Usability Evaluation of Randomized Keypad

Young Sam Ryu
Assistant Professor
Ingram School of Engineering
Texas State University-San Marcos
601 University Dr.
San Marcos, TX 78666
USA

yryu@txstate.edu

Do Hyong Koh
Graduate Student
Department of Computer Science
Texas State University-San Marcos
601 University Dr.
San Marcos, TX 78666
USA

dk1132@txstate.edu

Brad L. Aday
Undergraduate Student
Ingram School of Engineering
Texas State University-San Marcos
601 University Dr.
San Marcos, TX 78666
USA

ba1096@txstate.edu

Xavier A. Gutierrez
Undergraduate Student
Ingram School of Engineering
Texas State University-San Marcos
601 University Dr.
San Marcos, TX 78666
USA

xg1003@txstate.edu

John D. Platt
Undergraduate Student
Ingram School of Engineering
Texas State University-San Marcos
601 University Dr.
San Marcos, TX 78666
USA

jp1365@txstate.edu

Abstract

In this work, the usability of a randomized numeric keypad was examined and compared to the usability of a conventional numeric keypad. The comparison used completion time measurements and the error rate of short (4-digit) and long (8-digit) PINs to contrast efficiency and accuracy of the keypads. The results showed that the average completion time with a randomized keypad is longer than with a conventional keypad. Additionally, the number of errors with a randomized keypad was significantly higher than with a conventional keypad, particularly when using long PINs. Accordingly, a randomized numeric keypad is more applicable to tasks with short (4-digit) PINs.

Keywords

usability, input method, randomized keypad, numeric keypad, touch screen

Usability Professionals' Association
Promoting usability concepts and techniques worldwide

Introduction

Numeric keypads are popular input methods for personal identification numbers (PINs) for many applications, including automated teller machines (ATM), security screening systems within financial organizations, point-of-sale systems, and home/car door locks. However, the threat of "shoulder surfing" or observing private information from the well-known layout of numeric keys has inspired the idea of randomizing the layout of keys (Collins, 1990; Hirsch, 1982, 1984; McIntyre et al., 2003; Rehm, 1985).

The proliferation of touch screen interfaces on modern devices such as ATMs has enabled the concept of the randomized keypad. However, very little is known about the overall usability of the randomized numeric keypad. Although research has focused on the prevention of shoulder surfing (Hoanca & Mock, 2005; Roth & Richter, 2006; Tan, Keyani, & Czerwinski, 2005) and the use of picture-based keypads (Komanduri & Hutchings, 2008), the usability of randomized keypads is not covered in the literature.

This study evaluated the overall usability of a randomized numeric keypad. The primary goal of this study was to investigate if users were able to complete the task of entering a PIN via a randomized keypad and to compare this activity to the use of a fixed keypad with a conventional layout. To conduct the test, completion time measurements and error-rate statistics for various tasks were collected via specialized test fixtures. This data provided a valuable metric for the efficiency and accuracy of the use of a randomized keypad. A secondary goal of this study was to investigate whether a randomized keypad enhanced the perceived security of the PIN-entry task, particularly for publicly located systems (ATMs, door locks, etc.). Pre-test and post-test surveys provided subjective data to measure the perceived level of security and user satisfaction.

Methods

The following sections discuss the design, participants, and task and procedures used in this study.

Design
The experiment was designed as a within-subject test, where dependent variables were the completion time and number of errors, and the independent variable was the type of keypad (conventional vs. randomized). Two separate experiments were conducted based on the length of the PIN to be entered. The first experiment used short (4-digit) PINs, and the second experiment used long (8-digit) PINs.

Participants
A total of 50 participants were recruited to participate in the experiment, 25 for each PIN length. The average age of participants was 23.24 years, with a standard deviation of 6.60 years. There were 6 female and 44 male participants, with 39 Caucasians, 10 Hispanics, and 1 Asian.

Equipment
The randomized keypad and conventional keypad were implemented in software, and appeared on a 19-inch touch screen monitor (Figure 1). The software created a randomly generated PIN on top of the keypad for each trial. When a participant typed the PIN and pushed *Enter*, the software recorded the completion time and the correctness of the entry. Participants could correct errors with the *Clear* button until they pressed *Enter*.

Task and Procedure
Each participant completed a pre-test and post-test questionnaire to collect demographic and subjective data about the user experience with the randomized keypad. Before data collection, each participant entered three PINs to become familiar with the touch screen and keypad. During data collection, each participant completed 20 PIN-entry trials with a conventional keypad followed by 40 PIN-entry trials with a randomized keypad. A PIN for each trial was generated randomly. The randomized keypad was used for 40 iterations to investigate training effects. After the user testing session, each participant completed a post-test questionnaire to collect subjective usability assessment data.
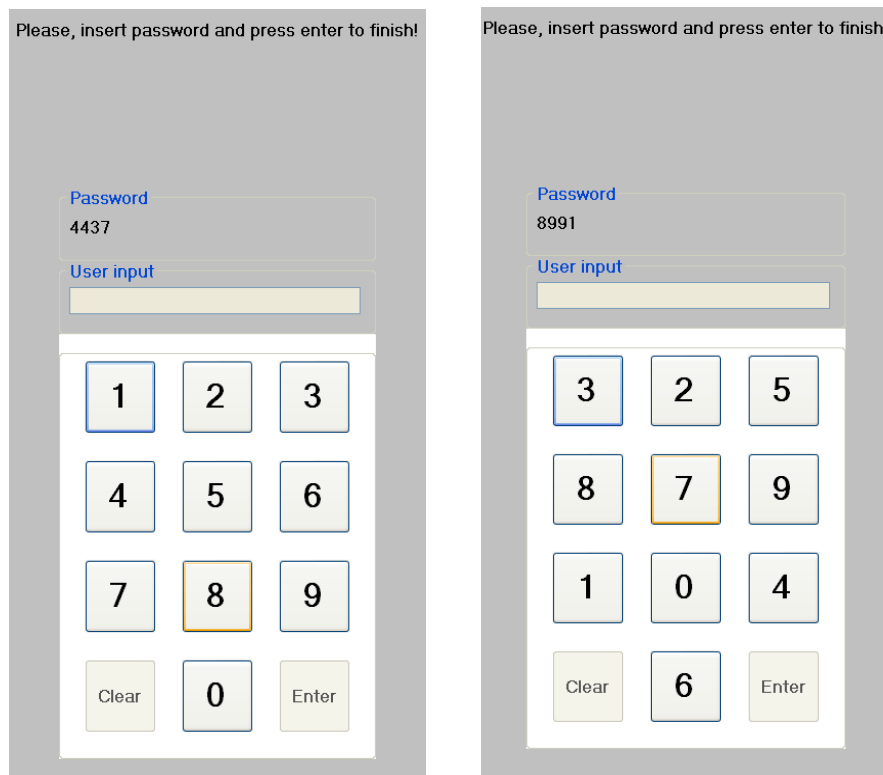
Figure 1. Conventional keypad (left) and randomized keypad on a touch screen

Results

The following sections discuss the results for the pre-test survey, the user test with 4-digit PINs, the user test with 8-digit PINs, the user test using mixed ANOVA, and the post-test survey.

Pre-test Survey

According to the pre-test survey for 50 participants, 48 out of 50 (96%) of the participants were familiar with touch-screen numeric keypads. Among the 48 participants, 44 participants were familiar with touch-screen numeric keypads at ATMs (91.70%), 34 at grocery store checkouts (70.80%), 14 at door locks (29.20%), 36 at touch-screen phones (75%), and 4 at others (such as, GPS and PC) (8.30%). Of those participants familiar with touch-screen keypads, the majority indicated use of the technology on a regular basis (36% "every day," 40% "a few times a week," 20% "a few times a month"). Thirty-four out of 50 participants (68%) reported they had felt insecure (yes or somewhat) when they typed in PINs using touch-screen keypads in public (Figure 2); both groups of participants (4 digits vs. 8 digits) showed the same percentage (68%, 17 out of 25 each). This percentage is significantly greater than 50% because the 95% adjusted-Wald binomial confidence interval for this percentage ranges from 54.13% to 79.30%, $p<.05$. Further, 34 out of 50 participants (66%) liked the idea of using a randomized keypad to provide more security (Figure 3); both groups of participants (4 digits vs. 8 digits) showed the same percentage (66%, 17 out of 25 each). Because the 95% adjusted-Wald binomial confidence interval for this percentage ranges from 52.11% to 77.61%, the percentage is significantly greater than 50%, $p<.05$.

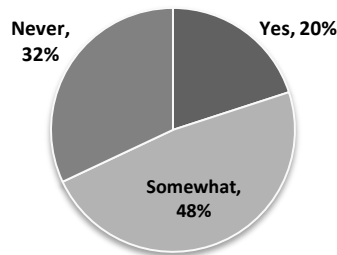**Have you ever felt insecure with typing passwords using a keypad in public?**

Figure 2. Pre-test survey question

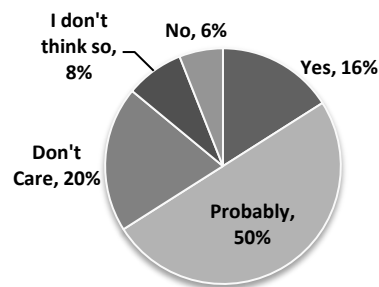**Do you think a randomized keypad would provide more security ? (Before Experiment)**

Figure 3. Pre-test survey question

User Test with 4-Digit PINs

A total of 25 participants conducted the task of entering a randomly generated 4-digit PIN. Each participant completed 20 trials of this task using a conventional keypad and 40 trials using a randomized keypad. The randomized keypad task used more iterations to assess the presence of any learning or training in the task.

The average completion time with the randomized keypad (4.598 seconds, SD=0.795) was significantly longer than the average completion time of the conventional keypad (3.405 seconds, *SD*=0.612), $F(1,24)=72.80$, $p<.001$ (Figure 4). However, there was no significant difference in completion time between the first 20 trials of the randomized task (4.634 seconds, *SD*=0.788) and the last 20 trials of the randomized task (4.563 seconds, *SD*=0.877), $F(1,24)=0.49$, $p=0.4894$ (Figure 5). Thus, no significant learning effect was observed in the use of a randomized keypad for entering a 4-digit PIN.
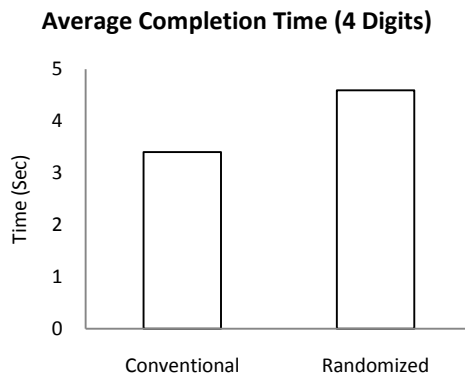
Average Completion Time (4 Digits)

Figure 4. Average completion time for the 4-digit PINs user test

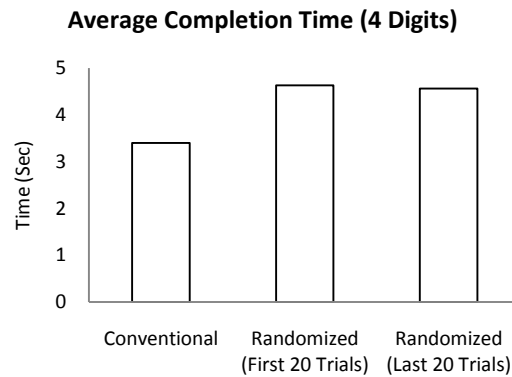**Average Completion Time (4 Digits)**

Figure 5. Average completion time for the 4-digit PINs user test

As part of the data gathering process, the error rate per trial was also measured. The average error rate for the conventional keypad (0.0440) was slightly higher than the error rate for the randomized keypad (0.0350) (Figure 6), but not significantly higher, $F(1,24)=1.45$, $p=0.2408$. The standard deviation of the error-rate data was very high (conventional keypad = 0.0545, randomized = 0.0445) as compared to the mean values. Also, there was no significant difference in error rate between the first 20 trials of the randomized task (0.04, $SD=0.07$) and the last 20 trials of the randomized task (0.03, $SD=0.04$), $F(1, 24)=1.20$, $p=0.2832$ (Figure 7).
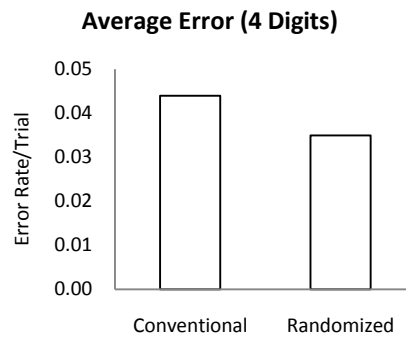
Average Error (4 Digits)

Figure 6. Average error rate for the 4-digit PINs user test
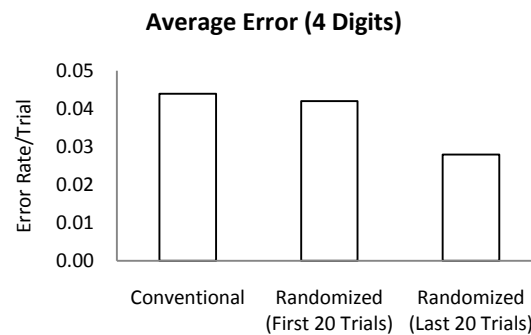
Average Error (4 Digits)

Figure 7. Average error rate for the 4-digit PINs user test

User Test with 8-Digit PINs

Data from the long (8-digit) PIN testing showed the average completion time with the randomized keypad (8.932 seconds, SD=1.546) was significantly longer than that of the conventional keypad (6.277 seconds, SD=1.480), $F(1,24)$=111.59, p<.001 (Figure 8). However, no significant difference was noted in completion times between the first 20 trials (9.060 seconds, SD=1.531) and the last 20 trials (8.80 seconds, SD=1.630) for the randomized keypad, $F(1,24)$=3.65, p=0.0681 (Figure 9). Thus, no significant learning effect was observed in the use of a randomized keypad for entering an 8-digit PIN.
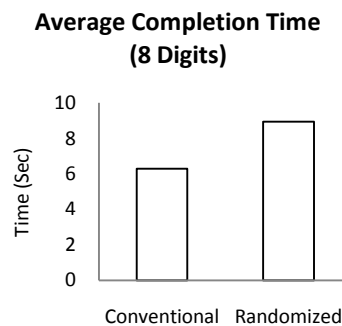
Average Completion Time (8 Digits)

Figure 8. Average completion time for the 8-digit PINs user test
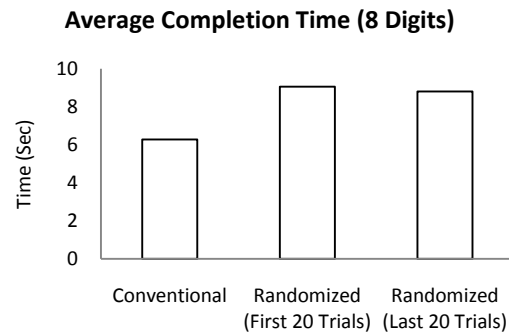
Average Completion Time (8 Digits)

Figure 9. Average completion time for the 8-digit PINs user test

Unlike the result from the 4-digit task, the average error rate for the conventional keypad (0.0380, *SD*=0.0525) was significantly lower than the error rate for the randomized keypad (0.0690, *SD*=0.0755), $F(1,24)=4.79$, $p=0.0386$ (Figure 10). However, there was no significant difference in error rate between the first 20 trials of the randomized task (0.0720, *SD*=0.0817) and the last 20 trials of the randomized task (0.0660, *SD*=0.0702), $F(1,24)=0.10$, $p=0.7549$ (Figure 11).
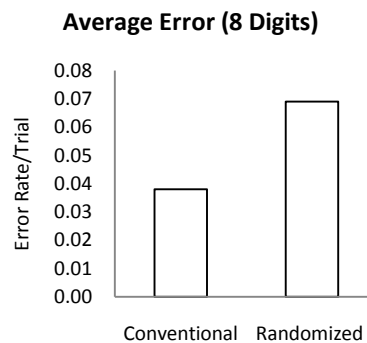
Average Error (8 Digits)

Figure 10. Average error rate for the 8-digit PINs user test
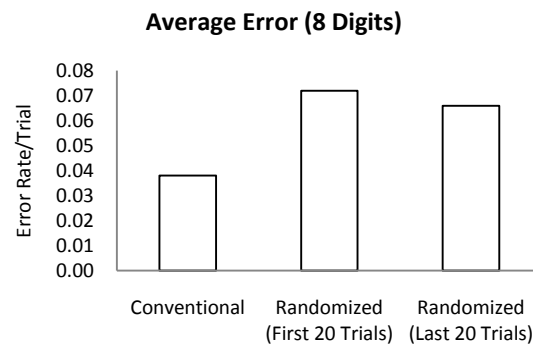
Average Error (8 Digits)

Figure 11. Average error rate for the 8-digit PINs user test

User Test— Mixed ANOVA

By combining the data from both the 4-digit group and the 8-digit group, mixed ANOVA was performed to assess any interaction effect according to the PIN length. The PIN length (4 digits vs. 8 digits) was the "between subject" variable, while the type of keypad (conventional vs. randomized) and trials (1st to 20th trials) were the "within subject" variables. To match 20 trials of the conventional keypad, only the first 20 out of 40 trials of the randomized keypad tests were taken for the mixed ANOVA.

Completion time

Obviously, the effect of PIN length on completion time was significant, $F(1,48)=146.47$, $p<0.0001$. The main effect of the type of keypad was significant as well, implying that task completion time took longer with a randomized keypad than with a conventional keypad, regardless of the PIN length, $F(1,48)=203.79$, $p<.0001$. Also, interaction was noted between the PIN length and the type of keypad, $F(1,48)=30.57$, $p<.001$. Figure 12 shows the interaction plot. This interaction can be interpreted as the increase in completion time between conventional and randomized keypads when users were required to type more digits for the longer PINs.
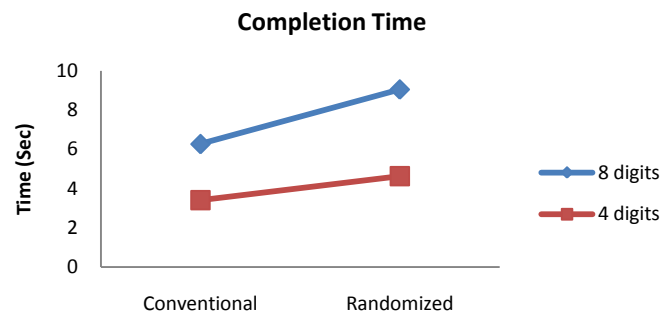

Figure 12. Interaction plot between the PIN length and type of keypad for completion time

The main effect and interaction effect caused by the trial were analyzed. The main effect of the trial was significant, $F(19,912)=3.43$, $p<0.0001$. This implies that the completion time variations among trials caused by different PINs were great. Also, there was a significant interaction between PIN length and trial, $F(19,912)=1.60$, $p=0.0495$. To interpret this interaction thoroughly, a one-way ANOVA was performed with the trial as the factor, according to each type of keypad (Table 1). According to the p-values in Table 1, the trial was not a significant factor for short (4-digit) PINs, while it was a significant factor for long (8-digit) PINs. Thus, the interaction between PIN length and trial can be attributed to greater completion time variations with longer PINs. In other words, different PINs account for significant proportion of completion time variations for longer PINs.

Table 1. One-way ANOVA result of completion time with the trial as the factor

	Type of keypad	F value	P-value
4-digit PINs	Conventional only	$F(19,456)=1.57$	$p=0.0594$
	Randomized only	$F(19,456)=1.20$	$p=0.2566$
	Both	$F(19,456)=1.46$	$P=0.0947$
8-digit PINs	Conventional only	$F(19,456)=2.19$	$p=0.0027$
	Randomized only	$F(19,456)=1.79$	$p=0.0215$
	Both	$F(19,456)=2.83$	$p<0.0001$

Error rate

No significant main effect due to PIN length was noted in the number of errors, $F(1,48)=0.63$, $p=0.4309$. Similarly, no significant main effect was noted due to the type of keypad, $F(1, 48)=2.23$, $p=0.1422$. However, the relatively low p-value, $F(1,48)=2.82$, $p=0.0997$, on the interaction between PIN length and keypad type can be explained by a significant effect of keypad type during 8-digits tasks, as indicated in the previous section, $F(1,24)=4.79$, $p=0.0386$ (Figure 13). No significant main effects of trial, $F(19,912)=1.23$, $p=0.2211$, were noted nor any other interaction effect.
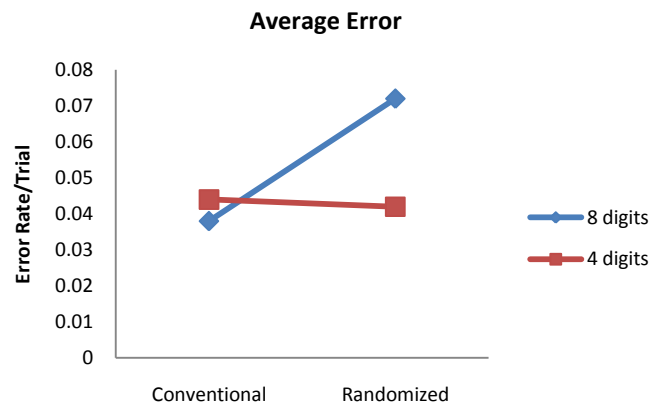

Figure 13. Interaction plot between the PIN length and type of keypad for average error rate

Post-test Survey

In the post-test survey, 18 of 25 (72%) participants with short (4-digit) PINs and 16 of 25 (64%) participants with long (8-digit) PINs believed (yes or probably) that the randomized keypad would provide more security. Thus, according to the Adjusted Wald Binomial confidence interval, the result from the short PINs (52%<95%*C.I.*<86%) was significant; however, the same metric applied to the longer PINs (44%<95%*C.I.*<79%) was not significant. Also, there was no significant difference in the beliefs between the two groups according to a Fisher's exact test, $p=0.7624$.

The randomized keypad was somewhat more difficult to use (yes or probably), with 7 of 25 (28%) participants and 10 of 25 (40%) participants expressing some difficulty in the short-PIN and long-PIN cases, respectively. According to the Adjusted Wald Binomial confidence interval, the result from the short-PINs case (14%<95%*C.I.*<47%) showed that the percentage of users having difficulty was significantly less than 50%, while the same metric applied in the long-PINs case (23%<95%*C.I.*<59%) was not significantly less than 50%. Also, there was no significant difference in the beliefs between the two groups according to Fisher's exact test, $p=0.5512$.

Recommendations

According to the pre-test survey, a majority of users had expressed insecurity when they typed PINs in public. This confirms a need for alternative input methods that reduce the feeling of insecurity. Also, a majority of users believed the use of a randomized keypad would improve the perceived security in the pre-test survey. However, according to the post-test survey, only users who performed short (4-digit) PIN tasks maintained this belief. Also, users found the randomized keypad more difficult to use for entering long PINs.

It was obvious that the randomized keypad required additional time to type PINs than the conventional keypad, regardless of the PIN length. However, the randomized keypad caused significantly more errors than the conventional keypad when entering long (8-digit) PINs. Thus, according to the subjective assessment in pre-test and post-test surveys and the objective

usability measures, the use of a randomized keypad would be recommended only for short (4-digit) PINs.

Future Research

In this study, the randomized keypad was configured with a fully randomized layout of the ten numeric keys in a standard, square orientation. However, a randomized keypad can be configured in many different ways. For example, in a square keypad orientation, rows can be randomized while columns are fixed. With more creative keypad designs, a circular numeric keypad could be used (i.e., rotary-dial) where the location of the *0* key is randomly selected while the ordering of the keys is not manipulated. Other orientations, such as a strip-shaped keypad or beehive-shaped keypad are feasible because touch-screen devices can produce arbitrary keypad shapes. Thus, the combination of different levels of randomization with novel keypad designs may result in surprising usability when evaluated by other researchers. The tradeoff between the complexity of randomness and the enhancement of perceived security is likely to depend on a combination of factors in keypad design and presentation.

The use of a randomly generated PIN for each trial might have reduced external validity of the study, because most people use familiar PINs that they have already memorized. Thus, the completion time with a conventional keypad and well-known PIN may be much faster than indicated by this study. However, the goals of this test included minimizing the training factor associated with the completion time. Thus, other researchers need to consider whether to use randomly generated PINs or a fixed PIN to evaluate keypads.

Conclusion

In this study, the usability of a randomized numeric keypad was evaluated. Via data collection and statistical techniques, the randomized keypad was compared to the conventional numeric keypad. Completion time and number of errors were used to objectively measure the efficiency and accuracy of each keypad in data-entry tasks of short (4-digit) and long (8-digit) PINs. In addition, subjective measurements including the perceived level of improved security and user satisfaction were assessed. The results indicated that the average completion time with a randomized keypad was longer than that with conventional keypad. This observation was consistent for both short and long PINs. The number of errors with a randomized keypad was significantly higher than with a conventional keypad when using long PINs, while the number of errors with a randomized keypad was not significantly higher than with a conventional keypad when using short PINs. According to the subjective assessment in pre-test and post-test surveys and the objective usability measures, a randomized numeric keypad was more applicable to tasks with short (4-digit) PINs.

Practitioner's Take Away

The following were the main findings of this study:

- Longer completion times should be expected when using a randomized numeric keypad versus a conventional keypad.
- The number of errors with a randomized keypad was significantly higher than with a conventional keypad when users typed longer PINs.
- The number of errors with a randomized keypad was not significantly higher than with a conventional keypad when users typed short PINs.
- A randomized numeric keypad is better suited to applications requiring short PINs.

References

Collins, E.R. (1990). Computer access security code system, United States Patent #4,926,481: United States Patent and Trademark Office, La Canada, CA.

Hirsch, S.B. (1982). Secure keyboard input terminal, United States Patent #4,333,090: United States Patent and Trademark Office, 305 Peck Dr., Beverly Hills, CA 90212.

Hirsch, S.B. (1984). Secure input system, United States Patent #4,479,112: United States Patent and Trademark Office, 305 Peck Dr., Beverly Hills, CA 90212.

Hoanca, B., & Mock, K. (2005). In Screen Oriented Technique for Reducing the Incidence of Shoulder Surfing. In *the Proceedings of the International Conference on Security and Management (SAM)*, Las Vegas, Nevada, USA.

Komanduri, S., & Hutchings, D.R. (2008). Order and entropy in picture PINs. *Proceedings of graphics interface*. Windsor, Ontario, Canada. Canadian Information Processing Society.

McIntyre, K.E., Sheets, J.F., Gougeon, D.A.J., Watson, C.W., Morlang, K.P., & Faoro, D. (2003, April). Method for secure pin entry on touch screen display, United States Patent #6,549,194: United States Patent and Trademark Office.

Rehm, W.J. (1985). Security means, United States Patent #4,502,048: United States Patent and Trademark Office, 22 Lomatta St., The Gap, Queensland, 4061, AU.

Roth, V., & Richter, K. (2006). How to fend off shoulder surfing. *Journal of Banking & Finance*, 30(6), 1727-1751.

Tan, D.S., Keyani, P., & Czerwinski, M. (2005). Spy-resistant keyboard: more secure password entry on public touch screen displays, *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*. Canberra, Australia. Computer-Human Interaction Special Interest Group (CHISIG) of Australia.

About the Authors

Young Sam Ryu
Young Sam Ryu is an Assistant Professor of Ingram School of Engineering at Texas State University-San Marcos. His research area includes human-computer interaction and usability.

Do Hyong Koh
Do Hyong Koh is a graduate student of Computer Science Department at Texas State University-San Marcos.

Brad L. Aday
Brad L. Aday is an industrial engineering undergraduate student of Ingram School of Engineering at Texas State University-San Marcos.

Xavier A. Gutierrez
Xavier A. Gutierrez is an industrial engineering undergraduate student of Ingram School of Engineering at Texas State University-San Marcos.

John D. Platt
John D. Platt is an industrial engineering undergraduate student of Ingram School of Engineering at Texas State University-San Marcos.