available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

User perceptions of security, convenience and usability for ebanking authentication tokens

*Catherine S. Weir**, *Gary Douglas*, *Martin Carruthers*[1], *Mervyn Jack*

The Centre for Communication Interface Research (CCIR), School of Engineering and Electronics, The University of Edinburgh, The King's Buildings, Edinburgh, Scotland EH9 3JL, United Kingdom

ARTICLE INFO

Article history:
Received 30 August 2007
Received in revised form
23 July 2008
Accepted 30 September 2008

Keywords:
Usability engineering
Internet Banking
Authentication
Security
Empirical study
Evaluation

ABSTRACT

This research compared three different two-factor methods of eBanking authentication. Three devices employing incremental security layers in the generation of one time passcodes (OTPs) were compared in a repeated-measures, controlled experiment with 50 eBanking customers. Attitudes towards usability and usage logs were taken for each experience. Comparisons of the devices in terms of overall quality, security and convenience as perceived by participants were also recorded. There were significant differences between all three methods in terms of usability measures, perceived quality, convenience and security ratings – with the perceived security ratings following a reverse order to the other measures. Almost two thirds of the participant sample chose the device they perceived the least secure as their preference. Participants were asked to use their preferred method again and tended to find their chosen device more usable. This research illustrates the usability-security trade off, where convenience, quality and usability are sacrificed when increasing layers of security are required. In their preferences, customers were driven by their attitudes towards usability and convenience rather than their perceptions of security.

1. Introduction

This paper describes an experiment to investigate customer perceptions of usability, convenience and security in two-factor (2-factor) authentication. 2-factor authentication is a security solution requiring the verification of two different modalities of authentication components. Typical components include: knowledge (e.g. passwords), possession (e.g. bankcard) and physical attributes (e.g. fingerprint). 2-factor solutions provide enhanced security by combining more than one authentication type, such that if a customer's password is compromised, the second factor will provide an extra barrier to fraudulent entry (O'Gorman, 2003).

The bank ATM has always made use of 2-factor authentication, via a bankcard (possession) and a secret PIN (knowledge). Other banking channels such as telephone and Internet banking (eBanking) are accessed remotely, without physical presence, using single-factor authentication. The most common factor employed is the password, although often dual passwords are required.

eBanking applications allow access to personal, private financial data through the Internet. Security concerns are important for customers and the Banks alike. Findings from studies in eBanking also have applications in the wider field of transaction security for eCommerce activities. Recent increases in online fraud have encouraged Banks to think about 2-factor solutions for their authentication procedures (Hiltgen et al., 2006). Further moves may extend this concept to other electronic transactions such as online shopping (Ives et al., 2004; Ranger, 2005). In selecting appropriate security

* *Corresponding author*. Tel.: +44 131 6502801.
E-mail address: cath@ccir.ed.ac.uk (C.S. Weir).
[1] Lloyds TSB plc.

0167-4048/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2008.09.008

procedures, Banks must research customer attitudes or the result could be a reduction in the use of low-cost eChannels due to usability issues (Knight, 2008).

Several publications have discussed the potential conflict between usability and security in the provision of authentication solutions (Besnard and Arief, 2004; Furnell, 2005; Johnston et al., 2003; Nodder, 2005; Renaud, 2005). Password mechanisms have also been widely studied from security and usability perspectives (Yan et al., 2005; Zviran and Haga, 1999). Many publications recommend designing new security technologies based on usability principles (Smetters and Grinter, 2002). Yet there are very few empirical studies examining the factors affecting usability, convenience and security (Piazzalunga et al., 2005). The usability of security is key for customer acceptance (O'Gorman, 2003; Schultz et al., 2001). The research presented in this paper makes use of the eBanking environment to explore customer perceptions of usability, security and convenience in authentication tasks.

1.1. *The usability and security of authentication interfaces*

The knowledge-based model of authentication has traditionally offered easy, quick and convenient access to computer systems (Tognazzini, 2005; Zviran and Haga, 1999). However, the relatively low security of password verification has caused fraud problems for the eBanking and eCommerce industry (Sinclair and Smith, 2005). Research has documented the limitations of the single-factor, knowledge-based model of authentication in the era of advanced computing and eCommerce: concerns range from human memory limitations when confronted with multiple authentication environments (Bishop, 2005; Morris and Thompson, 1979; Sasse and Flechais, 2005), to the range of information protected by password access – from highly sensitive banking applications to online newspaper access (Ives et al., 2004).

The personal and confidential nature of the financial data held in an eBanking service make the balance between security and usability a main concern. A customer's decision to use eBanking services (Sohail and Shanmugham, 2003) or engage in eCommerce activities (Eastin, 2002) is influenced by these issues. Security and trust are common themes for research, yet few studies have examined both usability and security from the customer perspective.

Convenience, control and efficiency are thought to be the main drivers for customers to bank online (Jayawardhena and Foley, 2000; Sohail and Shanmugham, 2003; Tan and Teo, 2000). The authentication process has a direct effect on security, the ease of use and convenience of a process. Therefore it is important to investigate the effect of different authentication systems on customer perceptions of an online service. Secure and usable authentication will be an important factor in adoption and expansion of the Internet channel for commerce and banking activities (Schultz et al., 2001).

In a recent publication (Nilsson et al., 2005), a security box generating a 'One Time Passcode' (OTP) was compared to the use of a fixed alphanumeric password in the authentication process for eBanking using a between-subjects methodology. The findings revealed that the security box was perceived to be significantly more trustworthy and secure than the password method. However, this study did not report usability levels or perceptions of convenience, further, participants did not make any direct comparisons between the two methods.

Recent guidelines from the US suggest the use of 2-factor authentication as they consider that the sole use of single-factor authentication "inadequate" for certain transactions (FFIEC, 2005), however the demands are not mandatory (Henry, 2006). There is also concern that 2-factor authentication is not sufficient (Schneier, 2005; Henry, 2006) to solve the array of internet transaction security threats. However, as Banks move away from static passwords, many are implementing 2-factor solutions and the impact on customers should be considered.

Further empirical research into usability and security of authentication is required to offer insights to guide new security technologies and interfaces. This paper presents a controlled experiment investigating the usability, convenience and security of several 2-factor authentication solutions for eBanking.

1.2. *Usability engineering*

Usability engineering is a process by which systems are built and tested with empirical methods to achieve efficiency, effectiveness and satisfaction for specified users performing specific goals in a particular environment (International Organization for Standardization, 1998). Usability experiments provide objective and subjective data for the design and engineering of technology solutions that are fit for purpose. The aim is to make appropriate trade-offs between satisfaction and performance measures (Hornbæk and Law, 2007; Hornbæk, 2006).

Usability engineering principles are becoming widely used in the creation of security solutions (Karat et al., 2005) and are suitable for investigating the potential for 2-factor authentication in a well-defined scenario such as eBanking. The experimental evaluation of usability used in this work is based on previous research (Weir et al., 2006) and uses a repeated-measures, within-subjects design.

Usability engineering emphasises data collected based on hands–on experience with alternative interfaces, involving representative users performing typical tasks in well-defined scenarios. The essence of a usability approach is to pose questions in the context of usage, after experiencing the process directly (Root and Draper, 1983). This is in contrast to the passive interview questions posed in market research.

In the experiment reported here, usability was measured in terms of efficiency, effectiveness and satisfaction. Efficiency was investigated by logging the time taken to complete each authentication method. The effectiveness of each authentication method was measured by task completion records and logs of help use. Both of these measures offer objective evaluations of usability. Directly after using each authentication process, participants completed attitude questionnaires as measures of their satisfaction (Preece et al., 1994). The questionnaire will be described in Section 2.5 and forms part of the subjective measure of usability.

In addition, after experiencing the alternative interfaces some additional subjective measures were recorded. Firstly

participants made a quality rating, recorded as a value on a 30-cm linear scale labelled "Worst" at 0 cm and "Best" at 30 cm. This quality rating allowed direct comparison of the alternative process options and was also recoded to indicate an explicit preference for one of the options (equal ratings were allowed). The 'Best to Worst' scale was also used to collect comparative judgements about the absolute and relative levels of perceived security associated with the alternate options. Finally, it was used to rate convenience for the three options in the same way. Statistical analysis using the analysis of variance (ANOVA) was used to make inferences about how various controlled factors relate to performance and satisfaction measures (Landauer, 1988).

2. Experiment details

This experiment was designed to investigate customer perceptions of usability, convenience and security in two-factor (2-factor) authentication. The experimental variables were controlled to examine incremental layers of procedure in a 2-factor authentication device and measure the usability, security and convenience perceptions for each after direct usage and in comparison.

2.1. *The experimental variables*

For controlled experimental comparison, variations between devices were kept to a minimum. Many token devices provided scope for a controlled comparison, particularly in terms of the number of steps required to generate an OTP. The three alternative devices and authentication solutions examined in the experiment are displayed in Table 1.

In terms of security, simple push-button tokens are considered inferior to smartcard and reader solutions, as smart cards ensure tamper resistance (Hiltgen et al., 2006). In addition, these simple tokens cannot be extended to include transaction inputs or challenge/response functions. The addition of the PIN allows the passcode generation to be further protected. In terms of the practical point of view, card readers can be generic, whereas unsecured tokens must be registered to the users and provided by the Banks directly. Card readers therefore offer a scalable solution using a single reader with different cards for different authentication purposes.

Table 1 – Comparison of hardware authentication devices examined in the experiment.

Method	Method of OTP Generation	Type of 2-Factor Solution
Push-Button Token (A)	User presses button on the device and a 6-digit access code is displayed	Know (Password) + Possess (Device)
Card-Activated Token (B)	User inserts their Bankcard into the card reader, presses button and a 6-digit access code is displayed	Know (Password) + Possess (Device) + Possess (Bankcard)
Chip and PIN-Secured Token (C)	User inserts their Bankcard into the card reader, enters card PIN and a 6-digit access code is displayed	Know (Password) + Possess (Device) + Possess (Bankcard) + Know (Card PIN)

The devices selected were all small and portable, as similar as possible to control extraneous variables, employing incremental security layers in the generation of the OTP. The authentication process in each case proceeded thus: Customer identification involved input of a User ID and was followed by provision of the password authentication (first factor), the second factor consisted of the input of an OTP generated by one of the three devices.

Fig. 1a and b illustrate the different devices and Fig. 1c shows a close up of the scroll wheel on the card-reader device.

Participants were presented with instructions on the access code demand page of the login process when they first used each device. This help page was also available via a 'help' text and device picture hyperlink at each subsequent prompt requesting an OTP. These instruction pages are shown in Figs. 2–4. The card-activated device was given to participating customers accompanied by a mock-up bankcard for the fictitious persona they were using to login to the eBanking session. The PIN-secured device was accompanied with a mock-up bankcard and a printed sheet showing the non-sequential PIN number (1458, used in every case) for use in the experiment. Participants were presented with one device for each method, with different colour devices being used in each alternate experiment room, assigned randomly.

The Bank's brand was only present on 'their' Bankcard (for two of the three conditions) and on the eBanking Website. The Website was only changed to allow for different device pictures and instructions to populate the OTP authentication steps. Thus each device presented required an incremental number of steps to obtain the OTP, whilst other variations were minimised.

The card-reader token is an example of a multipurpose device, and as such had several modes - two of which were OTP generation, both were explored in the experiment as methods B and C.

2.2. *Research questions and hypotheses*

The research aimed to investigate perceptions of usability, security and convenience of different methods of generating OTPs. In order to maximise exposure to the three alternate devices, each device was used twice in each experience in the scenario of performing the login and a transaction task. As an addendum to the experiment, learning effects were also considered. Participants were asked to use their preferred device again in a final set of tasks.

The null hypotheses tested were:

Hypothesis $H1_0$: The three methods would not differ in terms of usability.
Hypothesis $H2_0$: The three methods would not differ in terms of perceived quality, security and convenience ratings.

Fig. 1 – a. The push-button token. b. The card-activated and PIN-secured token (with Bankcard). c. The scroll wheel PIN-entry mechanism.

Hypothesis $H3_0$: The three methods would not differ in terms of preference rankings.
Hypothesis $H4_0$: The three methods would not differ in terms of time taken to complete the OTP input at login and confirmation.
Hypothesis $H5_0$: The three methods would not differ in terms of use of the help facility.
Hypothesis $H6_0$: In the final set of tasks that perceived usability scores, time to login and help usage for a participants' preferred method would not differ from their first use of that device.

2.3. *Participants*

Participants were recruited as customers of the sponsoring Bank who currently use their eBanking service. They were balanced by age (35 years and under/36 years and over) and gender. Fifty customers took part in the experiment.

2.4. *Tasks*

Tasks in each of the first-use conditions were to log on to eBanking, find the current account balance then perform a simple transaction. The tasks were designed to be quick and simple to complete. As such the three transaction tasks selected were:

- Deleting a Direct Debit (DD)
- Deleting a Standing Order (STO)
- Deleting a future payment

The transactions were varied in each experience to engage the participant in the tasks. The transaction tasks all required selecting the task item from the correct page and clicking the 'Delete' button. They were presented with the confirmation page and a prompt 'enter a new access code' with one of the following instructions:

- ''Press the button on your device'' [Device A]
- ''Insert card, select 'CODE' and press the left button'' [Device B]
- ''Insert card, select 'ECODE' and enter your PIN'' [Device C]

Next to the prompt a device picture and a text link both gave access to a page containing the instructions. Participants were also asked to log off to complete the experience.

In order that each device be used correctly, twice, by each participant in their session, a three-strikes policy was incorporated manually. If the participant failed to login or confirm after three attempts the researcher stepped in to ask them to read the instruction page. This modelled a potential help call and device reset policy, with the participant being trained to use the device. Whenever this occurred it was logged as assisted completion as this scenario would be undesirable in practice.

The final task sheet (fourth condition) required similarly to login and perform another simple transaction - to modify the details of a named Standing Order, then log off. This involved slightly more work than the straightforward deletion tasks but offered another different task to avoid habituation within the experimental task. However, this meant that only the login task timings could be compared between first use and preferred method reuse in the fourth condition.

2.5. *Usability measurements*

Efficiency was measured at two stages of the experiment, timing (automatically) both instances of the authentication steps requiring OTP generation. This offered one result for the login step (1st use) and a second for the transaction authentication step (2nd use).

Fig. 2 – The instruction page for the push-button token.

Effectiveness was measured in terms of task completion and use of the help pages was automatically logged.

Satisfaction was measured using attitude questionnaires, posed after each experience. The attitude questionnaire focused on usability issues based on previous work (Weir et al., 2006, 2007) and employed a 7-point Likert scale. There is much discussion in psychology of the benefits (or not) of using a neutral point on the scale. The 7-point scale is advocated as it is the optimum scale for reliability; in addition it allows a better approximation to true scalar data for parametric analysis using the ANOVA (Nunnally, 1978; Kline, 2000). The questionnaire is shown in the Appendix.

Several items were introduced to the usability statement set for this experiment. They were designed to gather data about participants perceptions of the different devices: whether it felt personal (where with and without bankcard and/or PIN may result in different ownership attitudes towards the methods); whether methods were convenient (similarly, as each method required incrementally more steps); attitudes towards the need for instructions; perceptions of security; and four questions about the cognitive and interactive qualities of the devices and details needed in each case.

The questionnaire was specifically about the device and process of authentication and was introduced by the facilitator: ''For this Questionnaire I'd like you to think about the device you just used to log on and confirm your banking transaction. Please state to what extent you agree or disagree with the following statements''.

2.6. *Other measurements*

After the three experiences a short interview captured reactions to the different authentication methods, likes, dislikes and suggestions for improvements. Participants were asked about their preferences by rating the three methods overall on the 0–30 point scale labelled 'worst' to 'best', with the results analysed as a quality rating and as a rank order of preference. Comparative ratings were also taken in terms of security and convenience, comparing each device used on the same 0–30 scale.

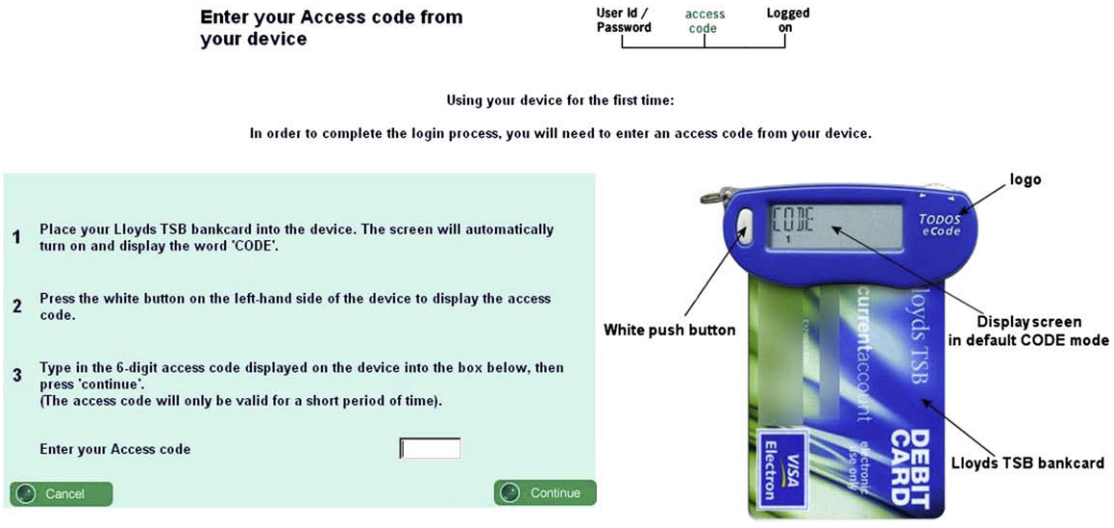

Fig. 3 – The instruction page for the card-activated token.

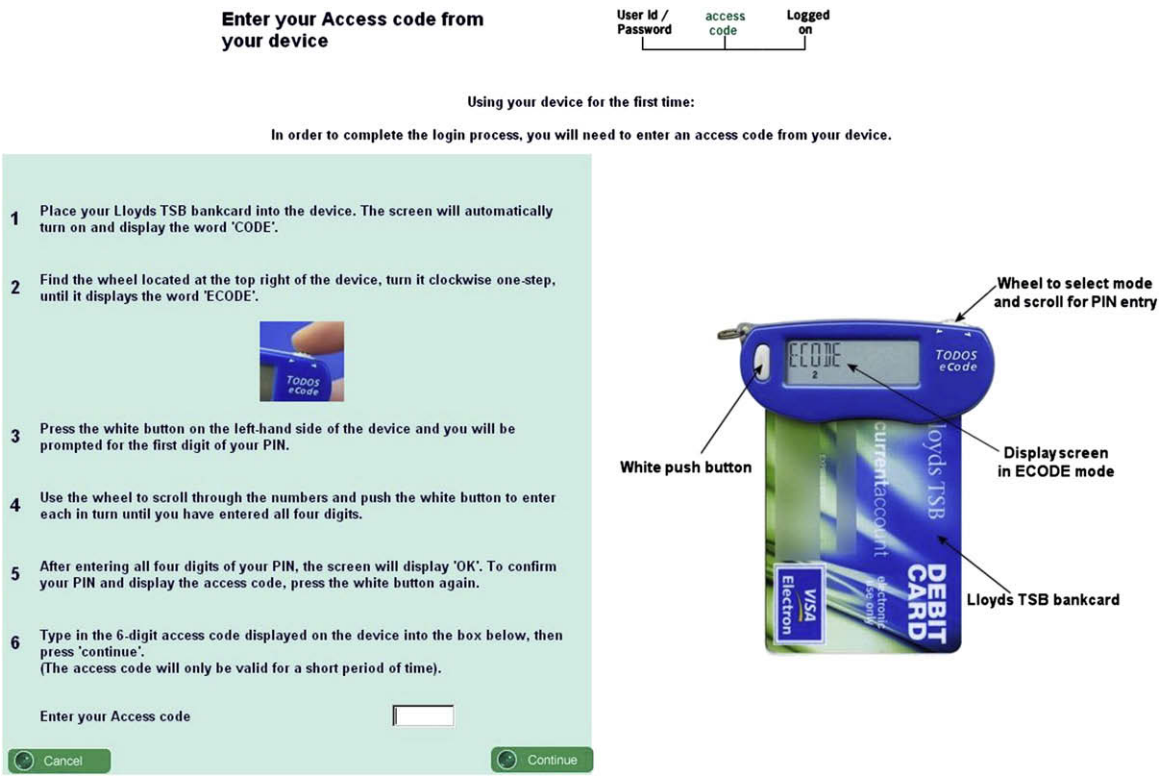

Fig. 4 – The instruction page for the PIN-secured token.

The fourth condition allowed participants another experience with their preferred method. The usability questionnaire was repeated after this experience with a few modifications: three questions (q1, q2 and q10) about the instructions were removed (as they were not shown by default to all participants), and replaced by one statement about device memorability, ''It was hard to remember what to do with this device''. This allowed some comparisons to be made in terms of device learnability - however there was no control over how many participants used each device in this condition.

2.7. *Experiment design summary*

Design: Three cell, repeated-measures, within-subjects.
Independent variables: Option A (Push-button Token); Option B (Card-Activated Token); Option C (Card and PIN-Activated Token).
Participant independent variables: Age (2 levels); Gender (2 levels); Order of experience (6 levels).
Dependent variables: Attitude questionnaires; Task completion counts; Quality ratings & preferences, Authentication step timings, help use.
Confounding variables: Researcher Bias (randomised); Room (randomised), Persona details and Tasks (balanced across designs), Device colour (randomised).
Other data: Researcher observations; Interview questions.
Sample size: 6 orders × 2 genders × 2 age groups × oversampling 2 = 48.
Honorarium: £30.
Session time: 1 hour.

After the experiment, participants were asked to reuse their preferred device to complete another set of tasks, with timing, attitude questionnaires and comments gathered about learnability.

3. Results

3.1. *Demographics*

The sample of 50 eBanking customers consisted of 25 participants in each of the age groups, half 35 years old or less, half 36 or older. They were also well balanced for gender, with 26 (52%) female. Participants used eBanking at home (46, 92%), at work (16, 32%) and elsewhere - at university or internet cafes (2, 4%). They logged on most often from home (39, 78%).

3.2. *Usability (efficiency) results*

Timings for the OTP generation and input steps reflect efficiency as a part of usability. Both the login task time and the subsequent transaction confirmation times were recorded, as shown in Table 2.

After eliminating order and colour (no effects found), a repeated-measures ANOVA with age and gender as the between-subject factors revealed a significant difference between the three methods in terms of efficiency $F = 126.1$; $df = 1.167$; $p < 0.001$ (Greenhouse-Geisser correction to the degrees of freedom, Mauchly's Test $p < 0.0001$). Post hoc pairwise comparisons (Bonferroni) indicate that there were

Table 2 – Efficiency (timing) results for the three methods.

Results	Push-button Token	Card-activated Token	PIN-secured Token
Mean Login Time (s)	25.20	36.70	117.8
Standard Deviation	10.76	20.09	52.33
Mean Confirmation Time (s)	32.00	40.87	79.67
Standard Deviation	18.58	24.67	44.23

significant differences between each pair of methods, $p < 0.0001$. PIN-secured OTP generation took significantly longer than either other method; and card-activation took significantly longer than the unsecured push-button OTP generation. There was a between-subjects age effect that approached significance, $p = 0.058$, indicating that older participants tended to take longer to Login than their younger counterparts using all three of the methods.

Similar analysis was performed for the transaction confirmation time. There were no significant differences in the task times for the three different tasks which had been carefully balanced across the three methods. Therefore the different transaction results were combined to compare between methods, as this was the experimental variable under scrutiny.

Again order and colour could be eliminated having no significant effects. The ANOVA examining age and gender as the between-subject factors indicates a very highly significant difference between the three methods for task confirmation time, $F = 49.162$; $df = 1.427$; $p < 0.001$ (Greenhouse-Geisser correction; Mauchly's Test $p < 0.0001$). Pairwise comparisons indicated that the push-button token took significantly less time than the card-activated token, $p = 0.020$ and the PIN-secured token, $p < 0.0001$. The card-activated token took significantly longer than the PIN-secured token, $p < 0.0001$.

Older participants took significantly longer than younger participants to complete task confirmation with all three methods, $p < 0.0001$.

3.3. *Usability (satisfaction) results*

The mean usability scores derived from usability questionnaires are shown in Table 3.

Order, colour and researcher were found not to have any significant effect on the data and were removed from further analysis. Using age and gender as the between-subject factors, there was a very highly significant difference between the three mean usability scores, $F = 81.040$; $df = 2$; $p < 0.0001$ for the three methods and no within-subject or between-subject effects. Pairwise comparisons showed significant differences between all three methods: Push-button token vs. Card-activated token, $p = 0.002$; Push-button token vs. PIN-secured token, $p < 0.0001$; Card-activated token vs. PIN-secured token, $p < 0.0001$.

Individual question responses for the three methods are shown in Table 4.

Repeated-measures ANOVA's were also used to compare each individual question by method with age and gender as the between-subject factors. The results of the ANOVA are shown in Table 5. Bonferroni pairwise comparisons indicating significant differences between the three methods and the direction of differences are included in the table.

The push-button token was characterised by positive attitudes (above 5 on the 7-point scale) towards usability for 28 (93%) out of 30 attributes. Participants gave lowest scores towards the need for instructions and personal ownership. The highest scoring attribute related to the speed of use.

The card-activated token was characterised by positive attitudes towards usability for 22 (73%) out of 30 attributes. Participants gave lowest scores towards knowing what to do next, matching expectations, degree of enjoyment, the need for instructions, and how much concentration was required to use the device. The highest scoring attributes related to the reliability, trust and security of the device.

Table 3 – Mean usability results for the three methods.

Method	Mean	Standard Deviation	N
Push-button token (A)	5.45	0.83	50
Card-activated token (B)	4.97	0.94	50
Chip and PIN-secured token (C)	3.74	1.01	50

Table 4 – Mean individual usability attribute scores for the three methods.

Attribute/Method	Push-button Token	Card-activated Token	PIN-secured Token
Concentration (Instructions)	5.26	5.00	2.78
Confusing Layout (Instructions)	5.60	5.06	3.76
Frustration	5.78	5.10	3.52
Flustered	5.64	4.90	3.50
Stress	5.74	5.04	3.58
Device Too Complicated	5.80	5.02	2.96
Degree of Control	5.42	4.96	3.72
Knew What To Do Next	5.42	4.08	2.88
Matched Expectations	5.00	4.36	3.52
Easy to Understand Instructions	5.88	5.42	4.06
Helpfulness	5.36	4.94	3.86
Size of Display	5.12	5.44	5.24
Ease of Reading Display	5.24	5.50	5.36
Would Use Device Again	5.70	5.24	3.78
Reliability	5.42	5.52	5.06
Speed of Using Device	6.02	5.22	2.92
Degree of Trust	5.02	5.24	5.16
Improvement Needed	5.04	4.68	3.06
User-friendliness	5.76	5.08	3.36
Attitude Towards Using	5.32	4.72	3.12
Appearance of Device	5.10	5.00	4.32
Degree of Enjoyment	5.34	4.48	3.22
Degree of Security	5.28	5.32	5.24
Personal Ownership	4.92	5.10	4.82
Need for Instructions	4.28	3.34	2.22
Degree of Convenience	5.76	4.76	3.46
Difficulty Keying In Details	5.72	5.52	4.14
Degree of Concentration using Device	5.46	4.64	2.72
Retrieving Details Straightforward	6.10	5.34	3.64
Ease Obtaining Code	6.06	5.10	3.16

Table 5 – Usability attributes means, pairwise comparisons by method, age & gender.

Attribute	A	B	C	p	B-S	Paired	p
Concentration/Instructions	5.26	5.00	2.78	<.001	–	A = B	>.05
						A > C	<.001
						B > C	<.001
Layout/Instructions	5.60	5.06	3.76	<.001	–	A > B	.041
						A > C	<.001
						B > C	<.001
Frustration	5.78	5.10	3.52	<.001	A/G $p = .024$	A > B	.002
						A > C	<.001
						B > C	<.001
Flustered	5.64	4.90	3.50	<.001	A/G $p = .037$	A > B	.004
						A > C	<.001
						B > C	<.001
Stressfulness	5.74	5.04	3.58	<.001	–	A > B	.001
						A > C	<.001
						B > C	<.001
Complication	5.80	5.02	2.96	<.001	A/G $p = .005$	A > B	.002
						A > C	<.001
						B > C	<.001
Control	5.42	4.96	3.72	<.001	G $p = .012$	A = B	>.05
						A > C	<.001
						B > C	<.001
Knew What To Do Next	5.42	4.08	2.88	<.001	–	A > B	<.001
						A > C	<.001
						B > C	<.001
Matched Expectations	5.00	4.36	3.52	<.001	A $p = .038$	A > B	.024
						A > C	<.001
						B > C	.010
Understand Instructions	5.88	5.42	4.06	<.001	–	A > B	.014
						A > C	<.001
						B > C	<.001
Helpfulness	5.36	4.94	3.86	<.001	–	A > B	.036
						A > C	<.001
						B > C	<.001
Size of Display	5.12	5.44	5.24	>.05	–	–	–
Readability of Display	5.24	5.50	5.36	>.05	–	–	–
Would Use Again	5.70	5.24	3.78	<.001	A/G $p = .027$	A = B	>.05
						A > C	<.001
						B > C	<.001
Device Reliability	5.42	5.52	5.06	>.05	–	–	–
Speed of Using	6.02	5.22	2.92	<.001	A/G $p = .030$	A > B	<.001
						A > C	<.001
						B > C	<.001
Trustworthiness	5.02	5.24	5.16	>.05	–	–	–
Improvement Needed	5.04	4.68	3.06	<.001	A $p = .043$	A = B	>.05
						A > C	<.001
						B > C	<.001
User-friendliness	5.76	5.08	3.36	<.001	A/G $p = .016$	A > B	.006
						A > C	<.001
						B > C	<.001
Attitude Using	5.32	4.72	3.12	<.001	G $p = .036$	A > B	.019
					A/G $p = .032$	A > C	<.001
						B > C	<.001
Appearance	5.10	5.00	4.32	.001	–	A = B	>.05
						A > C	.003
						B > C	.011
Enjoyment	5.34	4.48	3.22	<.001	–	A > B	.004
						A > C	<.001
						B > C	<.001
Security	5.28	5.32	5.24	>.05	–	–	–
Personal	4.92	5.10	4.82	>.05	–	–	–
Needed Instructions	4.28	3.34	2.22	<.001	–	A > B	.006
						A > C	<.001
						B > C	<.001

Table 5 (*continued*)

Attribute	A	B	C	*p*	B-S	Paired	*p*
Convenience	5.76	4.76	3.46	<.001	–	A > B	<.001
						A > C	<.001
						B > C	<.001
Keying in Details	5.72	5.52	4.14	<.001	–	A = B	>.05
						A > C	<.001
						B > C	<.001
Concentration	5.46	4.64	2.72	<.001	–	A > B	.001
						A > C	<.001
						B > C	<.001
Retrieving Details	6.10	5.34	3.64	<.001	–	A > B	.001
						A > C	<.001
						B > C	<.001
Obtaining Code	6.06	5.10	3.16	<.001	–	A > B	<.001
						A > C	<.001
						B > C	<.001

Notes: No Within-subjects interactions with the main effect were revealed in the Repeated-Measures ANOVA, B-S: Between-subject effects in the Repeated-Measures ANOVA, Paired: Post Hoc Bonferroni pairwise comparisons, A: Age, G: Gender, A/G: Age*Gender.

The chip and PIN-secured token was characterised by positive attitudes towards usability for only 6 (20%) out of 30 attributes. Participants gave lowest scores towards the amount of concentration required for reading instructions, how complicated the device was, knowing what to do next, speed of using the device, needing instructions and the degree of concentration required to use the device. The highest scoring attributes related to the reliability, trust and security of the device. The size and legibility of the display was also perceived positively.

The reliability of the usability attitude questionnaire, Cronbach's alpha, was measured. For the 30 item questionnaire alpha = 0.970 and the standardised item alpha = 0.971. These values indicate that reliability was good, alpha values with items deleted indicated that all items in the statement set contributed positively to the questionnaire.

3.4. *Usability (effectiveness) results*

All participants completed the two tasks with each device. Some errors were noted for the different devices. Typically, most other errors were typographic (i.e. in keying in the User ID number) and easily recovered by participants.

No errors were observed in the use of the push-button activated OTP device [A]. Participants were observed to complete the tasks quickly and easily. A few needed to re-orient the device. The graphical logo on the device (and the absence of text) made it a little difficult to know which way up to look at the screen.

With the card-activated token, few problems or errors were observed, but two participants made the error of failing to place the card in the slot before pressing the button. In both cases this was corrected by the participant referring themselves to the on-screen instructions, no assistance was required.

More errors were made with the PIN-secured token: Nine participants required assistance to complete authentication. Half of these problems were that participants entered the card PIN into the form instead of using it to get a code from the device. Others were seen to use the default CODE mode without PIN entry (the resulting code was not accepted due to the capture of the first two digits). Facilitators noted that several participants complained about the PIN-entry mechanism, and disliked having to read instructions to get the correct OTP.

The majority did not read or follow the instructions straight away. The main confusion in the use of this device was in entering the PIN; some kept pressing the push-button, not realising that the PIN number was in the process of being entered – they were often at the second digit before they realised what was happening. Nobody noticed that the digits could be deleted (backspaced). Instead of trying to make corrections, participants were often noted to remove the card which reset the device so they could start again.

Reluctance or annoyance at having to fall back on instructions was noted often. The length of the instruction page put some off before they tried to use the device. Lengthy instructions were generally needed in order to generate the PIN-secured OTP although all but 2 participants (96%) were able to use the device a second time without referring to the instructions.

Analysis of number of clicks logged to the help page for the three methods during task confirmation showed that no help was used for the push-button or card-activated token. Only 2 participants (4%) clicked on Help when using the PIN-secured token.

3.5. *Comparative ratings and preferences*

The mean quality ratings as a measure of overall preference, followed by the ratings for the three methods of authentication in terms of convenience and security specifically were recorded. The results are shown in Table 6.

Again, no effect was found for order of experience or colour of device, so these variables were removed from the rest of the analysis. The ratings were analysed using age and gender as the between-subjects factors in the repeated-measures ANOVA. The quality ratings for the three methods were found

Table 6 – Rating results for the three methods.

Rating Scale	Push-button Token	Card-activated Token	PIN-secured Token
Quality	21.40	17.32	10.88
Convenience	24.94	15.95	8.44
Security	14.65	18.94	22.52

to be significantly different, $F = 25.5$; $df = 1.45$; $p < 0.0001$ (Greenhouse-Geisser correction; Mauchly's Test $p < 0.0001$). There were no significant within-subject effects, but a highly significant between-subjects gender effect, $p = 0.003$. The data indicates that male participants gave significantly higher overall rating scores than their female counterparts. Pairwise comparisons showed significant differences between: Push-button token and Card-activated token, $p = 0.016$; Card-activated token and PIN-Secured token, $p < 0.0001$; Push-button token and PIN-Secured token, $p < 0.0001$.

These overall quality ratings were also analysed as a rank order of preference. There was a significant difference in the rank orders of preference for the three methods, $p < 0.001$. Pairwise comparisons indicate that the PIN-secured token was ranked significantly below both the other methods, $p < 0.001$ for both comparisons. There were no effects on the method preference by age or gender.

Perceived convenience ratings were analysed in the same way as the quality results, again there were no effects for order of experience or colour, so age and gender were used in the final analysis. The convenience ratings for the three methods were found to be significantly different, $F = 141.26$; $df = 1.56$; $p < 0.0001$ (Greenhouse-Geisser correction; Mauchly's Test $p = 0.001$). There were no significant within-subject effects. There was a significant between-subjects gender effect, $p = 0.035$, with female participants giving all three methods slightly lower overall convenience ratings than their male counterparts. Pairwise comparisons showed significant differences between all three methods: Push-button token and Card-activated token, $p < 0.0001$; Card-activated token and PIN-Secured token, $p < 0.0001$; Push-button token and PIN-Secured token, $p < 0.0001$.

The perceived security ratings were analysed in the same way as the other rating data, first eliminating order and colour. The security ratings for the three methods were found to be significantly different, $F = 21.84$; $df = 1.59$; $p < 0.0001$ (Greenhouse-Geisser correction; Mauchly's Test $p = 0.002$). There were no significant within-subject or between-subject effects for age or gender. Pairwise comparisons indicated significant differences between the three methods: Push-button token and Card-activated token, $p < 0.0001$; Card-activated token and PIN-Secured token, $p = 0.010$; Push-button token and PIN-Secured token, $p < 0.0001$.

3.6. *Relationship between usability and security*

Comparing the four different subjective scores (the usability questionnaire mean and the perceived preferences in terms of overall quality, convenience and security) on a standardised scale, where each is a fraction of the maximum value obtainable on that scale, the relationships between the different aspects of attitude can be examined, see Fig. 5.

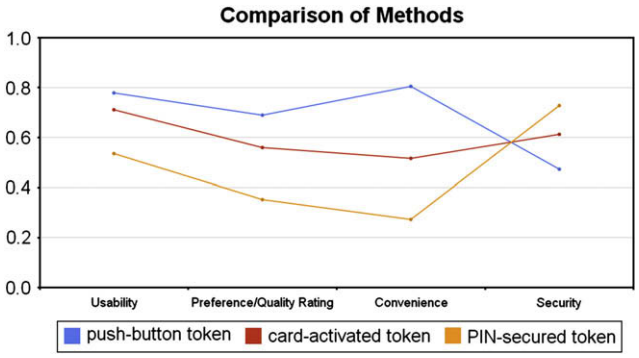

Fig. 5 – Usability, preference, convenience, security ratings for the three devices.

Pearson correlations were also computed for the three devices, between each pair of scores as shown in Table 7.

Usability and preference were very highly correlated for the push-button token and PIN-secured token, less significantly so for the card-activated token, although still significant. Convenience and preference were very highly correlated in all three cases. Convenience and perceived security were not at all related for the push-button or card-activated methods. For the PIN-secured method, this relationship was weak but positive (significant at $p = 0.05$) but usability scores in this case did not correlate with security.

This confirms that the alternative devices and corresponding number of steps in obtaining OTPs were perceived quite differently in terms of convenience. Usability was high for the push-button method, and preference and convenience ratings were also highly positive. Perceived security ratings were not as positive, actually just below 50% on the scale. For the PIN-secured method, usability scores were low, similarly for preference and convenience, yet perceived security levels were rated as the highest.

The perceived security, convenience and usability relationships were not stable across all three designs. This is a complex picture that will require further study with security processes to gauge how these concepts can be combined to provide the best compromise for the user and the Bank. Experimenting with biometric methods, passwords and alternate methods of OTP generation would be helpful now to explore these relationships as experienced through different authentication approaches.

3.7. *Reuse condition*

The numbers selecting each of the three methods to reuse (their preference) is shown in Table 8. Most participants chose the push-button token. One person failed to respond to the final usability questionnaire; therefore the results are based on the remaining forty-nine participants.

After reuse, usability perceptions were compared to first use for the mean of the matched 27 usability questions, and similarly for each individual item, login times and help use. For the push-button token, there were substantial numbers available for the analysis and some between-subject factors, such as age and gender were included in the analysis. For the

Table 7 – Correlations between measures of usability, preference, security and convenience for the three methods.

	Usability Score	Quality Rating	Convenience	Security Rating
Push-button token				
Usability	1	0.516, 0.266, <.001	0.432, 0.186, .002	0.467, 0.218, .001
Quality Rating	0.516, 0.266, <.001	1	0.461, 0.212, .001	0.467, 0.218, .001
Convenience	0.432, 0.186, .002	0.461, 0.212, .001	1	0.160, 0.026 (NS)
Security	0.467, 0.218, .001	0.467, 0.218, .001	0.160, 0.026 (NS)	1
Card-activated token				
Usability	1	0. 376, 0.141, .007	0. 334, 0.111, .018	0. 338, 0.115, .019
Quality Rating	0.376, 0.141, .007	1	0.704, 0.495, <.001	0.310, 0.096, .032
Convenience	0.334, 0.111, .018	0.704, 0.495, <.001	1	0.213, 0.053 (NS)
Security	0.338, 0.115, .019	0.310, 0.096, .032	0.213, 0.053 (NS)	1
PIN-secured token				
Usability	1	0.502, 0.141, <.001	0.455, 0.186, .001	0.194, 0.218 (NS)
Quality Rating	0.502, 0.266, <.001	1	0.686, 0.212, <.001	0.480, 0.218, .001
Convenience	0.455, 0.186, .001	0.686, 0.212, <.001	1	0.341, 0.026, .018
Security	0.194, 0.218 (NS)	0.480, 0.218, .001	0.341, 0.026, .018	1

Notes: Displayed Data: Pearsons' r, R^2, p; NS: not significant at $p = .05$.

other two methods the analysis was carried out without the between-subject factors due to the small sample sizes.

Usability attitudes were found to be more favourable after preferred method reuse.

3.7.1. *The push-button token*

Significant differences were found in attitudes to several individual items in the questionnaire, all with attitudes becoming more favourable after preferred device reuse. The time taken to complete the OTP step at login was found to be significantly different during reuse and age group had a significant effect, $p < .001$. Older participants took significantly longer than the younger group. No use of the help link in obtaining an OTP was logged. The results are summarised in Table 9.

3.7.2. *The card-activated token*

Significant differences were found in attitudes to nine individual items in the questionnaire, all with attitudes becoming more favourable after preferred device reuse. The time taken to complete the OTP step at login was found to be significantly different during reuse. One person used the help link in obtaining an OTP during login. The results are summarised in Table 10.

3.7.3. *The chip and PIN-secured token*

Significant differences were found in attitudes to nine individual items in the questionnaire, all with attitudes becoming more favourable after preferred device reuse. The time taken to complete the OTP step at login was found to be significantly different during reuse. No use of the help link in obtaining an OTP was logged for this method. The results are summarised in Table 11.

Table 8 – Mean usability scores and significant differences in reuse of preferred method.

Preferred Method	N	Initial Usability	Reuse Usability	Significance (p)
Push-button token	31	5.53	5.70	.036
Card-activated token	9	4.36	5.42	.024
PIN-secured token	9	3.49	5.41	<.001

3.7.4. *Memorability*

The ease of remembering what to do with the method was analysed with device-type as the between-subjects factor in a univariate ANOVA. The results are shown in Table 12.

Participants found their preferred method easy to remember how to use, no matter which device they selected. There were also no effects for age, gender or colour of device. All three devices were found to be equally memorable, one of many attributes that affects overall usability (Nielsen, 1993).

Participants made some comments during their final experience with their preferred method. Comments about the push-button token focused on its ease of use, ''I started using this [device] without looking at the instructions so that's quite good, it must be easy''. The card-activated token was appreciated from a different perspective, ''This [device] is more gadgety, which I like.'' For the chip and PIN-secured token, the comments were more focused on security, ''This [device] is more time-consuming, but more secure.''

3.8. *Qualitative*

Participants made comments on their likes, dislikes and suggestions. Some typical responses are presented here:

Table 9 – Scores for significantly different measures in reuse of push-button token.

Usability Metric	Initial Usability	Reuse Usability	Significance (p)
Control	5.48	5.84	.045
Reliability	5.52	6.03	.009
Trustworthy	5.23	5.81	.051*
Attitude towards using	5.39	5.84	.029
Security	5.29	5.84	.045
Time to Login (seconds)	26.7	13.9	<.001

Notes: * Not significant at $p < .05$, but a marginal result.

Table 10 – Scores for significantly different measures in reuse of card-activated token.

Usability Metric	Initial Usability	Reuse Usability	Significance (p)
Flustered	4.33	5.78	.012
Stressfulness	4.78	5.78	.053*
Knew what to do next	3.33	5.56	.005
Matched expectations	3.44	4.89	.038
Use again	3.89	5.44	.033
Improvement needed	3.67	5.22	.038
Attitude towards using	3.56	5.22	.020
Security	4.89	5.67	.023
Instructions needed	3.44	4.78	.050*
Convenience	3.56	5.22	.020
Enjoyment	3.44	5.67	.010
Time (login)	39.9	21.4	.005

Notes: * Not significant at $p < .05$, but a marginal result.

3.8.1. *The push-button token*

The device was often praised for convenience, simplicity and size - being small and easy to carry. The process of obtaining and using the OTP was very straightforward.

The main dislikes and suggestions made in reference to this method included comments that the screen was too small, doubts about the level of security offered and the inconvenience of having something to carry around, with a single purpose and which could be lost.

3.8.2. *The card-activated token*

This device was often praised for speed and security. People felt that the method was more personal as the device made use of their own card. It was frequently mentioned that the process was not difficult, and that the device size and screen were easier to read than the smaller push-button token.

The main dislikes and suggestions were that it was a bit more fiddly and bulkier to carry around. Some commented that it was a fairly lengthy process taking longer than they would like to spend.

Table 11 – Scores for significantly different attributes in reuse of PIN-secured token.

Usability Metric	Initial Usability	Reuse Usability	Significance (p)
Frustration	3.33	5.89	.001
Flustered	3.67	5.67	.015
Stressfulness	3.78	5.67	.020
Complication	3.00	5.56	.002
Control	4.00	5.56	.033
Knew what to do next	3.22	5.33	.004
Matched expectations	3.11	5.00	.012
Helpful	3.33	5.00	.017
Use again	3.22	6.22	<.001
Quick	2.33	5.78	<.001
Improvement needed	2.56	4.67	.002
User-friendliness	3.11	5.78	.010
Attitude towards using	2.78	5.56	.002
Appearance	3.11	5.56	.010
Enjoyment	2.89	5.56	.004
Instructions needed	1.67	5.44	<.001
Convenience	2.89	6.00	.002
Keying in	3.89	5.78	.015
Concentration	3.00	5.44	.002
Retrieval of details	3.78	5.67	.037
Ease obtaining code	3.11	6.11	<.001
Time (login)	105.7	36.5	<.001

Table 12 – Ease of remembering, by method used, final (reuse) experience.

Preferred Method	N	Mean Score	Standard Deviation
Push-button token	31	5.61	1.308
Card-activated token	9	5.78	0.833
PIN-secured token	9	5.33	1.936

3.8.3. *The PIN-secured token*

This version of the device was offered far fewer likes, but some positive features noted were the better security putting people at ease with their use of Online banking. The fact that it was much harder for anyone else to use, making it feel more personal.

However, there were many dislikes and suggestions for improvements to this device method, considered to be ''over the top'' in terms of security, a hassle, taking too much time, confusing and hard to remember all the steps. People described the device as laborious. They particularly mentioned that it would become tedious if required for all their transactions. The wheel as a method for PIN entry was considered too fiddly and complicated. A few mentioned that a larger device which always sat by the computer would be better, no need to carry it around (although this would not suit all customers).

3.9. *Summary of data analysis*

Hypothesis $H1_0$: The three methods would not differ in terms of perceived usability. There was sufficient evidence to reject this hypothesis. There was a significant differences in mean usability (satisfaction) scores for the different authentication methods, both $p < 0.001$. Usability scores for the push-button token were significantly higher than the other two methods, $p < 0.001$, and that the card-activated token was significantly higher than the PIN-secured token.

Hypothesis $H2_0$: The three methods would not differ in terms of perceived quality, security and convenience ratings. There was sufficient evidence to reject this hypothesis. There was a significant main effect for authentication method on quality (preference) ratings in the same pattern as the usability scores, only with a less strong result (still significant $p = 0.016$) between the push-button token and card-activated token methods. Convenience ratings showed the same significant differences as the usability scores, with the same pairwise comparisons. Perceived security ratings were also significantly different for the three authentication methods, but showed the opposite pattern to other data, the PIN-secured token rated highest, the push-button token method lowest.

Hypothesis $H3_0$: The three methods would not differ in terms of preference rankings. There was sufficient evidence to

reject this hypothesis. There was a significant bias towards preference of the push-button token from the three methods.

Hypothesis $H4_0$: The three methods would not differ in terms of time taken to complete the OTP input at login and confirmation. There was sufficient evidence to reject this hypothesis. Each incremental step in the generation of an OTP caused slower authentication, with very lengthy times recorded for the PIN-secured token.

Hypothesis $H5_0$: The three methods would not differ in terms of use of the help facility. There was insufficient data to test this hypothesis. Help was not used at all in the second authentication task for the push-button or card-activated token methods. It was used to refer to in using the PIN-secured method in a second task, but infrequently, providing too small a sample (2 participants, 4% of the cohort).

Hypothesis $H6_0$: In the final set of tasks that perceived usability scores, time to login and help usage for a participants' preferred method would not differ from their first use of that device. There was sufficient evidence to reject this hypothesis. In all cases, participants found their preferred device more usable in this experience, took less time to use and found it easy to remember what to do without needing help.

4. Discussion

Participants didn't comprehend the security of the OTP method in general, some security information in its introduction would be valuable in this regard.

The push-button token was most preferred and judged to be the most convenient of the three methods. In contrast, the perceived security ratings showed the converse, with participants relating the incremental steps used for the card-reader device to be analogous to increased security. However, participants rated all three devices as secure to use in the eBanking tasks.

All data indicated that customers concerns for security in eBanking do not override their desire for convenience and usability. Participants prefer simple methods which appear to offer adequate security from their point of view. Until their perceptions of threat increase to a point where they are more concerned about the eBanking authentication process, they are unlikely to welcome complex security mechanisms.

This study presents empirical evidence to illustrate the perceived security-usability trade off from the customers' viewpoint. Alternate devices with different interaction designs could be used to investigate this further. Designing intuitive methods for secure OTP generation could use the more common card-reader design, such as the calculator-style. This design should be more familiar as it matches Chip and PIN terminals in store and authentication at the ATM.

4.1. *Push-button token*

The push-button token was intuitive and quick to use. A few people were unsure how to orient the device. Text on the device could provide this cue, perhaps in the form of branding. The card-activated token was less intuitive. Participants did not like having to refer to instructions, similarly to the PIN-secured method. However, in this case the instructions were quite straightforward, few steps were required and the device was quickly mastered.

4.2. *PIN-secured token*

There were a variety of usability problems with the Chip and PIN method. The mean usability score for the device was in the lower half of the scale due to the range of attributes awarded low scores. Participants remarked on the awkwardness of the scroll wheel mechanism for PIN entry and mode selection on the device. It was akin to setting the time on a digital watch - time-consuming, requiring a series of scrolls and button pushes. The multiple modes on the device added a layer of complexity, making the device considerably more arduous and contributing to its low usability rating. Participants didn't feel they needed any of the differing modes. There was no indication that they had any knowledge about their potential use. Until 2-factor authentication becomes much more widely used in eCommerce, Banking and elsewhere, customers see little need for a multi-purpose device.

Observations indicated that although the scroll wheel allowed PIN entry on a small device, the operation required referring to the instructions onscreen, which was generally disliked. It is well established that people scan on the web. Participants have often commented that they disliked having to read instruction in eBanking (Weir et al., 2006). Choice of device should emphasise usability and intuitive operation rather than seeking very high perceived security if this requires a complex mechanism.

Nine (18%) participants had to be offered assistance to use PIN-secured token correctly. Assistance modelled a helpdesk call and device reset. No such assistance was required for the other two options. If such levels of help-desk calls are representative, this is likely to be unacceptable to the Bank.

4.3. *Timing*

Increasing number of steps required to obtain the OTP increased the time taken to complete the authentication steps with the devices, as would be expected, however this result was particularly apparent for older participants. Using the card and interacting with the devices seemed to be more confusing for older users. This may explain why they needed longer to complete the authentication steps than their younger counterparts.

Participants report subjective differences in efficiency (in the attitude questionnaire) which were reflected in the actual times logged. First and second use timing data taken here has offered some additional insight into usability and convenience in authentication; device use was found to be highly susceptible to practice effects, thus calling into question the use of efficiency metrics of usability for first-use experiences.

4.4. *Preferred device reuse*

In all cases, when using their preferred device again, participants rated the experience higher in terms of usability than their first use. Most participants reused the push-button token. This method should be highly acceptable to eBanking

customers as it was simple to use intuitively and remember for future use.

Fewer participants (9, 18% in each case) used the other methods again. They also reported very much higher usability during this experience. This suggests that the authentication methods were both easily learned. However, there may also be the effect of cognitive dissonance - which would predict a raising of attitudes to reflect a preference belief (Szajna and Scamell, 1993). Although it is not possible to comment on which may have had more effect, it would be interesting to explore whether higher attitudes can be achieved if a choice (amongst a small set of options) is offered to the customer.

Regardless of which method they preferred and reused, they found the methods easy to remember how to use. They made far fewer errors and were not likely to need the instructions (in the help pages).

Perceptions of efficiency and actual times also showed positive effects of practice in reuse. Higher usability scores and faster timings in reuse of the push-button token indicated learnability and good prospects for customer acceptance in the marketplace.

5. Conclusions

This experiment explored the relationship between perceptions of usability, preference, security and convenience in authentication devices. The push-button token was simple, convenient, considered secure in use for eBanking and was intuitive to pick up and use. The card-activated token, working in the default mode was quickly learned and fairly usable, convenient and secure. The chip and PIN method was less favourably regarded. The PIN-entry mechanism and need for instructions made it much slower to use – this was compounded by the fact that customers saw no obvious benefits to the additional steps and layers of security in obtaining the passcode.

Participants chose their preference following usability and convenience attitudes rather than what they perceived to be increasingly secure. In the attitude questionnaire following each method, they were all scored equally acceptable in terms of security in use.

While the majority of customers were not prepared to trade convenience and usability for what they considered increased security, a minority were willing. Nine (18%) chose each of the other methods as their preference. Participants found their preferred method much better in terms of usability and speed when using it a second time. This could be due to a variety of effects including selection bias, cognitive dissonance and practice effects. The further investigation of this effect could be turned to advantage. If offering a choice between a few different authentication methods can lead to compliance, this could be harnessed in the eBanking domain as well as other applications. As satisfaction is theorised to be a determinant of future adoption, any increased satisfaction associated with choice may be more important in the online self-service context, where use is voluntary. These relationships would be of great benefit to explore.

This investigation has enhanced the understanding of the trade off between requirements for high perceived security, convenience and usability in eBanking usage. At present customers see security as largely a concern of the Bank. Their preferences for authentication methods entirely followed usability and convenience concerns. Controlling for device security layers, limited desire was found for a highly secure 2-factor solution using a card reader and PIN-security for OTP delivery (at least for this device design). However, this was within three relatively high actual levels of security provided. The augmentation of a knowledge-based password method with a push-button token (possession) was rated as the most usable, convenient and preferred authentication method of the three. Further, for chip and PIN to be adopted, a much more usable device and PIN interaction mode needs to be developed. A potential compromise would be the use of a multipurpose device activated by individual cards for the different eCommerce and eBanking tasks requiring authentication steps, such as the APACS standard card-reader specification (APACS, 2008).

5.1. *Future work*

In future, more practice handling and using such devices within the experimental setting might provide more realistic data, reducing the potential strain and bias of first-time use. It would be of interest to examine how first-use usability and security perceptions change with more experience. Other areas of interest would be to examine the multi-purpose functionality of some card-reader devices, and obtain measures of their utility in the eBanking environment and beyond. Several devices such as these are available, including the APACS standard card readers (Reavley, 2005). Future work should include some of these alternatives.

More investigation of how perceptions of usability, security and convenience are related is still required. Differences in the usability of the methods in this experiment cannot be attributed entirely to the device security process, as there was also the requirement for mode and PIN input in one of the devices selected. The mechanism of a scroll wheel required concentration and dexterity in its use and practice appeared to be a major factor in perceptions of its usability. The initial difficulties in the experience of these small card-reader devices may put off future efforts. In addition, objective measures of timing were highly subject to practice effects, limiting the value of collecting such performance data in first-time usage sessions. Performance data collected by way of task completion and help requirements was more illuminating at this early stage.

Limitations to the laboratory approach were minimised by targeting users of the service as participants, offering them realistic experiences with alternative interfaces, using representative tasks in the context of a detailed scenario. However, further field based research would also be appropriate to study external factors in relation to usability, security and eBanking usage. To that end would encourage the development of methodologies which allow sensitive applications such as eBanking and aspects relating to security to be studied closely in the field without compromising participants' privacy.

5.2. *Recommendations to practitioners*

The empirical results, observations and reflections from this study point to the following recommendations for a usable, secure authentication process in eBanking:

1. OTP generating devices are a short term possibility.
   - Portable, small, single use tokens suit a wide range of users (although accessibility issues would need to be considered).
   - Card-activation techniques should be considered if security requirements dominate, these would make the most of a single device for multiple purposes.
   - Provide some details on the security of the device from the outset to inform customers and engage them in security issues.
2. Individual devices are not a long-term solution.
   - Highly portable Chip and PIN devices may be too complex. Currently, the devices are not easy or intuitive enough for a diverse user base.
   - eBanking customers do not see the need for 2-factor devices at present.
3. Offer a choice of device.
   - This research suggests a possible attitude boost can be achieved by offering a choice between two or three options.
4. Extend 2-factor to the transaction authentication process.
   - Increased use of 2-factor methods in eCommerce would create utility for multiple-mode tokens.
5. Investigate usability in device design.
   - Users have yet to be persuaded of the need for stronger authentication in eBanking tasks. Usability needs to be considered to ensure continued use of low-cost channels.

Appendix. Usability of security questionnaire

1. Reading the instructions for the device took a lot of concentration
2. I found the layout of the instructions for this device confusing
3. Using the device was very frustrating
4. I got flustered when using this device
5. I felt under stress while using the device
6. Using this device was too complicated
7. I felt in control when using this device
8. When using this device I didn't always know what to do next
9. This device did not match my expectations
10. The instructions for using this device were easy to understand
11. I felt that this device was helpful
12. The display on this device was too small
13. The display on this device was difficult to read
14. I would be happy to use this device again
15. I felt this device was reliable
16. Using this device was quick
17. I found this device trustworthy
18. I feel that this device needs a lot of improvement
19. I found this device 'user-friendly'
20. I liked using this device
21. I liked the appearance of this device
22. I did not enjoy using this device
23. Using this device felt secure
24. I thought this device was personal
25. I needed instructions to use this device
26. I thought this device was convenient
27. Keying in the details from the device was difficult
28. I had to concentrate hard to use the device
29. Retrieving the details I needed from this device was straightforward
30. Knowing how to get the code from the device was easy.

REFERENCES

APACS. Remote card authentication. Available from: http://www.apacs.org.uk/payments_industry/new_technology2.html; 2008 (accessed 22.07.08).

Besnard D, Arief B. Computer security impaired by legitimate users. Computers and Security 2004;23(3):253–64.

Bishop M. Psychological acceptability revisited. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 1–11 [chapter 1].

Eastin M. Diffusion of e-commerce: an analysis of the adoption of four e-commerce activities. Telematics and Informatics 2002; 19(3):251–67.

FFIEC. FFIEC releases guidance on authentication in internet banking environment, press release. Available from: http://www.ffiec.gov/press/pr101205.htm; 2005 (accessed 22.07.08).

Furnell S. Why users cannot use security. Computers and Security 2005;24:274–9.

Henry PA. Two-factor authentication - a look behind the headlines. Network Security 2006;2006(4):18–9.

Hiltgen A, Kramp T, Weigold T. Secure internet banking authentication. IEEE Security and Privacy 2006:21–9.

Hornbæk K. Current practice in measuring usability: challenges to usability studies and research. International Journal of Human-Computer Studies 2006;64(2):79–102.

Hornbæk K, Law EL. Meta-analysis of correlations among usability measures, CHI 2007; 2007. pp. 617–26.

International Organization for Standardization, ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs) part II: guidance on usability; 1998.

Ives B, Walsh KR, Schneider H. The domino effect of password reuse. Communications of the ACM 2004;47(4):75–8.

Jayawardhena C, Foley P. Changes in the banking sector – the case of internet banking in the UK. Journal Internet Research: Networking and Policy 2000;10(1):19–30.

Johnston J, Eloff JHP, Labuschagne L. Security and human computer interfaces. Computers and Security 2003;22(8): 675–84.

Karat CM, Brodie C, Karat J. Usability design and evaluation for privacy and security solutions. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 47–74 [chapter 4].

Kline P. The handbook of psychological testing. Routledge; 2000.

Knight W. The price of love. Infosecurity 2008;5(1):30–3.

Landauer TK. Research methods in human computer interaction. In: Helander M, editor. Handbook of human computer interaction. Amsterdam: NorthHolland; 1988. p. 905–28.

Morris R, Thompson K. Password security: a case history. Communications of the ACM 1979;22(11):594–7.

Nielsen, J. Usability Engineering. Academic Press; 1993.

Nilsson M, Adams A, Herd S. Building security and trust in online banking. In: Extended abstracts on human factors in computing systems (CHI '05). New York, NY: ACM Press; 2005. pp. 1701–04.
Nodder C. Users and trust: a microsoft case study. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 589–606 [chapter 29].
Nunnally JC. Psychometric theory. McGraw-Hill; 1978.
O'Gorman L. Comparing passwords, tokens and biometrics for user authentication. Proceedings of the IEEE 2003;91(12):2021–40.
Piazzalunga U, Savaneschi P, Coffetti P. The usability of security devices. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 221–42 [chapter 12].
Preece J, Rogers Y, Sharp H, Benyon D, Holland S, Carey T. Human-computer interaction. Wokingham, UK: Addison-Wesley; 1994.
Ranger S. Chip and PIN heads for cyberspace, silicon.com financial services news, CNET networks, UK. Available from: http://www.silicon.com/financialservices/0,3800010322,39152706,00.htm; 2005 (accessed 31.01.07).
Reavley N. Securing online banking. Card Technology Today 2005;17(10):12–3.
Renaud K. Evaluating authentication mechanisms. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 103–28 [chapter 6].
Root RW, Draper S. Questionnaires as a software evaluation tool, CHI 83; 1983. pp. 83–7.
Sasse MA, Flechais I. Usable security: why do we need it? how do we get it?. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 13–30 [chapter 2].
Schneier B. Two-factor authentication: too little, too late. Communications of the ACM 2005;48(4):136.
Schultz EE, Proctor RW, Lien M-C, Savendy G. Usability and security: an appraisal of usability issues in information security methods. Computers and Security 2001;20(7):620–34.
Sinclair S, Smith SW. The TIPPI point: towards trustworthy interfaces. IEEE Security and Privacy 2005:68–71.
Smetters DK, Grinter RE. Moving from the design of usable security technologies to the design of useful secure applications, ACM workshop: new security paradigms workshop. ACM Press; 2002. pp. 82–9.
Sohail MS, Shanmugham B. E-banking and customer preference in malaysia: an empirical investigation. Information Sciences 2003;150:207–17.
Szajna B, Scamell RW. The effects of information system user expectations on their performance and perceptions. MIS Quarterly 1993;17(4):493–516.
Tan M, Teo TSH. Factors Influencing the Adoption of Internet Banking. Journal of Association for Information Systems 2000;1(5):1–42.
Tognazzini B. Design for usability. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 31–46 [chapter 3].
Weir CS, Anderson JA, Jack MA. On the role of metaphor and language in design of third party payments in eBanking: usability and quality. International Journal of Human-Computer Studies 2006;64(8):771–85.
Weir CS, McKay I, Jack MA. Functionality and usability in design for eStatements in eBanking services. Interacting with Computers 2007;19(2):241–56.
Yan J, Blackwell A, Anderson R, Grant A. The memorability and security of passwords. In: Cranor, Garfinkel, editors. Security and usability. O'Reilly; 2005. p. 129–42 [chapter 7].
Zviran M, Haga WJ. Password security: an empirical study. Journal of Management of Information Systems 1999;15(4):161–85.

Catherine Weir joined the Centre for Communication Interface Research at The University of Edinburgh as Research Associate. She currently holds an MSc in Information Technology and has been working in the field of Human Computer Interaction (HCI), Usability Engineering and Interface Design for nine years. Her research interests include experiment design for investigations into how people interact with the Internet and eBanking applications, developing measures of customer satisfaction and usability for Web services, and providing usable security for authentication purposes in online and telephony environments. She has just completed a PhD on her research in eBanking usability.

Gary Douglas completed a PhD at the University of Edinburgh with a background in Mathematical Physics. He works at the Centre for Communication Interface Research at The University of Edinburgh as a Research Fellow. Over the last 8 years working in usability research, has been involved with designing studies of eBanking web portals, mobile services (including SMS/MMS and mobile banking), contact centre technologies, online self-help financial tools, eCRM solutions, biometrics and 2-factor authentication methods. His work also includes development of experimental interfaces for research purposes.

Professor Mervyn A. Jack is the Director of the Centre for Communication Interface Research, part of the School of Engineering and Electronics at the University of Edinburgh. He specialises in dialogue and Internet usability engineering, optimising use and creating improved user interface designs for mass market access to automated services such as telephone and Internet banking. He has been actively researching in the area for some 20 years. Professor Jack is a Fellow of the Royal Society of Edinburgh, a Fellow of the Institution of Electrical Engineers, and is holder of a Research Fellowship from Lloyds TSB Group. He is an author of some 240 papers and three textbooks.