"I don't see why I would ever want to use it": Analyzing the Usability of Popular Smartphone Password Managers

Sunyoung Seiler-Hwang*
University of Mannheim
Mannheim, Germany
sseilerh@mail.uni-mannheim.de

Patricia Arias-Cabarcos*
University of Mannheim
Mannheim, Germany
pariasca@mail.uni-mannheim.de

Andrés Marín
University Carlos III of Madrid
Leganés, Madrid, Spain
amarin@it.uc3m.es

Florina Almenares
University Carlos III of Madrid
Leganés, Madrid, Spain
florina@it.uc3m.es

Daniel Díaz-Sánchez
University Carlos III of Madrid
Leganés, Madrid, Spain
dds@it.uc3m.es

Christian Becker
University of Mannheim
Mannheim, Germany
christian.becker@uni-mannheim.de

ABSTRACT

Passwords are an often unavoidable authentication mechanism, despite the availability of additional alternative means. In the case of smartphones, usability problems are aggravated because interaction happens through small screens and multilayer keyboards. While password managers (PMs) can improve this situation and contribute to hardening security, their adoption is far from widespread. To understand the underlying reasons, we conducted the first empirical usability study of mobile PMs, covering both quantitative and qualitative evaluations. Our findings show that popular PMs are barely acceptable according to the standard System Usability Scale, and that there are three key areas for improvement: integration with external applications, security, and user guidance and interaction. We build on the collected evidence to suggest recommendations that can fill this gap.

KEYWORDS

usable security, password managers, authentication, user study

ACM Reference Format:
Sunyoung Seiler-Hwang, Patricia Arias-Cabarcos, Andrés Marín, Florina Almenares, Daniel Díaz-Sánchez, and Christian Becker. 2019. "I don't see why I would ever want to use it": Analyzing the Usability of Popular Smartphone Password Managers. In *2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19), November 11–15, 2019, London, United Kingdom.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3319535.3354192

1 INTRODUCTION

Passwords, despite their well-known security and usability issues [2, 18, 28, 57], are the most common form of authentication on the web today and in the foreseeable future [12, 25]. Furthermore, as we are increasingly using smartphones to perform activities that were previously done in laptop or desktop computers [1] (e.g., purchases, banking), we also need to frequently deal with passwords in this mobile environment, where usability problems are aggravated. Studies show that it is inconvenient and frustrating to use textual passwords on mobile devices because keyboards are often too small for many fingers, have different layouts in different devices, and need to be shifted to enter special characters [41, 42]. This lack of user-friendliness leads people to resort to weaker passwords when using mobile devices [29, 42], degrading overall security.

Password Managers (PM's) are a safe and convenient tool to improve password usability. They work as digital "wallets" that store all credentials, assisting users during the login process and helping them with password creation. This way, the only thing to remember is a master password that protects the whole wallet. PMs are recommended by security experts [27] and proved to have a positive impact in the strength of the generated passwords [40]. However, in spite of their obvious advantages, the adoption rate of PMs is only modest. Two recent studies [3, 52] report an adoption rate of 17,6% and 16,7% in 2016 and 2018, respectively. But if we narrow down the scope to mobile PM applications, the usage is even lower (6.8% of the respondents in [3]). Usability has been pointed out as a potential cause for the lack of adoption [3], but no work so far focused on analyzing PM usability in smartphones, an environment for which best practices and guidelines for interaction are very different [7, 21]. Therefore, with the goal to understand usability problems and elicit recommendations to foster adoption, we conducted an empirical study of four popular smartphone PMs, each evaluated by 20 users. The contribution of our work are as follows:

- (1) **First empirical study to evaluate the usability of mobile password managers**.
- (2) **Quantitative and qualitative evaluation of usability features.** We designed our survey based on the PACMAD [23] model, a tool specifically oriented to evaluate the usability of mobile applications. Besides qualitative questions, PMs are also quantitatively evaluated through the standard System Usability Scale (SUS) [15], which allows for objective comparison with other authentication alternatives.

*Co-first authors.

[1] http://gs.statcounter.com/press/mobile%2Dand%2Dtablet%2Dinternet%2Dusage%2Dexceeds%2Ddesktop%2Dfor%2Dfirst%2Dtime-worldwide

CCS '19, November 11–15, 2019, London, United Kingdom
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6747-9/19/11…$15.00
https://doi.org/10.1145/3319535.3354192

The reported quantitative scores are average and considered barely acceptable. These data, complemented with user responses, shed light to elicit recommendations.

- (3) **Recommendations for improving usability.** We identify three key areas for improvement: integration with external applications, security, and user guidance and interaction. Based on our observations and participant's feedback, we provide concrete suggestions to enhance usability on these areas.

Apart from the recommendations, we contribute to the literature with other insightful findings. Previous works identified unawareness as a strong reason leading to rejection of password managers [3, 43]. In our study, we observed that despite the majority of participants claimed to know what a PM is, very few were actually using one. However, after trying a PM themselves, around half of the users manifested their intention to continue with the application. This may suggest that there is a need to bridge the gap between awareness and trial by raising the interest of users on PMs. One such strategy is increasing the usability to higher levels. According to SUS-based analyses [14], when usability scores are above the 80th percentile, users act as *"net promoters"* and interest is spread by word-of-mouth recommendations, encouraging wider adoption.

The remainder of the paper is organized as follows. Section 2 describes our methodology, including study and tasks design, as well as limitations. We then report and analyze the results of the user study in Section 3, which serve as the basis to elicit recommendations for PMs improvement in Section 4. Next, Section 5 contextualizes our contribution within the related work on password managers. And finally, Section 6 closes the paper with the main conclusions.

2 METHODOLOGY

Since no work so far has put the focus on understanding the usability of smartphone PMs, we address this gap here through a user study dedicated to evaluate this specific subset of managers, with the aim to provide insights for further improvement. We chose a sample of four managers that are the most popular for the two mobile operating systems with higher usage share, Android and iOS. Based on their recommendations and number of downloads on Android market and Apple store (see Table 1), the selected PMs are 1password [1], Dashlane [17], Keeper [32], and LastPass [36]. In the following, we describe the methodology for the usability study.

2.1 Study Design

Evaluation framework. The usability of the four mobile PMs was evaluated qualitatively and quantitatively.

For the overall quantitative analysis, we base on the standard System Usability Scale (SUS), which has been widely used[2] and proved valid and reliable to measure usability [44] [34]. SUS yields an usability score between 0-100 after asking participants to determine their agreement to a set of 10 statements. According to Bangor *et al.* [9], products with SUS scores above 70 are at least acceptable, better products are located in the high 70s to upper 80s, and anything in the 90s is exceptional, based on the results of 206 studies collected over 10 years. An A-F grading scale for SUS scores can be also used, as it is shown in Figure 1. We will report adjective-based ratings along with SUS scores to provide readers with a better intuition of each PM's usability. Furthermore, the great advantage of SUS is that it allows for direct comparison with other studies using the same metric.

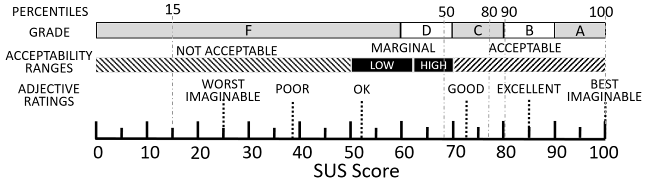


Figure 1: System Usability Scale: raw metrics and mappings to different interpretation scales [8].

Apart from SUS, we use the PACMAD (People At the Center of Mobile Application Development) [23] usability model. Developed by Harrison *et al.*, PACMAD combines significant attributes from different usability models to create a more comprehensive evaluation framework specific for mobile devices. The model evaluates mobile applications against seven attributes: *Effectiveness*, *Efficiency*, *Satisfaction*, *Learnability*, *Memorability*, *Errors*, and *Cognitive Load*. Table 2 shows the definition of these attributes, how we evaluate them as suggested by Harrison *et al.*, and a mapping to the concrete questionnaire items used in our study, when applicable. *Efficiency* is calculated through automated time measurements. Furthermore, we partially base on SUS questions to measure *Learnability*, and on the NASA Task Load Index (TLX)[24][3] for the *Cognitive Load*. For the rest of attributes that can be measured through questionnaires, we have defined a new set of fixed choice and open-ended questions. All these questions and tasks were designed following the guidelines for usability testing from Rubin *et al* [46].

Structure and User Recruitment. The study is organized in three parts. First, participants are presented with a *Pre-study Questionnaire* to gather demographic data and information about their previous knowledge on PMs. Next, we provide them with a brief definition of a what a password manager is and instruct them to install and use one of the four PM apps selected for the study, following a set of pre-defined tasks (see Section 2.2). After each task, participants must respond to a brief *After-task Questionnaire* designed to measure the usability attributes in Table 2, and to understand how they use PMs. Finally, we ask them to fill a *Post-Study Questionnaire* to obtain the quantitative SUS score they assign to the PM, and to get further insights on the overall usability of the application and users' adoption attitudes. Before publishing this 3-part survey, we ran a pilot with 4 people, asking them to provide feedback. This experience helped us in understanding the expected completion time, and the obtained feedback was useful to make some minor modifications to the survey flow. The final version can be found in Appendix B .

[2] Since SUS was developed by Brooke [15] in 1996, more than 2300 individual surveys were conducted using SUS in over 200 studies by 2008 [9]

[3] The NASA TLX is a multi-dimensional scale designed to obtain workload estimates from one or more operators while they are performing a task or immediately afterwards.

Table 1: Summary of features for the selected mobile PMs as of Sep 20, 2018.

	Dashlane	LastPass	Keeper	1Password
Founding Date	Jul 6, 2009	Apr 30, 2008	2011	2005
Rating	4.6/5 (Google Play) 4.7/5 (App Store)	4.6/5 (Google Play) 4.5/5 (App Store)	4.2/5 (Google Play) 4.9/5 (App Store)	4.2/5 (Google Play) 4.5/5 (App Store)
Downloads	75,199	115,522	74,738	25,354
Cost	Free / Premium (3.3 EUR/month)	Free trial / Premium (2 USD/month)	Free (limited) / Premium (2.5 USD/month)	Free / Premium (2.99 USD/month)
Mobile OS	Android, iOS	Android, iOS, Windows Phone	Android, iOS, Windows Phone	Android, iOS
Available Versions	Smartphone, PC, Tablet, Smart watch			

Table 2: Summary of PACMAD usability model attributes, evaluation techniques, and mapping to questionnaire items.

Attribute	Definition	Evaluation	Questionnaire Item
Effectiveness	Ability of a user to complete a task in a specified context	Whether or not participants can complete a task	Q9, Q16, Q21, Q27, Q37, Q51
Efficiency	Ability of a user to complete a task with speed and accuracy	Time to complete a given task	Automated Measurement
Satisfaction	Perceived level of comfort and pleasantness afforded to the user through the use of the application	Questions about likes, dislikes, and intention of continued use, examined through thematic analysis	Q11, Q12, Q18, Q19, Q23, Q24, Q29, Q30, Q39, Q40, Q44, Q45, Q53, Q54, PQ1
Learnability	Ease with which a user can gain proficiency with an application	SUS Learnability questions	SUS04, SUS10
Memorability	Ability of a user to retain how to use an application effectively	Asking participants to perform a series of tasks and after a period of inactivity, asking them again to perform similar tasks	N/A[1]
Errors	How well a user can complete the desired tasks without errors	Observing the nature of errors and the frequency with which they occur	Q13, Q20, Q25, Q31, Q41, Q46, Q55
Cognitive Load	Amount of cognitive processing required by the user to use the application	NASA Task Load Index	Q10, Q17, Q22, Q28, Q38, Q44, Q52

[1] We did not measure *Memorability*, as it would have required an additional follow-up study.

With regard to the number of users required for testing, though it was long believed that 5 participants were enough [58] to identify usability problems, more recent studies suggest numbers bigger than 10 to achieve better results [26]. We follow this guideline in our survey, recruiting 20 participants per PM, which is the average sample size used in security usability studies [47]. Accordingly, upon acceptance to take part in the survey, participants were shown a list of PMs from which to choose, with the restriction not to select the same manager they were already using. When a PM reached 20 users, it was removed from the list. For recruitment, we used the crowdsourcing platform Amazon Mechanical (MTurk)[4], which provides a commercial marketplace for so-called *"Human Intelligence Tasks"* or HITs. Our survey, recruitment process, payment, and communication with participants were designed in accordance to *Amazon's Acceptable Use Policy*[5] and following best practices for responsible research with crowds [50]. Before beginning the session, participants signed a consent form that explained the purpose of the study and how we would treat their data, and we informed them that participation was voluntary and the questionnaire could be abandoned at any time. Furthermore, though the study was announced in MTurk, the questionnaires were hosted in LimeSurvey[6], whose servers are located in Germany and comply with the European privacy regulations.

2.2 Tasks Design

Our questionnaire is structured around a set of tasks based on the pioneering study by Chiasson *et al.* that analyzed two emerging PMs in 2006 [16], but adapted to the smartphone scenario. The seven tasks, which cover the main functionalities of a PM, are described below:

Task 1: Initialization. Install and register to the PM application.
Rationale: This is the first step needed by the participants to set up the mobile PM on their devices.
Task 2: Account migration. Store an existing web account with its associated password on the manager[7].
Rationale: This task simulates the most common user action after downloading the app.
Task 3: Login. Start a session on a website for which the PM has an stored password.

[4] https://www.mturk.com/

[5] https://www.mturk.com/acceptable-use-policy

[6] https://www.limesurvey.org/

[7] Users were provided with pre-defined accounts to avoid them using personal information

Rationale: This task tests the core functionality of a PM.

Task 4: Account creation and usage. Create a new web account using the smartphone. Save the account and the new password in the PM application and log-in to the website[8].

Rationale: As people use mobile devices more frequently, they also use these devices to create accounts.

Task 5: Interaction with native apps. Download the native application of the website chosen in Task 4, then log-in to the app using the PM.

Rationale: Tasks 4 and 5 attempt to simulate common consumer behaviors: according to statistics, consumers usually become aware of, and engage in, services via mobile websites first since websites are easier to reach than native apps, which need to be downloaded. Then as people get more involved with the services, they prefer to download and use the native app version of the services due to better user experience, speed, extra features, and special offers like additional discounts [30].

Task 6: Password change. Participants were asked to change the password of the account created in Task 4 using the PM app and its password generator tool, and then perform a login with the changed password.

Rationale: After continued use of the PM, users will probably need to change passwords for different reasons, such as compliance with expiration policies or because of publicly reported breaches.

Task 7: Security settings. Participants were asked to change the security configurations of the PM to reflect their preferences.

Rationale: This task was included to gain insights on how user configurations and usability decisions affect security.

2.3 Limitations

In terms of the study sample, although the user population on MTurk is relatively diverse, workers are mainly considered "WEIRD", i.e., coming from Western, Educated, Industrialized, Rich, and Democratic countries [33]. We tried to eliminate this bias by opening the survey to any country, but our demographics still skew towards the United States and India, as it is common in many MTurk-based studies in the field of usable security [40, 45, 52]. We also acknowledge that our sample exhibits a higher level of education on average, so it is not representative of the general population. Studies with users having more diverse backgrounds are required to get a more complete picture of PM usability issues.

Regarding our user study, we advertised it without concealing its purpose in order to obtain fully informed consent. This comes with a risk of biasing the participants, as knowing the goal of the survey could make them prone to behave and answer as they think the experimenter wants, rather than behaving as they would do naturally. It is also possible that the participants that decided to take part in the study did it because the topic is familiar or motivating for them, excluding users that are not interested on PMs, whose opinions could have added further value to the collected data. Another inherent limitation is that participants have tested the mobile PMs in an artificial setting: they simulated real-life PM usage by following the instructions; hence there might be discrepancies between the test results and reality. Furthermore, our study is based on remote testing, and self-reporting techniques, whose implications must be considered. Using self-reporting could have captured the opinions of the participants in an inaccurate or limited way, especially when English was not the participants' first language. Besides, since it was not possible to observe participants' non-verbal cues, there was less information to interpret their responses.

Finally, we analyzed a reduced set of mobile PMs chosen based on popularity, which provides a limited view of the mobile PM market. Since we did not force specific versions of the operating systems or PMs, usability perceptions might vary due to differences in the implemented features across platforms. Future large-scale studies, covering a high number of applications and versions, are desirable for a more comprehensive view. Furthermore, field studies that follow participants while using a PM during a long period of time, can reveal usability issues that only arise after continued every-day use.

While our data may be impacted by these limitations, we believe that our study fills a gap in the current literature, by extending the understanding about smartphone PMs usability.

3 USABILITY STUDY

The study was posted on MTurk with a compensation of 7$[9] on October 22, 2018, without imposing any restriction on who could undertake the HIT apart from being older than 18 years of age and having a smartphone with Android or iOS. Individuals were only allowed to participate once. Three days after the posting date, we had collected 110 responses, from which 80 where finally used after filtering 30 surveys that were answered carelessly or too quickly with respect to the expected completion time.

On average, participants took 58 minutes to finish the survey, and the median session duration was 44 minutes.

3.1 Participant Demographics

Table 3 presents an overview of the demographics of our participants both globally and per password manager. It can be seen that per PM samples are similar to each other and show the same trends as the overall sample.

Overall, our participant number is not balanced with respect to gender, being composed by a 72.5%(58) of males, a 26.25%(21) of females, and a 1.25%(1) of users that preferred not to answer. Also, our sample covers an age range from 18 to over 45 years old, where the number of participants skews to younger users (75% are younger than 35) as can be commonly observed in usable security research, including other previous studies on password managers [16, 31]. In terms of education, 8.75%(7) had a high school degree or equivalent, 5%(4) received technical or vocational training, 21.25%(17) had some college education, 56.25%(45) had a college degree, and 8.75% (7) a graduate degree (master or professional). Furthermore, most of our participants (95%) come from USA or India, with only a 5%(4) of users based in other countries (Italy, Mexico, Sweden, and

[8] A list of 3 websites was provided to limit the choices of the participants: https://doodle.com, https://memrise.com, https://.goodreads.com

[9] Following the guidelines for academic requesters formulated by MTurk workers [59], we based our payment on the US federal minimum wage (7.25$ per hour), adjusted to the estimated time to complete the survey (slightly under 60 minutes).

Table 3: Demographics of participants in the usability study.

	1password	Dashlane	Keeper	LastPass	Total
Number of participants	20	20	20	20	80
Gender					
Male	60%	80%	65%	85%	72.5%
Female	40%	20%	30%	15%	26.25%
No Answer	0%	0%	5%	0%	1.25%
Age					
18 - 24	10%	25%	15%	15%	16.25%
25 - 34	75%	45%	45%	70%	58.75%
35 - 44	15%	15%	30%	0%	15%
>=45	0%	15%	10%	15%	10%
Education					
High School	10%	10%	10%	5%	8.75%
Some college, no degree	25%	25%	10%	25%	21.25%
Tech/Voc. training	0%	10%	5%	5%	5%
Bachelor Degree	60%	50%	60%	55%	56.25%
Master Degree	5%	0%	15%	5%	6.25%
Professional Degree	0%	5%	0%	5%	2.5%
Country					
USA	55%	70%	85%	75%	71.25%
India	40%	15%	15%	25%	23.75%
Other	5%	15%	0%	0%	5%
OS					
Android	75%	65%	70%	75%	71.25%
iOs	25%	35%	30%	25%	28.75%
PM Awareness					
Know	80%	80%	85%	90%	83.75%
Use	20%	20%	10%	20%	17.5%

Germany). With regard to the operating systems, our participants predominantly (71.25%) used Android over iOS.

Previous Knowledge. As we were also interested on the adoption rate of password managers, we added several questions related to awareness of this technology. We first asked participants if they knew what a PM is and if they used one. Then, for those self-defined as users, we further inquired which specific PM and version (e.g., smartphone, desktop) they used. Interestingly, though an 83.75%(67) of the respondents claimed to know about PMs, just a 17.5%(14) reported actually using one. In their responses, participants mentioned browser-based PMs (28,6%), as well as 3rd-party applications (71,4%), including several other PMs different from those analyzed in this study (i.e., Roboform, KeePass, Sticky Password, and Kaspersky). However, when going deeper into the results of the study, we noticed that self-reported usage was not completely accurate, since an extra 6,25%(6) of the participants argued that they would not continue using the tested PM because their passwords were already stored by their browser or iPhone. This suggests that different users have different understandings about which type of tools can be considered a PM.

3.2 PACMAD Attributes

3.2.1 Effectiveness. All four mobile PMs had above 90% success rate on average for all tasks combined, which can be interpreted as a good effectiveness. Task 5, which asked the participants to download a native app of a service and log-in using a previously stored password, had the lowest success rate. Participants expected to be able to directly use the PM auto-fill feature on the native application, but instead it was required to previously grant certain system-level permissions outside the application or to be aware that the auto-fill functionality does not work for every application. This knowledge gap is a plausible explanation for the low success rate of this particular task. Another lower value that stands out is the success rate for Task 3 using 1Password, which suggests that the login flow for websites with this PM might have usability issues.

Results are summarized in Table 4, and we analyze the underlying reasons for failure in Section 3.2.5.

Table 4: Effectiveness of PMs measured as the success rate in completing each of the seven tested tasks (T1-T7).

	T1	T2	T3	T4	T5	T6	T7
Dashlane	100%	100%	90%	90%	85%	90%	100%
Keeper	100%	100%	95%	90%	70%	85%	100
Lastpass	100%	100%	95%	100%	80%	90%	95%
1Password	100%	100%	75%	95%	85%	90%	100%

3.2.2 Efficiency. Efficiency results are summarized in Figure 2. The average time per task was consistently similar for the 4 analyzed PMs, and the differences proved to be not statistically significant after an ANOVA test (see Appendix A). Participants took longer to finish Task 1, which included download, installation, and registration; and Tasks 4 (account creation and use) and 6 (password change), which required actions on both the PM and a second application, having to switch between them. Though we do not have a benchmark against which to compare the efficiency, analyzing the additional information obtained from the open-ended questions, we observed a similar number of negative comments (e.g., "*took*

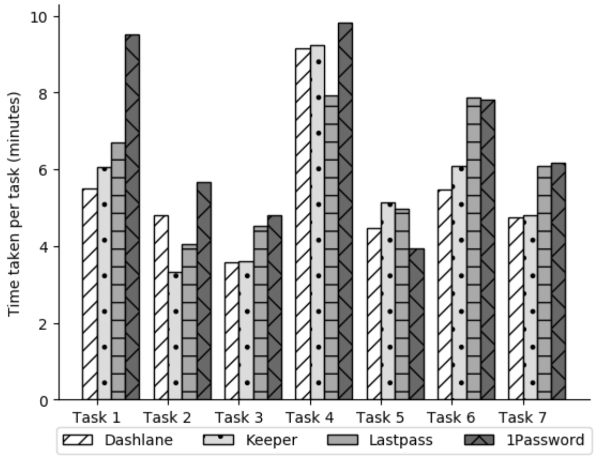


Figure 2: PM's Efficiency measured through the time taken to complete each task.

too long", "*felt clunky*", "*too many steps*") and positive comments (e.g.,"*I liked the quickness*") about perceived performance, which indicates that there is room for improvement on efficiency for better usability.

3.2.3 *Satisfaction.*

We evaluated satisfaction through open-ended questions. First, by asking what are the aspects of the PMs users like and dislike associated to the different tasks. Second, by querying them about continued intention to use the PM after the survey. To analyze the open-ended responses, we conducted thematic analysis [6], a common approach for exploring qualitative data in human-computer interaction and usable security [37, 52]. First, two of the authors initially elicited and agreed on high common themes after individual reading of the collected responses. Then, one of these researchers developed the initial codebook, which was further reviewed in detail and refined by the other.

Likes and dislikes. From the collected answers, we observed that respondents were generally satisfied with the PMs. The responses for the **positive aspects** presented two common themes, namely effectiveness and simplicity:

- Effectiveness: participants were pleased if they could perform the tasks successfully. Especially for Tasks 4 and 5, many participants responded that they liked that they could auto-fill using the PM.
- Simplicity: it was also important for the participants that it was easy to execute these functionalities. Responses like *"easy to use"*, *"easy to navigate"*, *"simple"* and *"straightforward"* all fall into this category. For example, participants liked that the PM offered a list of websites as they were entering the name of the website (*"website pre-selection"*) because it made the information input faster. Simplicity is a key element of efficiency. Participants liked it when a task was easily done because they did not have to spend much time and effort on it.

The responses for **negative aspects** had four core messages, namely lack of guidance, lack of features, mistrust, and performance:

- Lack of guidance: participants repeatedly complained that the lack of appropriate instructions, tutorials or help pages resulted in the prolonged length of time they needed to finish the tasks. Participants reported several cases, in which they simply did not know how to achieve their goals because the instructions were missing. A few participants were unfamiliar with the concept of a PM and had security concerns that stemmed from not understanding how a PM works.
- Lack of features: participants often had higher expectations of what a PM should be able to do. For them, a PM should have ideally functioned in a certain way and they did not like that it did not. Participants complained that the PMs did not: collect and save login details as a user logs in or signs up for a service, launch apps for login, update passwords automatically when a user has changed the password in the PM, and enable users to use the password generator directly on the website or app while creating an account.
- Mistrust: some participants simply did not like the concept of a PM. Participants worried that their passwords would be lost if they lost access to the PM. They were therefore reluctant to use the password generator in Task 6 because they were afraid to use passwords that were not memorable. Some participants were also concerned that it would be difficult to log in from a new or shared device that did not have the PM installed.
- Performance: performance issues are especially critical for PMs since some participants were already skeptical of the security of the app. When participants experienced performance issues, they found the PM to be less reliable.

Continued intention to use the PM. About half of the participants replied that they would continue to use the PMs after the study. In the follow-up question that asked why the participant decided on continued use, the reasons for **positive answers** were mainly convenience and enhanced security. They stated, for example: *"now that I have used one [PM], I think it is easier to use this app than trying to remember a bunch of passwords and writing them down has its risks."*, *"I think I may give it a try because it seems convenient and easy to use"*, *"This app will make me feel much more secure in the future."*, *"I found it very useful and easy to use. It makes me feel more secure with my passwords."*, *"I think it will save my time."*, *"This makes life on my mobile phone so much easier"*, and *"it seems fairly easy to use and a good idea for better security"*.

The reasons for **negative answers** could be grouped into three categories - usability issues, no perceived need, and already use another PM:

- Usability issues were reported by 40% of the participants: *"It doesn't work as described"*, *"It's a very complicated app that is not user friendly and I don't see why I would ever want to use it honestly."*, *"I don't know if I did the auto-fill incorrectly or if I didn't set something up correctly but it just wouldn't auto-fill any of the passwords."*, *"Great concept and really like how secure I felt when adding credentials, but the user interface and how to access some things like password generator needs some work."*, *"It just flat out doesn't work on my phone. I have a modern phone running the latest Android software, so there's no reason why it shouldn't work."*, *"Too much hassle for me"*, *"looks too complicated"*, *"I did not feel this app made things simple for me"* and *"It does not seem to actually do anything. It does not recommend passwords, automatically remember my credentials, or fill them in for me."*.
- No perceived need was given as a reason for not using a PM by 40% of the participants, explaining e.g., *"I don't need the extra help with my passwords."*, *"I prefer to have [passwords] written down paper myself."*, *" I don't have very much that is important hidden behind simple passwords"*, *"I personally use only 3 different passwords across various sites I know it's one of those 3. This [PM] is unnecessary for my life."*, *"I don't mind if my other passwords are hacked or something to be honest"* and *"I just don't need it since most of my passwords are generally the same variation and easy to remember."*
- *"Already use another PM"* was another reason provided by 26% of the participants that chose not to continue using the manager under evaluation.

3.2.4 Learnability. As explained by Lewis and Sauro [38], the SUS scale is actually composed of two dimensions: usability and learnability. According to their factor analysis, two items of the SUS questionnaire can be treated as an independent scale to measure Learnability: item 4 (*"I think that I would need the support of a technical person to be able to use this system"*), and item 10 (*"I needed to learn a lot of things before I could get going with this system*"). We used these two items to assess the Learnability dimension, as they are related to the ability of users to quickly understand and use an application without help [13].

Regarding the obtained scores, Dashlane had the highest learnability value (M = 75.6, SD = 27.65), followed by Keeper (M = 74.4, SD = 22.75), LastPass (M = 73.8, SD = 27.18) and 1Password (M = 51.3, SD = 29.77). ANOVA results showed statistically significant differences between the average learnability scores (F(3,76)=3.76, p=.014). We therefore ran a Tukey HSD test to determine where the differences occur, which confirmed that the learnability of Dashlane, LastPass, and Keeper is similar, while the difference of these values with 1Password's lower score is statistically significant. Furthermore, the learnability of the first three PMs can be considered *"acceptable"*, since it slightly crosses the threshold of 70 points on the SUS scale, but 1Password enters the *"not acceptable"* category. These results follow the same pattern as the global SUS scores reported in 3.3.

3.2.5 Errors. This dimension evaluates how well a user can complete the desired tasks without errors. In our study, we measured this dimension by asking those participants that failed a task what was the reason for failure. Since the success rates of the tasks were high, it can be inferred that the PMs are generally not error-prone. At the same time, there were only limited types of errors , which suggests that some specific aspects of the PMs' operation where easy for participants to misunderstand or make mistakes. We have identified three error categories: 1) related to the *auto-fill* feature, 2) related to *password generation and update*, and 3) related to *performance.* Most of the issues reported by participants fall in the auto-fill category, for example:

> *"I can't figure out how to login [...]"* (P43)
>
> *"I couldn't get Keeper to interact with the app."* (P38)
>
> *"...(PM) does not actually autofill or log you into any websites on its own, which makes me wonder what the point of it is. I made a Doodle account, and I can log into it but only if I do so manually, in which case, why do I need this app?"* (P66)

From the failure reasons given by users and their comments regarding what did not like about the PMs, it can be observed that the PMs did not offer any assistance when participants struggled with features that are difficult to understand or configure, like the auto-fill. Also, in many cases, as the complaint by participant P38 above reflects, the PMs do not meet the expectations or mental models of the users. It should be possible for the users to quickly deduce why a feature did not work as expected and change their behaviors or re-configure the PM to use it correctly. Therefore, we find that there is room for improvement to better assist the users, and we will further discuss this aspect in Section 4.

3.2.6 Cognitive Load. The subjective cognitive load measurements, reported in Table 5, are similar for all four PMs and there is no statistically significant difference between them (see Appendix A), except for Task 1 (F(3,76)=3.25, p=.026) and Task 6 (F(3,67)=3.3, p=.025).

Table 5: Cognitive load associated to tasks T1-T7 using the different PMs, on a scale from very low(1) to very high(5).

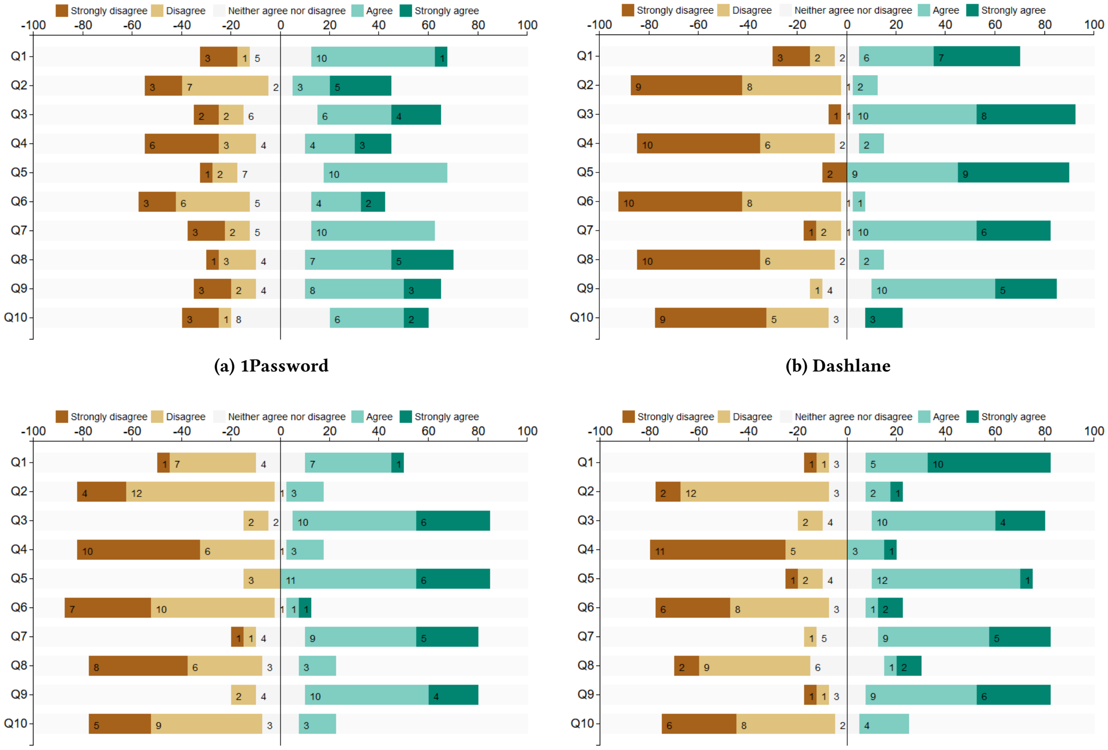
	T1	T2	T3	T4	T5	T6	T7
Dashlane	1.70	2.30	2.17	2.27	2.24	2.78	1.90
Keeper	1.45	2.00	1.84	2.39	2.43	2.21	1.75
Lastpass	2.20	2.10	2.47	2.45	2.38	3.17	1.84
1Password	2.35	2.35	2.73	2.68	2.24	2.83	1.85

Participants answered the question *"How mentally demanding was the task?"* on a 5-points Likert scale ranging from *very low (1)* to *very high (5).* While results were not very low overall, they were on average below the central bar for all tasks, showing that PM's were not particularly demanding to use. Tasks 1 and 7, focused on the installation and configuration of the applications, were rated with the lowest average value, possibly because those are familiar tasks for smartphone users and familiarity might decrease load [49]. In turn, Task 6 has the highest average cognitive load, which could be explained because changing the password and performing a new login required to perform several steps and context switching between the PM and the website. Users found this task confusing and it was, in fact, one of the most common sources of errors. To give an illustrative example, this is how one of the participants describes the cognitive struggle with this task:

> *"the password generator did not generate the password inside of my Memrise app. Instead, I was forced to generate the password inside of the LastPass app. Then I copied and pasted the new password into the Memrise app. Finally, I went to the LastPass app to change the password, but it would not paste the newly generated password. Now I have a big mess to clean up because the app would not generate the password inside of the Memrise app."* (P22)

3.3 SUS Scores

Dashlane got the highest average SUS score, equal to 76.5 (± 17.89), followed by Keeper with 71 (± 16.98), LastPass with 69 (± 19.66), and 1Password with 52.6 (± 21.83). Figure 3 shows participants' responses on 5-point Likert-scales. Since the SUS questionnaire alternates positive and negative wording of its items, those graphs with a more clear diverging looking reflect higher overall scores. We ran a one-way ANOVA test on these data that detected statistically significant differences among the four PMs (F(3,76)=5.73, p=.001), folowed by a Tukey HSD test that confirmed 1Password's mean SUS as significantly lower than the SUS scores of Dashlane, Keeper, and Lastpass. In order to evaluate the impact of demographics on the usability scores, we ran multiple linear regression tests per manager, where gender, age, education, country, OS, previous PM knowledge, and PM use were independent variables, and



(a) 1Password

(b) Dashlane

(c) Keeper

(d) LastPass

Figure 3: Distributions of participants' answers to the SUS Questionaire

SUS was the dependent variable. The full models–with all predictor variables– for 1Password (F(7,12)=1.134, p=.709, with R^2=.275), Dashlane (F(7,11)=.649, p=.288, with R^2=476.), Keeper (F(7,11)=1.425, p=.288, with R^2=.476), and LastPass (F(7,12)=.499, p=.818, with R^2=.226), are not statistically significant. In order to evaluate the impact of demographics on the usability scores, we ran multiple linear regression tests per manager, where gender, age, education, country, OS, previous PM knowledge, and PM use were independent variables, and SUS was the dependent variable. The full models–with all predictor variables– for 1Password (F(7,12)=1.134, p=.709, with R^2=.275), Dashlane (F(7,11)=.649, p=.288, with R^2=476.), Keeper (F(7,11)=1.425, p=.288, with R^2=.476), and LastPass (F(7,12)=.499, p=.818, with R^2=.226), are not statistically significant.

Based on comparisons to other systems and contextual descriptions provided by Bangor *et al.* [8] and Sauro *et al.* [48], (see Figure 1), the SUS scores for Dashlane and Keeper are considered "acceptable", because they slightly cross the score of 70, receiving a C grade. Lastpass, just one point under the 70 border, is considered "marginal high" and gets a D grade. In the case of 1Password, it enters the "low marginal"/F-grade category with a score that is very close to the "not acceptable" threshold of 50 points.

Overall, looking at the small sample of analyzed applications, PMs appear as software tools that can be subjectively considered "ok" or "good", but far from being "excellent" [8]. To contextualize these evaluations, it is interesting to refer to similar empirical tests of authentication mechanisms. In this regard, very few works so far have used SUS as standard measure of usability [45, 47, 54, 55], and just Trewin *et al.* [55] focused on authentication mechanisms in smartphones. Their study, a comparison of three biometrics (voice, face and gesture) against traditional passwords, concluded that the latter were the most usable with a SUS of 78, still a higher score than that of current popular PMs. In order to break this barrier and offer a better alternative for users, the SUS of PMs must be increased. The data we collected will be further analyzed in the following sections to offer recommendations on how to improve usability.

3.4 Further Observations

In our survey, we included some additional questions to explore the usability of a design aspect that is fundamentally different from desktop PMs, i.e., the introduction of a web browser within the manager. We also aimed at getting further insights on password generator usage, preferred features, and on how users configure the application to balance security and usability. Our findings are explained here.

3.4.1 In-app browser. We asked participants about the browser they used for performing a login to a website the first and the second time it was required (i.e., during Tasks 3 and 4). For the first login, 43% of the respondents used the PM in-app browser, while this percentage decreased to a 18,75% the second time they had to access a website. Though we do not have information about the reasons supporting these choices, a possible explanation can be that participants started using the in-app browser during Task 3 (login with stored password), because they were already navigating within the PM and it was the quickest option, but they decided to drop this browser for Task 4 (account creation and usage) because they did not like it, or because they are used to their favorite browser. Several answers to the open-ended questions confirm that users perceived the in-app browser as a limitation. Additionally, the other 57% of participants that did not use the in-app browser, might have done it due to a lack of awareness. The selection of the built-in browser has implications for usability because integration with other browsers is limited, which means that features like auto-fill and password generation might not work and the user either gets frustrated or has to perform a set of further steps to complete the login task.

3.4.2 Password generation. After the task to create an account and save the credentials in the PM (Task 4), we asked participants if they had used the automated password generation functionality and why. Participants who used it (on average a 28%) answered that *"it was easier than thinking of a new password [on my own] that fits the website's requirements"*, *"it produces a safe password"*, and *"I wanted to try it out"*. In turn, those who did not use a password generator, reported reasons that fall into three categories; (1) lack of awareness: *"it [the PM] didn't ask me"*, *"I didn't even think to use it"*, and *"didn't know how"*, (2) lack of interest: *"I had a password in mind already"*, or *"I want to create a password I can remember"*, and (3) distrust of PM: *"I don't trust [the PM] - if it fails I will never be able to remember the password"* and *"I wanted a password I could use even when I don't use the app [PM]"*. It can be observed that this type of participants generally believed that they should be able to memorize the passwords themselves and hence were reluctant to use the password generator.

Another observation is that passwords created for Task 6, where participants were instructed to use the password generator tool, were longer (1 to 5 additional characters) than those created for Task 4. This increase is aligned with the findings in [40] that PMs that provide users with password creation features positively influence the overall password strength.

In summary, better integration can improve awareness of the existence and utility of password generator tools within PMs, and thus foster the creation of stronger passwords.

3.4.3 Preferred features. PMs are more than a simple storage solution for passwords and so they provide additional features to enhance security and usability in many ways. To find out how users value these extra features, we created a list of 8 options and instructed participants to pick 6 and rank them by order of preference. The ultimate purpose of the question is to help mobile PM developers prioritize which features to add next or perhaps, remove. The list of features presented to the users was composed by the following items:

- Alerts about password breaches
- Automated log-in process using auto-fill feature
- Multi-factor authentication (MFA)
- Password generator
- Automated password updates
- Security analysis and feedback
- Password sharing
- Synchronization across different devices

The survey results, represented in Figure 4, showed that participants clearly prioritized the auto-fill functionality over all the other features. Other popular functionalities were the password generator and the feature that updates passwords automatically. The feature deemed least useful was password sharing.

3.4.4 Security settings. We designed Task 7 to collect information about which security settings the participants would choose for their PMs and why. Configuration results are graphically shown in Figure 5, and the underlying reasons can be categorized into three groups: (1) to make the app more usable, (2) to make the app more secure, and (3) to minimize setup effort. Participants whose answers belong to the first group wanted to use fingerprint instead of a master password to avoid typing it every time they wanted to use the PM. Some of them disabled the lock-on-exit feature saying that *"it is annoying to constantly have to be unlocking the app when I'm going back and forth"*. Some also disabled the "clear clipboard" feature saying that *"I use my clipboard and don't want it cleared"*. Those participants in the second category, the secure-aware, wanted the PM to lock itself after some idle time to secure their information. They also chose to hide passwords so that they are safe when *"someone looks over my shoulder"*. Finally, participants whose goal was to minimize setup effort, simply chose to leave the settings as they were, saying that *"these [settings] were preset and I kept what was recommended"* and *"the settings were set and I didn't see a reason to mess with them"*.

Given that a fraction of users (a 7.5% of the participants in our study) rely on established pre-configuration, a good practice for PMs is to make default settings the most secure choice. Additionally, some of the configurable options are questionable from a security perspective, such as the "remember master password" function offered by LastPass.

We also observed that one of the most changed features was the possibility of unlocking the PM with a fingerprint instead of typing the master password. Comments associated to this functionality were positive about the increased convenience and smoothness of the method and included words expressing the *"coolness"* of the experience. This positive perception triggered by the enjoyment of using technology was also noted in [3, 47], so it is an interesting path to explore for increasing usability.

4 RECOMMENDATIONS

After conducting the study and analyzing the results, several issues stood out. Based on the acquired knowledge, we emphasize the following recommendations to improve password managers and foster adoption:

4.1 User Guidance and Interaction

During the study, participants very frequently reported lack of guidance and information on how to use PM functionalities and showed gaps in knowledge that resulted in errors, unfulfilled expectations, or frustration. The easiest way to increase usability in this regard is to implement better tutorials, instructions, explanations,

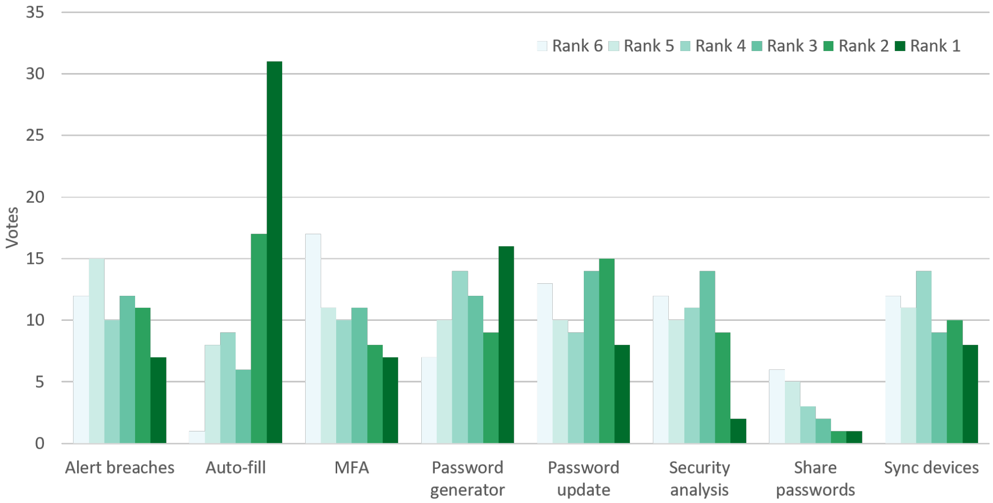


Figure 4: Ranking of additional PM features

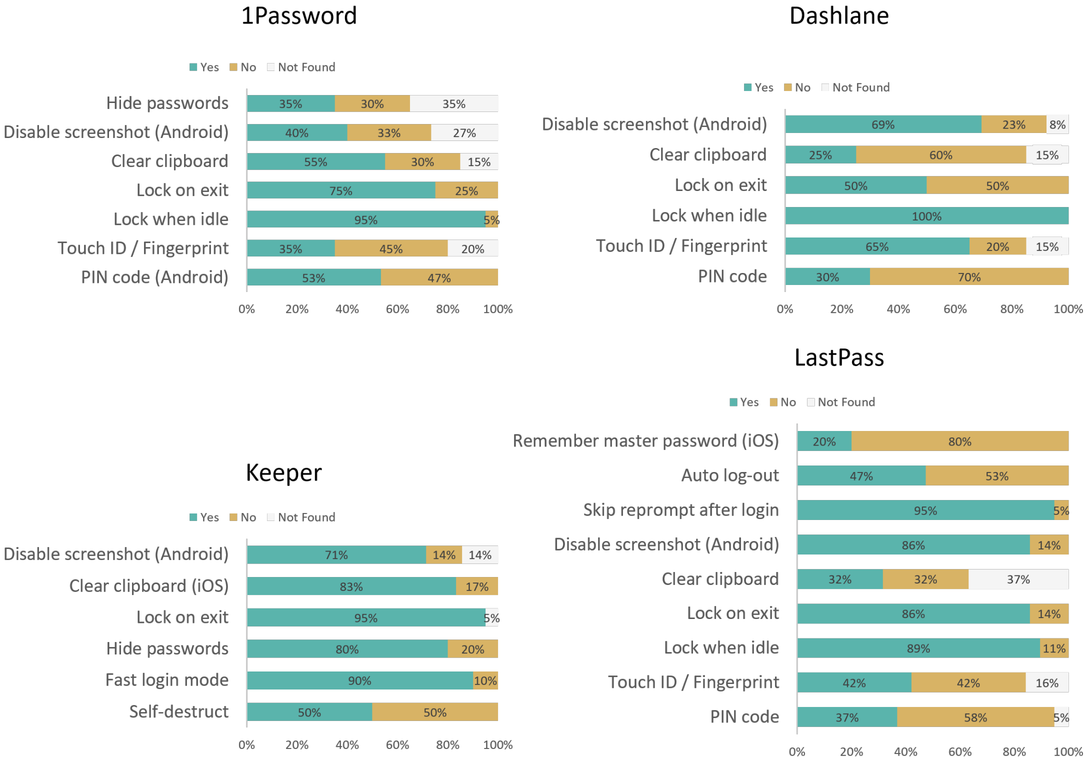


Figure 5: Security settings configured by participants for each PM

and help menus. Participants raised concerns about how difficult was to find and learn how to use the password generator, many of them learning by trial and error. Most of the complaints appeared when configuring the application settings, as users could not understand what exactly some configuration options mean and what are their practical implications on security and/or convenience. Several users also expressed feeling unsure of what was happening or if the task they were completing was successful or not, which points out to the necessity of including more feedback messages even when everything is working. Another key reported problem related to lack of guidance emerged right after installation, when setting up the master password:

> *"It [the PM app] didn't fully explain what a master password was up front before I signed up one."* (P73)
>
> *"The first master password I entered had fulfilled all the requirements, but the app didn't take it because it was too simple. It didn't prompt me why it was too simple. I had to figure out by myself why it was too simple."* (P76)
>
> *" It's hard to come up with something that is difficult enough but memorable for me. "* (P43)

In summary, PMs should: (1) educate the users on the concept of a PM-how it works and why it is secure-, and the concept of master password–what it is, why it is important to have a strong master password, and how to choose one–, (2) provide instructions on how to use basic functionalities such as where to save the passwords, where to find the password generator, and how to turn on features like auto-fill; and, (3) explain what the different options in the security settings mean. The format of the tutorials could be more creative adapting to different types of users (beginners vs advanced) and be naturally integrated with the interface so it is readily available when required but does not interfere with the user experience. One participant proposed video tutorials, and others reported that being suggested to configure the settings before starting using the app would help to have a smoother usage experience from the beginning, e.g., knowing that you can use biometrics right away. Finally, it is important that guidance support is designed to increase trust of the users on the PM, as its lack is one of the strong reasons for rejection.

With regard to user interaction, we see two aspects for improvement:

- Performance: it appeared consistently as an important factor that impacts the perceived usability of the PMs and should be considered by designers. Users need PMs that offer an improved experience with respect to manually using passwords, i.e., faster and less error prone. When functionalities like auto-fill were working as expected, participants praised the convenience and the fact that they "*don't have to think a lot*". Previous research on authentication [47] has demonstrated how designing for minimal interaction can significantly increase SUS scores.
- Additional features: while the main purpose of a PM is to store passwords, it is important to have additional features to enhance both usability and security. Features like password generator, auto-fill, and device synchronization are core and need to be well implemented to satisfy user mental models. Besides these features, participants in our study also appreciated as useful the password update feature, having alerts about database breaches, and synchronization across devices. Furthermore, based on previous research [47] and user responses (almost nobody said that the PMs were enjoyable), a recommended path to improve the usability of PMs is to modify current features or include additional ones that introduce a *"coolness factor"*, appealing to the hedonic motivations of users [35] as adoption factor. Further research is thus required on how to design PMs to be more "enjoyable". Another important factor not to forget when designing PM features is accessibility. For example, though many of the participants in our study replaced the master password with fingerprint-based authentication for convenience, one of them did exactly the opposite also for convenience, as a skin problem would result in constant failed fingerprint logins. Therefore, it is recommended to design features that can be configured to be convenient for all potential users.

The discussed problems of lack of guidance and lack of engagement in the user interaction, were also identified as the weak usability points in a recent study focused on desktop/laptop PMs [5]. Thus, there are similarities in the problems faced by both types of PMs.

4.2 Integration

Mobile development guidelines recommend applications to minimize text input, as this is one of the main usability issues [21]. Our findings are in line with these guidelines: the feature that the majority of participants picked to be the most useful was the auto-fill feature. From the various responses, we observed that many participants placed high importance on being able to log in easier by not having to type or copy-paste the passwords. But auto-fill was also the most problematic feature, as its integration with applications and browsers is not always well implemented or possible. We recommend to improve the interfacing between PMs and 3rd parties, which requires not only an effort from PM designers, but also from mobile OSs to offer better integration APIs and from applications to be more PM-friendly. A better guidance could also lead users to understand how to configure and use the auto-fill feature to get the most out of it, e.g., if users are timely advised to use the PM embedded browser, the auto-fill experience will be more pleasant, as they are designed to be perfectly integrated with the manager. Similarly, another feature that requires better integration is the password generator so it can be directly used in the password input field of websites.

4.3 Security

Mistrust was identified as a strong reason for rejecting smartphone PMs in our study. Better security (and better communication about security) can increase user trust and foster adoption. In this dimension, we observed that password policies for the master password are variable for the different PMs, being very relaxed in some cases, with some users worried that weak passwords were accepted. It is recommended that PMs have strong password policies for master passwords, as this is the protection mean for the whole user password database, and that they offer configurable degrees of protection (e.g., MFA) that can be tailored to the user needs. Additionally, password strength feedback should be provided to nudge users towards choosing better passwords [56]. Recent research in this direction has explored the accuracy of password meters [20], and provides evidence on which to base design choices.

It is also important to consider the impact of options that could weaken (perceived) security, such as the features to remember master password and the option to disable auto-lock so that the users would never be logged out of the PM. Participant P77's perception of the PM, for example, was negatively impacted by one of these features:

> *"I don't like that they even offer to remember people's master passwords. I feel like that simply shouldn't be in there."* (P77)

In our study, we found users with different mentalities, ones more concerned about security and others more concerned about convenience. However, when using a password manager, many configuration decisions have to be made that imply trade-offs between security and convenience. To be more useful, PMs could be designed to understand the security necessities of the user, and offer personalized assistance on how to configure the manager to achieve her security goals with the highest level of usability. Vulnerabilities are different when using a PM on a shared family device, on the smartphone used for work, or if only storing low-value accounts. In this sense, apart from getting explicit feedback from the user about her security concerns to customize the configuration, an interesting line to explore is adaptive context-based security, were the settings react to the situation inferred by sensors. This way, for example, auto-lock can be activated/deactivated transparently depending on user location, or authentication to the PM can vary intelligently (e.g., fingerprint while walking vs master password when static) [4].

Finally, we recommend that default settings of PMs are carefully chosen to be secure. The motivation is that our study showed that some users will accept those defaults, trusting that the PM must have provided an optimal configuration.

5 RELATED WORK

The security of PMs has been analyzed in the literature from different perspectives. Bojinov *et al.* [11] proposed an architecture for building theft-resistant password managers based on introducing decoy sets of passwords in the credential database, so online attacks become harder. Belenko and Sklyarov [10] analyzed mobile password managers, concluding that many of these apps failed to provide the claimed level of credential database protection. Subsequent works [19, 22, 39, 51, 53, 60], uncovered vulnerabilities on cloud and browser-based password managers, providing suggestions to help improve secure design. Among the identified problems, we find e.g., insufficient credential database security, feasibility of XSS attacks, and insecure auto-fill policies.

Fewer works, however, have looked at PMs through the lens of usability. Chiasson *et al.* [16] conducted the first user study comparing two browser-based managers, which reported significant usability problems with the interfaces, some of them leading to security vulnerabilities. Karole *et al.*, in [31], comparatively evaluated three types of PMs: online, mobile, and portable USB managers. Their user study showed that, among the three categories, phone-based managers are the preferred choice by users. The authors point out that more research on usability of mobile PMs is required, as the study reveals that user expectations were not fulfilled. More recently, Arias-Cabarcos *et al.* [5] analyzed the usability of desktop password managers, concluding that users positively rate PMs with regard to efficiency, effectiveness, and error tolerance, but there is room for improvement to make them more engaging and easy to learn. Alkadi and Renaud [3] looked into the reasons for adoption of smartphone PMs by analyzing their reviews in Google and Apple app stores, complemented with a user survey. They highlighted usability (specially *"no perceived usefulness"*) as one of the reasons that deter acceptance and, in line with [5], report that users find mobile PMs difficult to understand and that *"enjoyability"* is a desirable feature. More recently, *Pearman et al.* [43] conducted an interview-based study to further investigate the mindsets underlying adoption and effective use of password managers. Since no work so far has put the focus on understanding the usability of smartphone PMs, we address this gap here through a user study dedicated to evaluate this specific subset of managers, with the aim to provide insights for further improvement.

6 CONCLUSIONS

We contribute to the literature on usable authentication with the first usability study of mobile password managers. The study reveals that usability issues that were present in desktop/laptop PMs [5], are also observed in mobile applications, namely: lack of guidance (difficult to learn), and lack of engagement. Additionally, the most problematic area in mobile PM usability is poor integration with other applications and browsers. This issue impacts the operation of core functionalities, like password auto-fill and password generator, which degrades both user experience and security [40]. We recommend to improve these aspects and also enhance security-related features and tutorial materials to increase perceived usability and trust on mobile PMs. An additional line of work to be explored for PMs to capture the needs of different users in different situations, is the automated and adaptive configuration of security features. Finally, our results have important implications for moving forward. Feedback from the study participant's confirms previous findings on adoption and rejection factors [43, 47]. On top of that, for the case of unawareness as rejection factor, we observed that having knowledge of what a PM is, is not enough for adoption. Users should be further motivated to install and try password managers. Therefore, future research is needed to bridge this gap between knowledge and active interest, in order to foster adoption.

ACKNOWLEDGMENTS

The work of P. Arias-Cabarcos has been supported through an Alexander von Humboldt Post-Doctoral Fellowship.

REFERENCES

[1] 1Password. 2019. *1Password website*. Accessed: 2019-02-22.

[2] Anne Adams and Martina Angela Sasse. 1999. Users are not the enemy. *Commun. ACM* 42, 12 (1999), 40–46.

[3] Nora Alkaldi and Karen Renaud. 2016. Why Do People Adopt, or Reject, Smartphone Password Managers?. In *1st European Workshop on Usable Security (EuroSec 2016)*. 1–14.

[4] Patricia Arias-Cabarcos, Christian Krupitzer, and Christian Becker. 2019. A survey on Adaptive Authentication. (2019). to appear.

[5] Patricia Arias-Cabarcos, Andrés Marín, Diego Palacios, Florina Almenárez, and Daniel Díaz-Sánchez. 2016. Comparing Password Management Software: Toward Usable and Secure Enterprise Authentication. *IT Professional* 18, 5 (2016), 34–40.

[6] Jodi Aronson. 1995. A pragmatic view of thematic analysis. *The qualitative report* 2, 1 (1995), 1–3.

[7] Rosnita Baharuddin, Dalbir Singh, and Rozilawati Razali. 2013. Usability dimensions for mobile applications-a review. *Res. J. Appl. Sci. Eng. Technol* 5, 6 (2013), 2225–2231.

[8] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[9] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[10] Andrey Belenko and Dmitry Sklyarov. 2012. "Secure Password Manager" and "Military-Grade Encryption" on Smartphones: Oh, Really?. In *Blackhat Europe*.

[11] Hristo Bojinov, Elie Bursztein, Xavier Boyen, and Dan Boneh. 2010. Kamouflage: Loss-resistant password management. In *European symposium on research in computer security*. Springer Berlin Heidelberg, Berlin, Heidelberg, 286–302.

[12] Joseph Bonneau, Cormac Herley, Francesco Mario Stajano, et al. 2014. Passwords and the evolution of imperfect authentication. *Commun. ACM* 58, 7 (2014), 78–87.

[13] Simone Borsci, Stefano Federici, Silvia Bacci, Michela Gnaldi, and Francesco Bartolucci. 2015. Assessing user satisfaction in the era of user experience: Comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. *International Journal of Human-Computer Interaction* 31, 8 (2015), 484–495.

[14] John Brooke. 2013. SUS: a retrospective. *Journal of usability studies* 8, 2 (2013), 29–40.

[15] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[16] Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. 2006. A Usability Study and Critique of Two Password Managers. In *15th USENIX Security Symposium (USENIX-SS'06)*. USENIX Association, Vancouver, B.C., Canada, 1–16.

[17] Dashlane. 2019. Dashlane website. https://www.dashlane.com/. Accessed: 2019-02-22.

[18] Dinei Florencio and Cormac Herley. 2007. A large-scale study of web password habits. In *16th international conference on World Wide Web*. ACM, New York, NY, USA, 657–666.

[19] Paolo Gasti and Kasper B Rasmussen. 2012. On the security of password manager database formats. In *European Symposium on Research in Computer Security*. Springer Berlin Heidelberg, Berlin, Heidelberg, 770–787.

[20] Maximilian Golla, Benedict Beuscher, and Markus Dürmuth. 2016. On the security of cracking-resistant password vaults. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, New York, NY, USA, 1230–1241.

[21] Jun Gong, Peter Tarasewich, et al. 2004. Guidelines for handheld mobile device interface design. In *Proceedings of DSI 2004 Annual Meeting*. 3751–3756.

[22] Raul Gonzalez, Eric Y Chen, and Collin Jackson. 2013. Automated password extraction attack on modern password managers. *arXiv preprint arXiv:1309.1416* (2013).

[23] Rachel Harrison, Derek Flood, and David Duce. 2013. Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science* 1, 1 (2013), 1.

[24] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *"Human Mental Workload"*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. Elsevier, 139–183.

[25] Cormac Herley and Paul Van Oorschot. 2012. A research agenda acknowledging the persistence of passwords. *IEEE Security & Privacy* 10, 1 (2012), 28–36.

[26] Wonil Hwang and Gavriel Salvendy. 2010. Number of people required for usability evaluation: the 10±2 rule. *Commun. ACM* 53, 5 (2010), 130–133.

[27] Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. "... no one can hack my mind": Comparing Expert and Non-Expert Security Practices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 327–346.

[28] Blake Ives, Kenneth R Walsh, and Helmut Schneider. 2004. The domino effect of password reuse. *Commun. ACM* 47, 4 (2004), 75–78.

[29] Markus Jakobsson, Elaine Shi, Philippe Golle, and Richard Chow. 2009. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX conference on Hot topics in security (Hosec'09)*. USENIX Association, Berkeley, CA, USA, 9–9.

[30] Jmango360. 2018. Mobile App versus Mobile Website Statistics. https://jmango360.com/wiki/mobile-app-vs-mobile-website-statistics/. Accessed: 2019-02-22.

[31] Ambarish Karole, Nitesh Saxena, and Nicolas Christin. 2010. A comparative usability evaluation of traditional password managers. In *International Conference on Information Security and Cryptology*. Springer, Springer-Verlag, Berlin, Heidelberg, 233–251.

[32] Keeper. 2019. Keeper website. https://keepersecurity.com/. Accessed: 2019-02-22.

[33] Melissa G Keith and Peter D Harms. 2016. Is Mechanical Turk the answer to our sampling woes? *Industrial and Organizational Psychology* 9, 1 (2016), 162–167.

[34] Philip Kortum and S Camille Peres. 2014. The relationship between system effectiveness and subjective usability scores using the System Usability Scale. *International Journal of Human-Computer Interaction* 30, 7 (2014), 575–584.

[35] Adelyn Lai Kit Kuan, Hui Ni Ann, Mohd Badri, Emeilee Nur Freida, and Kia Yee Tang. 2014. UTAUT2 influencing the behavioural intention to adopt mobile applications.

[36] Lastpass. 2019. Lastpass website. https://www.lastpass.com/. Accessed: 2019-02-22.

[37] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.

[38] James R Lewis and Jeff Sauro. 2009. The factor structure of the system usability scale. In *International conference on human centered design*, Masaaki Kurosu (Ed.). Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 94–103.

[39] Zhiwei Li, Warren He, Devdatta Akhawe, and Dawn Song. 2014. The Emperor's New Password Manager: Security Analysis of Web-based Password Managers.. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 465–479.

[40] Sanam Ghorbani Lyastani, Michael Schilling, Sascha Fahl, Michael Backes, and Sven Bugiel. 2018. Better managed than memorized? Studying the Impact of Managers on Password Strength and Reuse. In *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, Baltimore, MD, 203–220.

[41] Mohammad Mannan and Paul C. van Oorschot. 2012. Passwords for Both Mobile and Desktop Computers: ObPwd for Firefox and Android. In *;login:, Vol. 37, No. 4*. The Advanced Computing Systems Association, 28–37.

[42] William Melicher, Darya Kurilova, Sean M Segreti, Pranshu Kalvani, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Michelle L Mazurek. 2016. Usability and security of text passwords on mobile devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 527–539.

[43] Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2019. Why people (don't) use password managers effectively. In *Fifteenth Symposium On Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 319–338.

[44] S Camille Peres, Tri Pham, and Ronald Phillips. 2013. Validation of the system usability scale (SUS) SUS in the wild. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 57. SAGE Publications Sage CA: Los Angeles, CA, 192–196.

[45] Joshua Reynolds, Trevor Smith, Ken Reese, Luke Dickinson, Scott Ruoti, and Kent Seamons. 2018. A Tale of Two Studies: The Best and Worst of YubiKey Usability. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, IEEE Computer Society, Los Alamitos, CA, USA, 872–888.

[46] Jeffrey Rubin and Dana Chisnell. 2008. *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons.

[47] Scott Ruoti, Brent Roberts, and Kent Seamons. 2015. Authentication melee: A usability analysis of seven web authentication systems. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 916–926.

[48] Jeff Sauro. 2011. Are Both Positive and Negative Items Necessary in Questionnaires? https://measuringu.com/positive-negative/. Accessed: 2019-02-22.

[49] Walter Schneider and Richard M Shiffrin. 1977. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review* 84, 1 (1977), 1.

[50] M. S. Silberman, B. Tomlinson, R. LaPlante, J. Ross, L. Irani, and A. Zaldivar. 2018. Responsible Research with Crowds: Pay Crowdworkers at Least Minimum Wage. *Commun. ACM* 61, 3 (Feb. 2018), 39–41. https://doi.org/10.1145/3180492

[51] David Silver, Suman Jana, Dan Boneh, Eric Yawei Chen, and Collin Jackson. 2014. Password Managers: Attacks and Defenses.. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 449–464.

[52] Elizabeth Stobert and Robert Biddle. 2018. The password life cycle. *ACM Transactions on Privacy and Security (TOPS)* 21, 3 (2018), 13.

[53] Ben Stock and Martin Johns. 2014. Protecting users against XSS-based password manager abuse. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*. ACM, New York, NY, USA, 183–194.

[54] Rana Tassabehji and Mumtaz A Kamala. 2012. Evaluating biometrics for online banking: The case for usability. *International Journal of Information Management* 32, 5 (2012), 489–494.

[55] Shari Trewin, Cal Swart, Larry Koved, Jacquelyn Martino, Kapil Singh, and Shay Ben-David. 2012. Biometric authentication on a mobile device: a study of user effort, error and task disruption. In *Proceedings of the 28th Annual Computer Security Applications Conference*. ACM, New York, NY, USA, 159–168.

[56] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, et al. 2017. Design and evaluation of a data-driven password meter. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3775–3786.

[57] Blase Ur, Fumiko Noma, Jonathan Bees, Sean M Segreti, Richard Shay, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2015. "I Added'!'at the End to Make It Secure": Observing Password Creation in the Lab. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 123–140.

[58] Robert A Virzi. 1992. Refining the test phase of usability evaluation: How many subjects is enough? *Human factors* 34, 4 (1992), 457–468.

[59] DYNAMO WIKI. [n.d.]. Guidelines for academic requesters (version2.0). http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters. Accessed: 2019-02-22.

[60] Rui Zhao, Chuan Yue, and Kun Sun. 2013. Vulnerability and risk analysis of two commercial browser and cloud based password managers. *ASE Science Journal* 1, 4 (2013), 1–15.

A STATISTICAL TESTS

Table 6 shows the ANOVA tests to check the significance of differences regarding PMs' mean efficiency. At 95% confidence interval,

Table 6: ANOVA results for time taken per task per PM.

Task 1

Source	Sum of Squares (SS)	df	Mean Square (MS)	F Value	*p*-Value: Sig.
Between-groups	41.5842	3	13.8614	0.79994	0.498
Within-groups	1230.299	71	17.3281		
Total	1271.883	74			

Task 2

Source	SS	df	MS	F Value	Sig.
Between-groups	58.4291	3	19.4764	1.81512	0.152
Within-groups	783.2941	73	10.7301		
Total	841.7232	76			

Task 3

Source	SS	df	MS	F Value	Sig.
Between-groups	22.1654	3	7.3885	0.64115	0.591
Within-groups	829.7103	72	11.5238		
Total	851.8757	75			

Task 4

Source	SS	df	MS	F Value	Sig.
Between-groups	36.9889	3	12.3296	0.33444	0.800
Within-groups	2691.285	73	36.8669		
Total	2728.274	76			

Task 5

Source	SS	df	MS	F Value	Sig.
Between-groups	17.0447	3	5.6816	0.41543	0.742
Within-groups	984.6925	72	13.6763		
Total	1001.737	75			

Task 6

Source	SS	df	MS	F Value	Sig.
Between-groups	86.6612	3	28.8871	0.69280	0.560
Within-groups	3043.811	73	41.696		
Total	3130.472	76			

Task 7

Source	SS	df	MS	F Value	Sig.
Between-groups	138.092	3	46.0307	1.28156	0.287
Within-groups	2621.988	73	35.9176		
Total	2760.08	76			

the null hypothesis that the time taken per PM per task is the same cannot be rejected for any of the tasks, so there is no statistically significant difference.

Table 7 shows the ANOVA test for *Cognitive Load*. At 95% confidence interval, the null hypothesis that the average cognitive load per task is the same among the four PMs could not be rejected for Tasks 2, 3, 4, 5, and 7. There is only a statistically significant difference for Tasks 1 (p = .026) and 6 (p = .025).

B SURVEY QUESTIONS

B.1 Pre-study Questionnaires

Q1 Which OS does your smartphone operate in?
□ Android
□ iOS

Q2 Do you know what a password manager (PM) is?
□ yes
□ no

Q3 Do you use a password manager?
□ yes
□ no

Q4 If yes, name of PM?

Q5 Which version of PM do you use?
□ PC
□ smartphone
□ tablet
□ other

Q6 What is your age?
□ 18-24 years old
□ 25-34 years old
□ 35-44 years old
□ above 45

Q7 What is your gender?
□ Male
□ Female
□ Other

Q8 What is the highest degree you have completed?(If currently enrolled, highest degree received.)
□ Less than a high school diploma
□ High school degree or equivalent
□ Some college credit, no degree
□ Technical/vocational training
□ Bachelor's degree
□ Master's degree
□ Professional degree
□ Doctorate degree

Table 7: ANOVA results for *Cognitive Load* per task per PM

Task 1

Source	Sum of Squares (SS)	df	Mean Square (MS)	F Value	*p*-Value: Sig.
Between-groups	10.65	3	3.55	3.25452	0.026
Within-groups	82.9	76	1.0908		
Total	93.55	79			

Task 2

Source	SS	df	MS	F Value	Sig.
Between-groups	1.6375	3	0.5458	0.45813	0.712
Within-groups	90.55	76	1.1914		
Total	92.1875	79			

Task 3

Source	SS	df	MS	F Value	Sig.
Between-groups	7.6697	3	2.5566	2.17659	0.099
Within-groups	78.6965	67	1.1746		
Total	86.3662	70			

Task 4

Source	SS	df	MS	F Value	Sig.
Between-groups	2.5136	3	0.8379	0.76432	0.518
Within-groups	77.833	71	1.0962		
Total	80.3467	74			

Task 5

Source	SS	df	MS	F Value	Sig.
Between-groups	0.4538	3	0.1513	0.13097	0.941
Within-groups	69.2962	60	1.1549		
Total	69.75	63			

Task 6

Source	SS	df	MS	F Value	Sig.
Between-groups	10.0397	3	3.3466	3.30338	0.025
Within-groups	67.8758	67	1.0131		
Total	77.9155	70			

Task 7

Source	SS	df	MS	F Value	Sig.
Between-groups	0.2344	3	0.0781	0.09668	0.962
Within-groups	60.6263	75	0.8084		
Total	60.8608	78			

B.2 After-task Questionnaires

We presented brief questionnaires after the participants conducted each of the tasks designed to evaluate the PMs. The first block of five questions is common for every task, oriented to test *Effectivenes*, *Satisfaction*, *Errors* and *Cognitive Load*; and they are followed by several task-specific questions designed to get further insights.

B.2.1 Task 1: Initialization.

Q9 Where you able to finish the task?
□ Yes
□ No

Q10 If yes, how mentally demanding was the task?

◻ Very Low ◻ Low ◻ Average ◻ High ◻ Very High

Q11 Name at least one aspect that you liked about the password manager app -design, interface, feature, etc- regarding this task.

Q12 Name at least one aspect that you didn't like about the password manager app -design, interface, feature, etc- regarding this task.

Q13 If you couldn't complete the task, why not?

Q14 How many characters (digits) does your master password have?

Q15 Your master password includes:

☐ Uppercase characters
☐ Lowercase characters
☐ Numbers
☐ Symbols

B.2.2 Task 2: Account migration. Questions Q16, Q17, Q18, Q19, and Q20 are the same as Q9, Q10, Q11, Q12, and Q13.

B.2.3 Task 3: Login. Questions Q21, Q22, Q23, Q24, and Q25 are the same as Q9, Q10, Q11, Q12, and Q13. The following task-specific question was added:

Q26 Which browser did you use to complete the task?

B.2.4 Task 4: Account creation and usage. Questions Q27, Q28, Q29, Q30, and Q31 are the same as Q9, Q10, Q11, Q12, and Q13. The list of task-specific questions follows:

Q32 Which browser did you use to complete the task?

Q33 Did you use password generator to change the password?

☐ yes
☐ no

Q34 Why did you/did you not use the password generator?

Q35 How many characters (digits) does your new password have?

Q36 Your new password includes:

☐ Uppercase characters
☐ Lowercase characters
☐ Numbers
☐ Symbols

B.2.5 Task 5: Interaction with native apps. Questions Q37, Q38, Q39, Q40, and Q41 are the same as Q9, Q10, Q11, Q12, and Q13.

B.2.6 Task 6: Password change. Questions Q42, Q43, Q44, Q45, and Q46 are the same as Q9, Q10, Q11, Q12, and Q13. The list of task-specific questions follows:

Q47 Which method of access did you use to change the password?

☐ Website (e.g., www.doodle.com)
☐ Native app (e.g., Doodle app you downloaded)

Q48 Why?

Q49 How many characters (digits) does your changed password have?

Q50 Your new password includes:

☐ Uppercase characters
☐ Lowercase characters
☐ Numbers
☐ Symbols

B.2.7 Task 7: Security settings. Questions Q51, Q52, Q53, Q54, and Q55 are the same as Q9, Q10, Q11, Q12, and Q13. There were no task-specific questions. Here, we added a set of questions to understand the PM security settings chosen by participants. Since settings differ depending on the PM, we list the questions asked per PM and OS version. Participants were also asked to elaborate on the reasons for their choices.

Dashlane for iOS.

SQ-A1 Did you enable "Clear Clipboard after 5 minutes"?

☐ yes
☐ no
☐ I didn't find this feature

SQ-A2 Did you enable 'Touch ID' to log in?

☐ yes
☐ no
☐ I didn't find this feature

SQ-A3 Did you enable 'PIN Code' to log in?

SQ-A4 How long is your Dashlane's 'Auto-lock timeout'?

SQ-A5 Did you enable 'Lock on exit'?

☐ yes
☐ no
☐ I didn't find this feature

Lastpass for iOS.

SQ-B1 Did you enable 'Touch ID' to log in?

☐ yes
☐ no
☐ I didn't find this feature

SQ-B2 Did you enable 'PIN Code' to log in?

☐ yes
☐ no
☐ I didn't find this feature

SQ-B3 How long is the timer for your Lastpass 'Lock Options'

SQ-B4 When should your LastPass 'Skip reprompt after login'?

SQ-B5 When does your LastPass 'Auto Logout'?

SQ-B6 When does your LastPass 'Clear Clipboard'?

SQ-B7 Did you enable 'Remember Master Password'?

☐ yes
☐ no
☐ I didn't find this feature

Keeper for iOS.

SQ-C1 When is your Keeper's 'Clipboard expiration'?

SQ-C2 When does your Keeper 'Auto-logout'?

SQ-C3 Did you 'Enable Self-Destruct'?

☐ yes
☐ no
☐ I didn't find this feature

SQ-C4 Did you enable 'Fast Login Mode'?

☐ yes
☐ no
☐ I didn't find this feature

SQ-C5 Did you enable 'Hide passwords'?

☐ yes
☐ no
☐ I didn't find this feature

1Password for iOS.

SQ-D1 Did you enable 'Lock on Exit'?

☐ yes
☐ no
☐ I didn't find this feature

SQ-D2 When does your 1Password 'Auto-Lock'?

SQ-D3 Did you enable 'Touch ID' to log in?

☐ yes
☐ no
☐ I didn't find this feature

SQ-D4 Did you enable 'Clear clipboard'?

☐ yes
☐ no
☐ I didn't find this feature

SQ-D5 Did you enable 'Conceal passwords'?

☐ yes
☐ no
☐ I didn't find this feature

Dashlane for Android.

SQ-A1a Did you enable 'Fingerprint' to log in?
☐ yes
☐ no
☐ I didn't find this feature

SQ-A2a Did you enable 'PIN code' to log in?
☐ yes
☐ no
☐ I didn't find this feature

SQ-A3a Did you enable 'Autolock'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-A4a What is your Dashlane's 'Auto-lock time'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-A5a Did you enable 'Allow screenshots'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-A6a Did you enable 'Clear clipboard'?
☐ yes
☐ no
☐ I didn't find this feature

Lastpass for Android.

SQ-B1a Did you enable 'Lock LastPass automatically'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-B2a When should your LastPass 'Lock when app is idle'?

SQ-B3a Did you enable 'Lock when screen is turned off'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-B4a Did you enable 'Fingerprint' to log in?
☐ yes
☐ no
☐ I didn't find this feature

SQ-B5a Did you enable 'PIN code' to log in?
☐ yes
☐ no
☐ I didn't find this feature

SQ-B6a When should your LastPass 'Skip reprompt after login'?

SQ-B7a When should your LastPass 'Log out when app is idle'?

SQ-B8a Did you enable 'Allow screenshots of this app'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-B9a Did you enable 'Fully clear clipboard history'?
☐ yes
☐ no
☐ I didn't find this feature

Keeper for Android.

SQ-C1a When does your Keeper 'Auto-logout'?

SQ-C2a Did you 'Enable Self-Destruct'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-C3a Did you enable 'Fast Login Mode'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-C4a Did you enable 'Hide passwords'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-C5a Did you allow screenshots?
☐ yes
☐ no
☐ I didn't find this feature

1Password for Android.

SQ-D1a Did you enable 'Fingerprint' to log in?
☐ yes
☐ no
☐ I didn't find this feature

SQ-D2a Did you enable 'PIN code' to log in?
☐ yes
☐ no
☐ I didn't find this feature

SQ-D3a Did you enable 'Lock on Exit'?
☐ yes
☐ no
☐ I didn't find this feature

SQ-D4a When does your 1Password 'Auto-Lock'

SQ-D5a When should your 1Password 'Clear Clipboard'

SQ-D6a Did you enable 'Conceal Passwords'
☐ yes
☐ no
☐ I didn't find this feature

SQ-D7a Did you allow screenshots?
☐ yes
☐ no
☐ I didn't find this feature

B.3 Post-study Questionnaire

The post-study starts with the SUS questionnaire, items SUS01-SUS10, answered with a 5-point Likert-Scale from *Strongly Disagree* to *Strongly Agree*. Additionally, we added questions PQ1 and PQ2 to understand users preferences and intention on continued use.

SUS01 I think that I would like to use this system frequently
SUS02 I found the system unnecessarily complex
SUS03 I thought the system was easy to use
SUS04 I think that I would need the support of a technical person to be able to use this system
SUS05 I found the various functions in this system were well integrated
SUS06 I thought there was too much inconsistency in this system
SUS07 I would imagine that most people would learn to use this system very quickly
SUS08 I found the system very cumbersome to use
SUS09 I felt very confident using the system
SUS10 I needed to learn a lot of things before I could get going with this system

PQ1 Will you continue to use this password manager app after the study? Please indicate your reasons

PQ2 Please rate the usefulness of the additional features of a password manager below, numbering each box in order of preference from 1 to 8.

□ Generates random, complex passwords
□ Able to share password with others
□ Updates passwords of different accounts automatically
□ Synchronization over different devices (computers, tablets)
□ Analyzes how secure you are and gives advice on how to increase your security
□ Automates log-in process using auto-fill feature
□ Alerts user about password breaches and other security problems
□ Multi-factor authentication to keep your data more secure