Comparative study of Two-Factor Authentication methods

Marcin Cuber, Keqin Feng, Rex Lau, Yuanlu Li, Siri Vinay

University College London

Email: {marcin.cuber12, keqin.feng.12, yiu.lau.12, yuanlu.li.12, siri.vinay.12}:@ucl.ac.uk

Abstract: There are currently many types of authentication methods in existence. Each method has its advantages and disadvantages when it comes to ensuring system security. Two-Factor Authentication (2FA) was introduced to many systems to add an extra layer of security that requires not only a username and password (single-factor authentication) but also an additional authentication step to verify the user further. But what about the additional consequences that come with the added step of authentication? Related research into 2FA suggested that it creates extra work for users and can impact overall usability of a login. This paper presents the study of two 2FA technologies: Duo Security (DS) and Google Authenticator (GA). They both provide authentication through the usage of applications that allow users to retrieve security tokens in different ways. We developed an e-learning platform to conduct our usability studies with six participants in order to compare various aspects of the two methods. Our primary goal is to compare the usability of these two methods by uncovering points of disruption to the user's primary task and the factors that generate friction. Our study therefore focused on comparing the two 2FA methods in terms of their ease of use, user satisfaction, user familiarity and the underlying problems that can detriment the overall usability of the two 2FA methods. We collected qualitative and quantitative data through two rounds of lab studies, feedback diaries and feedback forms, and found that DS was more usable in terms of time taken to log in and the number of successful attempts made while logging in. However, we found the GA had fewer errors while logging in than DS. There were generally similar qualitative results for both technologies where participants felt that 2FA was unnecessary for an e-learning platform, however would be more willing to use it for other purposes such as banking, email and/or social media.

Keywords: Two-Factor Authentication; Google Authenticator; Duo Security; usability.

I. INTRODUCTION

1.1 Overview

In this paper, we present a usability study of Two-Factor Authentication (2FA) mobile methods, specifically in the context of e-learning. We chose to implement an e-learning system as opposed to an e-commerce or banking site since it allowed us to design tasks that were quick and easy to complete, and did not require the user to share any of their private information. We conducted two series of face-to-face interviews; an initial enrolment and a final debrief session. The initial enrolment session introduced our participants to the study without revealing the primary goal of the study. The initial enrolment session was followed by remote use sessions (Section III) for which participants used our e-learning platform at home. Users were asked to log-in once a day to complete a primary task, which enabled them to get become more familiarised with using the both the system and the 2FA method they were given. There were five primary tasks set for each user, of which three were completed remotely. They were asked to maintain a feedback diary for each remote session, where they could keep a log of how they found the session. The diary form was completed by each participant at the end of each task, so we obtained three diaries for the remote use sessions and two diaries for the lab sessions. The final debrief session had the same format as the initial lab session, but also an additional series of questions where we obtained further feedback on the remote use sessions and finally revealed our study goal. We analysed data both qualitatively and quantitatively and uncovered the positive and negative aspects of each method. We found that participants were generally not familiar with mobile 2FA (apart from one participant) but were more familiar of physical tokens that they use for e-banking. One out of six participants preferred the use of physical tokens because they perceived it to be more secure. On the comparative side of the study we found that users spent, on average, less time using Duo Security (DS) for enrolment and for login than they did for Google Authenticator (GA). We also found that there is a decrease in the average time taken for the user to login from the first and final session. However there were fewer errors that occurred when using GA than during the usage of DS.

1.2 Contributions

Our study derives a basis for future analysis of 2FA mobile methods. It obtains insight on actual authentication interactions, as we interview actual users of 2FA and ask them to complete tasks on our e-learning platform. By analysing the data collected from semi-structured interviews and feedback diaries (Section IV), we found many factors that affect the usability of each method, such as overall user satisfaction, how the two methods perform in terms of reliability and execution (i.e. the technical implementation) of each 2FA method, as well as user familiarity playing a part in improving usability. Though we could not conclude which method was better, nor place a value on how usable each method was, our research successfully explored a number of factors that affect the usability, as well as uncover further factors that were not considered in previous research. The findings from the research form a set of preliminary results for further studies to expand upon and inform other researchers what factors can be considered in similar studies.

1.3 Goals

The major goal was to investigate the usability of two particular 2FA methods by uncovering points of disruption to the user's primary task and the factors that generate friction. To reach this goal we had to start by finding out the level of ‘ease of use’ of each 2F authentication. We did this by focusing on two aspects that we considered define the ‘ease of use’ of a method: the time taken to complete the authentication (so the actual log-in time) and subjective feedback from participants. We obtained the log-in times by measuring the time taken in first and last lab session (through evaluation of user video recordings [Section III]). Since we could not obtain the log-in times for the remote tasks, we designed questions within the debrief task that would allow us to get a subjective view on the time taken which could then be analysed later. The next step to reach our goal was to see how familiarisation affects the usability of the 2FA, since regular usage of a 2FA method can affect how easy and quick it is to use. Our goal also prompted us to identify any errors or issues when using the two methods (which would identify the factors that generate friction). The identification of such errors were implemented as questions within each feedback dairy as well as further questions that asked the participant to provide scores of the task and login experience.

II. Related Work

This section aims to explore related work that has been carried out in obtaining results on the usability and security on 2FA. We have explored various research papers that range from looking specifically at the security and usability of different types of 2FA to some that look at user behaviour towards passwords and the different layers of authentication generally.

2.1 Usability of 2FA

There are many types of 2FA methods, including tokens, cards, mobile phones and biometrics forming as a medium of additional authentication. While the additional authentication measure has been proven to increase the security of a system, it still seems to be compromised everyday due to poorly chosen passwords from users despite the provided guidelines and a seemingly straight forward process. However, these guidelines and rules to inform users may not succeed due to a variety of factors, and many other aspects of the system are overlooked, that essentially boil down to its usability. To fully understand the factors that cause friction between the user and the system, various research findings will be presented and analysed.

2.2 Usability of Passwords

During our research, we felt it was necessary to explore the idea of 2FA by first identifying the underlying issues with single-factor authentication. As pointed out by Sasse et al [1], users are often referred to as the "*weakest link in the security chain*", however this is suggested to be untrue since users should not be held responsible for any vulnerabilities they may present through the selection of their passwords. The paper shows that usability measures and attention to user behaviour is often overlooked in many security designs. Sasse et al describe four different user studies in order to understand problems encountered with usability of security mechanisms. The paper summarises that knowledge-based authentication is the way forward, given it is implemented well.

This similar theme of password selection is also looked at Hoonakker et al, where a large study is carried out to obtain and examine user behaviour towards passwords, and in addressing human factors behind why users are the "*weakest link*". Hoonakker et al [2] conclude by suggesting alternative authentication that has higher potential of addressing human as well as security factors, namely, graphical passwords having a potential of succeeding since human beings are better at remembering images over text. Furthermore, they also mention 2FA performing equally, if not better, than single password authentication, while still raising usability as a question, even more prominently than before.

Additionally, Yan et al [3] present a similar research idea to Hoonakker et al [2], however focus further on memorability of password and their perceived security through further empirical results. Yan et al describe the trade-off involved with passwords: easy to remember passwords are more susceptible to attacks, whereas stronger passwords are harder to remember. This essentially relates back to the balance of security and usability of the system. Results showed that the control group had far more passwords cracked, however more importantly; they showed that both random and pseudorandom passwords were equally as strong as each other and pseudorandom passwords were equally as easy to remember as user chosen passwords.

2.3 Exploration of 2FA

The security offered through a simple username and passwords seems to fall short as hacking technologies become more sophisticated and diverse as time progresses. Single-factor authentication has become more vulnerable to several types of online and offline attacks including malware, brute force attacks, replay attacks, key loggers, Trojans etc. As a solution to withstand such a wide array of threats and attacks, many security designers have considered using multi-factor authentication – in particular, 2FA.

2.3.1 Research on existing 2FA schemes

There is a wide range of 2FA technologies that are implemented through One-Time Passwords (OTPs), such as SMS-based, time-synchronised, physical tokens, mobile-based authentication, etc. that researchers have examined and compared in order to evaluate what methods are best in terms of both the usability and security factors. For example, Adams et al [4] produced one of the earlier papers that focused more towards usability of 2FA, and how its forms a trade-off with security. The study included a survey followed by detailed interviews with the participants. Adams et al address various issues involved with having multiple passwords, memorability, and writing down of passwords. Results from the study clearly outlined that users lack a sense of understanding as to why passwords have such strict policies. They suggested that users that do not fully understand these factors would often end up forming their own insecure rules and importance of security. Adams et al emphasise how important it is to not blame the users, but the security mechanism in place, and how it fails to fully encapsulate the needs of human users. They raise crucial points of security engineers and users having a gap of communication, and that educating both parties is the right solution to target the problem.

Similarly, De Cristofaro et al [5] analysed the usability of 2FA by evaluating three popular technologies used for 2FA: code generated by security tokens, one-time PINs and smartphone based apps that generated dynamic codes. The paper focuses on three major metrics: the ease of use of the technology, the cognitive efforts required by each individual users and the trustworthiness of each technology. Results showed that 2FA enforced for work is mostly forced upon people with the use of security tokens, whereas 2FA for personal use is mainly voluntary and is done mostly through email or SMS technology. They also showed that there is a correlation between the user's perception of the usability and their individual behaviours, and not actually the technology being used. The contrasting points raised in the paper proved that there is still insufficient research that has been done on usability of 2FA methods.

Focusing on a more specific type of 2FA method, Morse et al [6] address the usability issues involved with smart card authentication. They describe a field study being carried out on 24 participants over the course of ten weeks where participants were using smart cards as a type of 2FA. Common issues such as users forgetting their smart cards, leaving them in the reader, having to carry it around everywhere etc. were raised, as well as positive feedback of a smart card being easier to use over passwords was also mentioned. The results suggested smart card authentication to be a potential alternative to password-based authentication, however did not compare the security aspects of both methods in much detail. Although this does not directly relate to the type of 2FA studied within our research, it shows that this type of authentication is a possible candidate for future research to come.

2.3.2 Research on relating 2FA technologies

Many researchers also looked at relating 2FA methods that included applications surrounding online banking, or e-banking. A paper by Weir et al [7] describes the comparison of three different two-factor methods of e-banking authentication and the user perceptions of the usability surrounding these factors. The three types of devices used were a push-button token generator, a bank card token generator and a chip and PIN-secured token generator. Results showed that the usability score for the push-button method was higher than the other two, there was a significant effect for authentication method on quality ratings in the same pattern as the usability scores for the first hypothesis. Thus the PIN-secured authentication proved to take the longest time. Weir et al establish a clearer idea of the relationship between the requirements for perceived security, convenience and usability in e-banking usage. Results showed that a knowledge-based password method with a push-button token was regarded as the most usable, convenient and preferred authentication method of the three.

Similar to this, Just et al [8] explore different types of multi-factor authentication used in ten UK online bank implementations. They outline how the observed usability and security issues can be identified and also improved through analysing screen points and change. The final results showed that certain types of implementations delay the feedback given by not taking the opportunity to provide it at the time of a screen change, and some which provide granular feedback do so too late in the authentication process. The results also showed that many of the online implementation adopted a weaker security by not enforcing full single password implementations, and actually opted for a weaker alternative of using partial passwords and PINs.

Weir et al also produced another paper [9] focusing on the concept of usable security and the preferences of 2FA in e-banking. The paper compared three different e-Banking authentication processes, a one-factor method and two alternative two-factor processes with a sample of participants with variable experience with e-Banking. The paper makes a very interesting point on the perception of security being interlinked with convenience, intrusion, control and clarity. It suggests that security is the not the main goal of a user's interaction with the e-Banking system, rather perceived convenience is the main motivator of e-Banking adoption. The results showed that the users perceived the single-factor method as more usable and perceived it to be more secure than single-factor mode of authentication. The study also shows that the added level of security with 2FA increases security but decreases usability and the right level of usability and security is hard to achieve. The paper concludes with the statement that convenience rather than usability or security may be the key issue in selecting an appropriate method for long-term online security needs.

Gunson et al [10] discuss a similar concept of multi-factor authentication, except the paper is specifically focused on the user perceptions of usability and security of multi-factor authentication in automated telephone banking. The study involves an experiment with 62 participants from the banking industry to use single-factor and 2FA in an automated telephone format. The results found that for timing, a single-factor authentication method is quicker than a 2FA. When concerned with usability, the numbers of first-attempt-failures were greater with 2FA than with single-factor authentication due to some confusion during the authentication process. The single-factor authentication was highly rated for convenience and ease of use; however the 2FA was rated highly for security. The difference in this paper, compared to the previous one, is that the users were experienced and from the banking industry while the paper earlier to this explored preferences of novice users who perceived the single-factor authentication to be more secure purely due to convenience.

2.4 Summary of Findings

There are many major challenges that were tackled with the research presented, as well as some that are still due to be tackled. Understanding a user's behaviour is vital and often easy factor to miss. It is harder for designers and engineers to be able to tackle every problem a user may encounter while still building something that is sufficiently secure. Likewise, it is difficult to fully educate every user about how security works and password guidelines are put into place. It is clear that there needs to be further user studies that look at existing authentication methods in detail. Usability of an authentication method has to be given equal importance just like its security aspect.

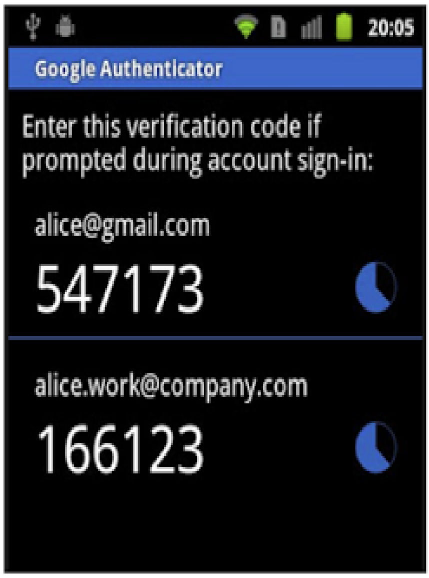
III. Study preparation & Design

We now present the details of the test platform, methodology and the design of our study.

3.1 Test Platform

The test platform was an essential part of our study which we used to test the usability of Google Authenticator (GA) and Duo Security (DS). GA is an application that generates time-based One-Time Passwords which we refer to as OTP security tokens. Users require a smartphone and the GA application in order to scan a QR-code (quick-response codes that are machine-readable form of barcodes) and to allow them to activate the 2FA method. After successful activation users will be prompted with a 6-digit OTP which is valid for 30 seconds. DS also requires a smartphone and the DS application where a user has to scan QR-code through which he will complete his registration. DS is different to GA in the sense that users can select the way in which they can receive the authentication; the most common is Duo Push authentication technology. With Duo Push users can easily approve a smartphone and authenticate with one tap via a push notification. While Duo Push is a key difference, it also supports other forms of 2FA, including phone calls and SMS passcodes.

Our participants had to download an application (i.e. Google Authenticator or Duo Mobile) through the Play Store or App Store depending on the mobile operating system used. Within the enrolment stage, it required them to open the corresponding application and to scan the unique QR-code. After completing this step, users were able to make full use of the 2FA technology. Figure 1 illustrates the example of OTP with timer next to it; users are required to input exactly the same code inside the login section. Figure 2 illustrates our second method where we use Duo Push. So, after we input the correct username and password (using our e-learning platform site), we open up the application in which we see the login request similar to one in Figure 2. We then tap the green approve button and our login procedure is completed with we have full access to the content of our test site.



20:05

Google Authenticator

Enter this verification code if prompted during account sign-in:

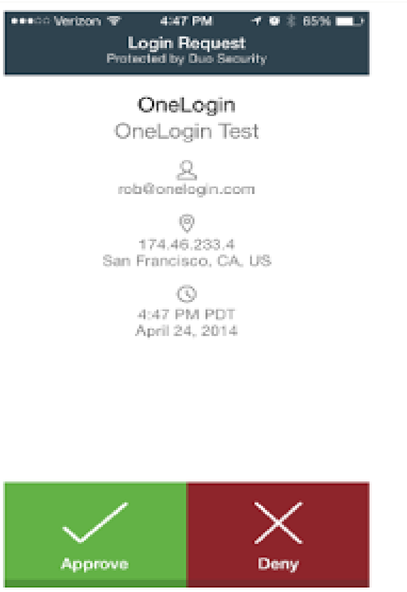
alice@gmail.com

547173

alice.work@company.com

166123

Figure 1- Google Authenticator application



Verizon 4:47 PM 85%

Login Request

Protected by Duo Security

OneLogin

OneLogin Test

rob@onelogin.com

174.46.233.4

San Francisco, CA, US

4:47 PM PDT

April 24, 2014

Figure 2- Duo Mobile application

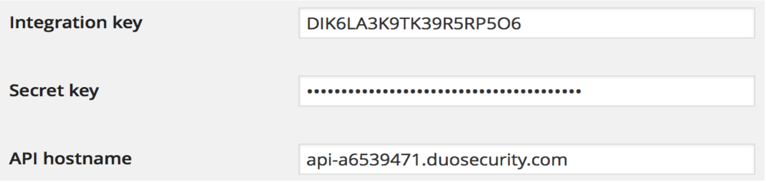
3.1.1 Back-end implementation

In order to test the usability of 2FA methods we used a private server to host our sites. Two separate sites were implemented to test GA and DS services. Software we used required PHP 5.4+, MySQL 5.5+ and mod_rewrite Apache module. Using FTP access, we installed back-end functionality using the WordPress blogging platform; and later installed our selected theme and plugins. To enable smoother and quicker loading of the site, we added code that enabled an advanced level CDN caching services. Additionally, to reduce bandwidth usage we optimised our sites by concatenating all scripts and styles, minifying and compressing them, adding expire headers, caching them, and moving styles to the page header, and scripts to the footer. This extra step was important to ensure the studies ran smooth and more importantly if more users were to log-in at the same time during the remote stage of the study there would be no lag and we would not encounter a problem with resources used (which are limited due to our chosen host provider). Our test sites used the same back-end settings and only the type of 2FA plugin differed which were essential to activate the second authentication factor.

To track users' activities we used a script which kept track of users' actions. We were interested in information such as login attempts, successful and unsuccessful logins. DS service also provided a tracking option so we were able to extract more specific information such as type of activation, IP address and phone number that was used. We ended up with three logging files which we exported to .csv files and later analysed.

3.1.2 Activation

To activate GA we had to enable it within the admin panel so all users could test it. In contrast, Duo Security required external account and activation details (Figure 3).



Integration key	DIK6LA3K9TK39R5RP5O6
Secret key	••
API hostname	api-a6539471.duosecurity.com

Figure 3- Duo Security activation details

DS 2FA requires a paid account (30 days trial version is available which we used) in order to allow our participants to register in the process of testing. This paid service offers more features than GA, so additionally to generating codes for authentication purposes; they offer push-button (Duo Push), SMS passcode and phone call authentication techniques.

3.1.3 Front-end implementation

For the study we did not concentrate on User Interface (UI), however our randomly selected theme made our sites user friendly and easy to use. Our sites are using a standard structure i.e. header, footer, menu with tabs, two columns; content section and section with user's options.

The structure of the registration (user enrolment step) was also manipulated for users to only enter relevant information and then go onto a section where they can enable the 2FA. To enable the use of GA, the user had to generate and scan a QR-code with their phone using the GA application, then update the profile. To enable DS, the user was actually forced to activate it inside the registration process so that the registration was not completed until they successfully completed all steps. These steps required the user to specify their phone's operating system, input a valid mobile number and scan QR-code with the DS application.
Both login sections had a common structure, similar to other services, where a username and password was required, and additionally a security token for GA or to select one of the options to authenticate for DS. GA had a login structure presented in Figure 4. Successful completion of login stage allowed users to move onto the second factor of authentication and then access the content of the site. This included tasks that user was required to complete throughout the study.

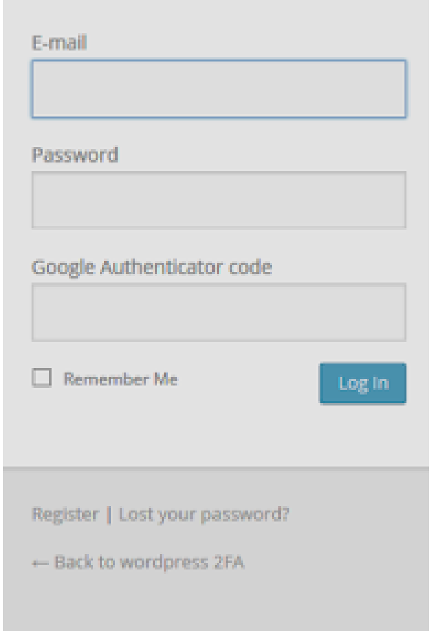


Figure 4- Login section of the website

3.2 Methodology

The research involved six participants who were invited to two user study sessions that took place in a lab: the initial enrolment session and the final debriefing session. There were also three sessions in between that did not take place within our labs; instead users were asked to access the system at home (or a place convenient to them) and complete certain tasks that were assigned to them. There are several factors that lead us to carry out two lab based user study sessions and three remote access sessions. They are looked at in detail below.

3.2.1 User Study

The primary reason behind accommodating user studies within the labs was to enable us to observe, in person, the user's initial and final experience with a 2FA login. It allowed us to setup a controlled environment where we could carefully monitor each user as they completed the tasks given to them, and also to gain more feedback from the users since we could ask further questions if required. In order to fully capture the user experience, we setup a video camera over the shoulder of the user that only showed us the user's interaction with our system on the computer. This way we could see everything that occurred during log-in time in the user's point of view and also effectively time each login session. The setup also included a Dictaphone in order to record the audio during the whole session, meaning that the video camera was only utilised while the user was using the computer, whereas the users' feedback and comments were recorded with the audio recorder.

3.2.2 Remote access

We decided to instruct users to access the system for three consecutive tasks between the first and last lab session. This was done for two main reasons: the first being that it allowed the user to complete the tasks in a more natural environment, where they would feel much more comfortable to log in, thus reflecting a more realistic situation of the use of 2FA. The second reason was so we can analyse patterns or trends that may occur when users utilised the same type of authentication on a regular basis. We instructed our participants to complete one task per day, so it felt more natural them and put lower workload on them. It aided the research in observing whether familiarity to the authentication method would impact the usability in any way, and if so, to what extent. The use of diary forms enabled us to collect the user's experience, albeit not to the same degree of detail as a lab session would provide.

3.2.3 Recruitment

The research was conducted in February/March 2015 and framed as a user study on an e-learning platform. Participants were recruited from a research participant pool at our university. Interested users were invited to an initial enrolment meeting and were asked to sign a consent form before start of the study. We collected the basic demographic information (age, gender, level of education) and asked participants whether or not they had previously used the specific 2FA method as well as for what purpose it was used.

3.2.4 Participants

Our participants were randomly assigned to either of the two experimental conditions. Half of them were testing DS (participants 1-3) and the rest GA (participants 4-6). If they had used the selected 2FA method before it did not affect anything as we did not reveal the goal of the study. Our users consisted of one female and five male participants. Mean age was 28 (range: 23-33), where two participants chose not to reveal their age. Out of the six participants, all of them had a postgraduate degree (a Masters or a PhD). One participant was a psychology graduate and the rest computer science graduates.

3.3 Study Stages

3.3.1 Initial enrolment

The initial stage of the study consisted of structured interviews, conducted in a lab setting with two researchers and one participant at a time. Interviews lasted 30 minutes on average. The goal of these interviews was twofold: (1) to see how they managed the registration and activation process, attitudes towards 2FA methods, and (2) measure times it took to log in, explore usability aspects, aiming to identify sources of common errors, misconceptions, and frustration. Interviews were structured around a basic set of open-ended questions allowing users to talk about their authentication habits and experience. When needed, the researcher followed up on a topic of interest raised by the participant with further questions. At the end, participants were introduced to, and briefed on, the next stage and invited to de-brief session.

3.3.2 Remote use

The remote stage involved users to complete three tasks in their own time, so they had to complete login process at least three times. To complete each of the tasks users were allowed to use any device or platform. Tasks involved watching educational videos and educational articles based on which users had to answer comprehensive questions. Each remote task took 15 minutes on average. The goal of the remote tasks was to see how quickly users adapt to 2FA methods, how long it took them to log-in, and detect how many times they logged in successfully or unsuccessfully.

3.3.3 Debriefing session

The last stage of study involved the same six participants from initial stage. Once again, we conducted semi-structured interviews – lasting 30 minutes on average – in a lab setting with two researchers and one participant at a time. We started with a final task in which users accessed our testing platform. This was followed by debriefing questions, where we asked participants to comment on the experience of using 2FA, on their authentication routine, and to comment on overall experience. Next, we asked them to talk about any other authentication methods they were familiar with, besides the context of e-learning, placing more emphasis on systems that they enjoy interacting with, allowing participants to convey what elements they consider vital for a successful authentication process. Lastly, we asked users to fill out NASA TLX form (Appendix A) to measure how they felt about tasks in terms of difficulty and pressure. In general, NASA TLX is a subjective assessment tool that rates perceived workload, which we used to assess tasks and system. The total workload is divided into six subscales and they serve as one part of the questionnaire: mental demand, physical demand, temporal demand, performance, effort and frustration.

During each session we asked each participant to complete a diary which allowed us to see how difficult the task was, their performance, how long they spent on it, and how satisfied they were with it.

IV. Analysis & results

4.1 Initial Enrolment and Login

The results obtained for the initial enrolment and login timings can be split into quantitative and qualitative results. Both types of data are now analysed in detail below.

4.1.1 Quantitative Results

The average initial enrolment and login times for each 2FA method is presented in Table 1. We found that the average time taken to enrol with GA was 12.4 seconds (s) longer than it took to enrol with DS. We also compared the average initial and final login timings for both 2FA methods and found that it actually took longer for participants to login with DS than GA, specifically 1.3s longer. However it was interesting to see that within the final log-in session, it took participants 4.6s longer to log in with GA than it did with those using DS. This leads onto looking at specifically the change from the first and final logins for both methods. The average decrease in time for GA between the two logins is 2.7s whereas for DS it is 8.6s. From these results we can see that average timing for enrolment was a lot higher for GA than DS. From the Authentication Diary Study [11], we can see from Table 15 (page 72) a login with a username and password would take an average of 14.8 seconds, however when compared to results obtained within this research, we found the average values to be as shown in Table 1. Although the initial log-in time for DS was actually higher than for GA, results showed that for the final log-in times were actually the other way around, with the time taken to log-in with DS decreasing significantly in relation to the decrease in time when using GA. This suggests that over time and regular usage, DS could potentially be more usable than GA since the time taken to complete the authentication methods decreases overall. Furthermore, the fact that log-in times decrease for both methods shows that familiarity with using that type of 2FA could potentially affect the usability of the 2FA method, since it may improve the ease of use, thus affecting the total time taken to login.

The total numbers of successful and unsuccessful logins for both authentication methods were also captured. Figure 5 shows the results obtained for both methods of 2FA. As shown, there were seven more successful logins for DS than GA. The results also show that there was only one total failure with DS, whereas there were seven failures with GA. Unsuccessful logins were identified to be errors on behalf of the system or user occurred during remote logins. This once again shows that DS may potentially be easier to use since there were fewer amounts of errors made during usage, when compared to GA.

We collected data from the server, which was turned into a graph to allow us to compare the GA and DS tests:

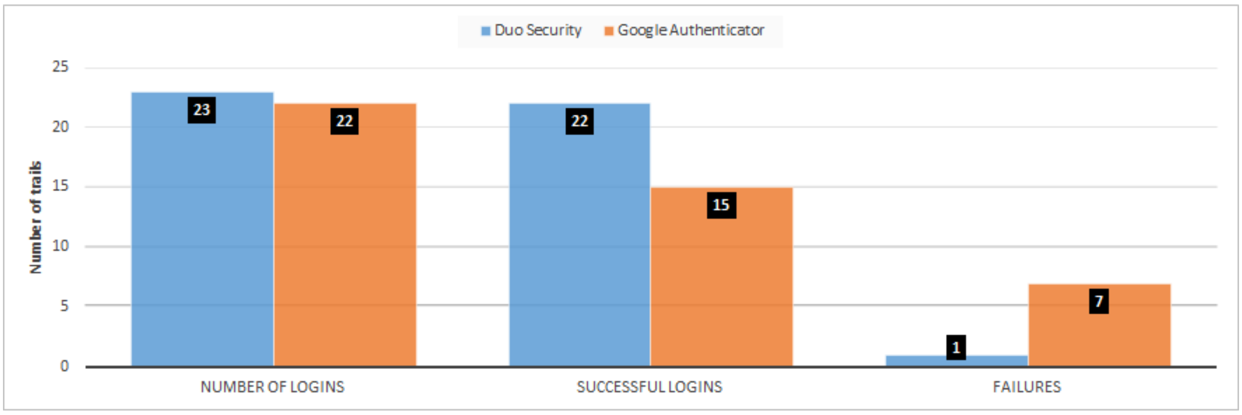


Figure 5- Number of successful logins/failures

4.1.2 Qualitative Results

User feedback from the studies suggested that the registration was mostly straightforward and easy to follow. All but one participant required a bit of extra assistance during the registration for GA. Participant 5 had difficulty with scanning the QR-code due to malfunctioning of the app, however after being prompted by us to keep trying, the participants was successful on the third try. This issue was relating to a technical problem that occurred with the actual application – therefore this is a potential factor on behalf of the execution of the 2FA that could affect actual usability experience of the technology.

In terms of the successful initial logins, all participants had no problems logging in for both GA and DS. However, one particular expressed confusion with the first use of the GA application. Participant 4 explained that while entering the 6-digit OTP, they "*thought that the lifetime span* [next to the six digits] *indicates the time needed to store the code somewhere*", and did not realise it was a timer indicating when the code expires. This is a specifically interesting point that could deplete the usability of GA since it means a fundamental confusion could occur with understand how the actual 2FA works. However, since this confusion only occurred with one participant, it is not significant enough to be able to conclude that DS performs better during initial login.

Table 1- Average enrolment and login times

	2FA setup time	Log in time (Initial)	Log in time (Final)
Average GA	69.7 seconds	33 seconds	30.3 seconds
Average DS	57.3 seconds	34.3 seconds	25.7 seconds

4.2 Diary forms

The diary forms gathered from all participants had varying results. Four out of six participants had no problems during their logins and task completion, however for two participants, certain technical problems occurred during their remote logins. For two of the participants using DS, the server would not accept their password even though it was entered correctly and also experienced the DS application crashing during one of the remote session, however the request for logging in was still valid and open once the application was re-launched. For the second participant, the app failed to send a request to approve the login from the system. Both these issues were related to the technical implementation of the application itself (NB: these attempts not registered as failed login attempts within our system), and similar to GA, could detriment the usability of the app gravely, should these problems occur more often among a larger group of participants. There were no such incidents with GA during consecutive logins after the initial one. This could potentially indicate that GA application is more robust than DS in terms of implementing the 2FA method. On the contrary, it is possible these errors were an anomaly, since we had a very small sample size, and only further studies could validate this should it be true.

Analysis on feedback scores from the Diary forms were averaged for both 2FA methods and presented in Figure 6. As the bar graph illustrates, the average perceived time required for the entire session including the login was slightly higher for GA than DS (by a score of 0.8), perhaps due to the fact that participants had to manually enter a 6-digit passcode, whereas as it was a single press of a button in the case of DS. However, the perceived average difficulty of the task was actually lower for GA than DS (by a score of 0.5). This is interesting and shows that even though it was quicker to complete a task using DS authentication as the login type, participants found it harder to complete the actual task.

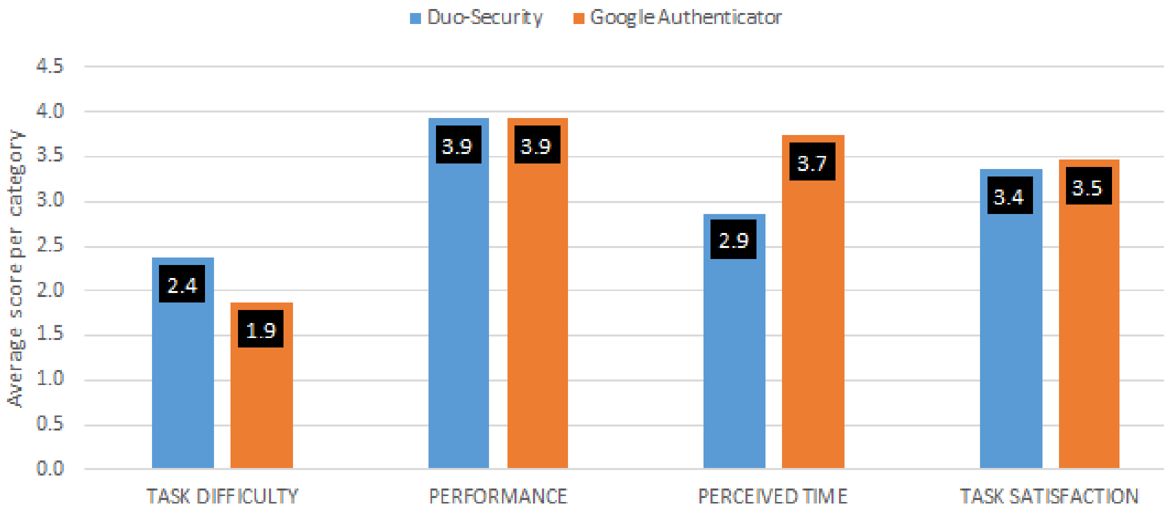


Figure 6- Dairy form feedback results

4.3 Debriefing Session

Carrying out the debriefing session enabled us to gain a lot more feedback and understanding on the usability of both 2FA methods. The results obtained were mostly qualitative and subjective data, unlike most of the results obtained previously during the dairy and initial login stages of the study.

4.3.1 NASA TLX

The overall results gained from the NASA TLX forms are shown in Figure 7 for both 2FA methods. For the performance, effort and frustration aspects of the whole study, the average scores are similar for the two 2FA methods, however for the physical demand required, DS seemed to have a higher score (meaning it was more physically demanding) and GA had a greater temporal demand (meaning the users felt more rushed during login). The reasons behind the latter of these two notable results could be explained due to the fact that the app for GA includes a 30 second timer, constantly telling the user how much longer they have to input the 6-digit passcode, which can be seen as pressuring and adding more stress to the user during the login. This may again affect the usability of the 2FA method, however to what extent, cannot be examined within the scope of this research study. The former notable result relating to the physical demand required when using DS is quite perplexing. Since the Duo Push allows a user to authenticate with a touch of a button as opposed to manually entering a passcode in GA, one could easily conclude that using DS requires *less* physical demand over GA. However, again, it is still possible this is an anomaly, or perhaps it could potentially be further point to research on in further studies.

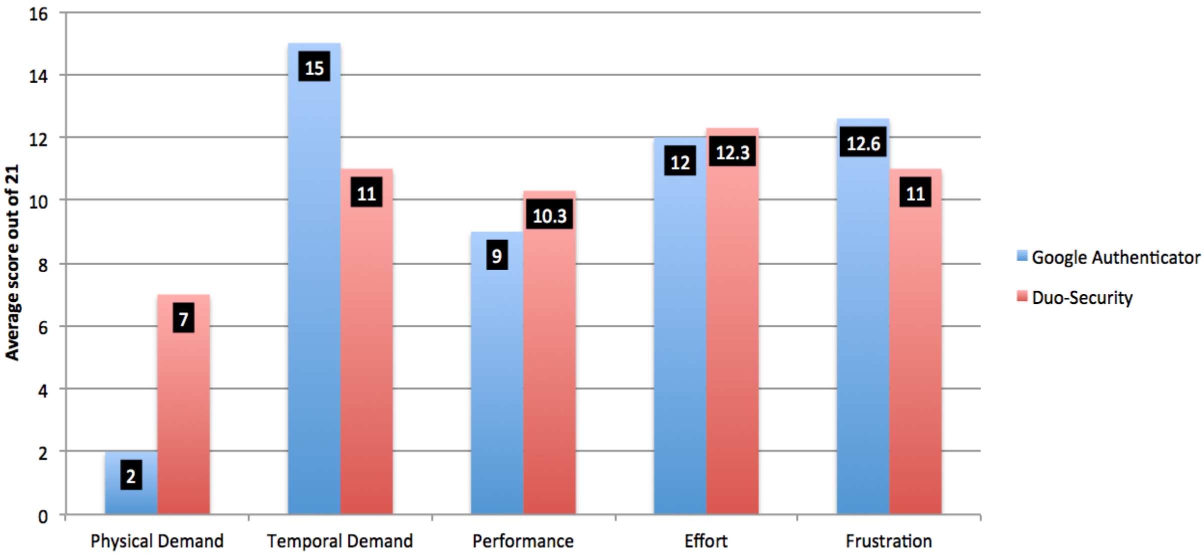


Figure 7- NASA TLX form results

4.3.2 Overall Feedback

The debrief questions (attached in appendix C) were designed to gain an understanding on the participant's overall experience of using the login, and specifically in identifying how usable they found the 2FA authentication as well as their level of satisfaction with using it. All participants for both GA and DS expressed the login and authentication method to be relatively easy and quick (30 seconds on average) to use. All participants were also familiar with 2FA, having used it for online banking, e-commerce, social media or email purposes.

Out of the three participants that used GA, participant 4 enjoyed using it as a 2FA method, and was very willing to use it again. However, participant 5 would not use 2FA at all if they did not have to. The participant explained there must be other ways to ensure the same level of security, and that it allows the security designers "*relax a little*" since they will be greatly dependant on the second factor to ensure the system still meets the security standards defined, meaning they can "*be lazy*" with implementing the first layer of authentication. The participant also mentioned "*having to switch your attention when using it can be detrimental to your task*" at hand. Participant 6 also said they would not be willing to use this type of 2FA for the purpose of using an e-learning system since it does not require the extra layer of security, however did say they would be willing to use it for any other purpose that requires protecting money or sensitive information such as online banking, PayPal, email etc. This shows that overall, GA showed good usability, but participants would still prefer a simple login, at least in the context of this research user study.

Out of the three participants that used DS, no one was familiar with this type of 2FA, and gave very similar feedback to GA. Participant 2 said they think the 2FA is an extra step that requires more effort, as well as being badly designed in many cases. Including DS, and physical tokens for banking, if the 2FA technology fails for some reason, there is no alternative step to continue with the login, providing a fundamental flaw that would essentially prohibit the user from carrying out their primary task, due to no fault of the user. The participant said that both 2FA methods are smartphone dependant, meaning that "*if your phone is out of battery or not with you at the time it will be useless*". The other two participants were more satisfied with the 2FA method where they expressed more willingness to use the 2FA method again in the future, but for other purposes that require the additional security layer rather than for a e-learning system.

An additional point that was noticed during the debriefing session was the idea of familiarity influencing the usability of 2FA in general. Though this is not fully confirmed, we feel that, as mentioned previously, it could have a potential impact on usability, and could be a further point to look at in future research conducted.

V. Conclusions

Though we had a small number of participants, there were several things that we discovered through the study. By observing user behaviour over regular uses throughout the study, we discovered that there may be a potential correlation between familiarities of usage of a 2FA method that affects how usable it is. We also learnt that overall, users did not find the extra layer of authentication to cause any complications while logging in, and neither did they feel it took a significant amount of extra time to their usual login timings. This was further supported by the fact that some participants thought having a smartphone-based 2FA makes the authentication type more usable since most people in this day and age carry their phones with them, while a few others were still concerned with the fact that smartphones were easy to hack into or be compromised.

Discussion

The results showed signs that DS may be more usable than GA in terms of the quantitative data obtained during login timings and successful and unsuccessful login attempts. However, for both the logins, the timings were too close together in duration to come to a full conclusion with the current participant sample size. Further studies with a larger amount of participants would help confirm or invalidate this. Also, certain results that we obtained (such as implementation errors with the DS application that affected the overall usability of the 2FA method and the perceived physical demand required for DS obtained within the NASA TLX form results) could or could not be anomalies within our study. Only additional studies with a larger sample size of participants could help to clear this uncertainty. Furthermore, there is no evidence that the errors that occurred during the remote login sessions actually occurred the way that the participants described. It may be that the participant may have simply misunderstood the error that occurred, and since there is no solid evidence to support this information, we were forced to make the assumption that all the data obtained from the remote sessions were valid in the scope of this study. Therefore, further research in the future would enable us to obtain more reliable data since a larger number of participants would give a clearer indication of what errors occurred and why they occurred.

Another aspect of the results obtained was the perceived time taken to complete the login as opposed to the actual time taken. We found from the results, that the perceived time taken was 30s to one minute whereas the actual average time taken was only around 30s. If we were to take the lower bound of the perceived estimation of time taken, then one could say that both these aspects were very similar. However, the comparison of one minute, opposed to the actual time of 30s is quite significant, and cannot be ignored. Further studies would be required to narrow down this estimated time that was perceived by the participants to identify whether the subjective and objective aspects of time taken for a log-in should be considered as potential factors that affect the usability. Once again, the scope of our study is not sufficient to address this.

Limitations

There were certain limitations within our research project that could not allow us to progress further with the research. The fact that there was a time limit on the entire project meant that we could not obtain a larger amount of participants to obtain more accurate results. Also, since we wished to capture a more natural and realistic usage situation with each user, we could not measure the exact login timings during the remote session logins, which would have further improved our results.

Future Work

Our research has been a preliminary study which will aid further research carried out in this area. We also believe that comparing two similar, mobile-based authentication methods enabled us to gain results that were more specific and narrow to the 2FA types, rather than the type of results that may have been obtained should we have chosen two completely different 2FA methods, say a physical token and a mobile-based 2FA. Developing a flexible platform that supports several types of 2FA methods also means extending the research in the future will be easier and more practical. Our e-learning platform can be expanded to support further types of 2FA including physical token like Yubikey. This will mean that future research to follow can build upon the platform we have designed to accommodate more types of 2FA for further expansion of this research.

In the future, to extend our research, we would need more participants in our experiments that are preferably from a wider range of age groups and educational backgrounds to confirm whether DS is more usable than GA in terms of the factors considered within the study. Moreover, our research only includes the study of only two different 2FA methods, and therefore it could be extended to accommodate more types of 2FA. The research could also be extended to continue the current research, but to also look at perceived security of each 2FA methods, and whether this in turn affects the usability as well. Further usability measures can also be looked at, such as a more detailed analysis of how familiarity exactly impacts the usability, as well as other measures of usability that have not been looked at within this study like convenience of the 2FA method for example.

To conclude, we believe that our research has opened up several points of interest which could be expanded in future research studies. We also believe that the results we have obtained have provided some preliminary evidence that can be used as basis of any future research to follow within this area of study.

Bibliography

[1] A. M. Sasse, S. Brostoff and D. Weirich, “Transforming the ‘weakest link’ - a human/computer interaction approach to usable and effective security,” 2001.

[2] P. Hoonakker, N. Bornoe and P. Carayon, “Password Authentication from a Human Factors Perspective: Results of a Survey among End-Users,” 2009.

[3] J. Yan, A. Blackwell, R. Anderson and A. Grant, “Password Memorability and Security,” 2004.

[4] A. Sasse and A. Adams, “Users are not the enemy,” 1999.

[5] E. De Cristofaro, H. Du, J. Freudiger and G. Norcie, “Two-Factor or not Two-Factor? A Comparative Usability Study of Two-Factor Authentication.,” 2013.

[6] C. Paul, E. Morse, A. Zhang, Y. Choong and M. Theofanos, “A Field Study of User Behaviour and Perceptions in Smartcard Authentication,” 2011.

[7] C. Weir, G. Douglas, M. Carruthers and M. Jack, “User perceptions of security, convenience and usability for e-banking authentication tokens,” 2009.

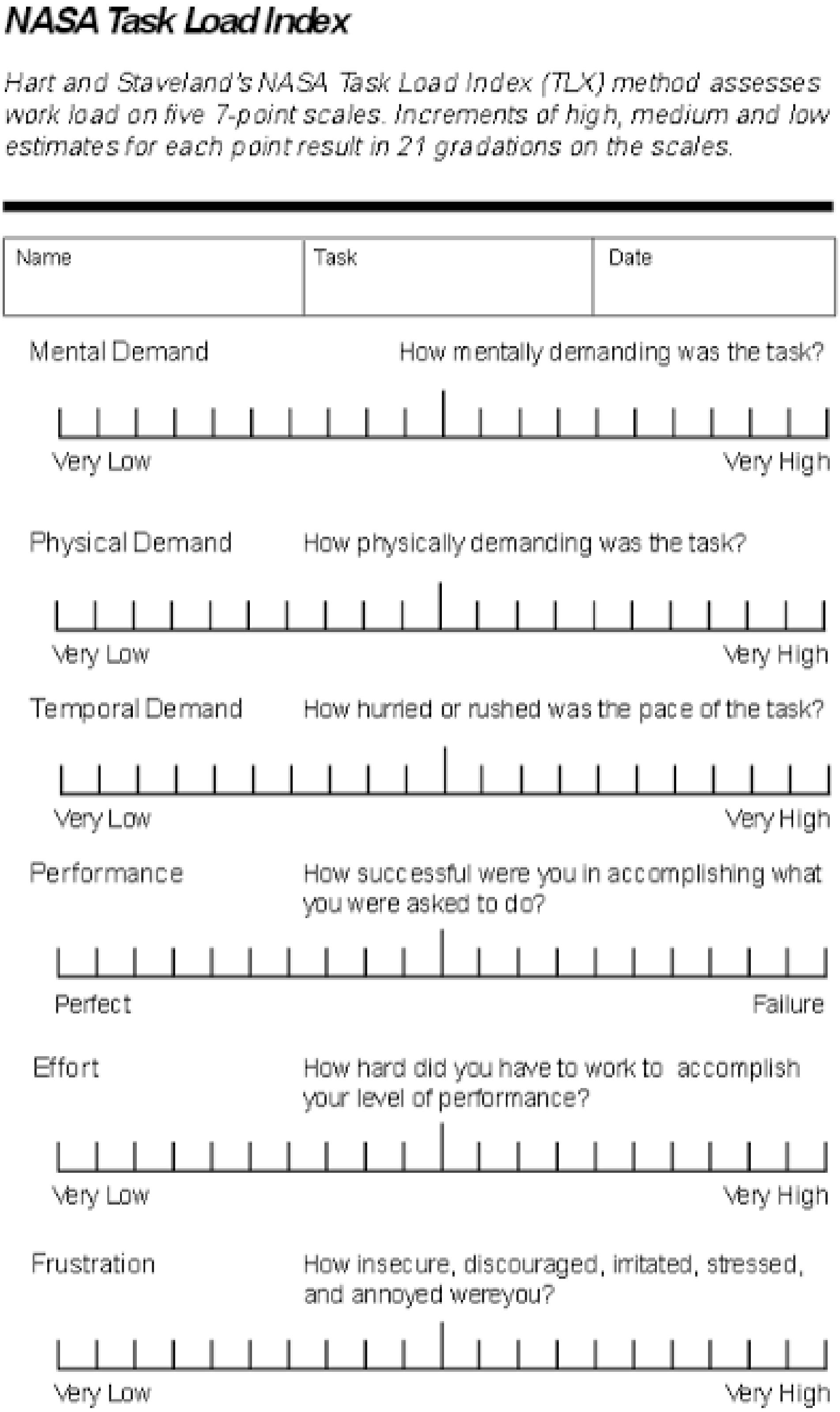
[8] M. Just and D. Aspinall, “On the security and usability of dual credential authentication in UK online banking.,” 2012.

[9] C. S. Weir, G. Douglas, T. Richardson and M. Jack, “Usable security: user preferences for authentication methods in eBanking and the effects of experience. Computers and Security,” 2010.

[10] N. Gunson, D. Marshall, H. Morton and M. Jack, “User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking,” 2011.

[11] M. Steves, D. Chisnell, A. Sasse, K. Krol, M. Theofanos and H. Wald, “Report: Authentication Diary Study,” *NISTIR 7983,* 2014.

Appendix A- NASA TLX form



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

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

Perfect — Failure

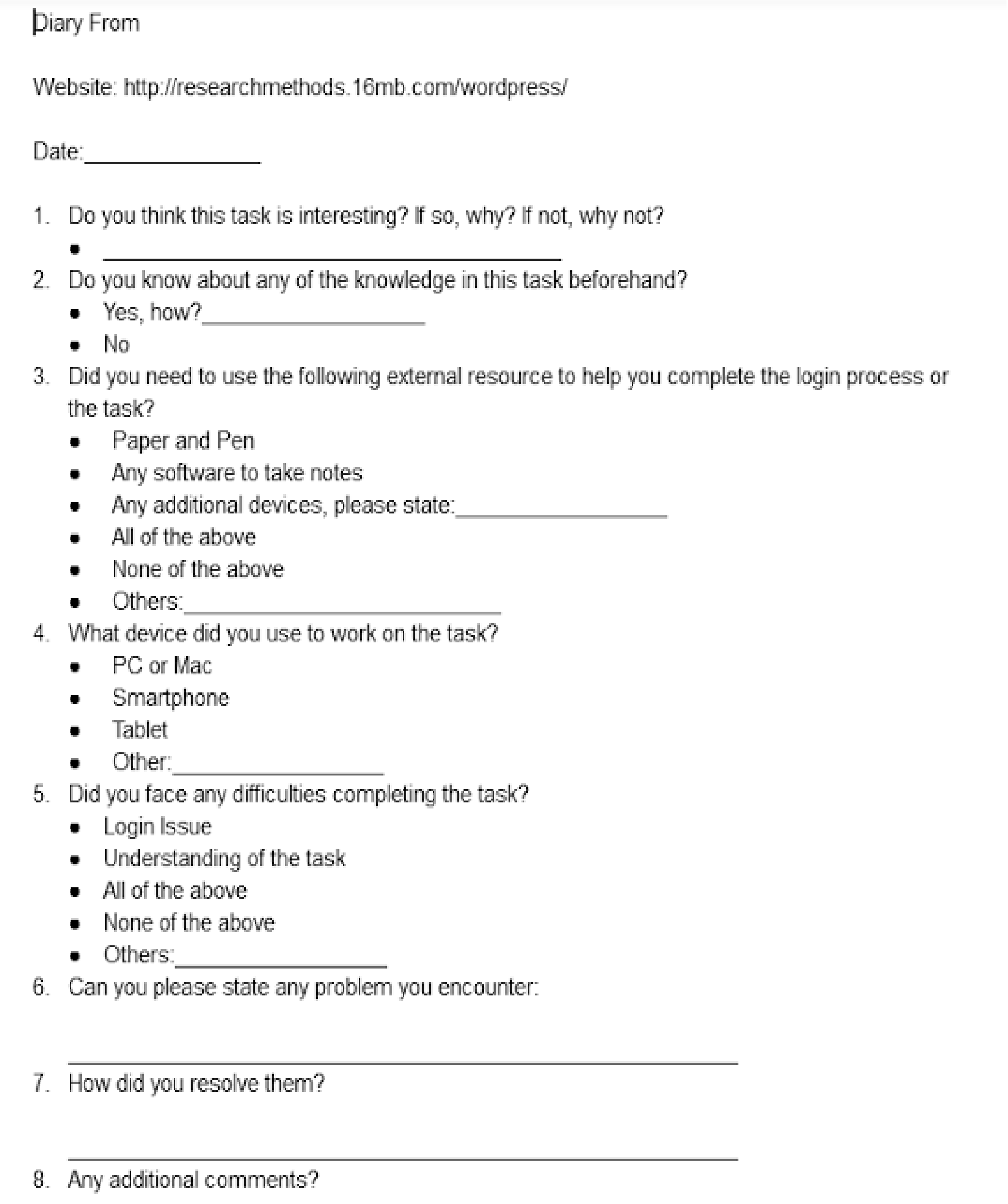
Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

APPENDIX B- DIARY AND FEEDBACK



Diary From

Website: http://researchmethods.16mb.com/wordpress/

Date:_______________

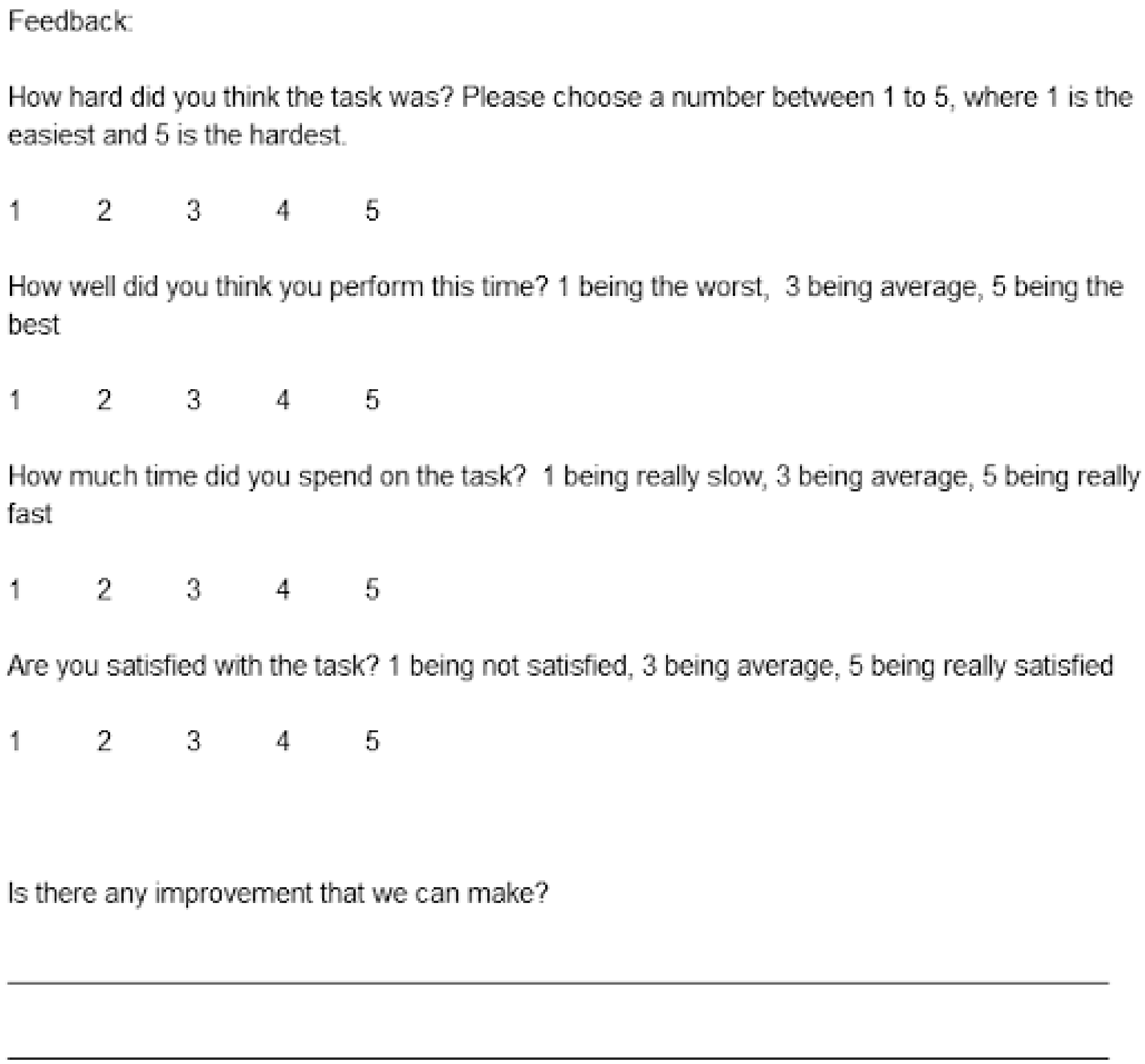
1. Do you think this task is interesting? If so, why? If not, why not?
   - ______________________________________
2. Do you know about any of the knowledge in this task beforehand?
   - Yes, how?___________________
   - No
3. Did you need to use the following external resource to help you complete the login process or the task?
   - Paper and Pen
   - Any software to take notes
   - Any additional devices, please state:__________________
   - All of the above
   - None of the above
   - Others:___________________________
4. What device did you use to work on the task?
   - PC or Mac
   - Smartphone
   - Tablet
   - Other:__________________
5. Did you face any difficulties completing the task?
   - Login Issue
   - Understanding of the task
   - All of the above
   - None of the above
   - Others:__________________
6. Can you please state any problem you encounter:

   ____

7. How did you resolve them?

   ____

8. Any additional comments?



Feedback:

How hard did you think the task was? Please choose a number between 1 to 5, where 1 is the easiest and 5 is the hardest.

1 2 3 4 5

How well did you think you perform this time? 1 being the worst, 3 being average, 5 being the best

1 2 3 4 5

How much time did you spend on the task? 1 being really slow, 3 being average, 5 being really fast

1 2 3 4 5

Are you satisfied with the task? 1 being not satisfied, 3 being average, 5 being really satisfied

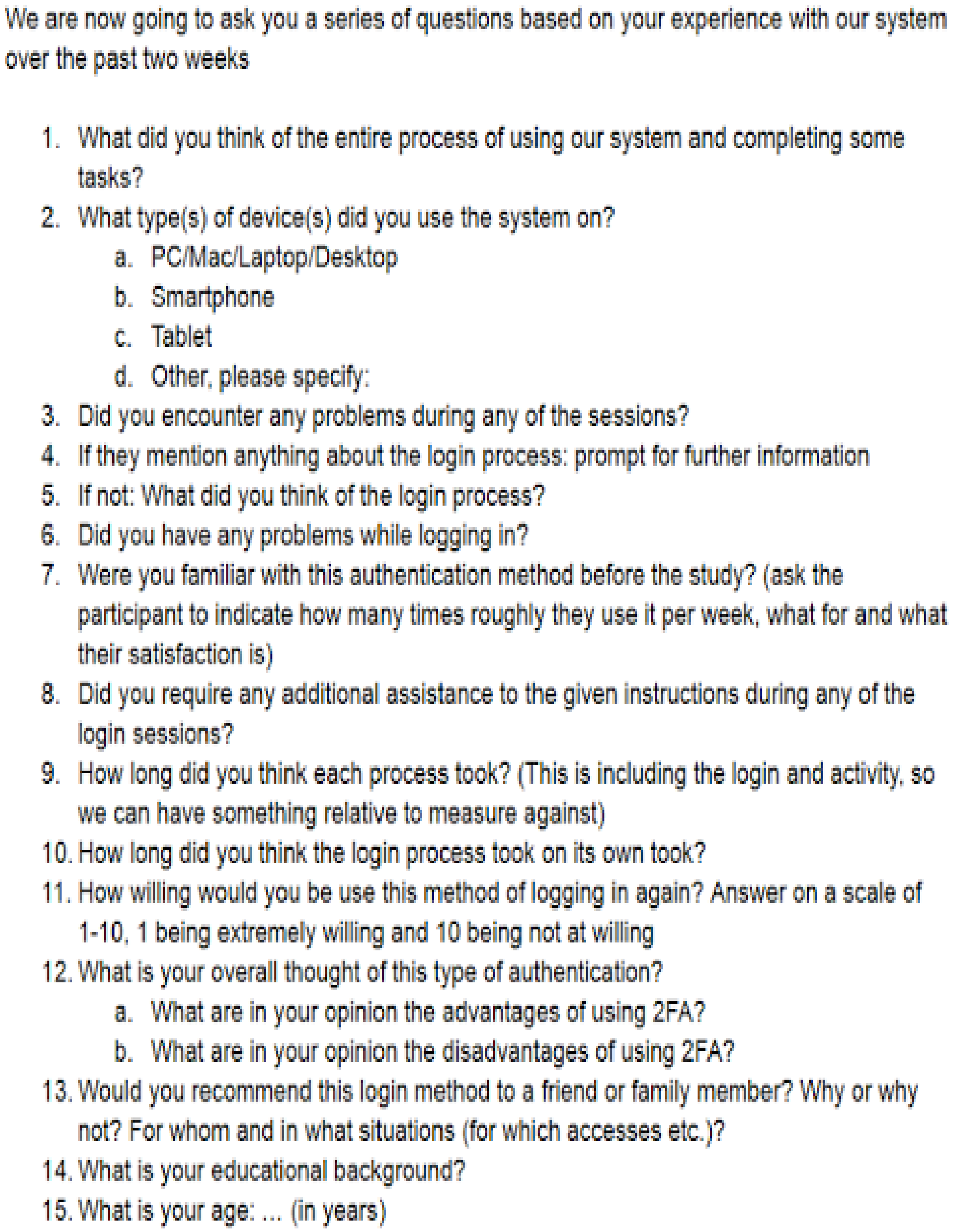
1 2 3 4 5

Is there any improvement that we can make?

____

____

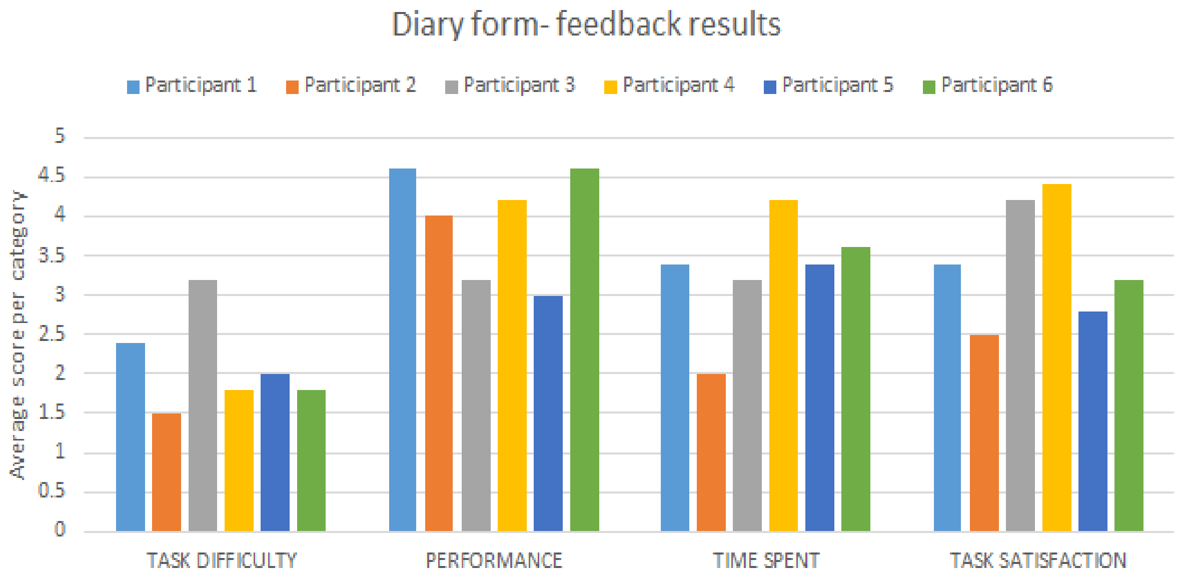
Appendix C- Final Questionnaire



We are now going to ask you a series of questions based on your experience with our system over the past two weeks

1. What did you think of the entire process of using our system and completing some tasks?
2. What type(s) of device(s) did you use the system on?
    a. PC/Mac/Laptop/Desktop
    b. Smartphone
    c. Tablet
    d. Other, please specify:
3. Did you encounter any problems during any of the sessions?
4. If they mention anything about the login process: prompt for further information
5. If not: What did you think of the login process?
6. Did you have any problems while logging in?
7. Were you familiar with this authentication method before the study? (ask the participant to indicate how many times roughly they use it per week, what for and what their satisfaction is)
8. Did you require any additional assistance to the given instructions during any of the login sessions?
9. How long did you think each process took? (This is including the login and activity, so we can have something relative to measure against)
10. How long did you think the login process took on its own took?
11. How willing would you be use this method of logging in again? Answer on a scale of 1-10, 1 being extremely willing and 10 being not at willing
12. What is your overall thought of this type of authentication?
    a. What are in your opinion the advantages of using 2FA?
    b. What are in your opinion the disadvantages of using 2FA?
13. Would you recommend this login method to a friend or family member? Why or why not? For whom and in what situations (for which accesses etc.)?
14. What is your educational background?
15. What is your age: ... (in years)

Appendix D- Individual participant diary feedback



Participant 01	how hard the task was	performance	time spent	satisfied with task
diary 1	1	5	5	5
diary 2	4	3	3	3
diary 3	1	5	5	5
diary 4	2	3	3	4
diary 5	1	5	5	5

Participant 02	how hard the task was	performance	time spent	satisfied with task
diary 1	2	3	3	4
diary 2	3	1	5	1
diary 3	2	5	2	3
diary 4	2	3	3	3
diary 5	1	3	4	3

Participant 03	how hard the task was	performance	time spent	satisfied with task
diary 1	2	4	3	3
diary 2	3	5	3	3
diary 3	1	5	4	3
diary 4	2	4	3	3
diary 5	1	5	5	4

Participant 11	how hard the task was	performance	time spent	satisfied with task
diary 1	2	5	3	3
diary 2	2	5	5	3
diary 3	3	5	3	5
diary 4	4	3	3	2
diary 5	1	5	3	4

Participant 12	how hard the task was	performance	time spent	satisfied with task
diary 1	2	5	2	3
diary 2				
diary 3				
diary 4				
diary 5	1	3	2	2

Participant 13	how hard the task was	performance	time spent	satisfied with task
diary 1	2	4	4	5
diary 2	4	3	3	4

diary 3	4	3	3	4
diary 4	3	3	3	4
diary 5	3	3	3	4

Appendix E- Individual participant NASA Scores

	Mental Demand (Out of 21)	Physical Demand	Temporal Demand	Performance	Effort	Frustration
01	17	2	19	19	19	15
02	13	2	15	6	7	13
03	10	2	11	2	10	10
11	10	3	7	5	7	4
12	21	11	15	13	14	14
13	15	7	11	13	16	15
average GA	13.3	2	15	9	12	12.6
average DS	15.3	7	11	10.3	12.3	11