"I found the text to be encouraging" - Evaluating Different Password Strength Calculator Designs

Rozalina Doneva
Karlsruhe Institute of Technology (KIT)
roz.doneva@gmail.com

Anne Hennig
Karlsruhe Institute of Technology (KIT)
anne.hennig@kit.edu

Peter Mayer
University of Southern Denmark (SDU)
mayer@imada.sdu.dk

***Abstract*—While passwordless authentication methods are on the rise, password-based authentication remains widely used in practice. In search of effective means to promote stronger password choices, we created and evaluated the effectiveness of six interactive password strength calculator designs with respect to usability, emotional affect, password strength, and password length, by conducting an online survey with 89 participants. The results showed that while all six designs increased password strength and length compared to the control group, the differences were not statistically significant. B ased o n t he mean values, fear-appeal nudges yielded results of similar strength to positive-feedback nudges. Still, positive feedback nudges resulted in slightly longer passwords, breaking with the paradigm that only fear appeals effectively support the creation of secure passwords. Furthermore, designs with additional information and guidance yielded longer and stronger passwords than those without, although the differences were not statistically significant. However, designs with additional information guidance exhibited significantly h igher u sability s cores, i ndicating t hat providing guidance not only has the potential to enhance password security effectively but also improves usability.**

I. Introduction

Despite a wide diversity in authentication, using a password still is the most common method for users to verify their identity [1]–[6]. Passwords offer the advantage of being a secure method given that certain security guidelines are followed [7], but suffer from common drawbacks, such as users selecting weak passwords, making them vulnerable to different types of attacks [8]. Password strength calculators, which dynamically indicate the password's strength as the user types it [4], [9], have proven to significantly increase the strength of passwords users create [10], [11]. Especially, it has been proposed that using fear appeal nudges with such calculators is more effective than merely visual nudges in the form of a simple strength meter or a colored feedback bar [4], [12]. However, Renaud et al. [13] argue that work on fear appeals might be based on weak evidence. Behavioral changes based on fear appeals may lead users to associate a topic with stress and frustration, which means they might not approach a topic constructively or avoid it altogether [14].

Furthermore, a few studies have investigated the use of alternative feedback nudges in enhancing password security. A motivational nudge was found to be at least equally effective as fear appeals nudge [15]. Gulenko [16] discussed that positive emotions, such as joy and interest, boost sociability, well-being, and constructive behavior, and may lead to the creation of stronger passwords. Furnell et al. [17] found that feedback, particularly through emoji and text-based feedback, improved password quality. These findings underscore that positive feedback nudges, rather than fear-based nudges, have the potential to effectively increase password security.

Recognizing that password authentication is a critical aspect of cybersecurity, which leaves a substantial amount of information vulnerable to users' password choices, it is crucial to develop effective strategies for encouraging users to adopt stronger, more secure passwords. Taking concerns about the use of fear appeals into account, this work compares different password-strength calculator designs, which integrate best practices reported in the existing literature, and specifically assesses the relevance of fear appeals in the cybersecurity context. In a user study with 89 participants, we compared the effectiveness of six different password-strength calculator designs, including designs that have yielded promising results, such as fear-appeal nudges, and novel approaches, such as positive-feedback nudges.

Specifically, we raised the following research questions:

- *RQ1: How do different password strength calculator designs differ regarding their effect on the security of users' passwords?*
- *RQ2: How do different password strength calculator designs differ regarding their emotional impact on users?*
- *RQ3: How do different password strength calculator designs differ regarding their usability?*
- *RQ4: Which effect does using positive feedback nudges in password strength calculators have compared to fear appeal nudges?*
- *RQ5: Which effect does using additional information and guidance in password strength calculators have compared to not providing additional information and guidance?*

By answering our research questions, we seek to determine whether fear is indeed the most effective motivator for password strength or whether alternative approaches can offer sim-

ilar – if not better – results. Additionally, our study evaluates the effectiveness of positive-feedback nudges in password-strength calculators relative to other designs. Furthermore, by comparing the effectiveness of designs with and without guidance and recommendations, we aim to determine the importance of including additional information in each design context. Ultimately, this work provides a comprehensive understanding of how various password-strength calculator designs affect password length and strength, usability, and emotional response, thereby informing the development of effective security practices.

II. Related Work

In the literature, various methods have been proposed to encourage users to create stronger passwords. In the following, several widely used strategies for improving password strength are discussed.

A. Providing Information and Guidance

Renaud and Zimmermann [18] tested the effectiveness of nine different password nudges; however, when compared to a control condition, none of the password nudges led to significantly stronger passwords. The authors found a significant limitation in their nudge designs: They did not actually improve users' understanding of what constitutes a strong password. Providing password information when a password is being created was, thus, deemed one key aspect. However, providing additional information should not raise ethical concerns, i.e., propose sanctions or use coercion. Furthermore, the additional information should be simple, clear, and not overwhelming, ensuring that users understand its implications [18]. In their subsequent work [19], the authors further defined that nudges should be transparent to users in two ways: (1) an intervention should be visible so that users are aware of it and its purpose, and (2) the reasons behind a nudge should be clear so that users understand why they are being guided toward a certain option. The authors proposed that combining nudges with additional information enhances transparency and acceptance [19].

B. Feedback Bars

Feedback Bars are often used in combination with other strategies to provide users with visual feedback on the strength of their created passwords. Thereby, they assist users in aligning their understanding with technical security standards and motivate them to strengthen their passwords without enforcing any security measures [20]. Ur et al. [10] investigated the effects of 14 different feedback bars, finding that in the absence of any intervention, users are inclined to create simple passwords. In contrast, each of the feedback bars tested in their study significantly increased the length of users' passwords [10]. In a later work, the authors investigated alternative visualizations of the commonly used feedback bars, such as color-changing bars with variations in text highlighting or an animated bunny dancing with different speeds, finding that even vastly different visualizations have little impact on password strength, and unusual feedback bars might negatively affect usability and increase the time to create a password [11]. In a systematic literature review, Zimmermann et al. [15] found that feedback bars are effective when they incorporate three key elements: (1) password strength feedback in textual or visual form, (2) visual nudges encouraging users to increase their password strength, and (3) additional information on what constitutes a strong password.

C. Fear Appeals and Cybersecurity

Another way to motivate users to create stronger passwords is by using fear appeal nudges. Amongst others, Vance et al. [4], [12], advocate for using interactive fear appeals to increase password strength. In a popular and often cited work [4], the authors evaluated the effectiveness of four treatments (Control, Password Strength Meter, Static Fear Appeal, and Interactive Fear Appeal) by conducting a user study, which showed that only the interactive fear appeal treatment led to significantly stronger passwords.

However, some researcher remain hesitant towards the effectiveness of fear appeals and their use (e.g., [13], [14], [21]–[25]). Marett et al. [23] warned that fear appeals can produce a "paralysing" effect, e.g., if a fearful message is overly extreme about potential damage, recipients might take actions that deviate from the recommended behavior, driven by a sense of hopelessness. Similarly, Renaud et al. [14] alerted that those who associate cybersecurity with fear, stress, and frustration are unlikely to approach a given cybersecurity-related subject constructively. Moreover, fear, as a short-lived emotion, may backfire in the long term, due to the negative affective states that it triggers [24]. When investigating different elements of fear appeals in the context of password hygiene, Dupuis et al. [22] found that fear appeals are most effective when combined with information on how to counter the threat and on the effectiveness of those countermeasures. Furthermore, Ruiter et al. [25] found that providing coping information to enhance perceptions of response effectiveness and self-efficacy is more crucial for encouraging protective action than presenting threats to raise risk perception and fear.

D. Fear Appeals Alternatives

Several studies propose more positive approaches that encourage users to create secure passwords [15]–[17], [26]. According to Gulenko [16], incorporating positive emotions in interventions can be achieved by using emoticons and positive messages. The author reports that users tend to create stronger and more secure passwords when experiencing positive emotions such as joy and interest during the password creation process [16]. In their systematic literature review, Zimmermann et al. [15] found a motivational nudge at least as effective as a fear appeals nudge. When investigating how different types of guidance and feedback can influence users' password choices, Furnell et al. [17] showed that providing interventions with emoji and text-based feedback yields the strongest passwords. Coopamootoo [26] investigated the role of empathy in mitigating frustration, suggesting that

incorporating empathy into security interventions can enhance password security and reduce users' frustration.

III. Methodology

To answer our research questions, we conducted an online survey using a between-subjects design with multiple implementation steps. First, we describe the general structure of our online survey (Section III-A). Then, we describe the zxcvbn algorithm that we used to measure strength and length of the passwords our participants entered, and explain its implementation (Section III-B). Next, we describe in more detail the six different password strength calculator designs (i.e., treatments) that we compared in our user study (Section III-C), and our pretest and recruitment process (Section III-D). Finally, we describe how we analyzed the data (Section III-E), and discuss potential ethical concerns and how we addressed them (Section III-F).

A. Survey

We used the online survey platform SoSciSurvey[1] for our study. Figure 1 visualizes all steps of the survey procedure. The survey can be found in Appendix B2. At first, the participants were greeted, given initial information about the study and informed that no personal data or plain text passwords would be stored. After the participants agreed to participate in the study, we introduced them to the task. We explained that we are evaluating a sign-up functionality for a website. We asked them to imagine creating an account. Participants were instructed to create a dummy password for their account. To reduce the likelihood that participants would use real passwords and ensure we observe genuine password-creation practices rather than reusing a preferred, already existing password, we clearly advised participants to create an entirely new, memorable password. We instructed participants to take as much time as needed to create a password before confirming their choice.

After the instruction, we asked two attention-check questions. Then, participants were asked to complete the Positive and Negative Affect Schedule (PANAS) scale [27] to measure their positive and negative affect before completing the task [Q1]. Afterwards, the password creation task followed. Each participant was randomly assigned to one of the treatments, i.e., only one of seven designs (see Section III-C for more information on our treatment groups).

Next, participants were asked to complete the PANAS scale again to measure their emotional state after task completion [Q2]. We used [Q1] and [Q2] to answer *RQ2*, whether we see any differences in users' emotional response with regard to different treatments. Then, participants were asked to rate the usability of the sign-up functionality [Q3], using the System Usability Scale (SUS) [28]. [Q3] was used to compare the usability of the different treatments and, thus, to answer *RQ3*. To gather feedback on participant interaction with the treatments, we asked them to briefly describe their interaction with the elements below the password field during the sign-up procedure [Q4]. Finally, we asked whether participants usually use password managers to remember passwords [Q5], as this might influence their password creation strategies, and checked for their technological affinity [Q6], using the Affinity for Technology Interaction (ATI) scale [29]. We used both, [Q5] and [Q6], as control variables. The answers to [Q4] were not systematically analyzed and are only presented anecdotally in the discussion.

Afterwards, we debriefed the participants about the actual purpose of the study (evaluating the effectiveness of different password strength calculators). We also asked if they wanted to have their data removed, now that they had the full information about the study. The survey ended by providing all participants with information on practices to increase the strength of a password. In accordance with the principle of data economy, we did not collect further demographic data of our participants (e.g., age, gender, or formal education), as that information was not needed to answer our research questions.

B. Strength Estimation Algorithm and Implementation

For the implementation of the treatments, we used HTML, JavaScript, and CSS. To determine the strength and length of the password a user entered during the sign-up procedure, each treatment used zxcvbn [30], a heuristic-based strength estimation algorithm that was also utilized in related studies evaluating password interventions [11], [15], [18], [21], [26], [31].

The algorithm has an easy-to-integrate design [32] and offers a more realistic evaluation of password complexity compared to other strength algorithms [26]. Furthermore, zxcvbn outputs values that represents the strength of a password, i.e., the number of guesses an attacker would need to break the password (guesses_log10), and the length of the password, i.e., the number of characters that were used. This allowed us to discard the password a user entered in plain text, as we could not be entirely sure that they would not enter a password that is close to an existing one. As soon as participants confirmed their password choices, the values for password strength and length were stored, allowing us to later evaluate the effectiveness of the different treatments regarding password security (*RQ1*).

C. Treatments

We designed a total of six different password strength calculator treatments that contained one of three types of nudges: (1) colored feedback bar, (2) fear appeal nudge, (3) positive feedback nudge, and one control treatment without any nudge. Half of the six nudge treatments also contained additional information and guidance on how to create a secure password [TXb], similar to Vance et al.'s interactive fear appeal design [4]. The other half of the treatment groups contained only the nudge and no additional information [TXa]. All of our treatments were interactive, i.e., changed their feedback based on what the users entered. Like Vance et al. [4], we did not enforce any policies for either treatment (e.g., by

[1]https://www.soscisurvey.de

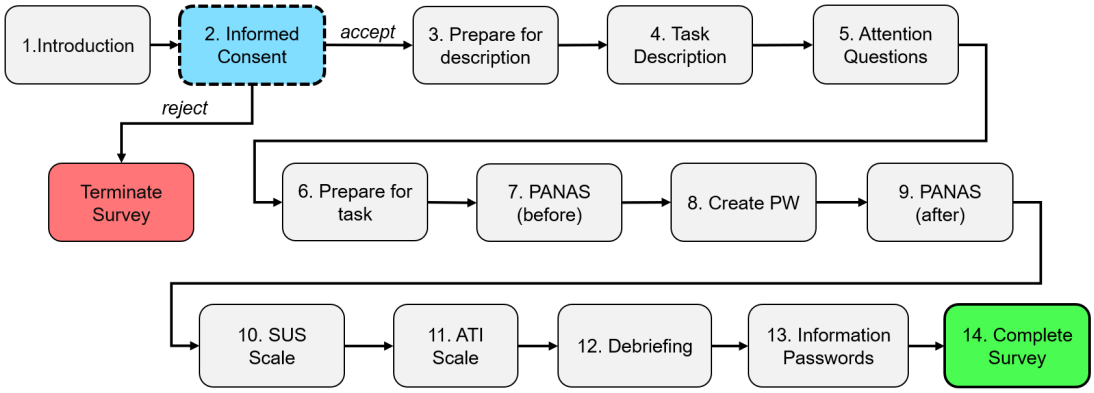


Fig. 1: Survey Procedure

not letting the user continue if not all recommendations were met). In total, we compared the following treatments: [T0] control treatment, [T1a] colored feedback bar - no guidance, [T1b] colored feedback bar - guidance, [T2a] fear appeal - no guidance, [T2b] fear appeal - guidance, [T3a] positive feedback - no guidance, [T3b] positive feedback - guidance. By comparing treatments [T2] and [T3], we answer *RQ4* and determine whether we observe significant differences between positive feedback nudges and fear appeal nudges.

a) Treatment Variants [TXb]: Additional Information and Guidance: Additional Information and guidance was provided in the form of an interactive checklist, similar to the interactive design used by Vance et al. [4] (see Figure 2). However, we updated the recommendations with up-to-date guidance from the National Institute of Standards and Technology (NIST) [33]. We implemented the guidelines in a more permissive manner, e.g., specifications such as minimal length were applied as recommendations instead of requirements for password creation. In the same manner, a comparison against commonly used values was done. However, instead of being required to select a different password, users were merely advised to change the current passwords if one of the recommendations was not met. Not imposing strict rules allowed us to observe genuine password choices and a fair comparison of password strength across all created treatments. Another guidance for password creation, which we included in our checklist, was derived from Kävrestad et al. [34], who recommended passphrases – passwords formed by linking four or more words – as those are both secure and memorable. Finally, we added an additional button ("Why?") next to each recommendation to enhance transparency about the recommendations used and to provide more detailed information to users while increasing interactivity. This idea was derived from Ur et al. [35], who used hyperlinks to display sources for their recommendations. As highlighted by Zimmermann et al. [19], transparency is crucial for users to accept such guidance. We compared the treatments with and without guidance to determine whether either of these conditions has a significant effect on our variables to answer *RQ5*.

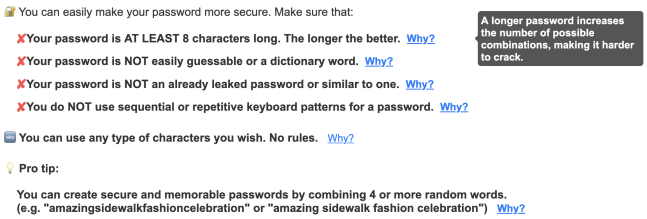


Fig. 2: Checklist with all Recommendations and the "Why?" Button Activated for Tips on Password Length.

b) Treatment 0: Control Treatment: Our control treatment, as shown in Figure 3, is a plain password entry field without any visible strength feedback or guidance on how to increase password security, providing users with minimal, basic functionality for creating a password and serving as a baseline for our treatment comparison.

Fig. 3: [T0] – Control Treatment

c) Treatment 1: Colored Feedback Bar: According to Zimmermann et al. [15], effective treatments that communicate password strength have three key elements: password strength feedback, visual nudges to encourage stronger passwords, and additional password creation guidance. We included the colored feedback bar treatment as the most straightforward design that incorporated the recommendations by Zimmermann et al. [15]. Password strength feedback is provided in both textual and visual form via a colour-changing progress bar, similar to the one used in Vance et al. [4], and a message which explains the corresponding strength category (i.e., "weak password" in the red area, "strong password" in the green area, see Figure 4 for an example).

Interactive recommendations in the form of a checklist, as described above, are only included in treatment [T1b]. The checklist is positioned below the password meter (see Appendix VI-A, Figure 7 for an example).

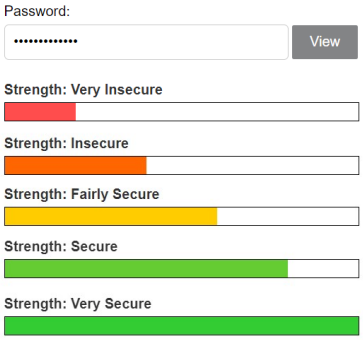


Fig. 4: [T1a] – Colored feedback bar without guidance. Depending on the user's input, the visual and textual feedback changes, and one of the feedback bars is displayed.

d) Treatment 2: Interactive Fear Appeals: The fear appeals treatments as shown in Figure 5 and 8 incorporate fundamental elements of Protection Motivation Theory (PMT) [36].

- *Perceived Susceptibility* – Users are informed that common, weak passwords can be easily guessed by hackers, by displaying the estimates for offline attacks with a slow hash that corresponds to 10,000 guesses per second.
- *Perceived Severity* – Users are confronted with negative consequences of compromised passwords, stating that hackers would be able to access other accounts that use a similar password.
- *Response Efficacy and Cost* – Users are shown general advice that stronger passwords are essential to prevent security breaches, without including detailed information on how to create a strong password.
- *Self-Efficacy* – Users are reassured that they have the means to avoid the potential threat by choosing a very secure password.

While the non-guidance variant of the treatment (see Figure 5) includes only general advice, e.g., stronger passwords are essential to prevent security breaches, the guidance treatment (see Appendix VI-A, Figure 8) provides detailed, interactive information on how to create stronger passwords.

Fig. 5: [T2a] – Interactive fear appeals nudge without guidance.

e) Treatment 3: Positive Feedback: The positive feedback treatment is based on the intervention design from Furnell et al. [17], which combines emoticons and textual messages aimed to reach the user on an emotional level. Before any password is typed or when the already-typed password is deleted, the user sees an empathetic message derived from Coopamootoo [26] to address potential feelings of frustration. This message provides brief information about the importance of password strength and assures users of their own capability to create a secure password. The strength of a typed-in password is visualized with an emoji and a textual message found in Furnell et al. [17] containing phrases to express praise, encouragement, and empathy. Again, treatment 3 is designed both without [T3a], see Figure 6, and with interactive recommendations in the form of a checklist [T3b], see Appendix VI-A, Figure 9.

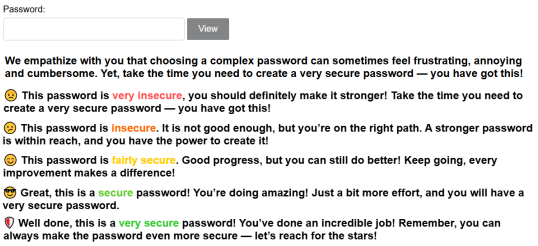


Fig. 6: [T3a] – Positive feedback nudge without guidance. While the user is typing, only one of the feedback statements is displayed, changing with the strength of the typed password.

D. Pre-Test and Recruitment

Before the survey was launched, a pre-test was conducted to improve clarity of our wording, eliminate potential ambiguities, and test technical functionality. The pretest was conducted with one senior IT security researcher, several researchers and students with an IT security background working at one of the author's institutions, and four volunteers with no prior knowledge of cybersecurity. After the pretest, some textual changes were made to provide clarity on the task description. Furthermore, the question on password manager usage was added.

After we implemented the changes suggested in the pre-study, we launched our survey. Participants were recruited using snowball sampling, including social media (i.e., personal and institutional Twitter, Mastodon, and LinkedIn accounts; see Attachment B1 for the social media posts we used), personal contacts, and word of mouth. The survey was launched on December 3, 2024, and was available online until December 15, 2024. All participants who took part in the survey were volunteers. No compensation was offered for their participation.

E. Data Analysis

We tested the effect of the different password strength calculator designs on password strength and length, emotional affect, and usability with one-way ANOVA to answer *RQ1*,

RQ2, and *RQ3*, and potential differences between the fear appeals nudge treatment and the positive feedback treatment, as well as between the treatments with and without guidance, with independent sample t-tests (*RQ4* & *RQ5*).

For one-way ANOVA and t-tests, our independent variable was password strength calculator design with seven nominal characteristics (control [T0], colored feedback bar no guidance [T1a], colored feedback bar guidance [T1b], interactive fear appeal no guidance [T2a], interactive fear appeals guidance [T2b], positive feedback no guidance [T3a], positive feedback guidance [T3b]). We defined three dependent variables: password security with two continuous characteristics (password strength and length); emotional affect with two continuous characteristics (PANAS positive diff, PANAS negative diff); and usability with one continuous characteristic (SUS). As described in Section III-A, we controlled for the password strength calculator design by randomly assigning participants to the designs.

Further, we tested which of our variables best predicted password strength or length using multiple linear regression. For our linear regressions, we used password strength and length as dependent variables, and password strength calculator design, emotional affect, usability, the use of a password manager two nominal characteristics (*use* for password creation vs *not use* for password creation), and technological affinity (one continuous characteristic) as independent variables (predictors).

We used SPSS for our data analysis. Before executing statistical tests on the collected data, we calculated the overall results of the PANAS positive, PANAS negative, ATI, and SUS scales for each participant, following the instructions by Watson et al. [27] for the PANAS scale, Franke et al. [29] for the ATI scale, and Brooke [28]) for the SUS scale. We tested the assumptions of each statistical test, and used non-parametric alternative where the assumptions were violated. All tests were conducted at an alpha level of .05, with post hoc Holm-Bonferroni corrections applied to counter alpha error cumulation where necessary.

We included an open-ended question to solicit feedback from participants. We anecdotally report relevant quotes from the feedback and include them in the discussion of our results. However, we did not systematically analyse these data.

F. Ethical Considerations

The study procedure was approved by the ethics committee of our university. We adhered to all ethical principles for research involving people published by our university. Specifically, all survey data, i.e., emotional affect or technological affinity, were collected anonymously. The SoSciSurvey entity we used for the online study is hosted on the servers of our university. Participants were informed about data processing and storage, and had to consent before taking part in the survey. Participants were not exposed to any physical or mental harm throughout the survey. We informed the participants about the voluntary nature of their participation and explained that they could leave the survey at any time without any disadvantage.

The study was framed as a test of sign-up functionalities, while we indeed evaluated the effectiveness of different password strength calculators. Therefore, we adhered to common practices for deception studies, i.e., debriefed participants immediately after the survey and offered that they could request to have their data deleted. In the debriefing, we explained the actual research goal and revealed full details about the nature of the deception and why we deemed it necessary for our study.

IV. Results

In total, 97 participants completed the survey. For further analysis, we excluded four participants who did not enter any password (e.g., because they tried to insert a password via copy & paste) and two participants who asked to have their data deleted after the debriefing. We further identified two extreme outliers in the exploratory data analysis for password strength. Since their strength values were notably higher than all other cases in the survey, we excluded them from the analysis.

Overall, our sample was rather average in technological affinity (M = 3.89, SD = 0.74) and less inclined to use a password manager. While 31.5% of participants reported they use a password manager to store and create their passwords, 28.1% use a password manager to store passwords but not to create them, and 40.4% do not use a password manager at all.

A. Descriptive Statistics

Judging from the descriptive analysis, none of our treatments clearly outperformed the others across all variables (see Table I for all results). The design using the colored feedback bar and providing additional information and guidance [T1b] scored highest with respect to password strength (M = 12.63, 95%-CI[10.49, 17.76]), and had the greatest negative difference in negative emotional affect (M = -0.19, 95%-CI[-0.40, 0.001]), which means it had the least negative impact on participants' emotional state. Whereas the positive feedback with guidance design [T3b] resulted in the longest passwords (M = 15.77 characters, 95%-CI[12.13, 19.40]) and the greatest difference in positive emotional affect compared to before (M = 0.07, 95%-CI[0.30, -0.30]), which means it had the greatest positive impact on participants' emotional state. Interestingly, this treatment also showed the largest positive difference in negative emotions (M = 0.06, 95%-CI[-0.25, 0.38]), indicating the greatest negative impact on participants' emotional state. The control treatment [T0] resulted in the least secure passwords, both in strength (M = 9.38, 95%-CI[6.42, 12.33]) and length (M = 11.50, 95%-CI[9.13, 13.87]). Finally, the fear appeals nudge with guidance treatment [T2b] scored highest in usability (M = 80.58, 95%-CI[72.09, 89.06]), while the fear appeals nudge without guidance treatment [T2a] scored lowest (M = 63.75, 95%-CI[56.38, 71.12]).

B. RQ1 – RQ3: Differences between the treatment groups

To provide answers to our first three research questions, i.e., whether different password strength calculator designs

				95% Confidence Interval	
	Variable	**Mean**	**Std. Deviation**	**Lower Bound**	**Upper Bound**
control group [T0]	negative emotions diff	-0.19	0.32	-0.40	0.01
colored feedback guidance [T1b]	strength	12.63	3.67	10.49	14.76
	positive emotions diff	-0.24	0.38	-0.46	-0.02
	negative emotions diff	-0.19	0.30	-0.37	-0.02
fear appeals guidance [T2b]	SUS	80.58	14.04	72.09	89.06
positive feedback guidance [T3b]	length	15.77	6.02	12.13	19.40
	positive emotions diff	0.07	0.38	0.30	-0.30
	negative emotions diff	0.06	0.52	-0.25	0.38

TABLE I: Treatments with variables that scored highest and lowest for the respective categories.

differ with respect to their effect on the security of users' passwords, the emotional impact on users, and their usability, we used one-way ANOVA. We compared the different password strength calculator designs with respect to differences in password strength and length (*R1*), as well as differences in emotional affect (*R2*) and usability (*R3*). Normal distribution was violated for length in treatment group [1a], strength in treatment group [T3b], the difference in negative emotions in the control treatment [T0], and the difference in positive emotions in treatment groups [T1a] and [T2a], but data were normally distributed for all other factors as assessed by the Shapiro-Wilk test ($\alpha = .05$). Although ANOVA has proven robust against violations of the assumption of normality [37]–[41], we used non-parametric Krukal-Wallis test to confirm our results. Homogeneity of variance (Levene's test, $p > .05$) was given for all factors.

The results showed no statistically significant differences between the groups for neither strength, $F(6, 82) = 0.84, p = .545$, nor length of the passwords, $F(6, 82) = 1.16, p = .338$, nor the emotional affect (see Table II for all results). However, we identified significant differences in usability with a large effect size, $F(6, 82) = 2.70, p = .019, \eta^2 = .165$. Post-hoc Games-Howell tests revealed significant differences in usability scores between the [T2a] and [T2b] treatment groups (fear appeal guidance vs. fear appeal no guidance), $p = .045$. Applying the Holm-Bonferroni correction unfortunately did not confirm these findings (see Table II). The results of non-parametric Kruskal-Wallis test confirmed these findings, showing that the treatment groups had neither an effect on strength ($H(6) = 6.65, p = .464$), length ($H(6) = 7.40, p = .286$), nor emotional affect (positive emotions: $H(6) = 6.09, p = .413$; negative emotions: $H(6) = 6.19, p = .403$). Again, treatment groups had a significant effect on usability, but the effect was very weak ($H(6) = 14.72, p = 0.023, \eta = .093$). Post-hoc tests showed that differences exist between several groups (see Table III), however, adjusted p-values with Bonferroni correction did not confirm the significant differences.

The different password strength calculator designs in our study did not affect password security or emotional affect. However, we identified that treatment [T2b] has significantly higher SUS scores than treatment [T2a].

C. RQ4: Positive Feedback Nudge vs. Fear Appeals Nudge

To answer which effect using positive feedback nudges in password strength calculator design has compared to using fear appeal nudges (*RQ4*), we conducted an independent samples t-test. We compared the fear appeals nudge treatments [T2a] and [T2b] against the positive feedback nudge treatments [T3a] and [T3b] with respect to password security, emotional affect, and usability. The normality assumption was violated for all factors except SUS score and password strength, as assessed by the Shapiro-Wilk test ($\alpha = .05$). However, t-tests have proven robust to violations of the normality assumption, even with sample sizes as small as 30 (which we exceeded by a large margin) [42]–[45]. Furthermore, homogeneity of variance (Levene's test, $p > .05$) was found, so we did not consider nonparametric alternatives. The t-test revealed no statistically significant differences across the groups for any of the variables $p < 0.05$ (see Table IV for detailed results).

There is no significant difference between the positive feedback nudge and the fear appeals nudge designs.

D. RQ5: Effect of Additional Information and Guidance

Our fifth research question asked whether using additional information and guidance in password strength calculators has an effect compared to not providing additional information and guidance. To answer *RQ5*, we compared the treatments without additional information and guidance ([T1a], [T2a], [T3a]) to the treatments with additional information and guidance ([T1b], [T2b], [T3b]). Again, assumption of normality

	stat.	df1	df2	p	η^2	adj. p
strength	0.84	6	82	.545	–	.963
length	1.16	6	82	.338	–	.963
SUS	2.70	6	82	.019*	.165	.095
positive emotions diff	1.54	6	36.28	.192	–	.768
negative emotions diff	1.19	6	82	.321	–	.963

TABLE II: Results of the One-Way ANOVA for all variables. Significant Differences are Indicated by *.

	stat.	p	adj. p
fear appeals no guidance vs. control	0.51	.613	1.000
fear appeals no guidance vs. feedback bar no guidance	0.84	.403	1.000
fear appeals no guidance vs. feedback bar guidance	1.78	.075	1.000
fear appeals no guidance vs. positive no guidance	-2.11	.035*	.742
fear appeals no guidance vs. positive guidance	-2.73	.006*	.134
fear appeals no guidance vs. fear appeal guidance	-2.81	.005*	.103
control vs. feedback bar no guidance	-0.32	.750	1.000
control vs. feedback bar guidance	-1.25	.210	1.000
control vs. positive no guidance	-1.60	.110	1.000
control vs. positive guidance	-2.21	.027*	.568
control vs. fear appeals guidance	-2.30	.022*	.454
feedback bar no guidance vs. feedback bar guidance	-0.95	.343	1.000
feedback bar no guidance vs. positive no guidance	-1.31	.190	1.000
feedback bar no guidance vs. positive guidance	-1.93	.053	1.000
feedback bar no guidance vs. fear appeals guidance	-2.02	.044*	.915
feedback bar guidance vs. positive no guidance	-0.41	.686	1.000
feedback bar guidance vs. positive guidance	-1.02	.309	1.000
feedback bar guidance vs. fear appeals guidance	-1.11	.269	1.000
positive no guidance vs. positive guidance	-0.58	.561	1.000
positive no guidance vs. fear appeals guidance	0.67	.505	1.00
positive guidance vs. fear appeals guidance	0.09	.930	1.000

TABLE III: Results of the post-hoc test with Bonferroni correction for usability. Significant Differences are Indicated by *.

	stat.	df1	p	d
strength	-0.03	48	.975	–
length	-0.70	48	.488	–
SUS	-1.40	48	.170	–
positive emotions diff	0.21	48	.839	–
negative emotions diff	0.31	48	.758	–

TABLE IV: Results of the Independent Samples T-Test for Positive Feedback Nudge Treatments [T2] and Fear Appeals Nudge Treatments [T3].

was violated for all factors except SuS score and password strength, as assessed by the Shapiro-Wilk test ($\alpha = .05$). However, as homogeneity of variance (Levene's test, $p > .05$) was found and our sample size was sufficiently large to assume that our data are approximately normally distributed according to the central limit theorem, we proceeded with parametric independent-samples t-test. Independent-samples t-tests showed no significant differences between the groups with respect to password strength and length, or emotional affect (see Table V). However, again, statistically significant

	stat.	df1	p	d
strength	-0.23	75	.819	–
length	-0.79	75	.432	–
SUS	-2.70	75	.009*	-0.616
positive emotions diff	-0.90	75	.372	–
negative emotions diff	-0.84	75	.401	–

TABLE V: Results of the Independent Samples T-Test for the Treatments With and Without Additional Information and Guidance. Significant Differences are Indicated with *.

differences with a medium effect size were observed for usability, $t(75) = -2.70, p = 0.004, d = -0.616$, with a mean difference of $-9.003(95\% - CI[-15.643, -2.364])$.

> Treatments with additional information and guidance have significantly higher SUS scores than treatments without.

E. General Model

We used linear regression to identify which of our variables (password strength calculator design, usability, emotional affect, use of password managers, and technological affinity) best explain password strength and length. We tested the assumptions using SPSS and identified one significant outlier for password strength, which we excluded from our analysis. All other assumptions (linear relationship, independence of observations, homoscedasticity, and normal distribution) were met. The R^2 for the overall model was .13 (adjusted $R^2 = -.02$), which indicates a low to moderate goodness-of-fit according to [46]. However, our predictors were not able to statistically significantly predict password strength, $F(7, 41) = .87, p = .540$.

For password length, we identified four possible outliers. However, as they were just above the threshold only for leverage, we did not exclude them from our analysis. Again, all other assumptions were met. The R^2 for the overall model was .18 (adjusted $R^2 = .12$), which indicates a moderate goodness-of-fit according to [46]. Unlike the model for password strength, our predictors were able to statistically significantly predict password length, $F(6, 82) = 2.90, p = .013$.

While the strength of a password cannot be well predicted by our predictors, this is possible for password length.

F. Feedback from Participants

As described in Section III-E, we added an open-ended question to our survey and asked users for feedback. In total, 68 participants provided some feedback, which was not always related to the password strength calculator design (i.e., two participants described which password they had chosen). We learned from the responses that participants often did not engage with the information we provided. P42 stated that they did not pay attention to the information, as they are accustomed to such tasks: "*Not at all [did I interact with the information]. This field is widely used across many web applications, so I am used to it. I usually pick quite strong passwords, so I don't pay much attention to it.*" Citing P20, it becomes clear that creating a password is not a task that requires users to allocate (a lot of) cognitive resources: "*I paid minimal attention, it required almost none of my mental capacity. I did it after work, so I might have been a bit distracted and I am not completely sure if I did everything properly.*"

In contrast, only a few participants described that they interacted actively with the design, which made them feel more confident in their password choices, e.g., *"I looked over them and made sure to cover the aspects which would help me have a secure password. When I saw that my password would be really secure, I felt confident about it*" (P54), or P80, who stated: "*the colors helped determine whether I need to make my password more complex and the sent[e]nces guided me what I chose wrong about them.*"

Some participants specifically commented on the design, e.g., P49 who provided feedback on the positive feedback treatment: "*I found the text to be encouraging*", or P58 who commented on the fear-appeals treatment: "*[T]he estimated time for breaking the PW was a [m]otivation to choose a good pw.*" Others mentioned that they were paying attention to the password strength calculator out of interest without clearly referring to it being helpful, e.g. P26: "*While I was trying on different passwords, I looked at the emojis change*" or P43 who described that the task: "*[...] was easy and funny.*"

V. Discussion

A. RQ1 & RQ2: Effect of Different Password Strength Calculator Designs on Password Security and Emotional Affect

The first research question investigated differences between our seven password strength calculator designs with respect to password security, i.e., password strength or length. Using one-way ANOVA, we could not identify any differences, not even between the treatment groups and the control group. This implies that either design was nearly equally (non-)effective in urging users to create secure passwords. However, we see differences in mean values, with the control group resulting in least secure passwords, while the colored feedback bar with additional information and guidance had the highest strength, and the positive feedback with guidance design had the highest length values. Inabilities to detect significant effects might stem from limitations in our methodology, e.g., the limited sample size (see Section V-E).

Furthermore, feedback from participants indicated that they often did not interact with the password strength calculator design. This is, on the one hand, positive, as users were not primed. On the other hand, this indicates that the choice of password was made independently of the password strength calculator.

Another reason might be that the distinction between the password strength calculator designs, and specifically the fear-appeal aspect of the fear-appeal treatments, was not sufficiently concrete or personally relevant [47]. Although Vance et al. [4] concluded that the effectiveness of their treatment was explicitly grounded in the fear appeals aspect, the conclusion might be based on weak evidence. The study compared only a fear-appeal treatment with a basic treatment. However, it can be assumed that other design factors of the "fear appeal" treatment, such as the additional information on password security provided by the authors, were more likely to be decisive for the high level of effectiveness. This is also supported by our findings with respect to the additional information and guidance treatments, which we will discuss in Section V-D.

So, while researching effective password strength calculator designs seems interesting from a research perspective, future studies should modify our methods to get more definitive answers to the research questions we posed. On a positive note, we could also not show that either of the treatments had significant negative effects on participants' emotional state. Again, mean values indicate that differences in positive emotions were, on average, higher for participants who used the password strength calculator design with the positive feedback nudge and additional information and guidance. Differences in negative emotions were lowest for the colored feedback bar design with additional information and guidance, and the control treatment.

Thus, we conclude that further research is needed using a larger and more diverse sample to validate our findings.

B. RQ3: Effect of Different Password Strength Calculator Designs on Usability

While we did not observe significant differences for password security or emotional affect, usability scores for the password strength calculator design with the fear appeals nudge and additional information and guidance were significantly higher than for the design only with the fear appeals nudge. This indicates that participants valued additional information and guidance, which we will discuss in more detail in Section V-D.

C. RQ4: Fear Appeals Nudge vs. Positive Feedback Nudge

The fourth research question investigated whether fear appeal nudges are the only effective method to urge users to create secure passwords, as suggested by Vance et al. [4]. Interestingly, our study could not identify that any of our

designs was most effective, as discussed in Section V-A. Similarly, we could also not show statistically significant differences in password strength, password length, usability, or emotional affect when solely comparing password strength calculator designs with fear appeal nudges to designs with positive feedback nudges.

While related work suggests that positive emotions, elicited by positive feedback nudges, can foster stronger and more secure or at least equally strong passwords [15], [16], we can only confirm that, for our study, the password strength calculator design using positive emotions and additional information and guidance scored highest for password length. Although no statistically significant differences were found, our results suggest that incorporating positive feedback nudges may contribute to longer and, consequently, more secure passwords – breaking the paradigm that only fear appeals nudges are effective in urging users to create secure passwords. Thus, we support Zimmermann et al.'s [15] conclusion that fear appeals may not be essential for urging users to create stronger passwords, and suggest that positive feedback nudges could serve as a viable alternative.

Interestingly, the password strength calculator design with fear appeal nudges and additional information and guidance scored highest in usability. Since the positive feedback design, which has additional guidance, is rated only slightly less usable, we assume that not the nudges were rated more usable, but rather the additional information and guidance (see Section V-D).

D. RQ5: Effect of Additional Information and Guidance

Finally, our fifth research question should explore whether providing additional information and guidance has an effect on password strength, password length, usability, or emotional affect. While our statistical analysis revealed no statistically significant difference in password strength or length, or emotional affect between the two groups, usability scores were significantly higher for password calculator designs providing additional information and guidance. This is also supported by P14, who stated: "I checked all 'Why?' popups as the information was interesting to me."

Thus, we conclude that providing additional information and guidance for users helps them to interact with a password strength calculator design. Furthermore, increasing usability by providing additional information and guidance did not negatively impact password strength and – for some treatments – even showed higher values for password security. This suggests a potential positive relationship between usability and security. Overall, further research is needed on whether additional information and guidance can also be effective in increasing password security.

E. Limitations

One of the main limitations of this work is the limited sample size. A priori power analysis indicated that at least 343 participants were required for a one-way ANOVA with seven groups, while we only included 89 participants in our analysis. This poses a risk that the affected tests are underpowered, potentially leading to a type II error, i.e., significant differences not being identified. Interestingly, related work from Ur et al. [10] found significant differences for even lower differences in mean values than occurred in our study. Furthermore, we learned from the responses to open-ended questions that participants often did not engage with the information we provided, which may explain why we did not see differences between our treatments. This indicates that further research is needed to validate our findings with a larger sample.

Not only did we not achieve the calculated sample size necessary, but all our participants were volunteers, which makes our study susceptible to sampling bias. It may be that, due to our sampling methods, we only recruited those interested in the topic. However, the mean value for technological affinity (ATI value) was 3.90 for all participants, indicating a moderate affinity for technologies. This suggests that we did not explicitly sample for IT (security) experts; however, our participants chose, in general, long and strong passwords. Again, further research is needed to validate our findings with a more diverse sample.

Another limitation of this work is that the sign-up procedure was conducted in a – kind of – lab setting, where participants created dummy passwords for a fictional account rather than real passwords for a real account. Future research should aim to evaluate such treatments in more realistic settings, possibly bypassing that participants are ignoring relevant information.

VI. Conclusion

Although password managers and password-less login functionalities could eventually make user-generated passwords obsolete, finding effective ways to increase the security of user-generated passwords is still meaningful. Sign-up functionalities requiring user-generated passwords are still widely used in practice [1]–[6]. Furthermore, our user study revealed that more than two-thirds of our participants still create their passwords manually, without using a password manager. For this reason, the goal of our research was to evaluate the effectiveness of different password strength calculator designs with respect to password security (i.e., password strength and length), usability, and emotional affect. In particular, we sought to compare fear appeals nudges with positive feedback nudges, and to evaluate the effect of additional information and guidance. In a user study with 89 participants, we could show that none of our password strength calculator designs was more effective with respect to password security, usability, or emotional affect. However, our results revealed that a colored feedback bar with additional information and guidance resulted in strongest passwords, while a positive feedback nudge with additional information and guidance resulted in longest passwords. Interestingly, the fear appeals nudge with additional information and guidance scored highest in usability. When comparing treatments with additional information and guidance against those without, we found the ones with additional information and guidance to score significantly higher in usability than those without. These findings suggest

that while additional information and guidance enhance usability, further research is needed to determine why none of the treatments significantly improved password security compared to the control group. We speculate that our sample was too small to identify significant effects. Future work should use our study design to verify our findings with a larger and more diverse sample.

Acknowledgment

This research is supported by funding from the topic Engineering Secure Systems, topic 46.23.01 Methods for Engineering Secure Systems, of the Helmholtz Association (HGF) and by KASTEL Security Research Labs.

References

[1] M. Zviran and W. J. Haga, "A comparison of password techniques for multilevel authentication mechanisms," *The Computer Journal*, vol. 36, no. 3, pp. 227–237, 1993.

[2] F. A. Alsulaiman and A. El Saddik, "Three-dimensional password for more secure authentication," *IEEE Transactions on Instrumentation and measurement*, vol. 57, no. 9, pp. 1929–1938, 2008.

[3] P. Jadhao and L. Dole, "Survey on authentication password techniques," *International journal of soft computing and Engineering (IJSCE)*, vol. 3, no. 2, pp. 67–68, 2013.

[4] A. Vance, D. Eargle, K. Ouimet, and D. Straub, "Enhancing password security through interactive fear appeals: A web-based field experiment," in *2013 46th Hawaii International Conference on System Sciences*, 2013, pp. 2988–2997.

[5] A. Bhanushali, B. Mange, H. Vyas, H. Bhanushali, and P. Bhogle, "Comparison of graphical password authentication techniques," *International Journal of Computer Applications*, vol. 116, no. 1, 2015.

[6] J. M. Kizza, *Authentication*. Cham: Springer International Publishing, 2024, pp. 215–238. [Online]. Available: {https://doi.org/10.1007/978-3-031-47549-8_10}

[7] N. A. Lal, S. Prasad, and M. Farik, "A review of authentication methods," *Int. J. Sci. Technol. Res*, vol. 5, no. 11, pp. 246–249, 2016.

[8] C. Herley, P. C. Van Oorschot, and A. S. Patrick, "Passwords: If we're so smart, why are we still using them?" in *Financial Cryptography and Data Security: 13th International Conference, FC 2009, Accra Beach, Barbados, February 23-26, 2009. Revised Selected Papers 13*. Springer, 2009, pp. 230–237.

[9] S. Furnell, "Assessing password guidance and enforcement on leading websites," *Computer Fraud & Security*, vol. 2011, p. 10–18, 12 2011.

[10] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, "How does your password measure up? the effect of strength meters on password creation," in *Proceedings of the 21st USENIX Conference on Security Symposium*, ser. Security'12. USA: USENIX Association, 2012, p. 5.

[11] M. Golla, B. Hahn, K. Selhausen, H. Hosseini, and M. Dürmuth, "Bars, badges, and high scores: On the impact of password strength visualizations," 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:196174492

[12] A. Vance, D. Eargle, D. L. Eggett, D. W. Straub, and K. Ouimet, "Do security fear appeals work when they interrupt tasks? a multi-method examination of password strength," *MIS Quarterly*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:231823845

[13] K. Renaud and M. Dupuis, "Cyber security fear appeals: unexpectedly complicated," in *Proceedings of the New Security Paradigms Workshop*, ser. NSPW '19. New York, NY, USA: Association for Computing Machinery, 2020, p. 42–56. [Online]. Available: https://doi.org/10.1145/3368860.3368864

[14] K. Renaud, V. Zimmermann, T. Schürmann, and C. Böhm, "Exploring cybersecurity-related emotions and finding that they are challenging to measure," *Humanities and Social Sciences Communications*, vol. 8, p. 75, 03 2021.

[15] V. Zimmermann, K. Marky, and K. Renaud, "Hybrid password meters for more secure passwords – a comprehensive study of password meters including nudges and password information," *Behaviour & Information Technology*, vol. 42, pp. 1–44, 03 2022.

[16] I. Gulenko, "Improving passwords: influence of emotions on security behaviour," *Information Management & Computer Security*, vol. 22, pp. 167–178, 06 2014.

[17] S. Furnell and R. Esmael, "Evaluating the effect of guidance and feedback upon password compliance," *Computer Fraud & Security*, vol. 2017, no. 1, pp. 5–10, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361372317300052

[18] K. Renaud and V. Zimmermann, "Nudging folks towards stronger password choices: providing certainty is the key," *Behavioural Public Policy*, vol. 3, pp. 1–31, 02 2018.

[19] V. Zimmermann and K. Renaud, "The nudge puzzle: Matching nudge interventions to cybersecurity decisions," vol. 28, no. 1, 01 2021. [Online]. Available: https://doi.org/10.1145/3429888

[20] V. Zimmermann, "From the quest to replace passwords towards supporting secure and usable password creation," Ph.D. dissertation, Technische Universität Darmstadt, Darmstadt, 2021. [Online]. Available: http://tuprints.ulb.tu-darmstadt.de/17425/

[21] T. Fordyce, S. Green, and T. Groß, "Investigation of the effect of fear and stress on password choice," in *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust*, ser. STAST '17. New York, NY, USA: Association for Computing Machinery, 2018, p. 3–15. [Online]. Available: https://doi.org/10.1145/3167996.3168000

[22] M. Dupuis, A. Jennings, and K. Renaud, "Scaring people is not enough: An examination of fear appeals within the context of promoting good password hygiene," in *Proceedings of the 22nd Annual Conference on Information Technology Education*, ser. SIGITE '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 35–40. [Online]. Available: https://doi.org/10.1145/3450329.3476862

[23] K. Marett, A. Vedadi, and A. Durcikova, "A quantitative textual analysis of three types of threat communication and subsequent maladaptive responses," *Computers & Security*, vol. 80, 09 2018.

[24] M. Dupuis, K. Renaud, and A. Jennings, "Fear might motivate secure password choices in the short term, but at what cost?" 01 2022.

[25] R. Ruiter, L. Kessels, G.-J. Peters, and G. Kok, "Sixty years of fear appeal research: Current state of the evidence," *International journal of psychology : Journal international de psychologie*, vol. 49, pp. 63–70, 04 2014.

[26] K. P. L. Coopamootoo, "Empathy as a response to frustration in password choice," in *Financial Cryptography and Data Security*, M. Bernhard, A. Bracciali, L. J. Camp, S. Matsuo, A. Maurushat, P. B. Rønne, and M. Sala, Eds. Cham: Springer International Publishing, 2020, pp. 177–191.

[27] D. Watson, L. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The panas scales," *Journal of Personality and Social Psychology*, vol. 54, pp. 1063–1070, 06 1988.

[28] J. Brooke, "Sus: A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, 11 1995.

[29] T. Franke, C. Attig, and D. Wessel, "A personal resource for technology interaction: development and validation of the affinity for technology interaction (ati) scale," *International Journal of Human–Computer Interaction*, vol. 35, no. 6, pp. 456–467, 2019.

[30] D. L. Wheeler, "zxcvbn: Low-Budget password strength estimation," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 157–173. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/wheeler

[31] L. Khan, K. P. L. Coopamootoo, and M. Ng, "Not annoying the user for better password choice: Effect of incidental anger emotion on password choice," in *HCI for Cybersecurity, Privacy and Trust*, A. Moallem, Ed. Cham: Springer International Publishing, 2020, pp. 143–161.

[32] M. Golla and M. Dürmuth, "On the accuracy of password strength meters," ser. CCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1567–1582. [Online]. Available: https://doi.org/10.1145/3243734.3243769

[33] D. Temoshok, J. Fenton, Y.-Y. Choong, N. Lefkovitz, A. Regenscheid, and J. Richer, "Digital identity guidelines: Authentication and lifecycle management," National Institute of Standards and Technology, U.S. Department of Commerce, Washington, D.C., Tech. Rep., 2022, initial Public Draft, Date Published: December 16, 2022, Comments Due: April 14, 2023 (public comment period is CLOSED), Planning Note (03/17/2023): The public comment period has been extended to April 14, 2023 (from March 24).

[34] J. Kävrestad, M. Lennartsson, M. Birath, and M. Nohlberg, "Constructing secure and memorable passwords," *Information & Computer Security*, vol. ahead-of-print, 06 2020.

[35] B. Ur, F. Alfieri, M. Aung, L. Bauer, N. Christin, J. Colnago, L. F. Cranor, H. Dixon, P. Emami Naeini, H. Habib, N. Johnson, and W. Melicher, "Design and evaluation of a data-driven password meter," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 3775–3786. [Online]. Available: https://doi.org/10.1145/3025453.3026050

[36] R. W. Rogers, "A protection motivation theory of fear appeals and attitude change1," *The Journal of Psychology*, vol. 91, no. 1, pp. 93–114, 1975.

[37] M. J. Blanca, R. Alarcón, R. Arnau, J.and Bono, and R. Bendayan, "Non-normal data: Is ANOVA still a valid option?" *Psicothema*, vol. 29, no. 4, pp. 552–557, 2017. [Online]. Available: https://doi.org/10.7334/psicothema2016.383

[38] L. M. Lix, J. C. Keselman, and H. J. Keselman, "Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test," *Review of Educational Research*, vol. 66, no. 4, pp. 579–619, 1996. [Online]. Available: https://doi.org/10.3102/00346543066004579

[39] E. Schmider, M. Ziegler, E. Danay, L. Beyer, and M. Bühner, "Is It Really Robust? Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption," *Methodology*, vol. 6, no. 4, pp. 147–151, 2010. [Online]. Available: https://doi.org/10.1027/1614-2241/a000016

[40] M. R. Harwell, E. N. Rubinstein, W. S. Hayes, and C. C. Olds, "Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases," *Journal of Educational Statistics*, vol. 17, no. 4, pp. 315–339, 1992. [Online]. Available: https://doi.org/10.3102/10769986017004315

[41] G. V. Glass, P. D. Peckham, and J. R. Sanders, "Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance," *Review of Educational Research*, vol. 42, no. 3, pp. 237–288, 1972. [Online]. Available: https://doi.org/10.3102/00346543042003237

[42] K. D. Kubinger, D. Rasch, and K. Moder, "Zur legende der voraussetzungen des t-tests für unabhängige stichproben," *Psychologische Rundschau*, vol. 60, no. 1, pp. 26–27, 2009. [Online]. Available: https://doi.org/10.1026/0033-3042.60.1.26

[43] R. R. Pagano, *Understanding statistics in the behavioral sciences (9th ed.)*. Australia, Belmont, CA: Thomson Wadsworth, 2010.

[44] D. Rasch and V. Guiard, "The robustness of parametric statistical methods," *Psychology Science*, vol. 46, pp. 175–208, 2004.

[45] R. R. Wilcox, *Introduction to robust estimation and hypothesis testing (3rd ed.)*. Amsterdam, Boston: Academic Press, 2012.

[46] J. Cohen, "Statistical power analysis for the behavioral sciences," 1988.

[47] S. W. Schuetz, P. B. Lowry, D. A. Pienta, and J. B. Thatcher, "The effectiveness of abstract versus concrete fear appeals in information security," *Journal of Management Information Systems*, vol. 37, no. 3, pp. 723–757, 2020. [Online]. Available: https://doi.org/10.1080/07421222.2020.1790187

APPENDIX

A. Treatments

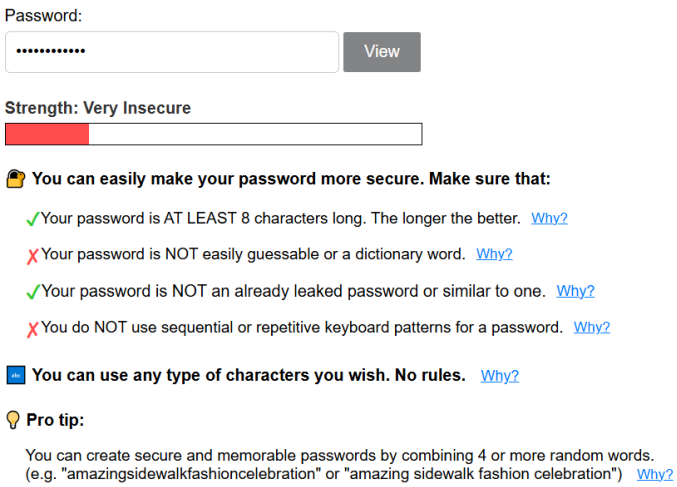


Fig. 7: [T1b] – Colored feedback bar with guidance. A very insecure password has been typed.

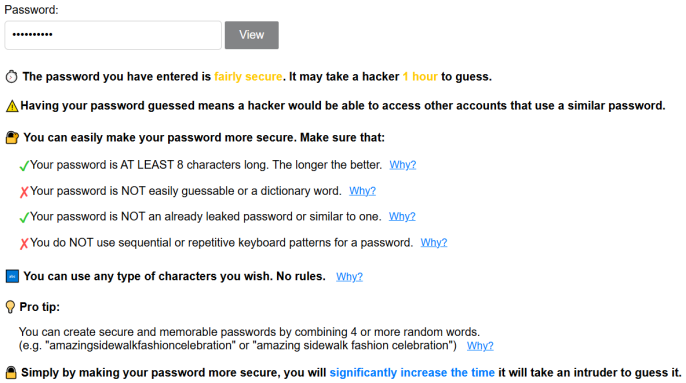


Fig. 8: [T2b] – Interactive fear appeals nudge with guidance. A fairly secure password has been typed.

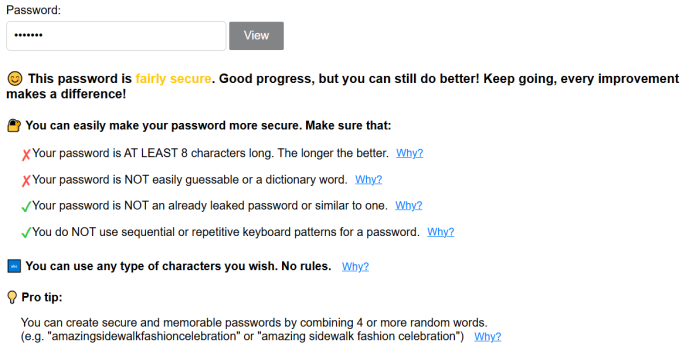
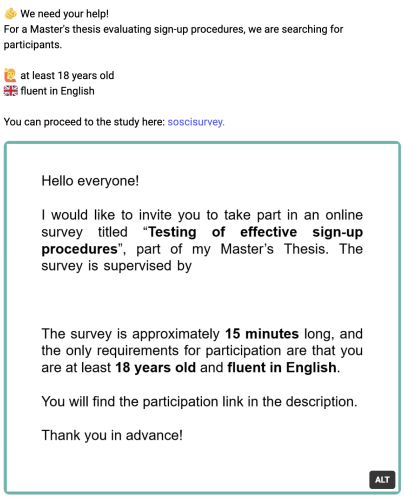


Fig. 9: [T3b] Positive feedback nudge with guidance. A fairly secure password has been typed.

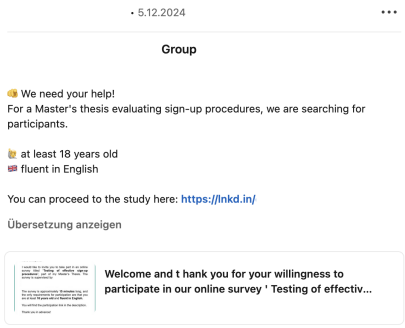
B. Survey

In this section, we provide the screenshots of the posts that we used to recruit participants Figure 10 (B1), as well as the survey questionnaire (B2).

1) Recruitment:



(a) Recruitment Post on Mastodon.



(b) Recruitment Post on LinkedIn.

Fig. 10: Recruitment Posts on Social Media.

2) Survey Questionnaire:

Welcome and thank you for your willingness to participate in our online survey "Testing of effective sign-up procedures"

Who are we?

My name is [anonymized, information about the study authors] and I am conducting this online survey as a part of my Master thesis, which is [anonymized, information about the institution].

What is the survey about?

The aim of our survey is to test web sign-up procedures. For this purpose, we will show you a section of a sign-up page and ask you to create a dummy password for a fictional account. Afterwards, we will ask you to fill out a questionnaire.

Are there any requirements for participating? How long is it? Can I quit anytime?

- To participate, you must be at least 18 years old and fluent in English.
- The survey will take approximately 15 minutes.
- No personal data, including passwords in plain text, will be collected.
- You can stop your participation anytime without giving any reason by closing the browser window.

Is there anything in particular I should be aware of?

There are some important things to pay attention to:

- The password you enter is not stored in plain text.
- However, please choose a dummy password — this means a password that you are not already using for one of your real accounts!
- If you have a particular method of creating a password, you could freely use that one to create the dummy password, as long as it doesn't closely resemble a password that you actively use for your real-world accounts.
- Shortly after starting the survey, you will receive a detailed task description.
- After the task is described, we will check whether you have understood the task description and study instructions.

How to interact with the survey?

- Once you are ready to move on to a next page, click on the "Next" button.

- There is no “Back” button as an option to return to the previous page.
- The survey is designed for desktop devices. Participation via a mobile device is still possible, yet some items might appear scattered. If you encounter any problems, or cannot continue with the survey, please contact the principal investigators.
- The survey is completed once you reach the final page.

If you proceed with this survey, we will inform you about the processing of your data on the next page and ask for your consent to participate. If you have any questions or comments about the study or the processing of your responses, you can contact the principal investigators of the study. Contact information is provided below.

Page break

Information on Data Processing and Purpose of the Study

The following information gives an overview of the processing of your survey data as part of this user study and your rights to withdraw any given consent.

- The study is conducted by researchers at [anonymized, information about the institution] as part of a Master’s thesis [anonymized, information about the institution].
- Your data are processed solely for the purpose of this thesis. The purpose of the thesis is to identify effective sign-up procedures.
- No personal data will be collected.
- None of the dummy passwords will be stored in plain text, meaning that the researchers would not be able to see the actual dummy password.
- All responses are anonymous and cannot be traced back to an individual.
- The results of the study will only be published in anonymized and/or aggregated form.

Declaration of Consent

I hereby confirm that I am over 18 years of age and have read and taken note of the above information. I have been informed about the purpose of the study and the processing of my responses. I consent to their processing as described above with my confirmation. The granting of consent is voluntary. There are no disadvantages if consent is refused or withdrawn. Consent can be withdrawn by terminating the survey at any time during the survey, with effect for the future. Effect for the future means that the withdrawal of consent does not affect the lawfulness of processing the data based on consent before its withdrawal. Since we do not store any personal data or send personalized links, we cannot trace your answers that you have made up to the point of withdrawal back to you and therefore cannot exclude them. Depending on when you terminate the survey, we might include some of your answers that you have provided up to this point in our analysis. Following the recommendation of the [anonymized, recommendation for safeguarding good scientific practice], all completed survey data and research results collected as part of the research project will be stored for up to ten years.

Please choose:
- I have read the informed consent and agree to participate in this study.
- I do not want to participate in this study.

Page break

Thank you for choosing to participate in this survey!

What is next?
- The task description and the instructions for completing the task are given on the next page.
- Before completing the task, you will be asked questions to ensure you have understood the task instructions correctly.

Please read the description of the tasks carefully!

Page break

Task Description

Let's begin!

We are in the process of designing a website that helps people organize their academic/work activities. Currently, we are working on the sign-up functionality for this website.

Imagine the following scenario:

This website is ready to be used. You are a student, currently attending many lectures, and you have decided to register and try it out. You have already clicked on the "Sign-Up" button and started creating your account. Likewise, you have already typed in your username and all other necessary information. There is just one last step before finishing your registration — **choosing a password for your account**.

Task description:

You will see a page with text "Please choose a password for your account:"
Your task will be to create a dummy password for your account using the provided functionality.

!!! Important notes !!!

Please DO:

- Come up with a completely new dummy password.
- Choose a password that you will be able to remember.
- Take as much time as you need.
- Click on the "Confirm Password and Continue" button once you have chosen the password.

Please DON'T:

- Enter a password that you already use to log into a real service or a similar password.

Page break

Attention Questions

Please answer the following questions:
(The attention questions aim to ensure you understand the task and pay attention to the survey.)

[AT01] What is your current task? [single choice]

- Looking at lecture notes.
- Choosing a password for a fictional account.
- Choosing a name for a website.

[AT02] What are the things you should NOT do while you complete the task? [multiple choice]

- Select a password that you use for one of your real accounts.
- Take too much time to complete the task.
- Select a password, similar to one of your real passwords.
- Choose an easy-to-remember password.

Page break

Thank you for your answers!

What's next?

- First, we will ask you to fill out a questionnaire.
- Then you will proceed to the main task. Remember: Your task will be to create a dummy password.
- After completing the task, we will ask you some more questions.
- Finally, you will be given additional insights about this survey and some useful information.

Page break

[Q1 – PANAS_1] Please indicate how do you feel at this moment: [very slightly or not at all, a little, moderately, quite a bit, extremely]

- interested
- distressed
- excited
- upset
- strong
- guilty
- scared
- hostile
- enthusiastic
- proud
- irritable
- alert
- ashamed
- inspired
- nervous
- determined
- attentive
- jittery
- active
- afraid

Page break

Please create a password for your account: [a random password strength calculator design is displayed]

Page break

Success! You have completed the task!

Now, please fill out the following questionnaires:

[Q2 – PANAS_2] Please indicate how do you feel at this moment: [very slightly or not at all, a little, moderately, quite a bit, extremely]

- interested
- distressed
- excited
- upset
- strong
- guilty
- scared
- hostile
- enthusiastic
- proud

- irritable
- alert
- ashamed
- inspired
- nervous
- determined
- attentive
- jittery
- active
- afraid

Page break

[Q3 – SUS] Please answer the following questions: [strongly disagree, disagree, neutral, agree, strongly agree]

- I think that I would like to use this password creation procedure frequently.
- I found the password creation procedure unnecessarily complex.
- I thought the password creation procedure was easy to use.
- I think that I would need the support of a technical person to be able to use this password creation procedure.
- I found the various functions in this password creation procedure were well integrated.
- I thought there was too much inconsistency in this password creation procedure.
- I would imagine that most people would learn to use this password creation procedure very quickly.
- I found the password creation procedure very cumbersome to use.
- I felt very confident using the password creation procedure.
- I needed to learn a lot of things before I could get going with this password creation procedure.

[Q4 – INTERACTION] Please tell us briefly about your interaction with the element(s) below the field in which you entered your dummy password: [open text]
(e.g. how much attention did you pay to them while you were choosing a password)

[Q5 – PW_USE] Do you generally use a password manager in real life?

- Yes, I use a password manager to create and to store my passwords.
- Yes, I use a password manager to create passwords, but not to store them.
- Yes, I use a password manager to store my passwords, but not to create them.
- No, I do not use a password manager.

Page break

In the next questions, we will ask you about your interaction with technical systems in general. The term "technical systems" refers to apps and other software applications, as well as entire digital devices (e.g., mobile phone, computer, TV, car navigation).

[Q6 – ATI] Please indicate the degree to which you agree/disagree with the following statements. [completely disagree, largely disagree, slightly disagree, slightly agree, largely agree, completely agree]

- I like to occupy myself in greater detail with technical systems.
- I like testing the functions of new technical systems.
- I predominantly deal with technical systems because I have to.
- When I have a new technical system in front of me, I try it out intensively.
- I enjoy spending time becoming acquainted with a new technical system.
- It is enough for me that a technical system works; I don't care how or why.
- I try to understand how a technical system exactly works.
- It is enough for me to know the basic functions of a technical system.
- I try to make full use of the capabilities of a technical system.

Page break

Debriefing Statement:

Thank you for participating in our survey "Testing of effective sign-up procedures", as part of a Master thesis project titled "Fear Appeals in Cybersecurity: How effective are they really?" made in collaboration with [anonymized, information about the institution].

The actual subject of our research is focused on investigating password creation procedures that would help people create and passwords that are resilient to attacks and easy-to-remember. Our research objective is to understand the incentives encouraging people to create stronger passwords. More specifically, the goal of this research is to investigate the effectiveness of fear appeals and other strategies such as password strength meters, positive feedback and information provision on password creation procedures. You were assigned one of 7 treatments, containing (or not) specific strategies that could potentially have an impact on the strength of selected passwords.

We are not developing a website. It was necessary for us researchers to provide such a scenario to illustrate a typical password creation context and to ensure that your actions and answers to questions accurately reflected your real-life password choices. The dummy password you entered is not stored anywhere and is not seen by anyone, except you. Rather than the password itself, for our data collection, we store two parameters indicating the strength and one parameter indicating the length of the chosen password.

Your participation in the study is important in helping researchers identify the best ways to encourage users to create stronger passwords. The final results of this study will be included in a master thesis document and a final publication. Your responses will not be available individually, and your participation will remain confidential. Since we did not collect any personal information, we would not be able to remove your responses from the dataset after the survey is completed.

If you, nevertheless, decide you do not want us to use your survey data, please let us know in the field below and we will delete your data from our dataset.

Lastly, we want to share some important password practices with you. Click the "Next" button to see them.

[DEL] Please write a comment here ONLY IF you want us to delete your data. [open text]
(leave the input field blank otherwise)

Page break

Important information regarding secure password practices:

Password Length: According to organizations such as the National Institute of Technology Guidelines (NIST) and The Federal Office for Information Security (BSI), a good password must be at least 8 characters long. The longer the password, the better.

Repetitive Patterns: Passwords which consist solely of sequential or repetitive characters, or keyboard patterns can be insecure even if they are long. For example, a password of length 100 that is the repetition of the letters "ha" ("hahahaha...") can still be insecure and easy to guess. For this reason, it is not a good practice to use only sequential and repetitive patterns as a password.

Dictionary Words: Your password should NOT be a dictionary word, or a word similar to a dictionary word. A dictionary word is a word that you will find in dictionaries that list the words of a language. Dictionary words and names (e.g. apple, Eve) can be very easily guessed and are not suitable password choices. Refrain from using such words as passwords.

Leaked Passwords: Similarly to dictionary words, you should NOT use passwords that have already been leaked. You can test if a password has been leaked at: https://haveibeenpwned.com/Passwords.

Personal Information and Service Names: Do not use a password that reflects any personal information that is accessible to others (e.g. your birthday, your pet name). Do not use the service name or variations of it for a password (e.g. facebook04).

Composition Rules: Composition rules have been commonly used to increase the difficulty of guessing user-chosen passwords (e.g. rules that force you to include at least one upper letter, number, special character, etc.). However, users respond in very predictable ways to the requirements imposed by composition rules. For this reason, forcing composition rules on users is not advisable.

Character types: Rather than being forced to comply with certain rules, you should be encouraged to include any desired type or combination of characters such as lowercase letters, uppercase letters, numbers, special characters and spaces, so that you can create a memorable password.

Memorability: Your password should be both secure and memorable. According to scientific research*, utilizing passphrases is an effective strategy for creating both secure and memorable passwords. You can create secure and memorable passwords by using passphrases with at least 4 random words.

* Kävrestad, J., Lennartsson, M., Birath, M. and Nohlberg, M. (2020), "Constructing secure and memorable passwords", Information and Computer Security, Vol. 28 No. 5, pp. 701-717.

Additional tips:

- Do not reveal your password to others.
- Do not reuse old passwords.
- Do not store your passwords in places accessible to others.
- Make passwords that are hard to guess but easy to remember.
- Use password managers to manage all your different passwords.

For additional information and materials on secure passwords, you can visit The Federal Office for Information Security (BSI)'s website at: [Link]

Page break

Thank you for your contribution! The survey is now complete!

If you wish to receive information regarding the outcome of our research in the future or if you have any questions regarding the survey, contact us via the email addresses below.

Your answers were transmitted. You may close this tab.

END

C. Statistical Analysis

In this section, we provide additional results on the statistical analysis. Table VI shows the results of the descriptive analysis with all variables.

				95% Confidence Interval	
	Variable	**Mean**	**Std. Deviation**	**Lower Bound**	**Upper Bound**
control group (T0)	strength	9.38	4.65	6.42	12.33
	length	11.50	3.73	9.13	13.87
	SUS	65.63	18.83	53.66	77.59
	positive emotions diff	-0.03	0.27	-0.19	0.15
	negative emotions diff	**-0.19**	0.32	-0.40	0.01
colored feedback no guidance [T1a]	strength	11.36	6.09	7.68	15.04
	length	13.46	5.85	9.92	17.00
	SUS	67.12	18.84	55.73	78.50
	positive emotions diff	-0.12	0.20	-0.23	0.00
	negative emotions diff	-0.16	0.36	-0.38	0.05
colored feedback guidance [T1b]	**strength**	**12.63**	3.67	10.49	14.76
	length	15.57	4.75	12.83	18.31
	SUS	74.29	11.66	67.55	81.02
	positive emotions diff	-0.24	0.38	-0.46	-0.02
	negative emotions diff	**-0.19**	0.30	-0.37	-0.02
fear appeals no guidance [T2a]	strength	12.58	4.34	9.82	15.33
	length	14.67	4.64	11.72	17.61
	SUS	63.75	11.60	56.38	71.12
	positive emotions diff	-0.08	0.21	-0.21	0.06
	negative emotions diff	-0.03	0.32	-0.23	0.18
fear appeals guidance [T2b]	strength	12.08	2.23	10.74	13.43
	length	14.23	3.54	12.09	16.37
	SUS	**80.58**	14.04	72.09	89.06
	positive emotions diff	0.05	0.24	0.20	-0.20
	negative emotions diff	0.02	0.27	-0.15	0.18
positive feedback no guidance [T3a]	strength	12.45	4.63	9.51	15.40
	length	14.92	4.40	12.12	17.71
	SUS	76.67	13.87	67.85	85.48
	positive emotions diff	-0.13	0.21	0.01	-0.50
	negative emotions diff	-0.14	0.29	-0.32	0.04
positive feedback guidance [T3b]	strength	12.27	4.53	9.53	15.00
	length	**15.77**	6.02	12.13	19.40
	SUS	79.81	14.77	70.88	88.73
	positive emotions diff	**0.07**	0.38	0.30	-0.30
	negative emotions diff	0.06	0.52	-0.25	0.38

TABLE VI: Descriptives with all variables. Those variables that scored highest for the different categories are highlighted in bold.