

How Hard Can It Be to Place a Ballot Into a Ballot Box? Usability of Ballot Boxes in Tamper Resistant Voting Systems

M. Grant Belton

Undergraduate,
Department of Cognitive
Science
Rice University
6340 Main St.
Houston, TX 77005
USA
gbelton@rice.edu

Philip Kortum

Assistant Professor
Rice University
Department of Psychology
6100 Main Street MS25
Houston, TX 77005
USA
pkortum@rice.edu

Claudia Acemyan

Postdoctoral Research Fellow
Rice University
Department of Psychology
6100 Main Street MS25
Houston, TX 77005
USA
claudiaz@rice.edu

Abstract

End-to-end verifiable voting methods are an emerging type of voting system, and a number of new designs are being actively developed. Many of these systems try to mirror current paper voting methods and use a paper ballot that can be scanned and then placed into a ballot box. Previous research has shown that having separate scanner and ballot box components lead to failures, as many users fail to either scan or place their ballot in the box, both of which are required to cast a vote that will be counted. In our study, we examined the usability of two different ballot box configurations designed to eliminate these kinds of errors. Results showed that both configurations were equally usable, but that specific design aspects of the scanner itself significantly affected the ability of voters to cast their votes, with only 37.5% of the voters able to do so. Based on the results of this research, implications for usable ballot box design are discussed.

Keywords

voting, casting ballots, scanning, system response time, security, end-to-end voting systems



Introduction

The act of voting is an essential part of participating in a democratic system. After the Florida butterfly ballot problem in the 2000 presidential election (Wand et al., 2001), new concerns arose about the general usability of U.S. voting systems. Since that time, significant research has been conducted to characterize the usability of current voting systems (Everett, Byrne & Greene, 2006; Green, Byrne & Everett, 2006; Byrne, Greene & Everett, 2007; Everett et al., 2008) and to more fully understand users' behaviors in voting and the kinds of errors they make, with the intention of mitigating these errors in future designs (Everett, 2007; Campbell & Byrne, 2009).

The most iconic component of a voting system is the ballot box. In a classic voting system, users fill out a piece of paper and drop it into a slot on the top of a ballot box. While simple and straightforward, the design is vulnerable to a host of issues like ballot stuffing, lost or stolen ballot boxes, and the inability to count votes in a timely manner—rendering them insufficient to meet current needs. New systems aim to address these concerns through the use of electronic vote recording. Current research is focusing on what is termed *end-to-end* (e2e) systems that often use paper ballots linked to an electronically filled form. Nonetheless, the user experience remains similar: voters fill out a ballot (on a computer in this case), obtain the printed form of their completed ballot, and place that paper in the ballot box.

This type of system allows for a very high level of security, but necessitates a more complicated ballot box that includes a scanner to record each ballot being cast. Previous work done by Acemyan, Kortum, Byrne, and Wallach (2014) demonstrated that visible scanning systems in which the scanner is completely separate from the ballot box are not usable by the general public. Users often ignored or bypassed the scanning mechanism and placed the ballot directly in the ballot box. In e2e systems, those votes are usually not counted because in scanning the ballot, the vote counting computer and ballot box communicate to verify each ballot as it is cast.

Our hypothesis was that hiding the scanner inside the ballot box would make it easier and more usable for the average voter, as it would most closely resemble and function like a traditional ballot box, with the benefit of seamlessly and unavoidably scanning the ballot as it is cast. Moreover, by concealing the scanner in the box, voters who are uncomfortable using devices like scanners might be put at ease.

Our team compared two integrated scanner and ballot box designs: an all-in-one ballot box with an integrated scanner and a two-piece system with a visually distinct scanner box sitting directly over the ballot box. These two configurations were chosen to investigate if completely hiding the scanner was more beneficial than a configuration that was visible to voters and accessible for scanner maintenance on polling day.

Methods

The following sections discuss the study design, participants, materials, and procedure used in the study.

Study Design

The study design was between-subjects to avoid any potential learning effects. The independent variable was the physical configuration of the ballot scanner, either as a separate box placed on top of the ballot box or as part of a single-piece ballot box with an integrated scanner combo. Other than the ballot box design, both voting systems were identical.

Per ISO 9241-11 (ISO, 1998), the dependent variables were task completion time (efficiency), the number and type of errors (effectiveness), and participants' satisfaction ratings for each of the devices using the System Usability Scale (SUS). The SUS was based on Brooke's usability scale (1996), modified to be specific to the voting system and with the word *cumbersome* on statement 8 changed to *awkward* (Bangor, Kortum, & Miller, 2008). The change of the word *system* to be more specific has been shown to have no impact on the reliability of the instrument (Sauro, 2011).

Participants

A total of 33 participants completed this study. The participants were Rice University undergraduates recruited through the university's participant pool. All were between the ages of 18 and 22 and residents of the United States (i.e., eligible voters). Each received credit towards a course requirement for their participation. All participants reported spending at least five hours per week on a computer. All were either familiar with or aware of the process of sheet-feed scanning.

Materials

The following sections contain information about the ballot boxes, ballots, and voting receipts used in the study.

Ballot boxes

This study centered on a pair of ballot boxes with integrated scanners. The prototype ballot boxes were constructed from standard corrugated cardboard. A single slot was cut in the top of each ballot box system, 9¼ inches long and ½ inch wide. Secured directly below the slot was a scanner that fed into the compartment below. All seams were sealed with clear packing tape.

The first ballot box consisted of a single white box with the scanner integrated into the actual box (Figure 1). This all-in-one (AiO) box was marked "Ballot Box" with a paper label affixed to the front. Exterior dimensions were 18 inches wide x 12 inches deep x 12 inches high. It was centered on top of a standard 29 inch high desk, with the front face of the box 7 inches back from the front edge of the desk. The bottom of the box remained unsealed to allow the experimenter to access the scanner controls and retrieve cast ballots, while maintaining a fully sealed appearance to participants.

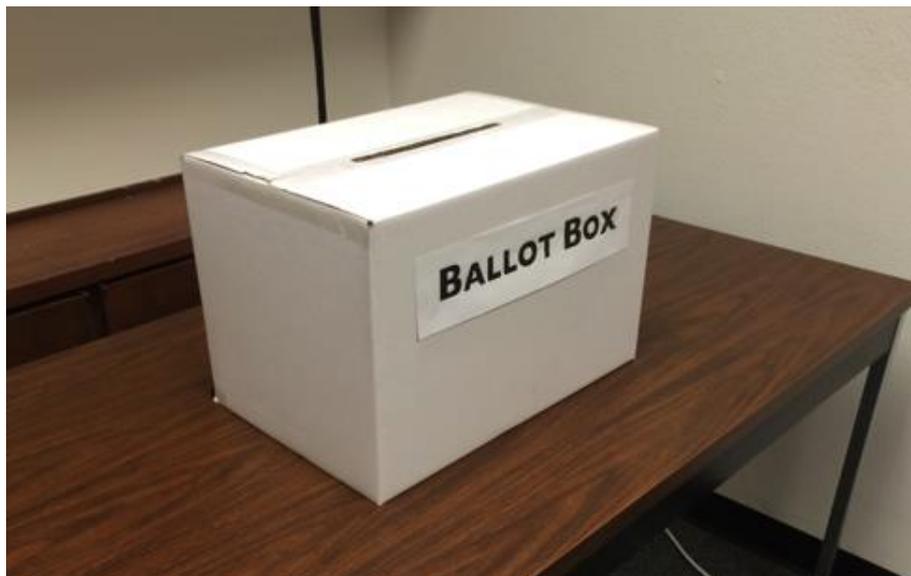


Figure 1. Ballot Box 1: all-in-one.

The second ballot box consisted of a pair of boxes stacked on top of one another and secured to one another by internal hidden clips. The top box contained the slot and scanner. Its coloring was natural cardboard to contrast with the box below. The bottom box was white and had a "Ballot Box" label, identical to the AiO label, placed on it. The 3 inch height of the top box was dictated by the size of the enclosed scanner. Overall dimensions of the two-box setup remained the same as the AiO ballot box. Placement on the desk was likewise identical. This configuration would allow a broken scanner to be replaced during polling without having to access the contents of the ballot box. Such a design would be more secure, and thus advantageous, to polling officials in charge of security.



Figure 2. Ballot Box 2: discrete scanner unit and ballot box.

The system that we are developing is intended to be constructed from commercial off-the-shelf (COTS) hardware wherever possible (Bell et al., 2013). We used a VuPoint Solutions Magic Wand scanner (Figure 3). We selected this highly rated commercial scanner because it would automatically feed and scan sheets of paper inserted by the user without the need to press buttons or otherwise interact with the scanner. This model scans a single side of the paper: the top of the paper when the scanner is placed in its designed configuration, or the side of the paper facing the user when installed in the ballot boxes, as we did. The scanner performed in the following manner: When a piece of paper was placed motionless over the sensor (indicated by the arrow) for approximately .5 seconds, it began to feed the paper into the scanner by means of intake feed rollers. After feeding a 0.25 inch of paper, it paused the feed for .5 seconds, checked again for the presence of paper over the sensor, then fed continuously until 1 second after the sensor was cleared. If, during the initial or secondary sensing event, the paper was in motion or not in contact with the sensor, the process was restarted from the beginning. The paper fed at a rate of just over 1inch per second, corresponding to 10 seconds for a letter-sized ballot.

Scanner performance had to be as reliable as we could achieve with the given hardware. Every effort was made to find the slot geometry that provided the best paper-sensor contact without interfering with the sensor's operation.

The scanner was secured by means of cardboard brackets that held it motionless in relation to the slot. A flap of the same material as the box was fitted inside the slot opposite the scanner feed mechanism in an attempt to guide ballots into contact with the intake rollers and paper sensor, as well as seal alternative points of ballot ingress.



Figure 3. Detail of installed scanner and slot, user perspective and cutaway.

Ballots

We used two functionally identical mock ballots, differentiated only by names and party affiliations to allow for consistency between this study and future studies in which users will use all portions of an e2e, voter verifiable system. At the top and bottom of each page was a barcode that, when scanned, identified the ballot to the system, which allowed the ballot-marking computer to confirm a given ballot as having been cast. For future study ballot specifications, ballots (Figure 4) will be single-sided.

Official Ballot
November 4, 2014, General Election
Harris County, Texas Precinct 101A

1 of 1 | # 884547 008 92014102-018-001-104 | 10/20/14 08:11:03 PM

President and Vice President Gordon Bera and Nathan Adrian REP	Presiding Judge Texas Supreme Court Place 3 DEM Tim Grady
United States Senator Cecile Cadieux REP	Presiding Judge Court of Criminal Appeals Place 2 REP Dan Plouffe
Representative in Congress Petro Brouss REP	District Attorney REP Cony Behrke
Governor Glen Travis Lott REP	County Treasurer REP Dean Caffee
Lieutenant Governor Shane Terrio REP	Sheriff LIB Jason Valle
Attorney General Tim Speight REP	County Tax Assessor CON Randy H. Clemons
Comptroller of Public Accounts Greg Converse DEM	Justice of the Peace Deborah Kamp
Commissioner of General Land Office Sam Sadler REP	County Judge Dan Atchley
Commissioner of Agriculture Roberto Aron DEM	Proposition 1 No
Railroad Commissioner Jillian Balas REP	Proposition 2 Yes
State Senator Ricardo Nigro REP	Proposition 3 No
State Representative District 134 Petra Bencomo REP	Proposition 4 Yes
Member State Board of Education District 2 Peter Varga REP	Proposition 5 Yes
	Proposition 6 No

Put 101A-BD-048602W_860-1142-4100-02/011eed56-BCD 345784 (p 1 of 1)

Figure 4. Sample Republican ballot.

Voting receipt

The voting receipt in this system served two purposes. First, it allowed voters to verify after the election that their vote was counted. Second, it served as a privacy measure for the voter: During the printing process, the receipt was placed on top of the completed ballot so that when voters left the privacy of the voting booth to cast their ballot, the selections on their ballot was not visible to bystanders.



Figure 5. Voting receipt.

Procedure

The study began with participants giving their IRB approved informed consent. Participants were then read instructions for the experiment informing them that they would be using one component of a multi-step voting system. In this study they were asked to pretend that they had just finished filling out their ballot on a touchscreen voting station, confirmed their selections by pressing a "Finish and Print Ballot" button onscreen, and a printer in the voting booth had just printed their ballot and receipt. In the output tray of the printer was the ballot, with the receipt on top.

Participants were instructed to go to the voting station and read the onscreen instructions consisting of the text, "To cast your ballot, place your printed ballot in the ballot box. Keep your receipt." The ballot box was located on a table opposite the voting station, approximately six feet away via an unobstructed path. Task time started when the participant either sat in the chair, started reading the screen, or reached for the ballot, whichever event occurred first.

Participants were told that because we were trying to understand how voting systems work, we would be unable to help them while casting their ballot. If they ran into trouble, the participants were instructed to, "Please do whatever you think you should do to cast your ballot." When they were done voting or wished to stop, they were instructed to simply say so. Task time was stopped when the participant signaled that they were done. This was determined either by the participant stating that they were finished or when they walked away from the ballot box, at which time the experimenter then asked if they were finished. They were then asked if they felt they had successfully cast their ballot. If the participant didn't feel that they had done so, or there was ambiguity, they were then asked to explain their thoughts about the process.

There was no specific time limit defined, and all participants either completed the task or gave up after a maximum of 2 minutes, 33 seconds.

The only time the experimenter stepped in was in cases where we felt the participant might cause damage to the experiment materials, saying, "Please be gentle, the ballot box is only made of cardboard," or in two cases requested that participants with long hair be careful to keep it away from the scanner's feed mechanism.

Feed attempts were classified as when a part of the ballot was inserted past the plane of the top of the box, and another feed attempt was tallied when either the orientation of feed was changed or the ballot was moved above the reference plane and reinserted.

Orientation of the ballot was recorded as each feed attempt was made. In the event of multiple orientations being tried, the final ballot orientation in the event of a successful feed was noted, as was the location of the ballot at the end of the test in the event of a failure. Specific patterns of folding were recorded for later analysis in determining ideal marking locations for folded ballots, and the ballots were saved for later examination. After the task portion was finished, participants were taken to an adjoining room and asked to fill out the SUS. They were debriefed about the study and offered the chance to ask questions.

Ballot boxes and ballots were switched out between trials to maintain an equal distribution of each combination as the study progressed.

Results

We used 32 of the 33 participants' data in the analysis. One participant's data (from the ballot box 2 condition) were excluded due to a mechanical failure of the scanner that occurred during the session, leaving 16 participants in each condition. There was no evidence to support differences in efficiency, ($t(30) = 1.28, p = .212$), effectiveness ($\chi^2(1, N = 32) = 1.2, p = .273$), or satisfaction ($t(30) = 1.18, p = .247$) across the two ballot box configurations. Refer to Table 1 for the results as a function of ballot box design. Because there were no statistically significant differences, results for effectiveness, efficiency, and satisfaction are reported for the entire data set.

Table 1. Effectiveness, Efficiency, and Satisfaction as a Function of Ballot Box Design

Ballot Box 1		
Effectiveness	Efficiency	Satisfaction
25% of ballots were cast	87 s	48.91
Ballot Box 2		
Effectiveness	Efficiency	Satisfaction
50% of ballots were cast	72 s	58.59

Note. effectiveness = percentage of cast votes, efficiency= mean time on task in seconds, and satisfaction = mean System Usability Scale rating.

Effectiveness

Participants were able to successfully cast a ballot only 37.5% of the time in the tested configuration (simplex scanning of a single-sided ballot). Failures included improper orientation of the ballot for successful scanning, folding of the ballot, and the inability to feed the ballot through the scanner.

Efficiency

The mean time to cast a vote was 79 seconds, with a standard deviation of 34 s, a 95% confidence interval from 67 s to 92 s, and a range of 26 s to 153 s. Differences in success with operating the feed mechanism strongly accounted for the disparity in task completion times.

Satisfaction

The mean SUS score used to measure satisfaction was 53.75, with a standard deviation of 23.34, a 95% confidence interval from 45 to 62, and a range of 2.5 to 90. This rating places it in the *low marginal* category as defined by Bangor, Kortum, and Miller (2009).

Discussion

This study did not provide evidence to support statistically reliable differences in the usability between the two tested ballot box configurations. We did however find a significant flaw in the usability of the operation of the devices that had not been previously anticipated.

It is clear that this study's failure rates in casting a ballot with these systems would be unacceptable had this been a real election. While the participants in this study are not fully representative of the U.S. voting population (although our participants were younger adults with fewer chances to vote, they were well educated, highly motivated and compensated, technologically sophisticated students), they are still eligible voters. A more representative sample of the voting population would clearly give a more accurate picture of the kinds of failure rates that would be observed in a real election.

Because previous research on standard ballot boxes has shown that nearly everyone is successful in putting a paper ballot into a standard (non-scanner) ballot box (Byrne et al., 2007), why did these tested systems fail so badly? The most common errors can be divided into two general categories: those related to the properties of the scanner, and those caused by the user's behavior. The first category had the highest rate of occurrence. While the appearance of the ballot boxes had no meaningful impact on the performance or satisfaction measures, in the process of testing our designs it became readily apparent that another design parameter played a huge role in the failures observed. The particular scanner chosen for the study, while highly rated and functional as a standalone device, became far less well suited to the task of scanning ballots when it was part of a ballot box.

Every participant tried more than once to feed his or her ballot. The timing of the feed mechanism was nearly perfect to elicit the worst possible reaction from the participant. In almost every case the person would place their ballot into the slot; wait a moment for it to feed in; and at the exact moment that the scanner actually began to feed the paper, the participant would pull the ballot away from the feed rollers; adjust the ballot in some fashion, then reinsert it. For those who did manage to leave it in place long enough to start the feed sequence, many pulled the ballot out of the rollers immediately after the feed paused or moved the unfed portion of the ballot towards themselves to view the intake rollers, thereby taking it out of contact with the sensor and stopping the feed sequence.

Participants who placed their ballots correctly and kept them motionless at the proper times were able to feed the ballots with minimal struggle. Those who moved the ballots around more after the first attempt at feeding it through the slot had much higher failure rates in casting their ballots. Numerous participants tried a number of times to feed the ballot into the scanner, failed, and either jammed the ballot into the single un-taped seam between the scanner and box ($n = 4$), or simply gave up and left the ballot outside of the box ($n = 7$).

Errors resulting from user behaviors, the second category, were influenced to some extent by the first, but not entirely. The primary user-generated error was that of folding the ballot, either before attempting to place it in the box ($n = 3$) or after one or more unsuccessful attempts at feeding it unfolded ($n = 21$, three participants ultimately unfolded the ballot and fed it successfully). Participants feeding the ballot with the print facing away from the scanner element caused a smaller subset of failed ballots to not be counted. However, even if we assume that the scanner had more advanced properties that would handle ballots inserted in the wrong orientation, success rates were still unacceptable. If the scanner were capable of simultaneous duplex scanning (allowing a user to put the ballot in forwards or backwards), success rates would have risen to 50%. If we assume further that the ballot design included a barcode on each corner of the ballot, front and back, so that a single-page ballot could be folded once in either or both lengthwise or widthwise configurations and still allow a duplex scanner to extract information contained on the ballot, success rates would have risen to only 65.6%. Only one participant failed to cast the ballot and instead cast the receipt, and one other participant folded both the receipt and ballot together and cast them simultaneously.

As a consequence of every participant attempting to feed the ballot at least once within the first 30 s, some ended up spending over two minutes struggling with the feed mechanism. Even with the complete feed time of approximately 10 s, the expected time on task would not exceed 40 s with a feed system that reliably fed on the first attempt.

The authors have identified a number of critical attributes in scanning hardware and the ballot box. For scanning systems integrated unobtrusively into the ballot box, certain conditions should be met. The feed mechanism should engage immediately upon paper being inserted into the ballot box and draw it into the box with a steady, predictable motion. The scanner should be able to scan both sides simultaneously to allow for variations in ballot orientation when placed in the machine. As ballots may exceed a single page in length, the ability to handle multiple sheets simultaneously will also be a relevant consideration. In addition, the system should be designed so that users are not required to have a mental model for the interior scanner mechanism in order to be able to cast their ballot. Instead, voters should be able to quickly insert their ballot in the box, which is one of the typical procedures to cast a vote in an election.

Many participants also clearly had ballot folding as part of their voting mental model. Some mechanism(s) should be implemented to handle folded ballots, either by accepting folded ballots with some way of extracting the necessary information or, if that is not technically viable, by informing the user not to fold the ballot, or creating ballot feeding mechanisms that encourage the mental model of an unfolded ballot.

In addition, the ballot box would ideally provide mechanisms to return non-castable materials (e.g., ballots that do not successfully scan, receipts, or other non-ballot papers) to the user and indicate to them that those materials were not accepted as valid voting material. This would allow folded ballots to be unfolded and successfully cast, as well as keeping the ballot box free of extraneous material.

While this study did not identify differences in the usability of the two tested ballot box configurations, a latent and unexpected design parameter associated with the scanners' operation dominated the failure modes for this testing. While important for the design of this specific system, it also points out the need to carefully test all aspects of complex systems to ensure that the systems are usable in their operational configurations.

Tips for Usability Practitioners

Our research identified a number of potential issues to keep in mind when designing a study of voting with a scanner and ballot box:

- Scanners should be integrated into the ballot box so voters cannot skip scanning their ballot.
- Scanners should be concealed from voters (i.e., placed inside the ballot box) because previous research has shown that having to operate or use a scanner deters and/or intimidates some voters.
- The exact configuration of the ballot box is less important than the operational parameters of the scanning mechanism, such as feed rate and the delay before the scanner starts when a ballot is inserted.
- Feed mechanisms should respond quickly to the presence of a ballot. Feeds should also be continuous once they begin.
- A physical means of supporting and guiding the ballot should be employed so that the user can relinquish control during the feed process.
- The scanning system should be able to handle multiple ballot orientations.

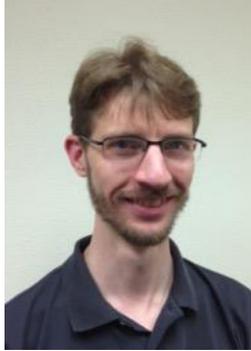
Acknowledgements

This research was funded in part by NSF-CNS-1408401. We would like to thank Dan S. Wallach, Michael D. Byrne, and the other members of the STAR-Vote team for their extensive input on the development of the system.

References

- Acemyan, C. Z., Kortum, P., Byrne, M. D., & Wallach, D. S. (2014). Usability of voter verifiable, end-to-end voting systems: Baseline data for Helios, Prêt à Voter, and Scantegrity II. *The USENIX Journal of Election Technology and Systems (JETTS)*, 2(3), 26–56.
- Bangor, A., Kortum, P., & Miller, J. A. (2008). The System Usability Scale (SUS): An empirical evaluation. *International Journal of Human-Computer Interaction*, 24(6), 574–594.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114–123.
- Bell, S., Benaloh, J., Byrne, M. D., DeBeauvoir, D., Eakin, B., Fisher, G. ... , & Winn, M. (2013). STAR-Vote: A secure, transparent, auditable, and reliable voting system. *The USENIX Journal of Election and Technology and Systems (JETTS)*, 1(1), 18-37.
- Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189–194). London, England: Taylor and Francis.
- Byrne, M. D., Greene, K. K., & Everett, S. P. (2007). Usability of voting systems: Baseline data for paper, punch cards, and lever machines. *Human Factors in Computing Systems: Proceedings of CHI 2007, USA*, 171–180.
- Campbell, B. A. & Byrne, M. D. (2009). Now do voters notice review screen anomalies? A look at voting system usability. *Proceedings of the Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE'09), USA*, 1–14.
- Everett, S. P. (2007). *The usability of electronic voting machines and how votes can be changed without detection* (Unpublished doctoral dissertation). Rice University, Houston, TX.
- Everett, S. P., Byrne, M. D., & Greene, K. K. (2006). Measuring the usability of paper ballots: Efficiency, effectiveness, and satisfaction. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, USA*, 2547–2551.
- Everett, S. P., Greene, K. K., Byrne, M. D., Wallach, D. S., Derr, K., Sandler, D., & Torous, T. (2008). Electronic voting machines versus traditional methods: Improved preference, similar performance. *Human Factors in Computing Systems: Proceedings of CHI 2008, USA*, 883–892.
- Greene, K. K., Byrne, M. D., & Everett, S. P. (2006). A comparison of usability between voting methods. *Proceedings of the Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE'06), USA*, 1–7.
- ISO, W. (1998). *I Ergonomic requirements for office work with visual display terminals (VDTs)*.
- Sauro, J. (2011). *A practical guide to the System Usability Scale: Background, benchmarks, & best practices*. Denver, CO: Measuring Usability LLC.
- Wand, J. N., Shotts, K. W., Sekhon, J. S., Mebane Jr, W. R., Herron, M. C., & Brady, H. E. (2001). The butterfly did it: The aberrant vote for Buchanan in Palm Beach County, Florida. *American Political Science Review*, 95(4), 793–810.

About the Authors



M. Grant Belton

Mr. Belton is an undergraduate in the Department of Cognitive Sciences. His current area of study focuses on improving the usability of emerging technologies for use in limited-resource settings.



Philip Kortum, PhD

Dr. Kortum is a faculty member in the Department of Psychology at Rice University in Houston, Texas. His research is focused on the development of user-centric systems, with an emphasis on voting and mobile computing systems.



Claudia Ziegler Acemyan, PhD

Dr. Acemyan is a post-doctoral research fellow in the Department of Psychology at Rice University. Currently, her research focuses on the usability and mental models of voting systems.