

User Input Device

Participants used three user input devices: 2-button, 3-button and 5-button (Figure 1). The layout of the buttons and their mappings to actions are described in the Introduction section of this document. The buttons (Enabling Devices Compact Switch #745) were connected to a switch device (X-keys XSI-38-US). The switch device was connected by 3.5 mm switch plugs of the buttons to the PC via a USB port. The switch inputs were mapped onto the desired keyboard inputs to control the Voting App.

Design

A repeated measures design was used. Each participant used each of the three input devices. Three equivalent ballots were created for use with each of the three input devices. The order in which the devices/ballots were used was counter balanced by Graeco-Latin square.

Input device (2-, 3-, or 5-button device) was the independent variable for all analyses of user performance, user preference, and eye tracking data. Repeated measures ANOVAs were conducted on each dependent variable.

Objective dependent variables were derived from event logs (user selections) and eye tracking data. Event logging variables included the following:

- *Race Time*: The average time spent on the races excluding the prompt pages (seconds). The initial instructions page, practice race, write-in, constitutional amendment, and ballot initiative were excluded, because behaviors varied widely among participants and ballots on these items.
- *Write-in Time*: The total time it took participants to write-in a candidate's name divided by the number of letters in the name (seconds/letter).
- *Click Time*: The time elapsed between selections of any on-screen buttons, including navigation control buttons and candidates' names (seconds).
- *Response Time*: The time elapsed between the onset of a race and the participant's selection of a candidate divided by the position of the candidate in the list (seconds/candidate).

Dependent variables for the eye tracking analysis included the following:

- *Controls Gaze Time*: The average amount of time participants looked at the on-screen control buttons (i.e., back, help, review, next; seconds).
- *Candidates Gaze Time*: The average amount of time the participants spent looking at the list of candidates (seconds).
- *Target Candidate Gaze Time*: The average amount of time participants spent looking at the target candidate (seconds). This is a subset of *Candidates Gaze Time*.

Subjective dependent variables included input device ranking, System Usability Scale ratings (Sauro, 2011), voting system usability ratings, and Wong-Baker pain scale ratings (originally published in Whaley & Wong's Nursing Care of Infants and Children).

Procedure

Participants completed the experiment in one-on-one sessions with an Experimenter, who was present throughout the session. Participants first signed an informed consent form and received a brief overview of the procedures from the experimenter. Then participants completed a short questionnaire on their arthritis symptoms and voting experience, and each participant's grip strength and pinch strength were measured. Grip strength assessment was performed with a Jamar Hydraulic Hand Dynamometer (5030J1 Serial #30303257). Users were instructed to grip the dynamometer as tightly as possible without experiencing excessive discomfort. Both left and right hands were tested. The pinch strength measurement was performed with a Jamar Hydraulic Pinch Gauge (7498-05 Serial # 60203139). Three pinch types (Tip, Key, and Palmar) were performed. For the tip pinch, users placed the tips of their index finger and thumb on opposing sides of the gauge. For the key pinch, users placed the lateral aspect of the middle phalanx of their index finger on the bottom of the gauge and their thumb on the top of the gauge. For the Palmar pinch, users placed their thumb on the top of the gauge and the pads of their index and middle fingers on the bottom of the gauge.

Next, the participant was seated comfortably in front of the computer and eye tracking system. The experimenter calibrated the eye tracking system and then launched the Voting App. The participant completed the first ballot while the experimenter observed. Immediately afterward, the participant completed a set of rating scales for the input device, which included a pain rating scale, the System Usability Scale (Sauro, 2011), and a usability scale created specifically for the Voting App. These procedures were repeated for the next two input devices.

After all three input devices had been used, the experimenter interviewed the participant. Each participant ranked the input devices based on their subjective preference. Participants explained their rankings and provided any additional comments. To conclude the session, participants were debriefed and compensated for their time.

Results

Arthritis Symptoms

All participants reported experiencing arthritis symptoms in their fingers, hands, or wrists on the day of the study. Figure 4 shows the reported distribution and severity of arthritis symptoms. Ten participants reported having only osteoarthritis/degenerative arthritis. One participant reported having osteoarthritis, gout, and systematic lupus erythematosus. Two participants reported having osteoarthritis and fibromyalgia. One participant reported having osteoarthritis and rheumatoid arthritis. One participant reported having only gout.

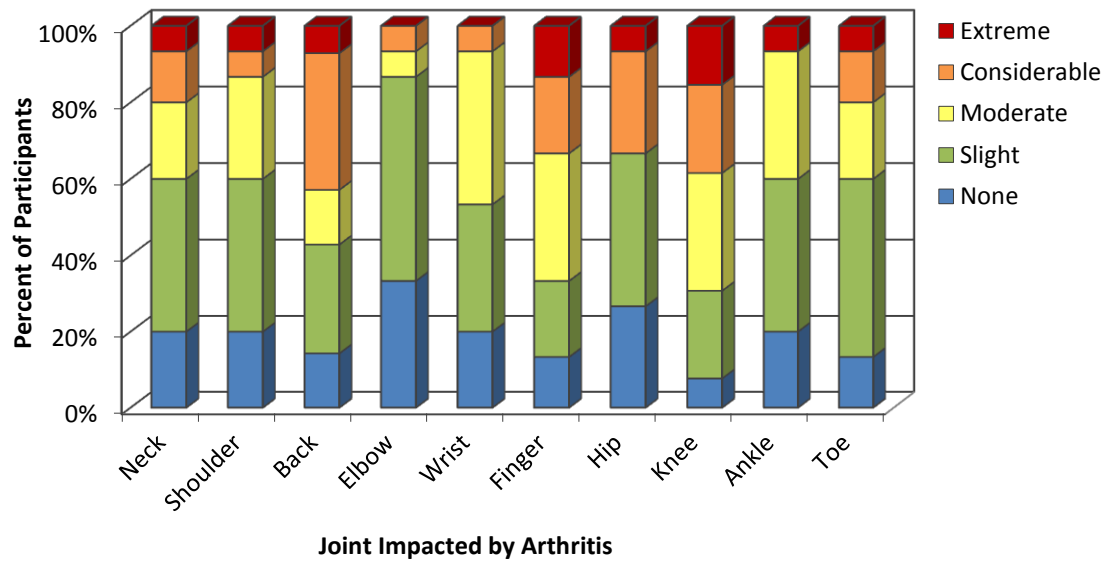


Figure 4. Distribution of arthritis symptoms.

Participants were asked to rate the arthritis-related difficulty that they had experienced using touchscreen devices, buttons, and voting systems (Table 1). Two participants reported having slight difficulty voting. Five participants reported having slight to moderate difficulty using tactile buttons. Four participants reported having slight difficulty using touchscreens.

Table 1. Arthritis user experience questionnaire.

Question	No Difficulty	Slight Difficulty	Moderate Difficulty	Extreme Difficulty
Do you have difficulty using a touch screen due to your arthritis?	13	2	0	0
Do you have difficulty pressing buttons due to your arthritis?	10	4	1	0
Did you have any difficulty voting at a polling place due to your arthritis?	10	4	0	0

Strength Assessment

Grip and pinch strength are related to the use of the button interfaces. It was important to characterize participants' hand strengths to identify any outliers (extremely strong or weak participants). Both the left and right hands were tested. Figure 5 shows the results of the grip strength assessment using a wide (open) grip and a narrow (closed) grip. Figure 6 shows the results of the pinch force assessment for the top, palmar, and key pinch. In both figures, the mean is represented by the filled diamond, median is represented by horizontal line, interquartile range (25th to 75th percentile) is the shaded box, plus or minus 1.5 interquartile range is represented by the vertical lines, and outliers are unfilled squares. None of the grip or pinch strength measurements showed any participants as outliers, which were defined as measurements three standard deviations above or below the mean.

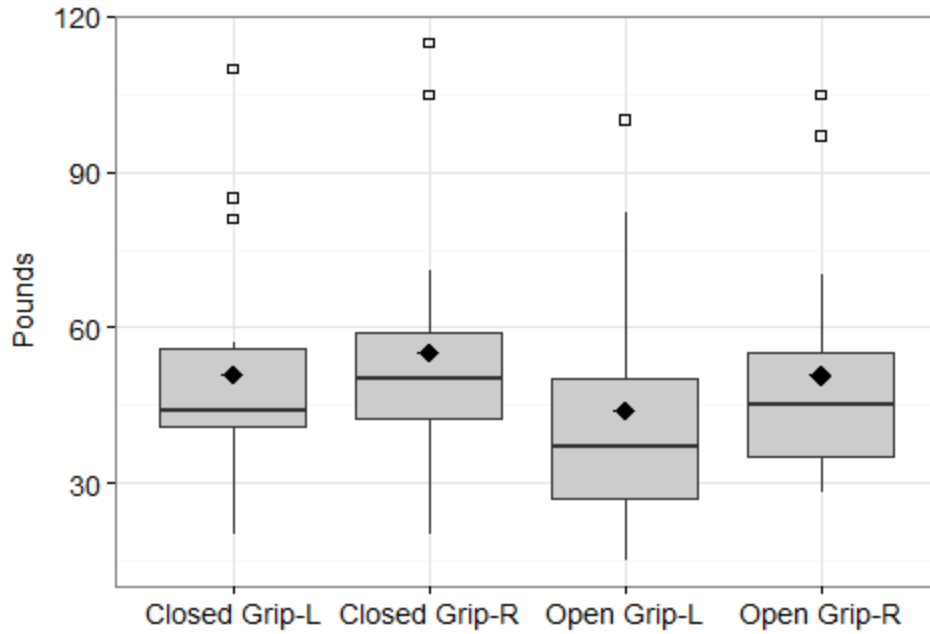


Figure 5. Boxplots of grip strength measurements. The boxplots represent medians, interquartile ranges (boxes), and 1.5 x the interquartile range (whiskers). The mean is represented by the filled diamond.

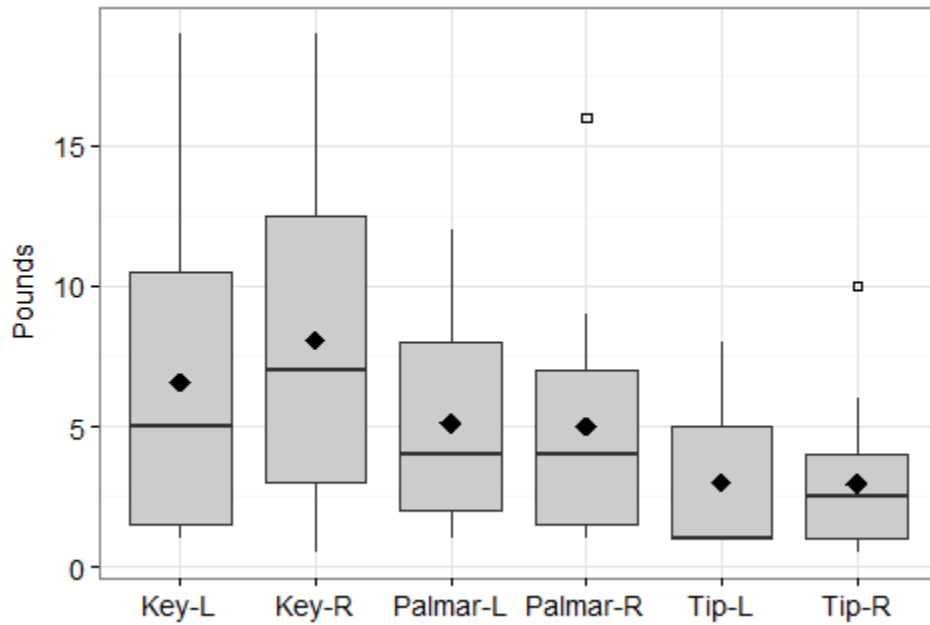


Figure 6. Boxplots of pinch strength measurements. The boxplots represent medians, interquartile ranges (boxes), and 1.5 x the interquartile range (whiskers). The mean is represented by the filled diamond.

Subjective Data

Rankings

The 5-button device was expected to afford faster voting times and receive higher user rankings (H1), but the data did not support the hypothesis. A Friedman rank test did not show significant difference in rankings among the three input devices, $\chi^2(2) = 0.13$, $p = .94$. Although six participants ranked the 5-button device highest, seven ranked it lowest (Table 2). Total Score in Table 2 was computed for each button by assigning scores of 3, 2, and 1 for favorite, second favorite, and least favorite, respectively, and summing across participants. Also, ratios of favorite-to-least-favorite for each device were computed by dividing the number of votes for favorite by the number of votes for least favorite. Based on both Total Scores and the Ratios, there was a slight but statistically insignificant preference for the 3-button device, although note that it received the fewest “favorite” rankings.

Table 2. Subjective rankings of preference for the three input devices. Cell values in the first three rows represent counts.

	2-button	3-button	5-button
Favorite	5	4	6
Second Favorite	5	8	2
Least Favorite	5	3	7
Total Score	30	31	29
Ratio Fav/Least Fav	1	1.3	0.86

It was expected that the 5-button device would enable users to vote faster with fewer button presses, and therefore it was expected that the 5-button device would be preferred. On the contrary, none of the participants reported that voting time was a determining factor in their preference rankings. Instead, responses from the follow-up interview indicated that some participants were confused by the 5-button device. Five of the fifteen participants reported that it was too complicated or confusing. The 2-button device was ranked the highest by five participants, all of whom reported that they favored its simplicity.

Usability Ratings

Participants also completed rating scales for each input device. Participants rated ease-of-use with the System Usability Scale and the Voting App Usability Scale. A total score was computed for each usability scale varied System Usability Scale ratings did not differ significantly among the input devices, $F(2,28) = 0.02$, $p = 0.98$, $\eta_p^2 = 0.002$, power = 0.053, nor did the ratings differ on the Voting App Usability Scale, $F(2,28) = .849$, $p = 0.438$, $\eta_p^2 = 0.057$, power = .181.

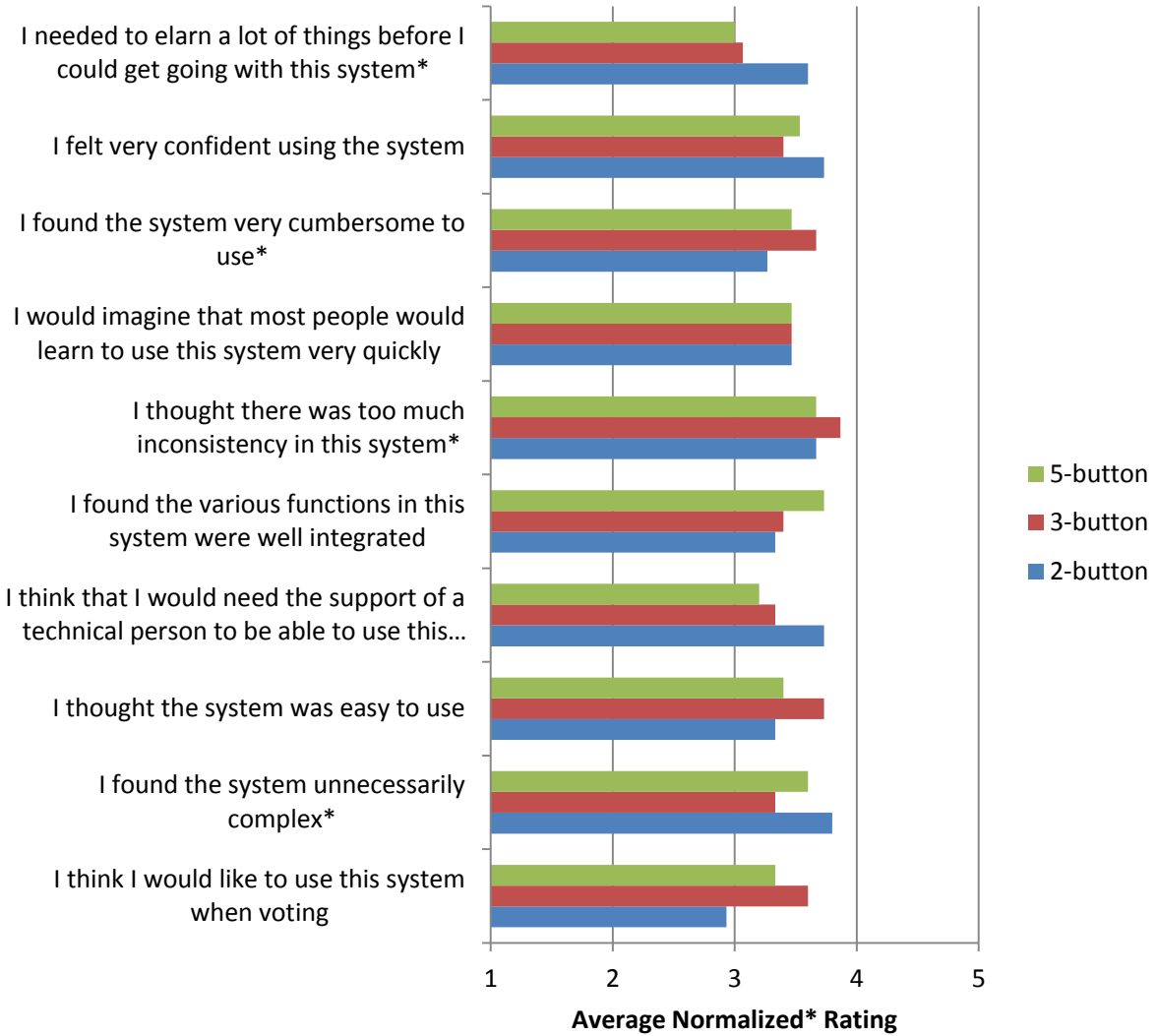


Figure 7. System Usability Scale responses for each input device. Response options ranged from 1 to 5. *For negative or undesirable statements (see starred statements), the ratings were normalized so that higher scores represented greater desirability.

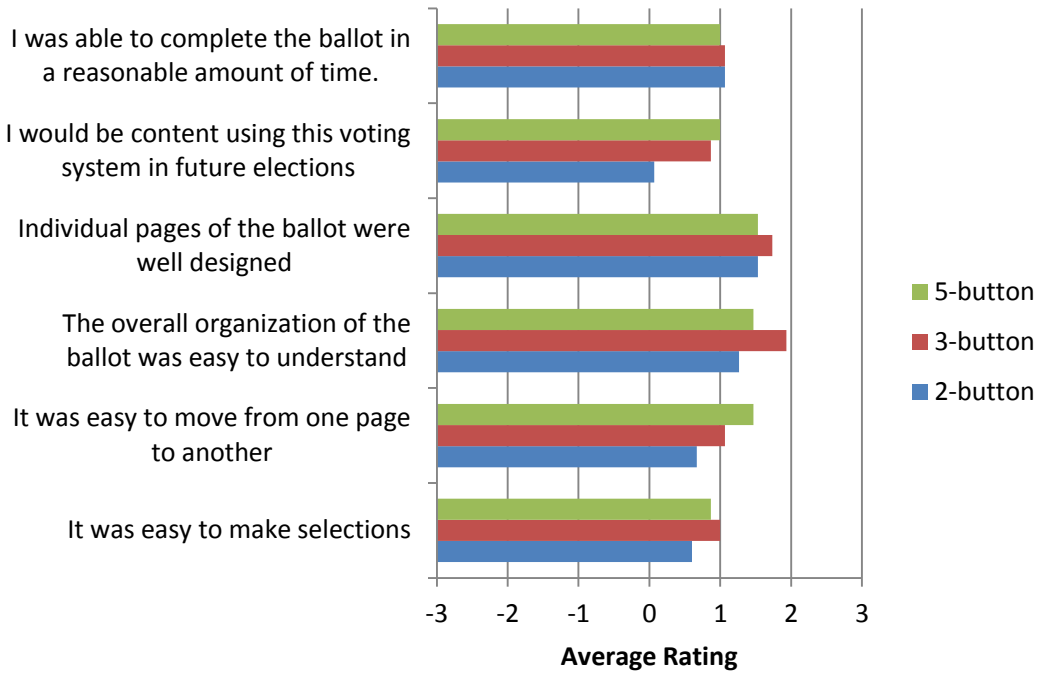


Figure 8. Voting App Usability Scale. Response options ranged from -3 to 3.

Pain

Participants used the Wong-Baker Faces Pain Rating Scale to indicate the intensity of pain they experienced while using each input device (Results shown in Table 3). It was expected that pain ratings would be affected by the number of button presses required (H2). Ratings were expected to be higher for the 2-button than the 3-button, and higher for the 3-button than the 5-button. Repeated measures ANOVA revealed that average pain ratings did not differ significantly among the different input device, $F(2,18) = 1.19$, $p = 0.33$, $\eta_p^2 = 0.117$, power = 0.227 (Figure 9).

Table 3. Pain ratings experienced while using each device. Values in the cells represent counts.

	No Hurt 0	Hurts little bit 1	Hurts little more 2	Hurts even more 3	Hurts whole lot 4	Hurts worst 5
2-button	5	2	3	1	0	0
3-button	6	1	3	0	0	0
5-button	7	3	0	1	0	0

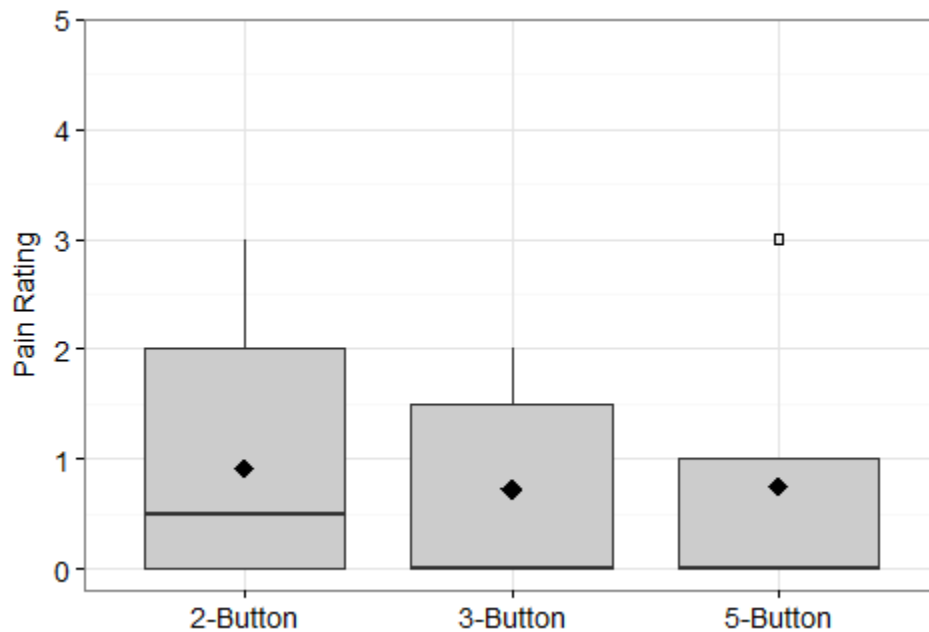


Figure 9. Pain ratings while using each input device. The scale ranged from 0 (least pain) to 5 (most pain).

Objective Data

Race Time

Race time represents the amount of time it took participants to vote in the 12 races (Figure 10). This is the sum of reaction time (time to select a candidate) and the subsequent epoch during which the participant navigated from the selected candidate to the on-screen “next” page button. It excluded the prompts before each race. The 5-button device was expected to be faster than the 3-button device, and the 3-button device was expected to be faster than the 2-button device (H1).

Mean race time was highest in the 3-button condition, but this was due to an extreme outlying data point in the 3-button condition, which was 4.7 standard deviations above the mean (Figure 10). Repeated measures ANOVA did not show significant differences among mean voting time for the three input devices, $F(2, 28) = 0.67$, $p = 0.52$, $\eta_p^2 = 0.046$, power = 0.151. The differences were not significant when the outlier was excluded, $F(2, 26) = 0.546$, $p = 0.586$, $\eta_p^2 = 0.04$, power = 0.13. Thus, the 5-button device did not afford the faster voting times that were expected. Four of the fifteen participants did not use the page-forward or page-backward buttons, and this might have contributed to the failure to find faster race times for the 5-button device. The results also failed to support the hypothesis that the 3-button device would be significantly faster than the 2-button device. It is possible that the amount of time saved by using the bi-directional navigation of the 3- and 5-button devices was not large enough to show a statistically significant benefit over the unidirectional 2-button device. Also, participants’ confusion about the additional button might have slowed their task completion times.

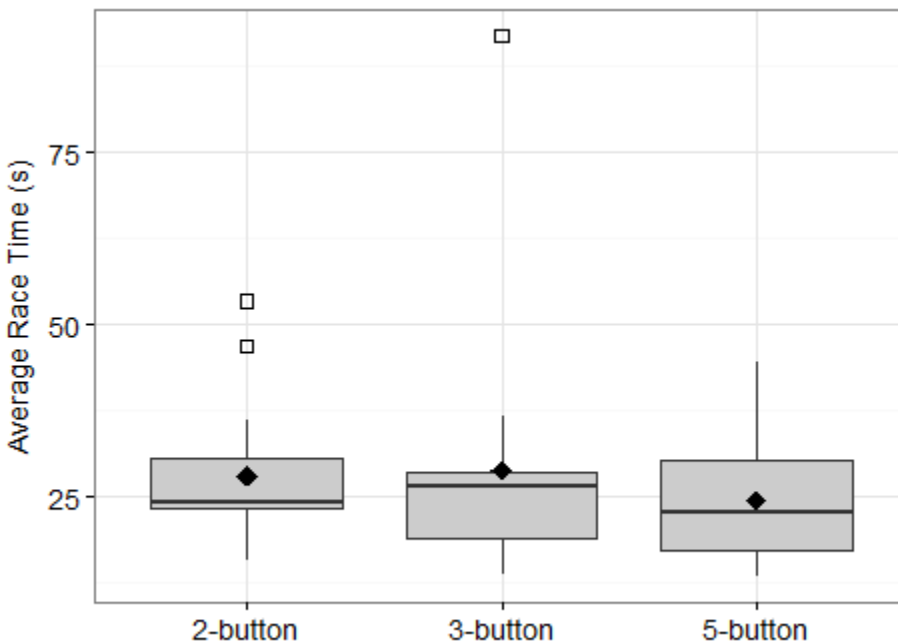


Figure 10. Race time as a function of input device, median and interquartile ranges.

Subtask Times

The analyses below examined three voting subtask times separately (write-in time; response time; click time). The three subtasks were expected to show slower times for the 2-button device; these included entering a candidate's name on the write-in race (*Write-in Time*), moving the onscreen cursor to a desired button (*Click Time*), and finding and selecting a candidate's name (*Response Time*). Each of these subtasks reflects navigation speed, which is dependent on visual search and cursor movement. Visual search speeds were not measured, and they were not expected to vary across input devices because the same display configuration was used on all three ballots. Cursor movement speed was expected to vary by input device.

Write-in Time

In the single write-in contest, participants were asked to write-in the name of a specified candidate (see Figure 11). Participants scrolled through the on-screen keyboard to enter the candidate's name. With the 2-button device, participants could not move the cursor backward through the alphabet, and instead had to go to the bottom and wrap back to the top. In contrast, the 3-button and 5-button devices included a back button for bi-directional navigation. It was expected that participants would be slower with the 2-button device when entering names whose letter orders did not match their alphabetical orders (refer to hypothesis H3). Specifically, it was expected that the 3-button and 5-button devices would be faster when entering names that could be entered with fewer button presses by using the back button. For example, to enter the name "Ed" the participant could press the tactile back button once to navigate from "e" to "d", but with a 2-button device they would have to press the tactile next button 25 times to wrap through the alphabet.

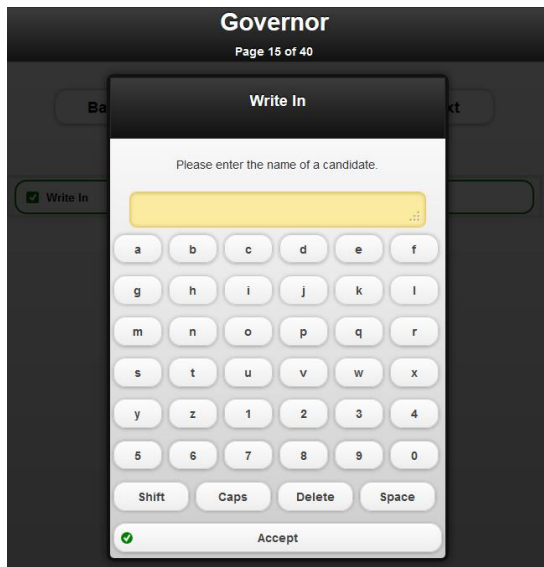


Figure 11. On-screen keyboard for entry of a write-in candidate's name.

Hierarchical linear regression was conducted to analyze the effect of input device on write-in time. A hierarchical model was used to control for (i.e., account for) the variation in the length of the names entered by participants. An interaction was expected between the input device and alphabet reversals, showing that alphabet reversals affected task time for the 2-button device more than the 3- and 5-button devices. The full regression model included input device, number of letters entered by the participants, and the number of times the participant could have reduced button presses by using the tactile back button (henceforth referred to as “alphabet reversals”). Variables were successively removed from the full model, testing the statistical significance of each. The order of removal was as follows: input device, alphabet reversals, and number of letters. Contrary to hypothesis H3, the interaction between input device and alphabet reversals was not significant, $F(2) = 2.16$, $\Delta R^2 = 0.09$, $p = 0.13$. The only significant variable was number of letters, $F(1) = 12.42$, $\Delta R^2 = 0.25$, $p = 0.001$.

Although the expected difference in write-in time among the input devices was not obtained, Figure 12 shows a pattern in the expected direction (i.e., average time for 2-button device was higher than 3-button and 5-button devices, and 3-button and 5-button devices averages were the same). More importantly, it can be shown that a 2-button device requires more button presses than the other input devices. Given a simple keyboard with only alphabetical characters and a spacebar, the minimum number of button presses to enter the name “BARACK OBAMA” with the 2-button device is 163. In contrast, the minimum number of button presses for the other input devices is 94. Although this difference might not equate to a large difference in time, it increases the physical demands on the user. Moreover, the above calculations assume that the user never overshoots a target. An overshoot with the 2-button device would require an additional 26 button presses to cycle through the alphabet, while an overshoot with the 3-button or 5-button device would require only a single additional button presses.

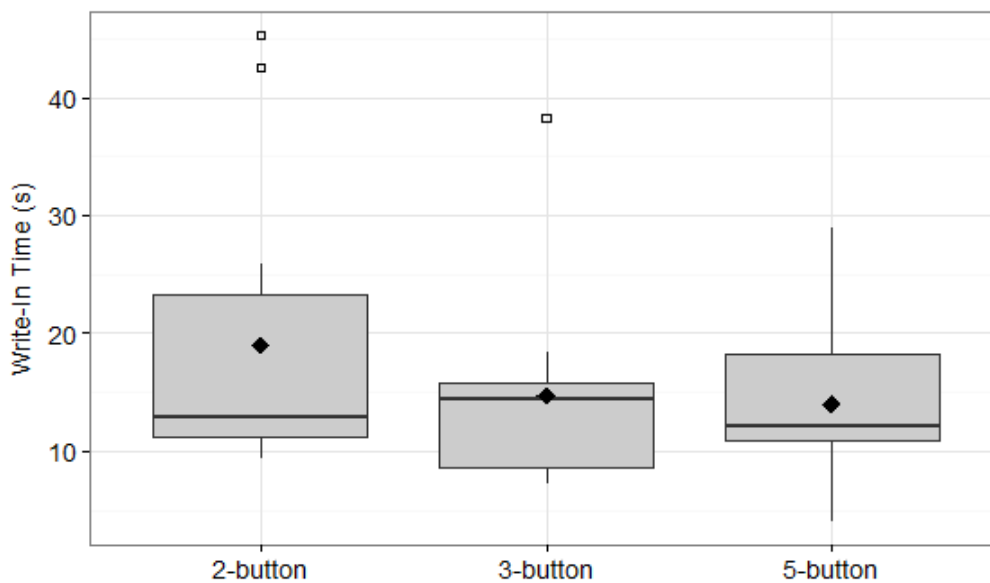


Figure 12. Mean write-in times normalized by the number of letters in the candidate’s name.

Click Time

Mean click time represents the average amount of time elapsed between selections of any on-screen buttons, including navigation control buttons, and candidate names. This is an indication of the amount of time needed to move the cursor to the desired button. Like navigation controls of the onscreen keyboard for the write-in race, the 2-button device was expected to yield a higher mean click time due to its unidirectional character that required participants to cycle through the bottom of the screen to return the cursor to the navigation controls at the top of the screen. However, repeated measures ANOVA indicated that there was no significant difference in mean click time among the three input devices, $F(2, 28) = 0.22$, $p = 0.81$, $\eta_p^2 = 0.015$, power = 0.08.

Response Time

A third measure of navigation speed was the average amount of time elapsed between the onset of a race (i.e., when the list of candidates appeared) and the participant's selection of a candidate. This response time was normalized by the target candidate's location in the list. There was no significant difference among mean normalized response time for the three input devices, $F(2,28) = 0.21$, $p = 0.81$, $\eta_p^2 = 0.015$, power= 0.08. This implies that once participants reach the candidates list, they really only need to use the advance tactile button and the select button to reach the candidates name, which requires only a 2-button device. The additional tactile back button is useful when the user makes a mistake and overshoots or when the user wants to navigate to the on-screen "next" button by moving backward instead of forward through the candidate list, which sometimes reduces button presses.

Eye Tracker Results

Controls Gaze Time

Participants were expected to look at the onscreen navigation controls less for the 5-button device than the other two devices (H4), because they could use the two tactile buttons for page-forward and page-backward instead of the onscreen buttons. Controls gaze time refers to the amount of time participants spent looking at the on-screen navigation control buttons at the top of the ballot (i.e., the on-screen back, help, review and next buttons). An average was computed across the pages of each ballot. Repeated measures ANOVA revealed that there was no significant difference in mean time spent looking at the control buttons among the three devices, $F(2,12) = 0.871$, $p = 0.443$, $\eta_p^2 = 0.127$, power = 0.166, although the pattern of means was in the expected direction (Figure 13). The pattern is consistent with the hypothesis that the 5-button device does not require as much gaze time on the on-screen navigation controls.

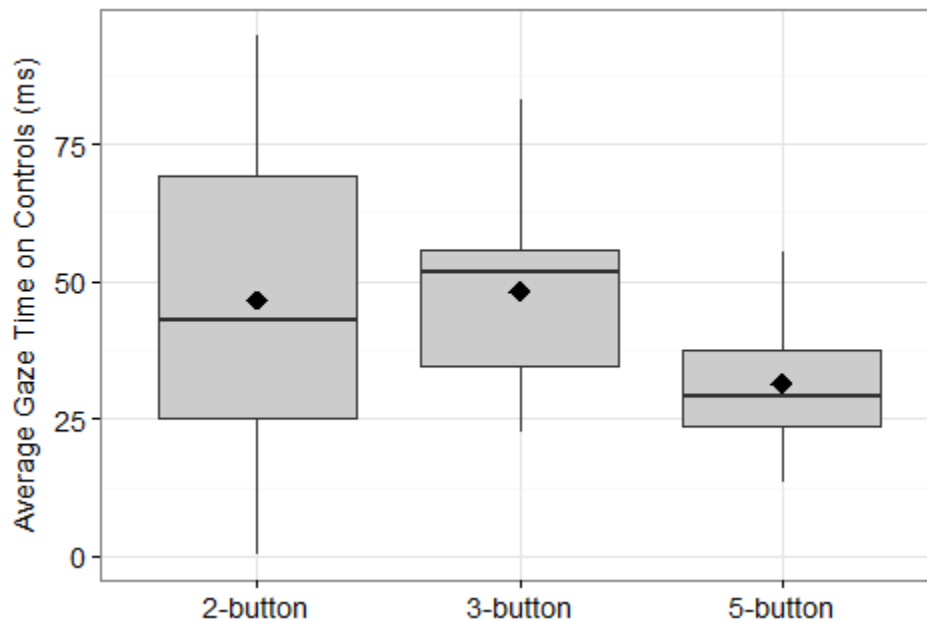


Figure 13. Mean time per page spent looking at the on-screen control buttons.

Candidates Gaze Time

Candidates gaze time refers to the amount of time participants spent looking at all candidate options on the ballot. An average was computed across all races with each ballot. Repeated measures ANOVA revealed that there was no significant difference in mean time spent looking at the selections among the input device types, $F(2,12) = 2.327$, $p = 0.14$, $\eta_p^2 = 0.279$, power = 0.381. However, there was an overall pattern that the candidate gaze time for the 2-button was longer than the 3-button and 5-button devices (Figure 14).

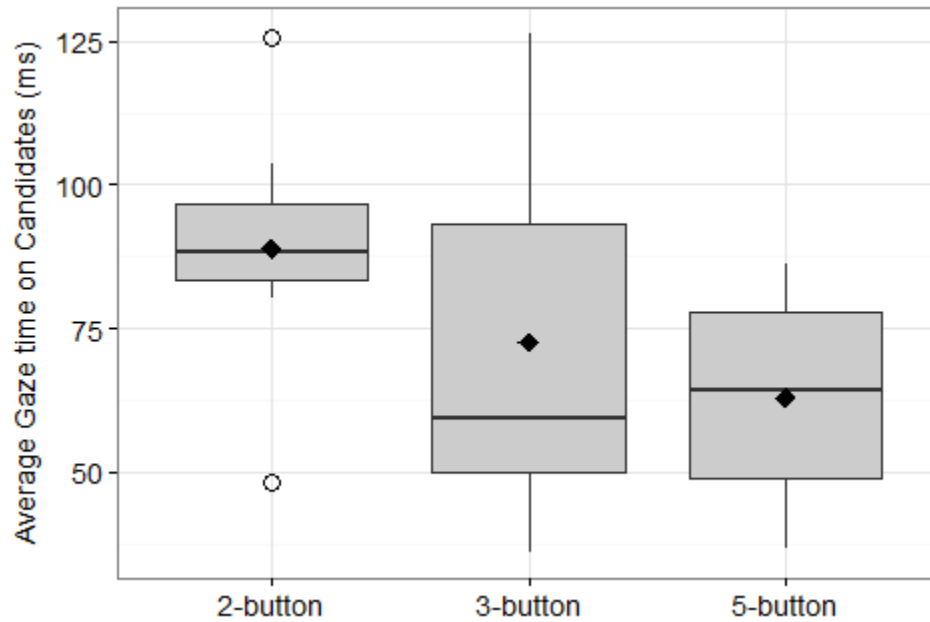


Figure 14. Mean time per race spent looking at all candidates.

Target Candidate Gaze Time

Target candidate gaze time refers to the amount of time participants spent looking at the target candidate. Times were averaged across all races for each ballot. Repeated measures ANOVA revealed that there was no significant difference in the mean time spent looking at the target candidate among the input devices, $F(2,12) = 0.742$, $p = 0.497$, $\eta_p^2 = 0.11$, power = 0.147.

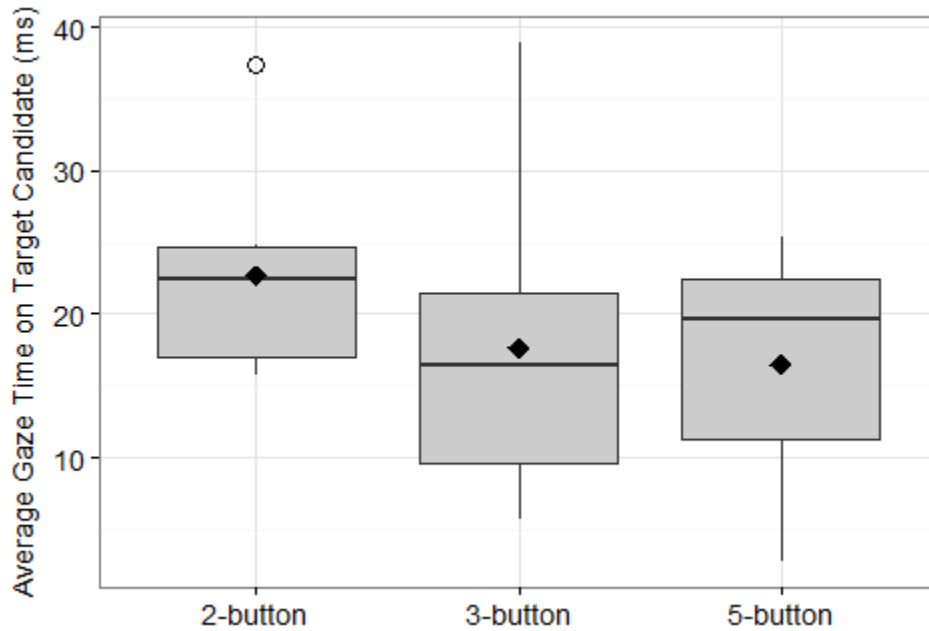


Figure 15. Mean time per race spent looking at target candidate.

Practice Effects

Participants were expected to improve their skills over time during the experimental session. Specifically, the order in which participants used the input devices was expected to affect task completion time and subjective rankings of ease-of-use (H5). Participants were expected to get progressively faster and more satisfied (higher ranking) over time. The primary measure of voting time, average race time, was used in this analysis.

Note that the order in which the input devices were used was counterbalanced across participants, so order of use could not favor any of the three input devices when averaging across participants. Due to counterbalancing of order, the analyses and figure below have no bearing on differences between the 2-button, 3-button, and 5-button devices. Instead, they examine the effect of practice on race time and subjective rankings.

Figure 16 shows the relationship among average race time (y-axis), rankings of the input devices (x-axis), and order of use (dot color/shape). As expected, the effect of order of use on ranking was significant, $F(2,28) = 6.72, p = .004, \eta_p^2 = .324$, indicating that rankings were higher for devices that were used later in the session. This is illustrated in the figure by the distribution of red points (first-used device) and blue points (third-used device). The first-used device was ranked highest by only one participant (the single red circle in the right-most group of points), and the third-used device was often ranked best.

Also as expected, the effect of order of use on average race time was significant; there was a negative correlation between order of use and average race time, indicating that participants became progressively faster¹, $t(42) = 2.47, p = .017, r = -.357$.

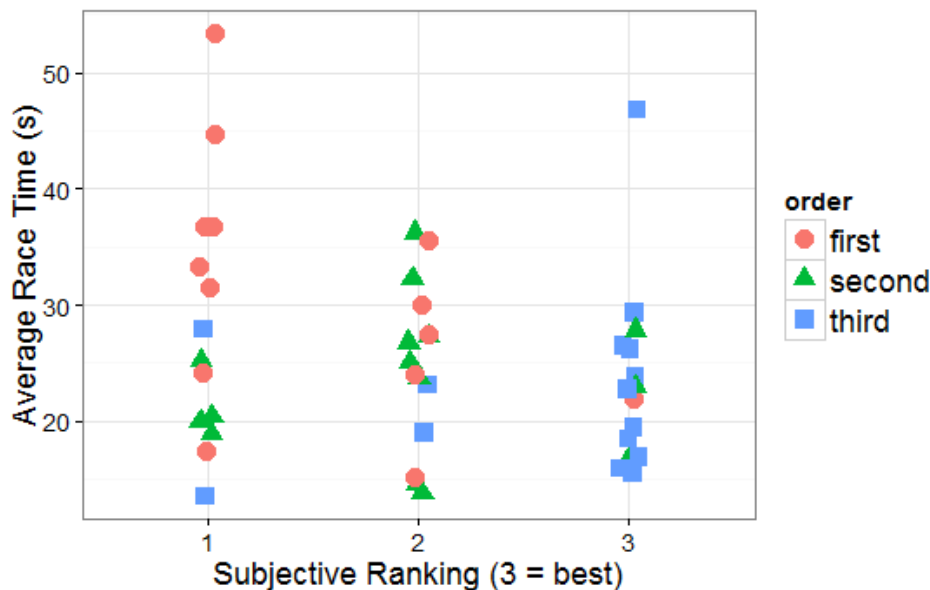


Figure 16. Race time as a function of device preference and order of use. Best fit line excludes the outlier from Figure 10, which is not shown in this figure.

¹ One outlier was omitted. It was 4.7 standard deviations above the mean race time.

Discussion

This study was designed to determine which of the three voting input devices was the best candidate for older adult users with arthritis in their wrists, hands, and fingers. Several of the variables showed trends that might indicate that the 5-button device was the best, these variables included, race time, write-in time, and pain ratings. However, usability ratings, device rankings, and user comments did not support the conclusion that the 5-button device was the best. Researcher observations and subjective reports by participants indicated that the 5-button device was confusing to some participants, and some participants did not take advantage of the two navigation control tactile buttons (page-forward and page-backward).

Participants in this study did not report that their arthritis symptoms caused excessive pain while using the input devices or prevented them from completing any tasks. An effect of input device on pain was not obtained, so there was no evidence to support the hypothesis that the 2-button device would cause more pain.

Some users experienced difficulty due to cognitive issues associated with the additional buttons on the 3-button and 5-button devices. Many of the participants in this study had difficulty deciding which button to press on the 5-button device, and they reported being confused by the buttons. This replicates and extends the findings of Harada et al. (2010) who found that older adults learned to use a smartphone-like application more easily when decision points were removed. It is also consistent with the Hick-Hyman law, which states that response time is dependent upon the complexity of the decision (Wickens, 2004).

Some of the older adults in this study may have been experiencing cognitive deficits associated with age-related cognitive decline (i.e., reduced working memory capacity, difficulty acquiring new skills or learning new procedures). This may partially explain why some participants had considerable difficulty with the input devices – particularly the 5-button device. During an exit interview, participants expressed their confusion in understanding the additional functions of the 5-button device, and in fact four of the participants used the 5-button device exactly the same as if it was the 3-button device (i.e., they did not use the additional two buttons for navigating between pages). Many participants used only two buttons on the 3-button device, or used two or three buttons on the 5-button device, which could account for the lack of significant differences between the input devices.

It was surprising that no statistically significant differences emerged between the 2-button and 3-button devices. The 2-button device's simplicity, which was preferred by some users, might have outweighed the benefits of having another button that moved the cursor backward on the 3-button device. All five of the participants who ranked the 2-button device as their favorite indicated that their rankings were based on the 2-button device's simplicity. This preference for simplicity might not generalize to younger populations.

There was an overall trend that indicated that task completion times were faster for the 3-button than the 2-button, although the differences were not statistically significant. The added benefit of having the

tactile back button for instances when the participant overshoots the target may outweigh the cost of including such an option.

The lack of statistical significance in many of the analyses may have been due to low statistical power, which was caused by the small sample size and small effect sizes. Power did not exceed 0.4 for any of the non-significant analyses – less than half of the typically desired power of 0.8.

Overall, the results revealed design tradeoffs. Trends in subjective pain reports and voting times suggested that the 3-button and 5-button devices were more efficient. In contrast, many participants reported a preference for the 2-button device due to its simplicity. The recommendation is to proceed with a 3-button device in future studies, because it represents a compromise between the simplicity of the 2-button device and efficiency of the 5-button device.

Acknowledgments

The authors would like to thank Dr. Cara Bailey Fausset and Ms. Hannah Jahant for their contributions in reviewing and editing this manuscript.

References

- Hall, T., & Snoot, L. W., (2013). Voting Machines: The Question of Equal Protection. In M. J. Streb (Ed.), *Law and Election Politics: The Rules of the Game*, 2nd ed. 4, 81 – 87, New Your, NY: *Lynn Renier Publishers*.
- Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. [Article]. *Cognitive Psychology*, 44(4), 339-387.
- Harada, E. T., Mori, K., & Taniue, N. (2010). Cognitive aging and the usability of IT-based equipment: Learning is the key. [Article]. *Japanese Psychological Research*, 52(3), 227-243.
- Harley, L.R., Mosley, S., Jones, A., Baranak, A., Kline, K., and Fain, B. (2013). A web based voting application software design. *ITIF Working Paper*.
- Harley, L. R., Kline, K., Ray, J., Bell, C., Baranak, A., Price, C., Hung, M., and Fain, B. (2013). The evaluation of a voting web based application. [Article]. *HCI International*.
- Herrnson, P. S., Niemi, R.G., Hanmer, M.J., Bederson, B.B., Conrad, F.G., Traugott, M.W. (2008). *Voting technology: The not-so-simple act of casting a ballot*. Washington, D.C.: Brookings Institute Press.
- Lawrence, R. C., Helmick, C. G., Arnett, F. C., Deyo, R. A., Felson, D. T., Giannini, E. H., . . . Wolfe, F. (1998). Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis & Rheumatism*, 41(5), 778-799.
- Morbidity and Mortality Weekly Report (2013). Prevalance of doctor-diagnosed arthritis and arthritis-attributable activity limitation – United States, 2010 – 2012, *Centers for Disease Control and Prevention MMWR* 62(44).
- Sauro, J. (2011). Measuring usability with the system usability scale Retrieved February 25, 2013, from <http://www.measuringusability.com/sus.php>
- Wickens, C. D., Lee, J.D, Liu, Y., Gordon Becker, S.E. (2004). *An Introduction to Human Factors Engineering* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.