EAI Endorsed Transactions
on Security and Safety

Research Article EAI.EU

Comparing Online Surveys for Cybersecurity: SONA and MTurk

Anne Wagner[1], Anna Bakas[1], Shelia Kennison[2], Eric Chan-Tin[1,*]

[1]Loyola University Chicago, Chicago, IL, USA
[2]Oklahoma State University, Stillwater, OK, USA

Abstract

People have many accounts and usually need to create a password for each. They tend to create insecure passwords and re-use passwords, which can lead to compromised data. This research examines if there is a link between personality type and password security among a variety of participants in two groups of participants: SONA and MTurk. Each participant in both surveys answered questions based on password security and their personality type. Our results show that participants in the MTurk survey were more likely to choose a strong password and to exhibit better security behaviors and knowledge than participants in the SONA survey. This is mostly attributed to the age difference. However, the distribution of the results was similar for both MTurk and SONA. In the second part of our study, we found that security behaviors actually went down – this could be due to the pandemic or indicative of a need for more regular messaging/training.

Received on 17 December 2021; accepted on 30 January 2022; published on 08 February 2022

Keywords: Survey, Password Strength, Personality, MTurk, SONA

doi:10.4108/eai.8-2-2022.173334

1. Introduction

With the increase in Internet usage and digital communication, online surveys have developed into a popular medium for researchers. Online surveys are beneficial because they provide access to populations groups that would otherwise be difficult to reach. Time is another benefit of online surveys. Organizations and researchers can reach thousands of people across wide geographic regions in a short amount of time. Online surveys are also cheaper because the need for paper and associated costs, such as printing and delivery, are eliminated [1]. However, online surveys have some issues, such as reliability of the data.

Many universities have a pool of students enrolled in introductory psychology courses. These students are asked to complete some surveys (either in-person or online) either as a requirement for the class or for extra credit. Each survey tells the students how much credit they will be receiving. This pool is managed by a system called SONA. Access to these students' surveys is generally called the SONA pool, or SONA for short. Since these are actual students enrolled in courses, the SONA survey is generally reliable as the students will usually complete the survey correctly. However, the age of the participants in the survey will be the typical undergraduate student age group, which is between 18 to 21 years old. In our SONA online survey, 90% of participants were between the ages of 18 and 21 years old and 10% of participants were older than 21 years old.

Online surveys can also be administered through crowdsourcing platforms such as Amazon Mechanical Turk (also known as MTurk) [2], Crowdflower [3], microworkers [4], and clickworkers [5]. Participants in these online surveys are compensated after they complete a survey, usually a couple of dollars depending how time-consuming the survey is. The crowdsourcing platforms have some protections in place to prevent a participant taking the same survey more than once. A wider and more diverse group of participants can be reached through these crowdsourcing platforms. In our data collection, 23% of participants were in their 20s, 39% in their 30s, 19% in their 40s, 11% in their 50s, 7% in their 60s, and 1% in their 70s. The youngest participant was 20

*Corresponding author. Email: chantin@cs.luc.edu

years old, and the oldest participant was 73 years old. However, inaccurate data is a potential issue. Many participants will employ bots to automatically complete the online surveys and still get compensated. These bots could select random answers to questions or be more advanced and attempt to understand the questions using natural language processing. Thus, the reliability of these surveys is not as high as SONA.

In this research, the same online survey is distributed to both SONA and MTurk. One of the goals of this research is to determine if there is a relationship between the personality self-schema of participants and these participants' password usage and management. The second goal is to determine if there is a difference in SONA and MTurk responses. This could lead to a deeper knowledge of whether online surveys are reliable. The online survey consisted of two parts; in the first part, each participant received a message about what constitutes a strong password. The goal of the messaging is to determine if a training message will help in improving password security behavior.

The motivation for this research is that if a correlation can be found between certain personality type(s) or self-schema(s) and insecure behaviors, then more targeted cybersecurity training can be performed on such people. More specifically, individualized password education based on personality can be designed. The use of two different survey platforms will lead to determining whether both surveys are needed, which can save time and effort.

From our results, we found that the SONA and MTurk surveys produced similar outcomes. The differences are attributed mostly to the age range between SONA participants and MTurk participants. Due to this difference, MTurk participants were more likely to create a strong password than SONA participants. Part 2 of our survey did not yield a positive outcome as more scores decreased rather than increased.

The contributions of this paper are as follows

- The differences in SONA and MTurk are due to age difference in the two groups. Combined together, both surveys could be generalized to the entire population.
- Age affects the creation of stronger passwords, higher security knowledge and behaviors. Older people are more exposed to security training and reminders due to their job and are likely to be more responsible because a mistake could have more consequences than a college student's mistake. Younger people should also be exposed to security training and education.
- There is a slight correlation between personality self-schemas and creation of a strong password. Some personality self-schemas, such as True Color orange, tended to create a stronger password.
- Better security knowledge tends to lead to creation of stronger passwords. Although this is expected, some highly knowledgeable participants also created weak passwords.

A short paper [6] of 2 pages was previously published. This full paper extends the previous work to include more analysis of different features and a 2-part analysis.

Section 2 provides an overview of the different personality traits types and gives a brief description of how password strength is calculated. Related work is also provided in Section 2. Our data collection procedure is outlined in Section 3. Results from the first part of our survey are given in Section 4 while results from the second part of our survey are shown in Section 5. A discussion of the research is provided in Section 6. Section 7 gives a summary and some avenues for future work.

2. Background

2.1. Personality Traits/Types

Big Five. The Big Five model is a popular taxonomy for classifying personality traits. It consists of five core personality traits: conscientiousness, agreeableness, neuroticism, openness, and extraversion. Unlike other models that place individuals into binary categories (i.e. introvert or extrovert), the Big Five model holds that each personality type is a spectrum. Individuals are placed on a scale, for instance, determining their level of conscientiousness [7, 8].

- Conscientiousness. Conscientiousness measures an individual's ability to regulate their impulse control. People who score high on conscientiousness can be described as reliable, organized, and methodical.
- Agreeableness. Agreeableness measures how individuals interact with others in their relationships. People who score high on agreeableness can be described as empathetic, helpful, and cooperative.
- Neuroticism. Neuroticism measures the emotional stability of an individual and how they are likely to perceive the world. People who score high on neuroticism can be described as insecure, irritable, and tense.
- Openness. Openness measures individuals' willingness to participate in new experiences and abstract intellectual activities. People who score high on openness can be described as creative, imaginative, and unconventional.

Table 1. The advantages and disadvantages of using SONA and crowdsourcing platforms such as MTurk, CrowdFlower, Clickworkers, and Microworkers.

SONA	Crowdsourcing
College-age students enrolled in college	More diverse in age, demographics, income. However, need to know some technology to use the platform
Participants are real people	Could be automated bots
Tend to be more reliable since the survey provides a credit for a course	Could complete the survey as fast as possible to get compensated

- Extraversion. Extraversion measures the level of comfort and assertiveness of individuals when placed in social situations. People who score high on extraversion can be described as outgoing, excitement-seeking, and sociable.

True Colors. True Colors is a model that assigns individuals with a color to represent their primary personality self-schema. True Colors utilizes four different colors to represent primary self-schemas: orange, gold, green, and blue. According to their website, in a global population, 27% of people identify with orange, 35% with gold, 23% with green, and 15% with blue. [9].

- Orange. People who are comfortable with taking risks and are more action-oriented tend to identify as "oranges."
- Gold. People who value structure and punctuality in their lives tend to identify as "golds."
- Green. People who think outside of the box and enjoy problem solving tend to identify as "greens."
- Blue. People who value sincerity and collaboration to form relationships tend to identify as "blues."

The Big Five personality traits or the True Colors personality self-schemas are usually obtained through a self-reported questionnaire. The validity and reliability of these self-reported personality self-schemas have been shown in [10, 11].

2.2. Password Strength

zxcvbn [12] was created by Dropbox for the purpose of rating the strength of passwords. It generates scores ranging from zero to four, with a zero score being considered "too guessable" and four being considered "very unguessable." To obtain these numbers, passwords are run against pattern matching, common names, popular English words, and common patterns such as dates and keyboard patterns. This tool is used to rate password strength as it is freely available and used by a well-known company.

2.3. Online Surveys and Crowdsourcing

Crowdsourcing [13] leverages the Internet workforce and allows organizations to obtain and analyze data on a large scale. There are many crowdsourcing platforms that exist with each serving different needs. Some are geared towards finding designers for projects while others are geared towards more simple, small, one-time tasks. MTurk [2] is one of the first crowdsourcing platforms that is run by Amazon. It is a platform that is made for simpler tasks such as surveys. Crowdsourcing is appealing to businesses and researchers because it brings in a diverse survey population, it is convenient, and is of relative low cost. Each participant completing a task is compensated with a few dollars. Other crowdsourcing platforms such as CrowdFlower [3], Microworkers [4], and Clickworkers [5] work similarly to MTurk.

Another platform that is available to many universities is the SONA system. The participants are college students enrolled in the introductory psychology courses. As part of their grade or as extra credit, these students are asked to complete a certain number of "credits" on the SONA system. A survey like ours could reward students with 1 "credit" upon completion.

Table 1 shows the advantages and disadvantages of each platform. Crowdsourcing includes MTurk, CrowdFlower, Microworkers, and Clickworkers. To mitigate the automated bot issue with crowdsourcing platforms, a CAPTCHA can be used. To improve the reliability of the crowdsourcing platforms, an attention check question can be used. That question would be similar to "Pick the third option below". Bots or inattentive participants might choose the wrong option and their answers can be discarded.

Due to the widespread use of the Internet, there are different online surveys that can be used [14, 15]. They all function in a similar way. They allow different types of questions, allow participants to save and complete the survey at a later time, prevent retaking of the same survey, and include a CAPTCHA to prevent bots.

2.4. Related Work

It has been shown that data collected from online surveys, such as SONA and MTurk, are mostly

equivalent to data collected in-person [13, 16, 17]. Social desirability has also been shown to be reduced with online surveys. When participants complete surveys in-person, they may feel compelled to respond in a more desirable manner. However, through SONA or MTurk (at home, over the internet), participants may feel less compelled to answer in a socially desirable manner, resulting in more accurate responses. It has also been shown that MTurk participants are slightly more demographically diverse [16, 18–20], which is something that we saw in our survey responses in terms of a wider age range for participants. However, the results from SONA and in-person surveys were similar [16]. Participation in MTurk surveys are affected by compensation rate and task length, but it is still an efficient way to distribute surveys. It has also been shown that data from MTurk surveys are at least as reliable as data obtained via traditional methods. Ultimately, MTurk is an efficient way to reach a large amount of people at a low cost.

Previous work has found some relationship between personality and cybersecurity behavior. For example, neurotic people tend to be more likely to fall for phishing e-mails [21] and open people tend to post more private information on social media. Other work [22] has found that when it comes to cybersecurity, extroverts tended to behave better. [23] found no relationship between personality trait and password strength. This research shows that there is some correlation between personality, password strength, and cybersecurity behaviors.

3. Data Collection

3.1. Survey

The survey is deployed online using Qualtrics and consists of two parts. The types of questions included in the survey are as followed: general questions relating to the participant's personality, password usage, social media usage, password behavior, and demographic information. Each participant was shown a message about how to create a strong password. They were then prompted to create a password that they would have to remember for the second part. The second part of the survey asked similar questions about password behavior and to enter the password from part 1. Each question of the survey could be skipped. About a month after completing the first part of the survey, each participant was asked to fill out the second part. One month was chosen as this was enough time for the participants to change their password behaviors based on the message. Only English-speaking participants were recruited from MTurk.

The surveys for both MTurk and SONA were similar except for the last question. For MTurk, participants were asked for their MTurk worker ID and for SONA, participants were asked for their university e-mail address. This information was deleted after part 2 data collection.

Participants were compensated for their involvement in the survey. SONA students were compensated with 1.5 course credits for completing part 1 and 0.5 course credits for completing part 2. In addition, if they completed both parts of the survey, they were entered into a drawing for a $50 gift card. MTurk participants were compensated $3 for completing part 1 and $1 for completing part 2 of the survey.

3.2. Summary of Data Collected

168 people participated in the SONA survey. For SONA, 73% of participants identified as women, 25% identified as men, 1% identified as Gender Non-Conforming / Non-Binary, and 1% identified as transgender. 62% of participants were under the age of twenty, and 38% of participants were in their 20s. 85 participants completed part 2 of the survey. 57% of participants were under the age of 20, 42% of participants were in their 20s, and 1% were in their 40s.

391 people participated in the MTurk survey. For MTurk, 44% of participants identified as women, 55% identified as men, and 1% identified as Gender Non-Conforming / Non-Binary. Additionally, 23% of participants were in their 20s, 39% were in their 30s, 19% were in their 40s, 11% were in their 50s, 7% were in their 60s, and 1% were in their 70s. 226 participants completed part 2 of the survey. Less than 1% of participants were less than 20, 24% were in their 20s, 37% were in their 30s, 19% were in their 40s, 11% were in their 50s, 8% were in their 60s, and less than 1% were older than 70.

The SONA data collection ran from February 2020 to May 2020 while the MTurk data collection ran from March 2020 to June 2020.

4. Results (Part 1)

In order to assess a participant's True Colors self-schema, the survey showed participants four images, as shown in Figure 1. Each participant was asked to select the image and description that most closely resembled them. The complete description of the text in each image is given next.

- Image A – I am warm, communicative, compassionate, and feeling. I need to search for the meaning and significance of life. I want to find ways to make my life count and matter, to become my own authentic self. Integrity, harmony, and honesty are very important to me. I feel that I am highly idealistic and spiritual by nature.
- Image B – I need to be responsible, dependable, helpful, and sensible. I want to fulfill my duties

Figure 1. The four images to determine a participant's True Colors self-schema. Image A says "I am warm, communicative, compassionate, and feeling". Image B says "I need to be responsible, dependable, helpful, and sensible". Image C says "I am versatile, wise, conceptual, and curious". Image D says "I am adventurous, skillful, competitive, and spontaneous". The full description of each image can be found within the text of the paper.

and obligations, to organize and to structure my life as I see fit. I am practical, sensible, and punctual, and I believe that people should earn their way through work and service to others.

- Image C – I am versatile, wise, conceptual, and curious. I need freedom to pursue knowledge and wisdom to develop competency by acquiring skills and capabilities. I think life is something to make sense of, to be understood, and explained.

- Image D – I am adventurous, skillful, competitive, and spontaneous. I need to be free to act on a moments notice, impulsively and spontaneously. I believe that life is to enjoy, so I thrive on fun, variety, and excitement. Living in the moment, I act on every opportunity.

4.1. Personality

Figure 2 shows the percentage of MTurk participants with each Big Five personality trait for each zxcvbn password score. Each participant is represented by their highest scoring Big Five personality trait which had a score higher than 36. This graph shows that those who scored highest in neuroticism also had the highest zxcvbn password scores. This could be because these types of people are more likely think negative and might think their accounts have been hacked and they need to create a strong password. Two other personality traits that also tend to create strong passwords are agreeableness and conscientiousness. The latter tend to be more responsible. Figure 3 shows

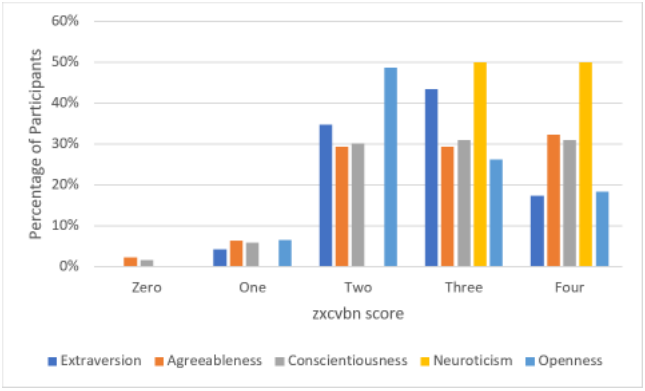

Figure 2. MTurk survey: zxcvbn password scores for each Big Five personality trait.

the percentage of SONA participants with each Big Five personality trait for each zxcvbn password score. Each participant is represented by their highest scoring Big Five personality trait which had a score higher than 36. From this graph, those who created a strong password with a zxcvbn score of 4 tended to score highest in openness. This could be because these types of people are more likely to be cooperative and creative and might think of unique ways to create a strong password. Another personality trait that also tend to create strong passwords are conscientiousness, which tend to be responsible people. There were not many participants who had the personality trait of neuroticism, which could explain the lack of neurotic participants with a password score of four.

Figure 4 shows the percentage of MTurk participants with each True Colors self-schema for each zxcvbn password score. Each participant is assigned a True Colors color based on their self-schema. From this graph, those who created the strongest passwords, ones with a zxcvbn score of four, identified as "Orange". These people tend to be adaptable and like being in charge. This could explain their tendency to pick a strong password. Figure 5 shows the relationship between SONA participants' True Colors self-schemas

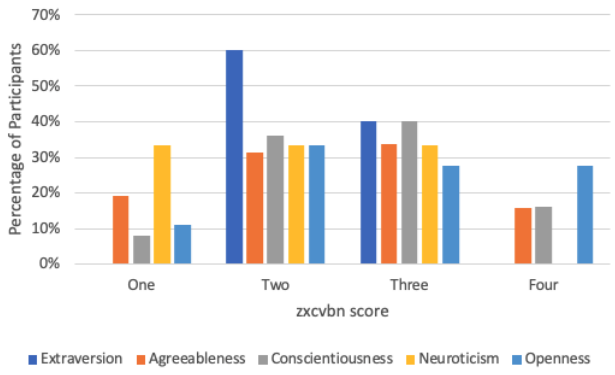

Figure 3. SONA survey: zxcvbn password scores for each Big Five personality trait.

and their zxcvbn password scores, breaking down the percentage of participants for each self-schema according to their corresponding password score. No participant created a password that was given a password score of zero. In order, each True Colors self-schema is represented: blue, gold, green, and orange. Figure 5 shows that there is a higher percentage of participants with a blue personality who created a password with a zxcvbn password score of three. 63% of participants with a blue self-schema created a password with scores of three and four versus 41% for gold self-schema, 50% for green self-schema, and 43% for orange self-schema.

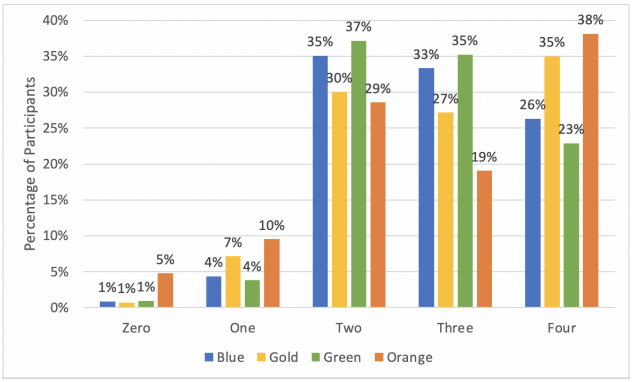

Figure 4. MTurk survey: True Colors personality self-schemas and zxcvbn password scores.

Figure 6 shows the percentage of participants for each True Color self-schema in MTurk vs SONA. While the majority of MTurk participants identified with Blue, Gold, or Green, very few identified with Orange. However, an Orange personality was slightly more common in the SONA survey of college students than the MTurk surveys of the general public. Individuals with the Orange personality self-schema tend to be more comfortable taking risks and being more action-oriented. This also describes most college-age students,

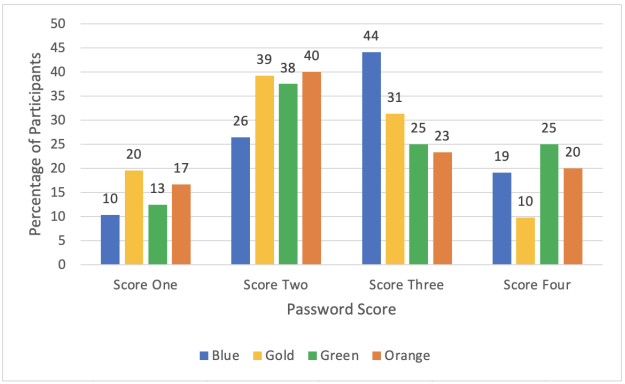

Figure 5. SONA survey: True Colors personality self-schemas and password strength according to zxcvbn.

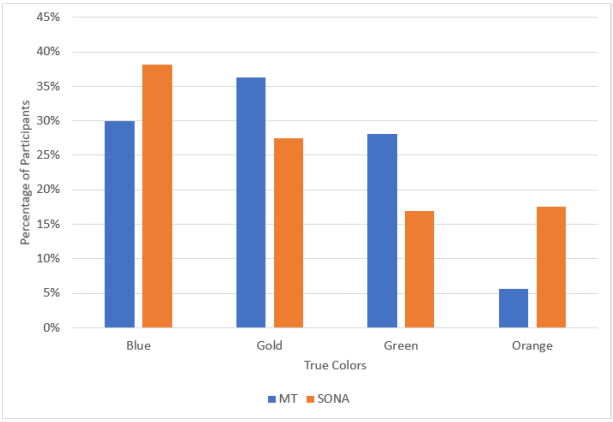

Figure 6. Participants with each True Color self-schema in MTurk (MT) vs SONA.

so that explains why SONA participants had more Orange personalities than MTurk.

4.2. Password Strength

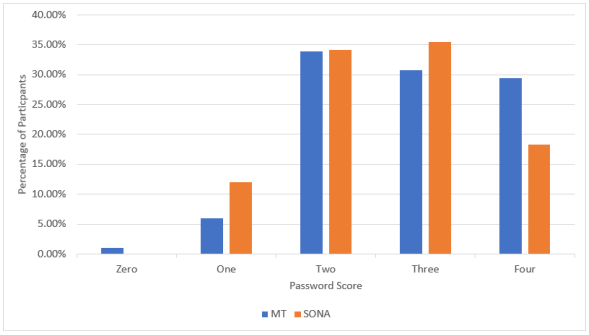

Figure 7. Password Scores in MTurk vs SONA.

Figure 7 shows the percentage of participants that received each zxcvbn score in both MTurk and SONA. The figure shows that a higher percentage of MTurk participants received a zxcvbn score of 4 than SONA

participants. This may be because of the age differences between participants in MTurk and SONA. SONA participants had an average age of 20, while MTurk participants had an average age of 39. The password score differences of these two groups could indicate a correlation between age and password strength. This could be because participants who are older are working adults and might receive more cybersecurity training at the workplace.

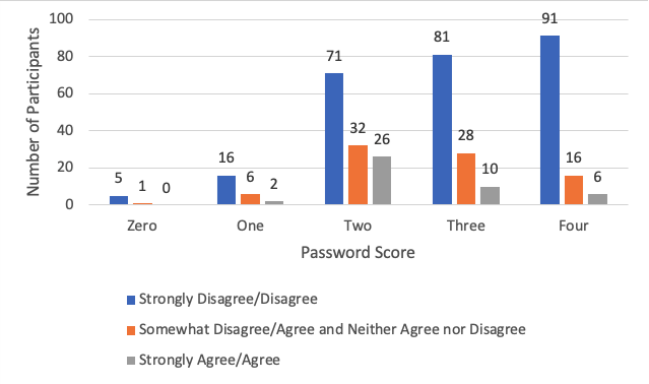

Figure 8. MTurk survey: Responses about whether it is acceptable to use social media passwords on work accounts in relation to zxcvbn password scores.

Figure 8 shows the relationship between MTurk responses about whether it is acceptable to use social media passwords on work accounts and participants' zxcvbn password scores. Participants were shown the statement "It's acceptable to use my social media passwords on my work accounts". They could respond with strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, and strongly agree. Figure 8 groups responses as follows: strongly disagree / disagree, somewhat disagree / somewhat agree / neither agree nor disagree, and strongly agree / agree. Figure 8 shows that participants who strongly disagreed or disagreed with this statement had stronger passwords than those who with lower password scores. 91 participants who received a zxcvbn password score of four responded to this statement with strongly disagree or disagree versus 71 participants who received a password score of two. This shows that participants who create a stronger password tended to know that passwords should not be re-used. We did not include the SONA results here since the response is about work accounts which most college-age students do not have.

Figure 9 shows the relationship between participants' password scores and their rating of the password "oklahomaStateUniversity". Participants could rate the password "oklahomaStateUniversity" on a scale from one to five, with one being "weak" and five being "strong". Figure 9 shows that participants who received a higher password score were more likely to rate "oklahomaStateUniversity" as a weak password versus those who received a lower password score of three or two. 38% of participants who received a password score of four rated "oklahomaStateUniversity" versus 34% of participants who received a password score of three.

Figure 10 shows the relationship between participants' password scores and their response to how likely they were to use the password "F63$uj2*" themselves. Participants could answer on a scale from one to five, with one being "not likely" and five being "likely". Figure 10 shows that participants who received a higher password score were more likely to use "F63$uj2*" as a password versus those who received a lower password score of two or one. 29% of participants who received a password score of four were likely to use "F63$uj2*" versus 20% of participants who received a password score of two. The results for the MTurk survey were similar – we showed only the SONA results since the first password "oklahomaStateUniversity" is only applicable to the students enrolled at that university.

Figure 11 shows the relationship between participants' responses about password security in relation to their zxcvbn password scores. Participants were asked to indicate how much they agreed or disagreed with the statement "having 'strong' passwords is a top priority for me." Participants could respond with strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, and strongly agree. Each response was assigned a number of points: strongly disagree = 0 points, disagree = 1 point, somewhat disagree = 2 points, neither agree nor disagree = 3 points, somewhat agree = 4 points, agree = 5 points, strongly agree = 6 points. To normalize the point totals, the total number of points for each password score was divided by the number of participants for each password score. The higher the point total for each password score, the more participants strongly agreed with the statement "having 'strong' passwords is a top priority for me" versus strongly disagree. In Figure 11, password score four has the highest point total out of all the password scores with a point total of 5.27. Thus, participants who created strong passwords tended to strongly agree that having strong passwords was a top priority for themselves. Although this is expected, it is interesting to see that some people who agreed with this statement also created a weak password (password score of zero).

4.3. Security Knowledge/Behavior

One method that our survey used to measure a participant's security knowledge and behavior was the Human Aspects of Information Security Questionnaire (HAIS-Q) [24]. The questionnaire consists of 9 questions:

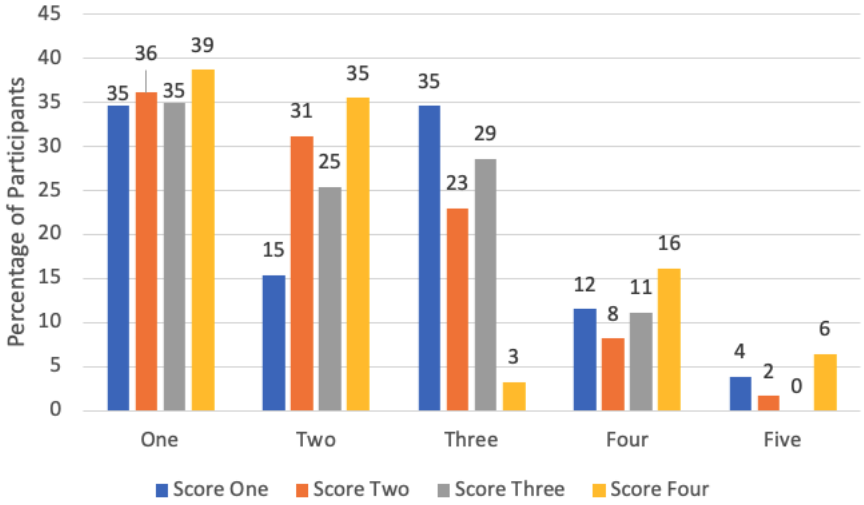

Figure 9. SONA survey: Rating of password "oklahomaStateUniversity" and password strength score. In this figure, the bar charts indicate the password score as measured by zxcvbn. The x-axis indicates the rating a participant gave for this password – 1 is "weak" and 5 is "strong".

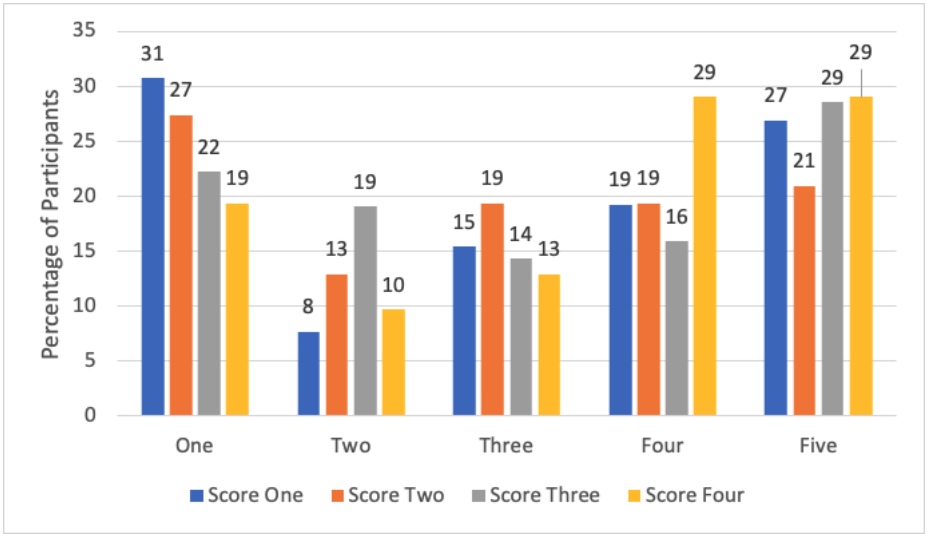

Figure 10. SONA survey: Likelihood of using the password "F63$uj2*" and password strength score. In this figure, the bar charts indicate the password score as measured by zxcvbn. The x-axis indicates how likely a participant says they are to use this password – 1 is "not likely" and 5 is "likely".

- It's acceptable to use my social media passwords on my work accounts.
- I am allowed to share my work passwords with colleagues.
- A mixture of letters, numbers, and symbols is necessary for work passwords.
- It's safe to use the same password for social media and work accounts.
- It's a bad idea to share my work passwords, even if a colleague asks for it.
- It's safe to have a work password with just letters.
- I use a different password for my social media and work accounts.
- I share my work passwords with colleagues.
- I use a combination of letters, numbers, and symbols in my work passwords.

Each question was rated on a 7-point Likert scale (1 - Strongly Disagree, 2 - Disagree, 3 - Somewhat Disagree, 4- Neither Agree nor Disagree, 5 - Somewhat Agree, 6 - Agree, 7 - Strongly Agree). The points were then added

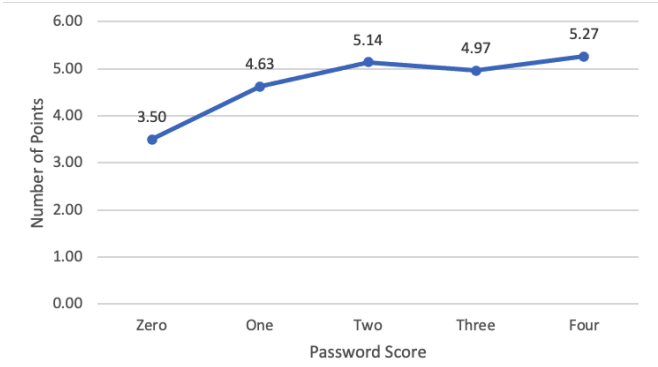

Figure 11. SONA survey: Responses about password security in relation to zxcvbn password scores.

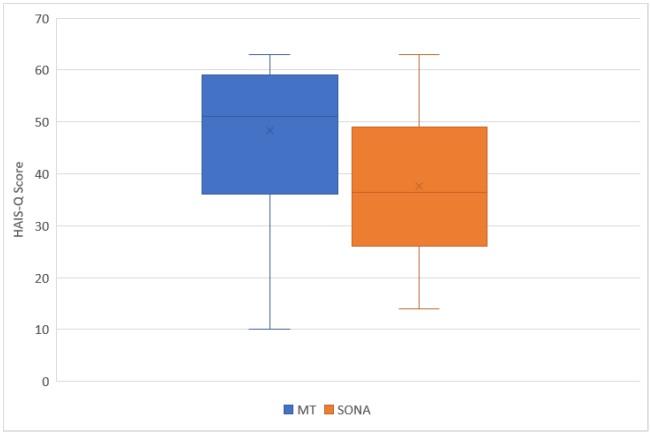

Figure 12. HAIS-Q scores for MTurk vs SONA.

up for the 9 questions for each participants. The lowest points possible was 9 points while the highest points possible was 72 points. For some of the questions (e.g. the first two questions), the points value were reversed. Figure 12 shows the distribution of HAIS-Q scores between MTurk and SONA. While the range of scores is similar, there is clearly a higher average score for MTurk participants. MTurk participants had a higher average age than SONA participants, which could imply that higher age can correlate to higher security knowledge and behavior. The average HAIS-Q score for MTurk participants was 48 and the median score was 51. The average HAIS-Q score for SONA participants was 37 and the median score was 36. The highest scores for both MTurk and SONA were 63, the lowest MTurk score was 10, and the lowest SONA score was 14.

Participants were also evaluated on their password security knowledge in each part of the survey through a series of six statements. These statements were:

- When creating a password, I always try to create one that is rated as "strong".
- Having "strong" passwords is a top priority for me.

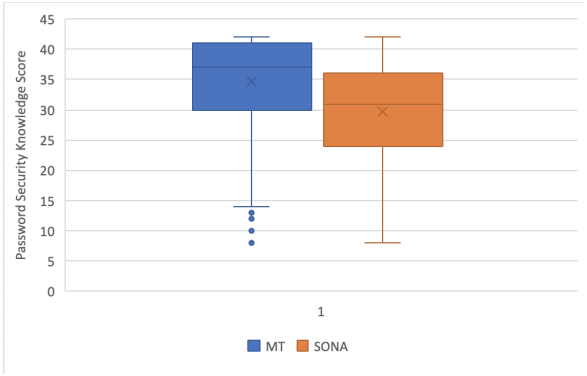

Figure 13. Password security knowledge scores for MTurk vs SONA.

- Usually, I do not make an effort to create passwords that are rated as "strong".
- If my password is rated as "weak," that is usually okay with me.
- I am concerned about the security of my passwords.
- I think that the danger from having weak passwords is exaggerated.

Similarly to HAIS-Q, each statement was rated on a 7-point Likert scale; participants rated each statement from "Strongly Disagree" to "Strongly Agree". The highest score a participant received was 42, and the lowest was 8 (minimum possible lowest was 6). Figure 13 shows that participants in the MTurk survey scored a slightly higher average score than participants in the SONA survey. The average score for MTurk participants was 35 and the median score was 37. The average score for SONA participants was 30 and the median score was 31. The highest scores for both MTurk and SONA were 42 and the lowest score for both was 8.

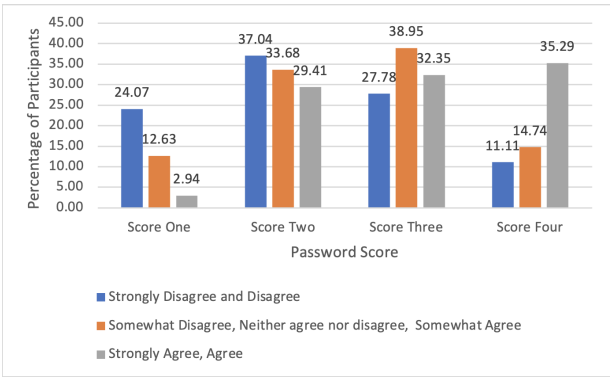

Figure 14. Responses about password security behavior practices in relation to zxcvbn password scores

Table 2. Percentage of participants of each age range in MTurk and SONA.

Age Range	MTurk	SONA
18-20	0%	79%
21-40	65%	21%
41-60	29%	0%
61+	6%	0%

Figure 14 shows the relationship between participants' responses about password security practices in relation to their zxcvbn password scores. The statement "I tend to change my passwords because I like to do something or some action to protect my account" was shown to participants. They could respond with strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, and strongly agree. Figure 14 groups responses as follows: strongly disagree / disagree, somewhat disagree / somewhat agree / neither agree nor disagree, and strongly agree / agree. Figure 14 shows that participants who responded with strongly agree or agree to statements about good security practices tended to create stronger passwords themselves.

4.4. Age

Participants in each survey were asked to provide their age at the time of the survey. Participants of the SONA survey had a lower average age than participants in the MTurk Survey, as shown in Table 2. This is expected as SONA participants are college-age students while MTurk tends to represent the general population.

5. Results (Part 2)

A month after the completion of part 1 of the survey, participants were asked to participate in part 2. For the SONA survey, there were 168 participants in part 1 and 85 participants in part 2. For MTurk, there were 391 participants in part 1 and 226 participants in part 2. This decrease in participation between part 1 and part 2 is normal for multi-part surveys [25]. Participants were asked at the beginning of each survey to create unique codes for themselves. Some participants may have forgotten or mistyped the unique codes that they created in part 1, which could have played a role in the lower number of participants in part 2. Some participants who had minor errors, such as 1 or 2 characters difference, in their unique codes in part 2 were able to be matched to their responses in part 1.

5.1. Password Strength

Figure 15 shows the variation in password scores between Part 1 and Part 2 for both MTurk and SONA.

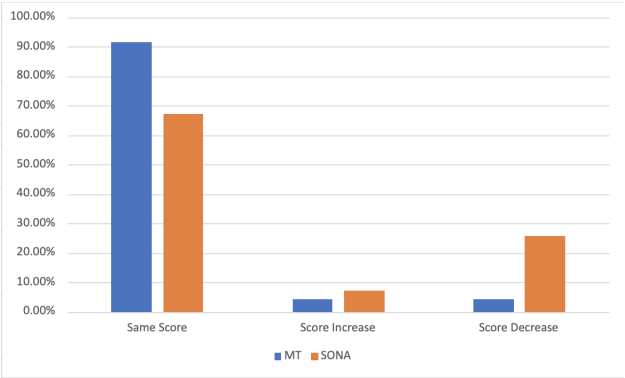

Figure 15. Difference in password strength scores between Part 1 and Part 2 for MTurk and SONA.

In Part 1, participants were asked to create a password with at least 8 characters to remember for Part 2. The goal of this was not to test the participants' memories, but to evaluate their ability to choose a strong password after a period of time. Out of the participants that were involved in both Part 1 and Part 2, 78.72% of MTurk participants wrote the same password that they created in Part 1, while 48.57% of SONA participants had the same password. Each of the passwords were scored using zxcvbn, and the participant's responses were compared from Part 1 to Part 2. In both SONA and MTurk, the majority of participant's password strength scores stayed the same in Part 1 and Part 2 (note that two different passwords might have the same strength). SONA scores, however, appeared to have a higher percentage of score decrease than the MTurk scores. The decrease in password strength score among SONA participants could be correlated to the pandemic. All of the SONA participants are college students, and therefore possibly experienced a change of environment that could have impacted the strength of their password scores. Due to this, we do not attempt to generalize the results of comparing MTurk and SONA for part 2 of the survey.

Password selection requirement scores were determined by 3 questions, which evaluated what characteristics the participants thought were included in a strong password. The participants were asked about how long a strong password should be, if a strong password should include uppercase letters, and if a strong password should include numbers. The lowest score was 0 and the highest was 3. Figure 16 shows the average password selection requirement scores for participants in both MTurk and SONA. The average score for MTurk participants in Part 1 was 3, while the average score for SONA participants in Part 1 was 2.53. The average score for MTurk participants in part 2 was 2.78, while the average score for SONA participants in part 2 was 2.45. While MTurk participants tended

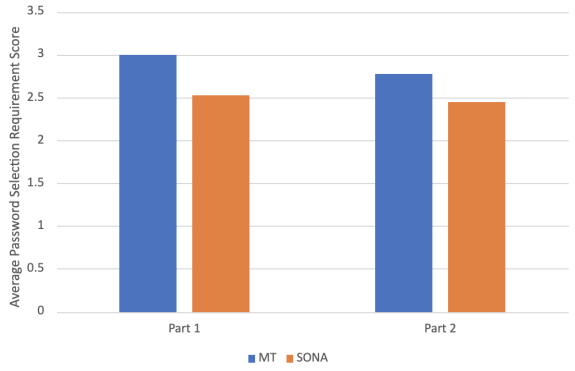

Figure 16. Password selection requirement Scores for for MTurk and SONA.

to have higher scores than SONA participants, both groups had a decrease in average score from Part 1 and Part 2. As stated earlier, this decrease in average score could have a correlation to the stress that the pandemic put on participants.

5.2. Security Knowledge/Behavior

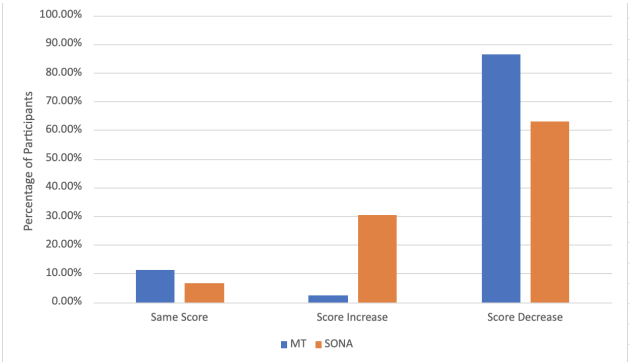

Figure 17. Differences in participants' HAIS-Q scores in Part 1 and Part 2.

In each part of the survey, participants were asked to respond to the HAIS-Q questions listed in Section 4. In both the MTurk and SONA surveys, most of the participants' scores in Part 2 were lower than their scores in Part 1. MTurk participants had a higher percentage of score increases than SONA participants. There could be different reasons for this. It could be that one month after part 1 is too long or as mentioned earlier, it could be a side effect of the pandemic.

6. Discussion

The use of both of the surveys allowed for a more diverse group of participants. SONA participants tended to be college-age students, meaning that they are considered to be part of the Gen-Z generation. The MTurk survey provided the opportunity to reach a wider age range, so that more perspectives could be accounted for. This helped give more reliable results for a wider variety of people. There were few participants younger than 20 years old in MTurk. Nearly 30% of MTurk participants received a zxcvbn score of 4, compared to less than 20% of SONA participants. In the future, maybe more surveys should be done through both a crowdsourcing platform such as MTurk, and SONA to reach all age groups.

The Big Five personality traits seemed to have an impact on password strength zxcvbn scores. People who had a Neurotic Big Five personality trait either received a 3 or 4 as their password score. However, there were very few participants who identified with the Neurotic personality type for both MTurk and SONA. To be more certain about this correlation, more participants with a Neurotic personality type would need to be surveyed. Participants with a True Colors self-schema of Orange tended to create a strong password compared to the other True Colors. There was a higher percentage of Orange participants in the SONA surveys, which could have been related to the fact that the participants of that survey were college students.

Other than age, education and socio-economic status could also have an effect. However, these were not recorded in the survey. We did not ask the participants how they remembered the password for the Part 2; this could be future work.

One factor that likely affected the results between part 1 and part 2 of the survey is the COVID-19 pandemic. The pandemic created unprecedented circumstances that increased stress and anxiety levels of participants. This could have impacted the responses of participants. In general, password strength scores and HAIS-Q scores decreased in part 2. This could mean that there is a relation between anxiety/stress and cybersecurity – this should be studied in future work. Although there is a decrease, it is small – for example the average score for SONA went down by 0.08 points or 3%. Since the number of participants in Part 2 was lower than for Part 1, it could be that only the participants with lower scores came back for Part 2 to obtain the incentive (credit and money).

7. Conclusion and Future Work

We distributed a two-parts survey through both SONA and MTurk. The survey involved questions about what participants perceive as strong password security, as well as their personality traits and self-schemas. For SONA, 168 participants completed part 1 and 85 participants completed part 2 of the survey. For MTurk, 391 participants completed part 1 and 226 participants completed part 2 of the survey.

The results show that MTurk participants tended to answer password practices better than SONA participants. This include higher HAIS-Q scores and higher password strength scores. This likely corresponds to the difference in age ranges of the participants. Another explanation is the difference in personality self-schemas and personality traits of the participants. There were more MTurk participants with the Gold and Green personality self-schemas than SONA participants, which also could have been the reason for higher password scores among the MTurk participants.

The relationship between personality self-schemas and password strength and security needs to be researched further. Our study showed there might be a correlation. Utilizing the strengths of the different personality self-schemas to create stronger password practices in professional settings could improve general cybersecurity for companies. We found a decrease in HAIS-Q score and password strength score in part 2 of the survey which occurred around the time when the pandemic was declared. The messaging and longitudinal study need to be researched during a time outside of the pandemic, which could be causing undue stress on participants of the study. This could be an interesting aspect to study as future work.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DGE 1918591 and 1919004.

References

[1] Wright, K.B. (2006) Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication* **10**(3): 00–00. doi:10.1111/j.1083-6101.2005.tb00259.x.

[2] Turk, A.M. (Accessed 2021), https://www.mturk.com/.

[3] CrowdFlower (Accessed 2021), https://appen.com/.

[4] Microworkers (Accessed 2021), https://www.microworkers.com/.

[5] Clickworkers (Accessed 2021), https://www.clickworker.com/.

[6] Wagner, A., Bakas, A., Kennison, S. and Chan-Tin, E. (2021) A comparison of sona and mturk for cybersecurity surveys. In *European Interdisciplinary Cybersecurity Conference*, EICC (New York, NY, USA: Association for Computing Machinery): 87–88. doi:10.1145/3487405.3487657, URL https://doi.org/10.1145/3487405.3487657.

[7] van Thiel, E. (accessed 2020), What are the big five personality test traits? - learn all about the theory, https://www.123test.com/big-five-personality-theory/.

[8] Lim, A.G., The big five personality traits, https://www.simplypsychology.org/big-five-personality.html.

[9] Colors, T. (accessed 2020), The four color personalities: True colors intl.: Personality assessment training, http://www.truecolorsintl.com/the-four-color-personalities/.

[10] International, T.C. (2013), https://truecolorsintl.com/wp-content/uploads/2013/05/Research-Validity-and-Reliability-I.pdf.

[11] Crews, T.B., Bodenhamer, J. and Weaver, T. (2010) Understanding true colors personality trait spectrums of hotel, restaurant, and tourism management students to enhance classroom instruction. *Journal of Teaching in Travel & Tourism* **10**(1): 22–41. doi:10.1080/15313220903558538, URL https://doi.org/10.1080/15313220903558538. https://doi.org/10.1080/15313220903558538.

[12] Dropbox (2017), zxcvbn, https://github.com/dropbox/zxcvbn.

[13] Behrend, T.S., Sharek, D.J., Meade, A.W. and Wiebe, E.N. (2011) The viability of crowdsourcing for survey research. *PsycEXTRA Dataset* doi:10.1037/e518362013-534.

[14] Qualtrics (Accessed 2021), https://www.qualtrics.com/.

[15] SurveyMonkey (Accessed 2021), https://www.surveymonkey.com/.

[16] Gamblin, B.W., Winslow, M.P., Lindsay, B., Newsom, A.W. and Kehn, A. (2017) Comparing in-person, sona, and mechanical turk measurements of three prejudice-relevant constructs. *Current Psychology* **36**(2): 217–224.

[17] Gosling, S.D., Vazire, S., Srivastava, S. and John, O.P. (2004) Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American psychologist* **59**(2): 93.

[18] Buhrmester, M. and Kwang, T. (2011), Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? on psychological science.

[19] Paolacci, G. and Chandler, J. (2014) Inside the turk: Understanding mechanical turk as a participant pool. *Current directions in psychological science* **23**(3): 184–188.

[20] Zhang, B. and Gearhart, S. (2020) Collecting online survey data: A comparison of data quality among a commercial panel & mturk. *Surv. Pract* **13**.

[21] Gratian, M., Bandi, S., Cukier, M., Dykstra, J. and Ginther, A. (2018) Correlating human traits and cyber security behavior intentions. *Computers & Security* **73**: 345 – 358. doi:https://doi.org/10.1016/j.cose.2017.11.015.

[22] Halevi, T., Lewis, J. and Memon, N. (2013) A pilot study of cyber security and privacy related behavior and personality traits. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion (New York, NY, USA: Association for Computing Machinery): 737–744. doi:10.1145/2487788.2488034.

[23] Maraj, A., Martin, M.V., Shane, M. and Mannan, M. (2019) On the null relationship between personality types and passwords. *2019 17th International Conference on Privacy, Security and Trust (PST)* .

[24] McCormac, A., Zwaans, T., Parsons, K., Calic, D., Butavicius, M. and Pattinson, M. (2017) Individual differences and information security awareness. *Computers in Human Behavior* **69**: 151–156.

[25] Wright, B. and Schwager, P.H. (2008) Online survey research: can response factors be improved? *Journal of Internet Commerce* **7**(2): 253–269.