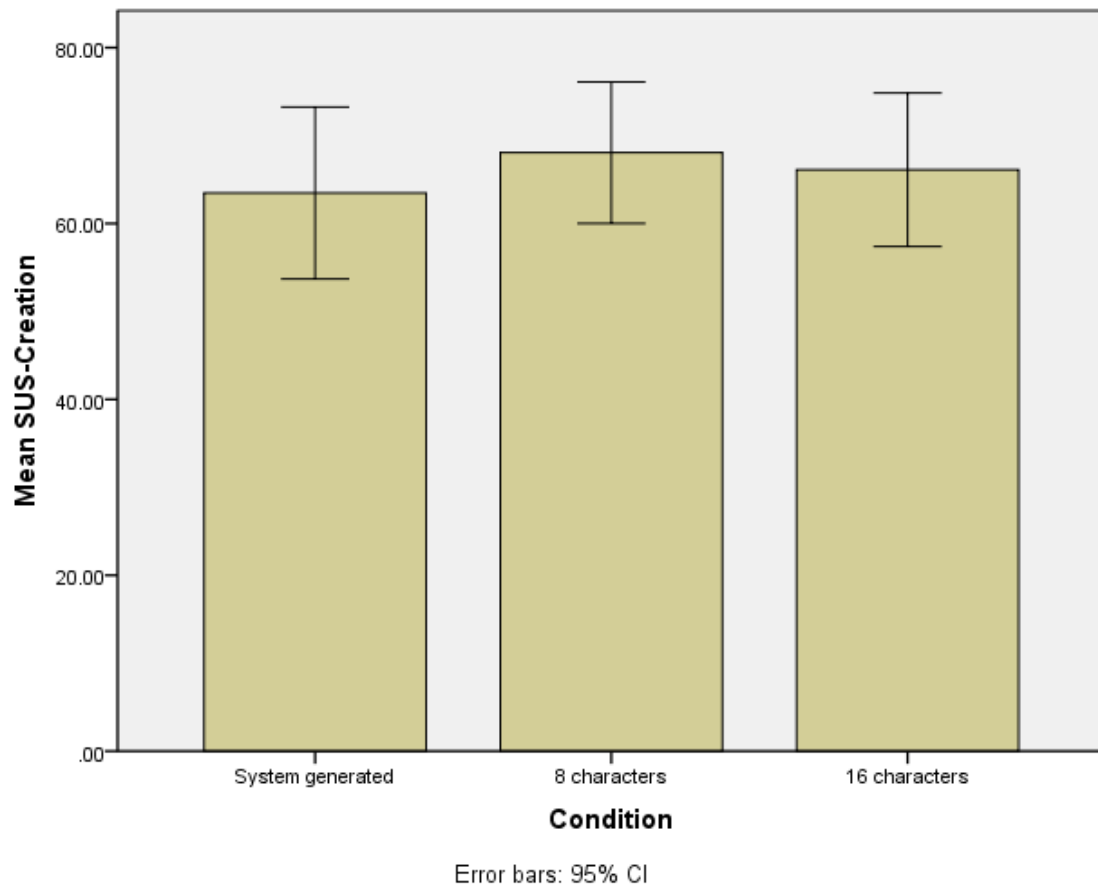


Figure 5.13: Mean SUS for creation task

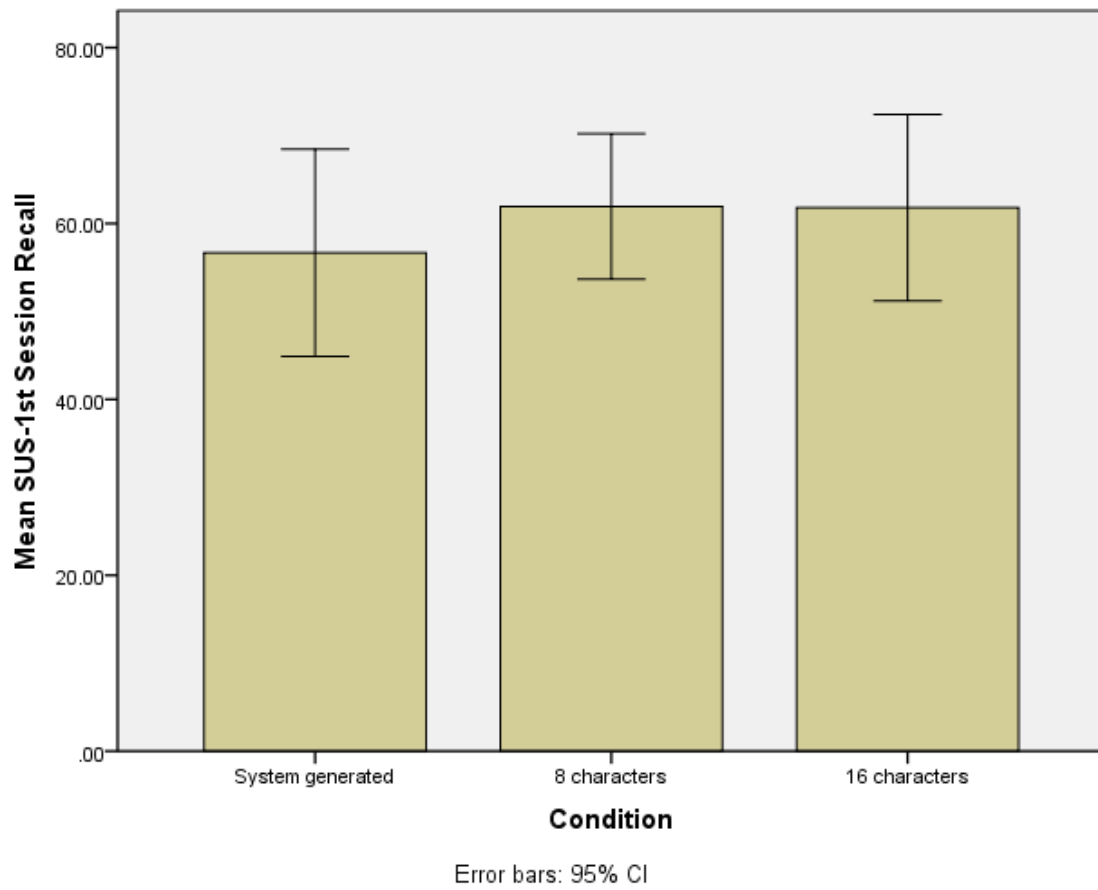


Figure 5.14: Mean SUS for second session recall

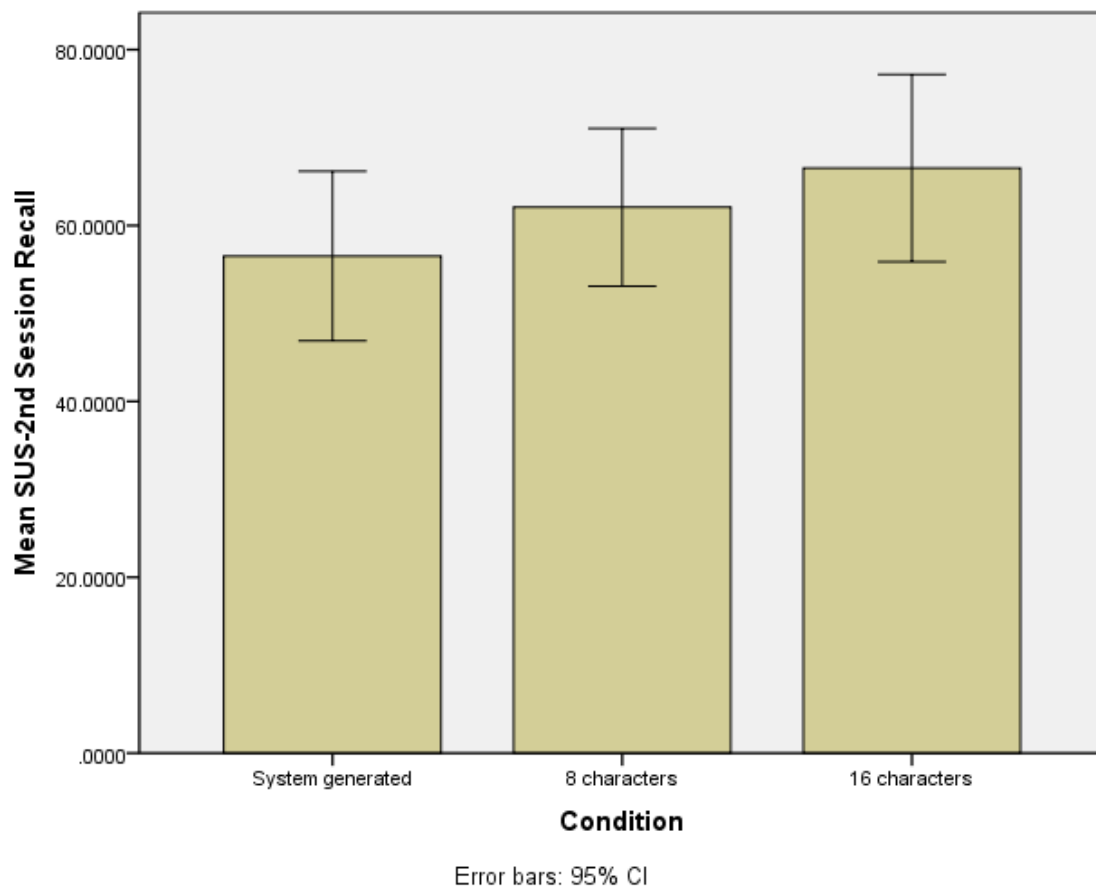


Figure 5.15: Mean SUS for second session recall

6. DISCUSSION

Overall, the results from this study supported one hypothesis, that it would take less time to create an account with the system-generated password than with the other two password conditions; however, the remaining hypotheses were not supported.

Significant differences were found in error rates for the creation of passwords, the time taken to recall the password, error rates during recall and temporal demand across password conditions. Across task sessions, the time taken to recall system-generated passwords, the error rates during recall, the performance index of the NASA-TLX and the SUS scores were found to be significantly different. The overall relative performance among password policy conditions for dependent variables during each task sessions is provided in Tables 6.1, 6.2 and 6.3:

Table 6.1: Relative performance of policy conditions during password account creation

| Dependent Variables | System-generated password | 8-character password | 16-character password |
|---|---------------------------|----------------------|-----------------------|
| Time take to create password account | Low | Medium | High |
| Password account creation error rates | Low | High | Medium |
| NASA TLX-Mental Demand: Password creation | Low | High | Medium |
| NASA TLX-Physical Demand: Password creation | Medium | High | Low |
| NASA TLX-Temporal Demand: Password creation | Low | High | Medium |
| NASA TLX-Performance: Password creation | Medium | Poor | Good |
| NASA TLX-Effort: Password creation | Low | High | Medium |
| NASA TLX-Frustration: Password creation | Low | High | Medium |
| SUS Score- Password creation | Low | High | Medium |

Table 6.2: Relative performance of policy conditions during first session recall

| Dependent Variables | System-generated password | 8-character password | 16-character password |
|--|---------------------------|----------------------|-----------------------|
| Time taken to recall passwords: First session | Low | Medium | High |
| Recall error rates in the first session | Low | High | Medium |
| NASA TLX-Mental Demand: First session recall | Low | High | Medium |
| NASA TLX-Physical Demand: First session recall | Low | High | Medium |
| NASA TLX-Temporal Demand: First session recall | Low | High | Medium |
| NASA TLX-Performance: First session recall | Good | Poor | Medium |
| NASA TLX-Effort: First session recall | Low | High | Medium |
| NASA TLX-Frustration: First session recall | Low | High | Medium |
| SUS Score- First session recall | Low | High | Medium |

Table 6.3: Relative performance of policy conditions during second session recall

| Dependent Variables | System-generated password | 8-character password | 16-character password |
|---|---------------------------|----------------------|-----------------------|
| Time taken to recall password: Second session | Low | High | Medium |
| Recall error rates in the second session | Medium | High | Low |
| Damerau-Levenshtein edit distances | High | Medium | Low |
| Jaro-Winkler proximities | Low | Medium | High |
| Unrecoverable password: Second session | High | Medium | Low |
| NASA TLX-Mental Demand: Second session recall | Medium | High | Low |
| NASA TLX-Physical Demand: Second session recall | Low | High | Medium |
| NASA TLX-Temporal Demand: Second session recall | Low | High | Medium |
| NASA TLX-Performance: Second session recall | Medium | Poor | Good |
| NASA TLX-Effort: Second session recall | Low | High | Medium |
| NASA TLX-Frustration: Second session recall | Low | High | Low |
| SUS Score- Second session recall | Low | Medium | High |

Password Account Creation

It took significantly less time to create an account with the system-generated password, followed by the 8-character and the 16-character passwords. The reason the latter two passwords took longer is likely because of the two stages required for account creation with user-generated passwords, i.e. the participants had to both create and memorize their passwords while complying with the restrictions presented. In contrast, the creation of the system-generated password account involved only memorization of an

assigned 6-character alphanumeric password. These findings are partially supported by the study conducted by Proctor et al. (2002), who found that the time taken to create passwords with only a minimum length restriction was less than for passwords of the same minimum length but with additional restrictions.

In addition, the participants who were assigned system-generated passwords committed no errors during the creation of their password accounts. While the participants who created 16-character passwords committed fewer errors than those creating 8-character passwords, this difference was not statistically significant. The participants required to generate their own passwords may have committed more errors because they failed to comprehend fully the restrictions imposed on their passwords during their first attempts to create one. Participants perhaps employed the creation password strategies they use in the wild; these may have conflicted with the restrictions imposed by this study. This explanation is further substantiated by the fact that 22 of the 36 participants in the 8-character and 16-character conditions stated at the end of the first session that they habitually used a specific strategy to create user-generated passwords.

The observation that the participants creating 16-character passwords committed fewer errors than those creating 8-character passwords, though not statistically significant, is consistent with that of Komanduri et al. (2011). In the Komanduri et al. (2011) study, the participants creating 8-character restricted passwords had difficulty determining whether their password was a dictionary word without entering it. As a result, they took more than one attempt to create it. The 16-character passwords did not

have any such restrictions in the Komanduri et al. (2011) study, and the participants took fewer attempts to create them.

The effect of password policy on temporal demand was significant. Post-hoc analysis revealed that 8-character passwords incurred higher temporal demand than the system-generated ones. This finding could be the result of the higher number of restrictions placed on the creation of 8-character passwords compared to the other two conditions. Participants creating these passwords may have felt under greater time pressure to complete the password creation task in a reasonable amount of time even though no time constraints were placed on them by the study.

Recall Task in First Session

When the recall data were analyzed across both sessions, the main effects of password policy and task session as well as their interaction were found to be significant. The interaction effect was further analyzed using simple effects analysis, which revealed that the system-generated passwords took less time to recall than either the 8-character or 16-character passwords in the 1st session. One of the reasons for this result is that all of the participant assigned system-generated passwords recalled their passwords successfully on their first attempt. This was not the case for the participants using self-generated passwords.

The fact that users of the system-generated password committed no recall errors in the first session suggested a potentially significant difference compared to the other conditions. To explore this possibility further, a between-subject one-way ANOVA of the

effect of password policy conditions on error rates in the first session was conducted. This analysis revealed statistically significant differences across conditions. However, this result may be an artifact caused by the zero variance in the recall error rates for the system-generated password.

Recall Task in Second Session

Four participants could not recall their system-generated passwords in the second session; three could not remember their 8-character passwords, and two could not remember their 16-character passwords. Although this trend supports the hypothesis that 16-character passwords would have the fewest unrecoverable passwords, this difference is not statistically significant. This finding is partially consistent with the results found by Komanduri et al. (2011), who determined that there was no statistically significant difference in the number participants who failed to recall their user-generated passwords across password conditions.

One of the reasons for the lack of significance could be the better-than-expected recall of system-generated passwords. Informal discussion with several participants suggested that they found six-character alphanumeric passwords similar in nature to their previous or current passwords. A second reason for the lack of a significant difference in the number of unrecovered passwords could be the lower-than-expected performance in the recall of 16-character passwords. Although these passwords could be composed of only lower-case letters, seventeen of the eighteen participants in this password condition created passwords that included combinations of upper-case letters, lower-case letters,

numbers or special characters. Recalling 16-character passwords of such complexity could be a difficult cognitive task, a conclusion supported by a study conducted by Zviran et al. (1993) in which the user-generated passwords composed of only lower-case letters were recalled more frequently than the ones composed of more than one character set. Ten participants who belonged to the 16-character password condition commented that they felt that their passwords were more secure when they included characters other than lower-case letters. These participants were probably not aware that passwords composed of 16 lower-case letters are secure. Additionally, nine of these eighteen participants commented that they found the 16-character minimum length to be overly long.

Difference Across Task Sessions

The simple effects analysis of the interaction effect for the time taken to recall also revealed that the system-generated passwords took less time to recall in the first session than in the second. A similar trend was observed for the 8-character user-generated passwords, but this difference was not significant. Additionally, there was also a significant difference between the error rates in the 1st and 2nd recall task sessions, presumably due to the degradation effect of time on memorability.

The Performance index of the NASA-TLX showed a significant effect for task sessions. This result suggests that the participants may have felt that it is significantly harder to create password accounts than recalling them after the distraction task in the first session. The participants may have felt that it was harder to create a password with

restrictions and then memorize the assigned or created password than to recall the passwords they created five minutes back, prior to the distraction task. The participants also indicated that they believed that their performance was significantly poorer during recall in the second session than in the first. This explanation is supported by the fact that the recall error rates for the first session were significantly lower than those for the second, possibly due to the degradation effect of time on memorability.

No statistical significance was found for ease-of-use across password policies. The descriptive statistics suggested that the usability of system-generated passwords was lower than that of the other password policies for account creation and for recall in the 1st and 2nd sessions. All of the SUS scores were in the range of 56 to 68, below the acceptable SUS score of 70 (Bangor, Kortum & Miller, 2008), indicating all three password policy conditions are marginally usable. Thus, the task of developing a usable password policy that is also secure requires further study. However, the SUS scores were significantly higher during the creation of the password than for the recall task in the same session. This finding may be associated with the lower error rates in the creation of a password than for recalling them in the same session.

Qualitative Analysis of Participant Comments

A total of fifty-three comments were recorded from sixty-two participants, including the eight participants who were replaced in the 1st session. However, a majority of the comments came from the participants in the 8-character and 16-character password

conditions. These comments were grouped using an open card sort method. As a result of this sort, the seven categories seen in Table 6.4 emerged:

Table 6.4: Categories of participant comments

| Categories | Number of comments by participants |
|--|------------------------------------|
| Password strategy | 22 |
| Secure password composition | 10 |
| Password length | 8 |
| Less secure password composition | 5 |
| Password composition for memorability | 4 |
| 16-character password case-sensitivity | 2 |
| Discomfort using 16-character password | 2 |

The category “Password strategy” had the highest number of comments, twenty-two, indicating that participants used strategies to create passwords. One of the participants commented, saying “I try to remember how I created my passwords and not what I created.” This category was followed by “Secure password composition” with ten comments, emphasizing that passwords created by combining numbers, lower- or upper-case letters or special characters were considered more secure. The participants in this category commented, saying “I prefer adding number and special characters to make my passwords more secure.” The participants were also of the opinion that 16-character passwords were too lengthy to create, a belief supported by eight comments under the category “Password length.” One of them commented on the password length of 16-

character password, saying “This is too long!” indicating bias towards shorter password length, perhaps 8-character passwords. In addition, the five comments in the “Less secure password composition” category revealed that the participants thought that the 16-character passwords composed of lower-case letters were less secure than the 8-character passwords. A participant commented on the 16-character password, saying “Although I knew I had an option of using only lower-case letters, I added number to make it more secure.” Four comments by participants suggested that including numbers with letters makes passwords easier to remember. One of the comments for this category was “I added numbers to my password to help make it more memorable.” Two comments suggested that the participants were unsure whether the 16-character password condition was case-sensitive or not. A participant was unsure of the case-sensitivity requirements and commented, “Do lower- and upper-case of the same letter count as two different characters?” Two other comments indicated the reluctance of the participants to use 16-character passwords. One of the participants commented, saying “I am not used to creating a 16-character password.”

This analysis perhaps indicated a bias of the participants towards the Clemson University password policies, which requires a password to have at least one number, a letter and a special character. These passwords must be at least 8-characters long with no spaces and no more than two repeated characters. This bias would explain the large number of comments, twenty-three, indicating the use of a strategy to create passwords.

Pareto analysis of the frequencies of these seven categories seen in Figure 6.1 revealed that the three most frequently occurring comment categories of comments were password strategy, secure password composition and password length.

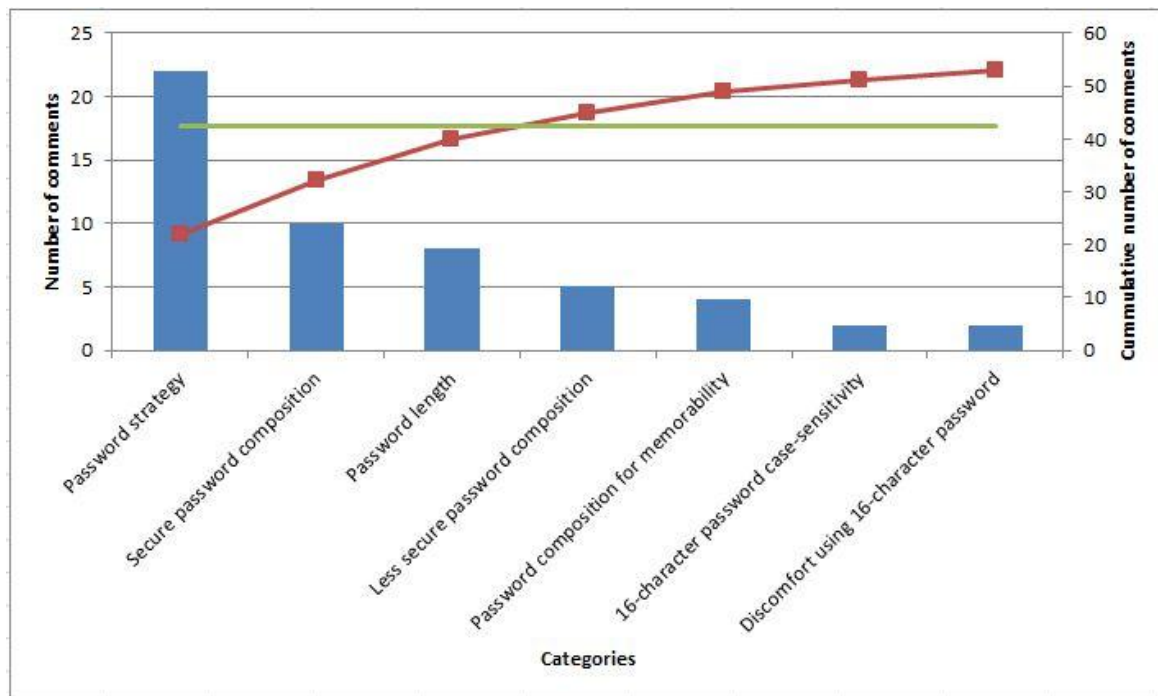


Figure 6.1: Pareto chart analysis

These three categories represented 80 percent of the total number of comments.

Analysis of User-Generated Passwords

An additional analysis of the user-generated passwords was conducted. The average length of the minimum 8-character passwords in this study was 12.05 characters. All of these passwords included lower-case letters, upper-case letters, number and special characters. The average length of the minimum 16-character passwords was 17.61

characters, with the average number of character sets included being 2.83. This extensive use of multiple character sets again suggests that the participants either did not comprehend the 16-character restrictions well or preferred adding more character sets rather than relying only on lower-case letters alone. Seventeen of the eighteen minimum 16-character passwords were composed of more than lower-case letters. Tables 6.5 and 6.6 list the passwords created by the participants:

Table 6.5: 8-character passwords

| | | |
|------------------------|------------------|---------------|
| Hitman#21 | Fr3nchfr!es | Zzyellow#9 |
| acegl#5I | Harry_Grewal1986 | Ummagumma90() |
| BeingHappyIsEasy123!@# | 24Te&01di | Rempasishar1@ |
| Ragam_endral_sahana1 | Q!w2e3r4 | IA@csmman123 |
| N1keSh*x | Demonicsages@786 | Zandubam@30 |
| FluorDaniel123# | ZaMBI@316 | iMd0ne++ |

Table 6.6: 16-character passwords

| | | |
|----------------------|-------------------------|---------------------|
| f@llSemester201! | Mechanical_2010@clemson | !@#caretAker123!@# |
| pocpoc1poc2poc3poc4 | m@thewph111p@ng@y11 | ThISvLoNGpWORD16 |
| samurai_09031987 | qwertyuilkjhgfds | clemsonpssjul11* |
| shanpa0320210917 | clemson2011mechanical | IwillmissClemson |
| H12A34S56H78I90M | cuid@iefreeman103 | Rgalgalclemson77&& |
| lijjo@fman0801sirvey | abcdefgh12345678 | Aniruddha_1985_5nov |

The 8-character and the 16-character password had an average of 2.11 and 3 chunks of information i.e. words/numbers/letters/special characters respectively. The participants appeared to use password creation based on their previous or current environment, condition and actions. For example, the password “f@llSemester201!” is based on the academic calendar and the “lijjo@fman0801sirvey” is based on the first and last name of participant, the location, the date of the study and the nature of the study, exhibiting the tendency of the users to create passwords reflecting the current period, environment or action. Passwords like “ThISvLoNGpWORD16”, “Mechanical_2010@clemson” and “IwillmissClemson” support the comments that participants use a strategy for creating passwords, either ones they learned in the past or the ones taught to them during the creation session. Other passwords like “qwertyuilkjhgfds” were composed of sequences of characters based on their placement on the keyboard. Except for the password “qwertyuilkjhgfds,” the 16-character passwords in Table 6.6 mirrored the comments of the participants in relation to including character sets other than lower-case letters.

7. CONCLUSION

This study compared the usability of three password conditions that assigned or helped users to generate passwords of approximately equal minimum security, evaluating the trade-off between the length and the complexity of the passwords. The most important conclusion of this study is that the performance of the 8-character password was weaker than that of the system-generated password during the creation of password accounts and was weaker than the 16-character password in the terms of long-term recall.

Compliance with the restrictions associated with 8-character passwords strengthen security, but creates a password that is complex in composition. Thus, with the increase in applications requiring 8-character password accounts, a user may experience cognitive load when recalling a password from among competing passwords of similar composition. However, if a 16-character password is created from a meaningful combination composed of preferably lower-case letters, it may be more memorable than 8-character passwords subject to multiple restrictions.

Currently, the designers of password applications put most of the responsibility for creating a secure password on the users, forcing them to comply with a variety of restrictions. The complexity of such passwords may increase their security, but such security can also be achieved by increasing the minimum length of the password and lowering the complexity of these passwords, reducing the cognitive load on users. Thus, efforts should be taken to educate users on the trade-off between the length and the complexity of user-generated password. A simpler and longer password can be as secure as a shorter but more complex one.

Designers should consider developing applications that aid users in creating longer but more meaningful passwords to reduce the cognitive load for the users. These applications could implement methods to produce 16-character passwords with meaningful combinations of letters, making the password more memorable to the user. However, care should be taken by the designers to avoid explicitly restricting users to lower-case letters only.

This study is a first step in exploring usable password conditions of approximately equal security. Below are suggestions for future research:

- Studies involving participants belonging to a wider range of demographics.
- Studies in the wild (real setting outside the laboratory) involving more participants.
- Studies on the effect of educating participants on the security of longer passwords composed of lower-case letters.
- Studies involving a longer time period between creation and recall tasks to validate the results of the long-term recall of passwords across conditions.

APPENDICES

Appendix A

Consent form for study participants

Information Concerning Participation in a Research Study
Clemson University

Evaluating the usability of system-generated and user-generated passwords of approximately equal security

Description of the Research and Your Participation

You are invited to participate in a research study conducted by Sourav Bhuyan under the direction of Dr. Joel Greenstein. The purpose of this research is to evaluate the usability of passwords, either assigned or created, having approximately 30 bits of entropy (a measure of the security of the password).

This study will take place over two sessions. Your participation in the first session will involve being introduced to the research, signing an informed consent form, completing a pre-test questionnaire, completing tasks according to the instructions from the researcher and completing post-test questionnaires. You will be asked to return after 5 to 7 days to complete a second set of tasks and answer post-test questionnaires. These post-test questionnaires consist of standardized satisfaction and workload surveys.

The amount of time required for your participation will be approximately thirty minutes for Session One and twenty minutes for Session Two.

Risks and Discomforts

There are no known risks associated with this research.

Potential Benefits

This research may help us to discover more usable and secure methods for generating passwords.

Protection of Confidentiality

We will do everything we can to protect your privacy. Collected data will be stored securely in 147 Freeman Hall and access will be limited to the investigators. Your identity will not be revealed in any publication that might result from this study.

In rare cases, a research study will be evaluated by an oversight agency, such as the Clemson University Institutional Review Board or the Federal Office for Human Research Protections, which would require that we share the information we collect from you. If this happens, the information will only be used to determine if we conducted this study properly and adequately protected your rights as a participant.

Voluntary Participation

Your participation in this research study is voluntary. You may choose not to participate and you may withdraw your consent at any time. You will not be penalized in any way should you decide not to participate or to withdraw from this study.

You may choose to stop taking part in this study after today. If you do, we will remove your information from the study. However, if we have already completed our research analysis, we will not be able to remove your information from the study.

Contact Information

If you have any questions or concerns about this study or any problems arise, please contact Dr. Joel Greenstein at Clemson University at 864-656-5649. If you have any questions or concerns about your rights as a research participant, please contact the Clemson University Office of Research Compliance (ORC) at 864-656-6460 or irb@clemson.edu. If you are outside the Upstate South Carolina area, please use the ORC's toll-free number, 866-297-3071.

Consent

**I have read this consent form and have been given the opportunity to ask questions.
I give my consent to participate in this study.**

Participant's signature: _____ Date: _____

A copy of this consent form will be given to you.

Appendix B

PRE-TEST QUESTIONNAIRE

GENERAL

Participant: _____ (This will be filled out by the test administrator.)

Age: _____

Gender: ☐ Male ☐ Female

EDUCATION

1. Please select your academic level:

☐ Undergraduate student

☐ Graduate student

☐ Other

(Please specify: _____)

2. List your major area of study: _____

COMPUTER EXPERIENCE

3. How long have you been using computers?

☐ < 1 year ☐ 1-2 years ☐ 3-5 years ☐ > 5 years (Please specify) _____

4. How long have you used passwords?

☐ < 1 year ☐ 1-2 years ☐ 3-5 years ☐ > 5 years (Please specify) _____

5. How many unique passwords do you have?

☐ 1 ☐ 2 ☐ 3 ☐ More than 3 (Please specify the number) _____

Appendix C

Methodologies for remembering passwords*

*Source: Guide to Enterprise Password Management (Draft), NIST Special Publication 800-118 (Draft)

- 1. Mnemonic Method:** A user selects a phrase and extracts a letter from each word (e.g., the first or second letter of each word), adding numbers or special characters or both.

Example:

| Phrase | Password |
|--|------------------|
| Please be my best valentine! | Pbmbval! |
| This is the worst car I have ever driven in my LIFE! | TitwcIhedimLIFE! |
| I am definitely your #1 fan. | Iady#1f. |

- 2. Altered Passphrases:** A user selects a phrase and alters it to form a derivation of that phrase.

Example:

| Passphrases | Alternate Passphrases |
|----------------------|-----------------------|
| to be or not to be | 2.be.0r.n0t@to0.bEE |
| Dressed to the nines | Dressed*2*the*9z |

- 3. Combining and Altering:** A user can combine two or three unrelated words and change various letters to numbers or special characters.

Example:

| Words | Password |
|---------------------|------------|
| “bank” and “camera” | B@nkC@mera |
| “mail” and “phone” | m4!lf0N3 |

Appendix D

System Usability Scale Questionnaire

I think that I would like to use this policy frequently.

Strongly disagree Undecided Strongly agree

I found this policy unnecessarily complex.

Strongly disagree Undecided Strongly agree

I thought this policy was easy to use.

Strongly disagree Undecided Strongly agree

I think that I would need assistance to be able to use this policy.

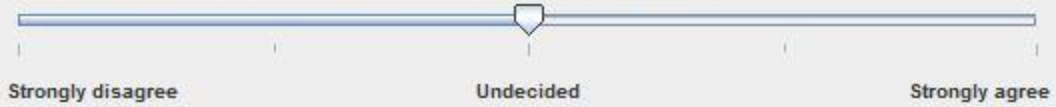
Strongly disagree Undecided Strongly agree

I found the various aspects of this policy were well integrated.

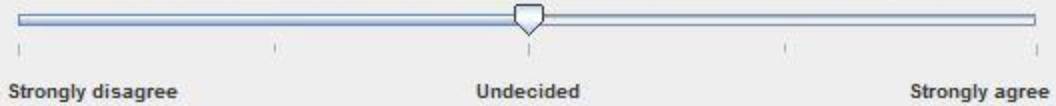
Strongly disagree Undecided Strongly agree

Next

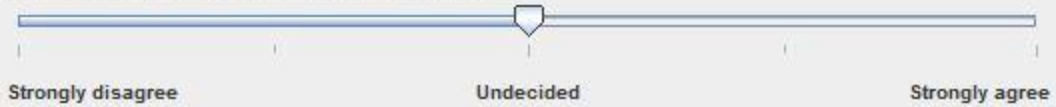
I thought there was too much inconsistency in this policy.



I would imagine that most people would learn to use this policy very quickly.



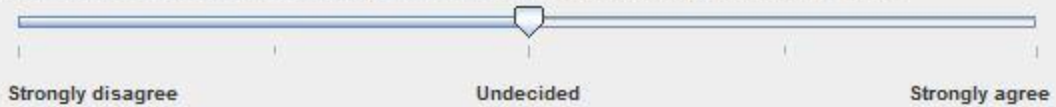
I found this policy very awkward to use.



I felt very confident using this policy.



I needed to learn a lot of things before I could get going with this policy.



Done

Appendix E

NASA-TLX questionnaire

How mentally demanding was the task?

Very light Neutral Very demanding

How physically demanding was the task?

Very light Neutral Very demanding

How hurried or rushed was the pace of the task?

Very comfortable Neutral Very hurried

How successful were you in accomplishing what you were asked to do?

Perfect Neutral Failure

How hard did you have to work to accomplish your level of performance?

Very easy Neutral Very hard

How insecure, discouraged, irritated, stressed, or annoyed were you?

Very relaxed Neutral Very frustrated

Next

Appendix F

Consent form for pilot study participants

Information Concerning Participation in a Research Study
Clemson University

Evaluating the usability of system-generated and user-generated passwords of approximately equal security

Description of the Research and Your Participation

You are invited to participate in a research study conducted by Sourav Bhuyan under the direction of Dr. Joel Greenstein. The purpose of this study is to record your preference for different passwords that are system-generated and assigned, having approximately 30 bits of entropy (a measure of security of the password).

Your participation will involve being introduced to the study, signing an informed consent form, completing demographic questions, and ranking the three passwords in terms of your preference.

The amount of time required for your participation will be approximately 15 minutes.

Risks and Discomforts

There are no known risks associated with this research.

Potential Benefits

This research may help us to discover more usable and secure methods of generating passwords.

Protection of Confidentiality

We will do everything we can to protect your privacy. Collected data will be stored securely in 147 Freeman Hall and access will be limited to the investigators. Your identity will not be revealed in any publication that might result from this study.

In rare cases, a research study will be evaluated by an oversight agency, such as the Clemson University Institutional Review Board or the Federal Office for Human Research Protections, which would require that we share the information we collect from you. If this happens, the information will only be used to determine if we conducted this study properly and adequately protected your rights as a participant.

Voluntary Participation

Your participation in this research study is voluntary. You may choose not to participate and you may withdraw your consent at any time. You will not be penalized in any way should you decide not to participate or to withdraw from this study.

You may choose to stop taking part in this study after today. If you do, we will remove your information from the study. However, if we have already completed our research analysis, we will not be able to remove your information from the study.

Contact Information

If you have any questions or concerns about this study or any problems arise, please contact Dr. Joel Greenstein at Clemson University at 864-656-5649. If you have any questions or concerns about your rights as a research participant, please contact the Clemson University Office of Research Compliance (ORC) at 864-656-6460 or irb@clemson.edu. If you are outside the Upstate South Carolina area, please use the ORC's toll-free number, 866-297-3071.

Consent

**I have read this consent form and have been given the opportunity to ask questions.
I give my consent to participate in this study.**

Participant's signature: _____ Date:

A copy of this consent form will be given to you.

Appendix G

PREFERENCE RANKING QUESTIONNAIRE

GENERAL

Participant: _____ (This will be filled out by the test administrator.)

Age: _____

Gender: ☐ Male ☐ Female

EDUCATION

1. Please select your academic level:

- ☐ Undergraduate student
- ☐ Graduate student (Master's or Ph.D.)
- ☐ Other

(Please specify: _____)

2. List your major area of study: _____

RANK THE PASSWORDS

Rank the passwords that you prefer the most to be assigned to you as #1 and least preferred as #3

1. Password - kholscx

Rank # _____

2. Password - djh45j

Rank # _____

3. Password - V#l9N

Rank # _____

REFERENCES

- Adams, A., & Sasse, M. A. (1999). Users are not the enemy. *Commun.ACM*, 42(12), 40-46. doi:<http://doi.acm.org/10.1145/322796.322806>
- Adams, A., Sasse, M. A., & Lunt, P. (1997). Making passwords secure and usable. Paper presented at the *Proceedings of HCI on People and Computers XII*, 1-19. Retrieved from <http://portal.acm.org/citation.cfm?id=646684.702633>
- Allendoerfer, K., & Pai, S. (2005). *Human factors considerations for passwords and other user identification techniques part 1: Field study, results and analysis (DOT/FAA/TC-05/20)*. Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center:
- Allendoerfer, K., & Pai, S. (2006). *Human factors considerations for passwords and other user identification techniques part 2: Field study, results and analysis (DOT/FAA/TC-06/09)*. Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center:
- Bangor, A., Kortum, P.T., Miller, J. T. (2008) An empirical evaluation of the system usability scale. *International Journal of Human Computer Interaction*, 24(6), 574-594. doi: 10.1080/10447310802205776
- Brostoff, S., & Sasse, M. A. (2000). Are passfaces more usable than passwords? A field trial investigation. Paper presented at the *Proceedings of HCI 2000*, Retrieved from http://www.cs.ucl.ac.uk/staff/S.Brostoff/index_files/brostoff_sasse_hci2000.pdf

- Burr, W. E., National Institute of Standards, & Technology. (2006). *Electronic authentication guideline [electronic resource] : Recommendations of the National Institute of Standards and Technology / William E. Burr, Donna F. Dodson, W. Timothy Polk* (Version 1.0.2. ed.) U.S. Dept. of Commerce, Technology Administration, National Institute of Standards and Technology, Gaithersburg, MD.
- Clarke, N., & Furnell, S. (2005). Biometrics - the promise versus the practice. *Computer Fraud & Security*, 2005(9), 12-16. Retrieved from [http://dx.doi.org/10.1016/S1361-3723\(05\)70253-0](http://dx.doi.org/10.1016/S1361-3723(05)70253-0)
- Conklin, A., Dietrich, G., & Walz, D. (2004). Password-based authentication: A system perspective. Paper presented at the *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 10.
- Dhamija, R., & Perrig, A. (2000). Deja vu: A user study using images for authentication. Paper presented at the *Proceedings of the 9th Conference on USENIX Security Symposium - Volume 9*, Denver, Colorado. 4-4. Retrieved from <http://portal.acm.org/citation.cfm?id=1251306.1251310>
- FIPS, National Institute of Standards, & Technology (1985). Password Usage. *Federal Information Processing Standard Publication (FIPS PUB)*.
- Florencio, D., & Herley, C. (2007). A large-scale study of web password habits. Paper presented at the *Proceedings of the 16th International Conference on World Wide*

Web, Banff, Alberta, Canada. 657-666.

doi:<http://doi.acm.org/10.1145/1242572.1242661>

Forget, A., Chiasson, S., & Biddle, R. (2007). Helping users create better passwords: Is this the right approach? Paper presented at the *Proceedings of the 3rd Symposium on Usable Privacy and Security*, Pittsburgh, Pennsylvania. 151-152.

doi:<http://doi.acm.org/10.1145/1280680.1280703>

Hart, S. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. Paper presented at the *50th Annual Meeting of the Human Factors and Ergonomics Society, HFES 2006, October 16, 2006 - October 20, 2006*, 904-908.

Herley, C. (2009). So long, and no thanks for the externalities: The rational rejection of security advice by users. Paper presented at the *New Security Paradigms Workshop 2009, NSPW 2009, September 8, 2009 - September 11, 2009*, 133-144. Retrieved from <http://dx.doi.org/10.1145/1719030.1719050>

Jeyaraman, S., & Topkara, U. (2005). Have the cake and eat it too - infusing usability into text-password based authentication systems. Paper presented at the *Proceedings of the 21st Annual Computer Security Applications Conference*, 473-482. doi:10.1109/CSAC.2005.28

Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., et al. (2011). Of passwords and people: Measuring the effect of password-composition policies. Paper presented at the *Proceedings of the 2011 Annual Conference on*

Human Factors in Computing Systems, Vancouver, BC, Canada. 2595-2604.

doi:<http://doi.acm.org/10.1145/1978942.1979321>

Kuo, C., Romanosky, S., & Cranor, L. F. (2006). Human selection of mnemonic phrase-based passwords. Paper presented at the *Proceedings of the Second Symposium on Usable Privacy and Security*, Pittsburgh, Pennsylvania. 67-78.

doi:<http://doi.acm.org/10.1145/1143120.1143129>

Proctor, R. W., Mei-ching Lien, Vu, K. L., Schultz, E. E., & Salvendy, G. (2002).

Improving computer security for authentication of users: Influence of proactive password restrictions. *Behavior Research Methods, Instruments, & Computers*, 34(2), 163-9.

Rovio Entertainment Ltd. (2009). Angry Birds (Version 1.1.2.1) [Google Chrome Application Software]. Retrieved from chrome.angrybirds.com

Sasse, M. A., Brostoff, S., & Weirich, D. (2001). Transforming the 'weakest link' — a human computer interaction approach to usable and effective security. *BT Technology Journal*, 19(3), 122-131. doi:10.1023/A:1011902718709

Vu, K. L., Proctor, R. W., Bhargav-Spantzel, A., Tai, B., Cook, J., & Schultz, E. E. (2007). Improving password security and memorability to protect personal and organizational information. *International Journal of Human-Computer Studies*, 65(8), 744-57. Retrieved from <http://dx.doi.org/10.1016/j.ijhcs.2007.03.007>

Yan, J., Blackwell, A., Anderson, R., & Grant, A. (2004). Password memorability and security: Empirical results. *IEEE Security & Privacy*, 2(5), 25-31. Retrieved from <http://dx.doi.org/10.1109/MSP.2004.81>

Zviran, M., & Haga, W. J. (1993). A comparison of password techniques for multilevel authentication mechanisms. *Computer Journal*, 36(3), 227-37.