Assessing the Usability of the Hart InterCivic eSlate During the 2016 Presidential Election

Claudia Ziegler Acemyan and Philip Kortum
Human Factors Research Lab, Department of Psychology, Rice University
6100 Main Street, MS-25, Houston, TX 77005, USA
{claudiaz, pkortum}@rice.edu

For the first time, the system usability of the eSlate--one of the most ubiquitous electronic voting machines used in US elections--was assessed in the field on election day 2016 using ISO 9241-11 metrics. Effectiveness and satisfaction measures were collected immediately after voters used the eSlate to vote at their assigned polling location. The findings showed that the perceived usability of the system in the polling station was judged to be quite good, despite previous research that indicated usability difficulties with the voting system. It is hypothesized that voters were making retrospective assessments of not just the eSlate, but also the supporting systems and environment in which they used it on election day. In particular, voters were likely taking into account their *entire* experience—including the help of poll workers and circumstances like short lines—versus just their experience with the machine itself. Practitioners and researchers should be aware that a system might not be completely usable but still be subjectively rated as usable due in part to other supporting systems and environmental variables.

INTRODUCTION

"No county in America uses more eSlate voting machines than Harris County" (Bernstein, 2008). While much has been written about potential security issues surrounding the eSlate (Bannet, Price, Rudys, Singer, & Wallach, 2004; Inguva, Rescorla, Shacham, & Wallach, 2007; Proebstel et al, 2007), the usability of the system (shown in Figure 1) is less clear as there have been conflicting reports. Advocates for the eSlate's usability predominantly include the system developer and voting officials. For example, Hart InterCivic posts on their website that voters rated the eSlate "extremely easy to use" in a survey that Hart claims was "independently administered" (2005). The offices of election officials in other counties have posted on their websites that they chose the eSlate for elections because the method is easy to use (e.g., San Mateo County, 2017). Yet, it is not clear how the officials came to their conclusions.

In contrast, there is evidence, both anecdotal and experimental, suggesting that the eSlate is not particularly usable. For example, there have been suggestions that the elderly need to practice with the eSlate before voting, because older people have a hard time using the voting machines--a probable consequence of the digital divide (e.g., Herrnson, 2008; Loy, 2006).

Anecdotal evidence comes from observed poll worker behavior during the Harris County 2016 presidential election. Specifically, the authors of this paper observed poll workers announcing to voters

Figure 1: A Hart InterCivic eSlate (Photo by Joseph Lorenzo Hall, used by permission CC by 3.0)

waiting in line that voters must press the large red button to cast a ballot; voters were warned that if they did not press the button and then see the American Flag wave on the eSlate screen, then their ballot would not be cast. The need to provide continuous warnings to voters waiting to enter a voting booth indicates that there may be a system usability problem. This problem is common enough that it has a name: the fleeing voter. Voters are completing their ballots and reviewing them; however, they are not completing the final step of casting their ballot by pressing the red button. This post completion error,

described by researchers nearly 20 years ago (Byrne & Bovair, 1997) can be traced directly to deficiencies in interface design that cause users to commit the error.

Third, in *Voting Technology: The not-So-Simple Act of Casting a Ballot*, Herrnson et al.'s (2009) findings revealed that older voters found the eSlate's wheel and button physically difficult to use. Fourth, and the strongest evidence for the eSlate's lack of usability, comes from a large usability study conducted with over 1500 voters in 2008 (Herrnson, 2008). In this study, Herrnson and colleagues (a group of political scientists) found that the eSlate was rated 4.1 on a 7-point scale for ease of use (58%), and 4.18/7.0 (60%) for comfort using the system. These ratings were the lowest of the scores for the six systems they tested, which included some of the most widely used electronic systems: Diebold AccuVote-TS, ES&S Model 100, Avante Vote-Trakker, Zoomable TS, and Nedap LibertyVote. Participants also reported that using the eSlate's wheel and buttons to complete their ballots required more work than the other electronic systems, making the process slow and cumbersome. Moreover, 35.9% of participants requested and received help to complete their voting task for the mock election. As can be seen from these examples, despite the system developer's and voting officials' claims that the eSlate is easy to use, all voters (i.e., everyone in the U.S. who is 18 years old or older who participates in elections) might not come to the same conclusion.

Despite numerous types of voting systems—such as paper ballots, punch cards (Byrne, Greene, & Everett, 2007), electronic voting machines (i.e., DREs [Everett et al. 2008]), newer end-to-end voting methods (Acemyan, Kortum, Byrne, & Wallach, 2014 and 2015a)—having been tested in mock elections to assess system usability, all of these voting systems have not been assessed during a real, large-scale election using widely accepted, rigorous measurements. More specifically, ISO 9241-11 recommends measuring a product's effectiveness (i.e., can users perform the task without making errors?), efficiency (i.e., how long does it take users to complete the task on the product?), and satisfaction (i.e., do users perceive the system to be usable) when testing system usability (1998). This lack of information on real voters in actual elections is potentially problematic, because it is not known if human performance and behavior might differ from controlled, experimental studies. Furthermore, it is beneficial to usability research to replicate research over time to confirm findings.

Counties are already spending a great deal of money on the eSlate and using the systems in elections despite potential usability issues that could lead to longer voter times (and thus longer lines), lost votes due to voters failing to complete the voting process, and mistakes made by voters while making selections on ballots (Acemyan, Kortum, Byrne, & Wallach, 2015a). In turn, these usability problems can lead to severe, negative ramifications such as altered election outcomes, disenfranchised voters who avoid using the system, and a loss of trust in the election results--all threats to the integrity of elections, and ultimately democratic processes (Acemyan et al., 2014).

In response to this problem, this study assesses in the field the usability of the eSlate at polling stations in Harris County, Texas on presidential election day 2016. The findings from this study can be shared with the voting community, usability practitioners, researchers, and election officials who are the ones developing, selecting, and setting policies for future election systems.

METHODS

Participants

Participants were 88 Harris County, Texas voters who cast ballots on November 8, 2016, which was presidential election day. Voters were recruited as they were walking out of their polling station. Data was collected at eight different polling locations across neighborhoods associated with various socioeconomic statuses. This sample was representative of Houston area voters.

Research Design and Measures

The study was a between-subjects design with a single condition. Each participant voted on the eSlate in their assigned polling station on election day.

Two dependent measures were collected per ISO 9241-11 suggested metrics for assessing usability: effectiveness (i.e., perceived errors) and satisfaction (i.e., subjective usability per Byrne, Greene, & Everett, 2007, to give one example). Both measures were collected through the administration of a short survey after participants voted.

To assess effectiveness, participants were asked to respond to the item, "I did not make any mistakes while voting" on a 5-point Likert scale, with 1 associated with having a complete lack of confidence and 5 being extremely confident. Errors were assessed through post-voting self-assessments because it is not possible to watch people vote and/or look at their ballots due to voter privacy requirements, which are key to the U.S. democratic process. Not being able to more accurately assess error rates (or collect voting times to measure efficiency per ISO 9241-11) is a disadvantage of performing voting usability assessments in the field

during an actual election. In addition, since the length of ballots vary across precincts within counties and across voting studies, it would be difficult to make meaningful comparisons for time to vote. With all this being said, one could argue that being able to assess the usability of a system during a large-scale election outweighs these types of disadvantages.

Voters' subjective usability assessments were captured using the Modified System Usability Scale , or SUS (Bangor, Kortum, & Miller, 2008). The SUS (Brooke, 1996) is a 10-item survey that measures the overall usability of a system. Because it is technology agnostic, it can be used to measure the usability of a wide variety of products, services, and systems (Kortum & Bangor, 2013). The measure has well established reliability and validity metrics (Bangor, Kortum, & Miller, 2008; Sauro, 2011), which makes it an ideal instrument for quickly and reliably capturing subjective usability assessments from users under a wide variety of conditions. The SUS has been used to assess voter satisfaction with system usability in numerous voting usability studies (e.g., Acemyan, Kortum, Byrne, & Wallach, 2014; Byrne, Greene, & Everett, 2007; and Everett et al., 2008), allowing systems to be easily compared.

Procedures

As voters were leaving their polling station after voting on election day, they were asked if they would complete a short survey about the system that they just used to vote. If they agreed, they completed an IRB-approved informed consent form, and then completed the survey, which was composed of the ten SUS items as well as the item about perceived errors. Afterwards, participants were thanked for their time and their questions were answered.

RESULTS

Voters' confidence in completing their ballot as intended was 4.43 on a 5-point Likert scale. This implies that voters did not think that they made many mistakes while voting. However, the score was not a perfect 5, meaning they had some misgivings about errors and there is potentially room for improvement.

As can be seen in Figure 2, subjective usability for the eSlate was worse than the previously tested VoteBox DRE, similar to the paper bubble ballot (but still not as good), and better than the lever machine and punch card (Everett et al., 2008). The eSlate's SUS score of 81.41 would be considered "excellent" according to Bangor Kortum and Miller's 2009 adjective rating scale, which was developed to help practitioners and researchers interpret the SUS scores for products.

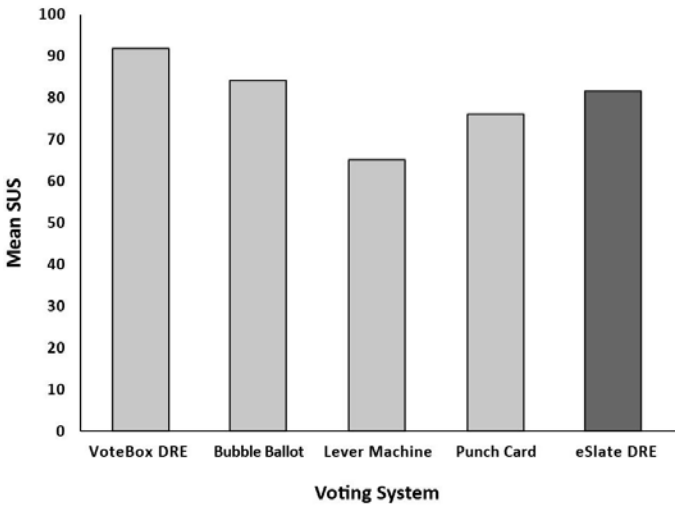

Figure 2. Mean System Usability Scale (SUS) score as a function of voting system; means for Votebox, bubble ballot, lever machine, and punch card methods were obtained from Everett et al., 2008

DISCUSSION

The results from this voting day field study indicate that the eSlate has relatively good subjective usability. Indeed, the eSlate performed the same or better than other voting systems like the bubble ballot and punch card. Even though the overall usability measurements gathered in this study are in the acceptable range, data from Herrnson et al. (2008) suggest that the eSlate has significant usability problems, which were not identified in this summative, post-use study.

There are many potential reasons why this might be the case. It is well known that when collecting subjective satisfaction measures, users might rate a system to be more usable than one might expect despite the fact that they are failing at tasks. One probable cause of this discrepancy is voters might be exhibiting classic demand characteristics; they are telling the experimenter what they think they want to hear, and thus inflating the usability rating of the system.

In the case of assessing a voting system used at a polling station, it is possible that voters are receiving assistance from poll workers while they complete their ballots. It is poll workers' job to ensure elections run smoothly, which includes eliminating bottlenecks that could cause lengthy waits for voters standing in line. Often this means that poll workers are actively looking out for voters and helping them whenever they run into

trouble while using the voting method. In turn, when a voter makes a retrospective assessment of the system, they likely take into account their entire experience, which includes the help of the poll worker and not just their experience with the machine. Therefore, the resulting usability assessment is in part a reflection of the responsiveness of the entire system to their difficulties, including the assistance of poll workers, and does not reflect the difficulties they might have had in using the machine itself. Similarly, if poll workers give instructions and helpful hints to voters waiting in line, such as the Harris county poll workers who admonished voters to remember to push the red "cast vote" button after making all of their selections, then the usability of the eSlate is not the only thing being assessed, but also the effectiveness of poll workers teaching voters how to overcome one of the major difficulties in using the system.

Another potential reason that such high scores were observed despite the eSlate having likely usability problems is that voters expect that the system will be easy to use (Raita & Oulasuirta, 2011). After all, voting systems are part of a reliable, trustworthy, and time tested voting process run successfully by the government over the last several hundred years. Further, the eSlate has been used in Harris County for over a decade. Given the situation, voters could think that the eSlate must be a usable system, since government officials chose it and have used it in past elections. This halo effect may inflate a user's assessment of the system's usability. At the same time, any noted system deficiencies may be attributed to being the users' fault, rather than the fault of a system that has been designed by an 'expert.'

While there are many explanations to account for inflated subjective usability scores, it is entirely possible that eSlate scores reported in this paper accurately represent the perceived usability of the system. If this is the case, then voters do feel that the system is sufficiently usable and that they were able to vote without making too many errors. This is a reasonable assessment because the eSlate is simply one part of the entire voting system--including components such as instructions, poll worker help, and previous experience, which performed admirably.

Would an electronic voting machine that is easier to use improve the user experience? Undoubtedly. However, it appears that within the ecosystem of the polling station, the usability of the eSlate was quite good and allowed the voters of Harris county to vote in the election.

As with any field study, some experimental control must usually be sacrificed in order to gain the most realistic data possible. This study is no exception. Because the data was collected in the field, during a major election, using using real voters, the effectiveness measure is less precise, and task times could not be collected. These limitations could be overcome in future usability studies on the eSlate by running the assessment in a controlled laboratory environment with well established, rigorous voting system testing protocols (see Acemyan et al., 2014; Byrne et al., 2007; and Everett et al., 2008).

CONCLUSION

Summative usability data for the eSlate from an election day, field study indicate that the system is perceived by voters to be quite usable in this setting. However, data from the laboratory and anecdotal evidence seems to suggest that the system still suffers from significant usability problems. Practitioners and researchers should be aware that a system might not be completely usable but still be subjectively rated as being usable. Therefore, usability researchers should consider these possible mismatches between field and laboratory findings when they evaluate systems.

ACKNOWLEDGEMENTS

We would like to thank Bob Stein, Michael Byrne, Dan Wallach and their fall 2016 *Election Systems* students for assisting with the data collection. This work was supported in part by NSF grant CNS-1409401.

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. *USENIX Journal of Election Technology and Systems*, *2*(3), 26-56.

Acemyan, C. Z., Kortum, P., Byrne, M. D., & Wallach, D. S. (2015a). From Error to Error: Why Voters Could not Cast a Ballot and Verify Their Vote With Helios, Prêt à Voter, and Scantegrity II. *USENIX Journal of Election Technology and Systems, 3(2)*. 1-25.

Acemyan, C. Z., Kortum, P., Byrne, M. D., & Wallach, D. S. (2015b). Users' mental models for three end-to-end voting systems: Helios, Prêt à Voter, and Scantegrity II. In *International Conference on Human Aspects of Information Security, Privacy, and Trust* (pp. 463-474). Springer International Publishing.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. 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.

Bannet, J., Price, D. W., Rudys, A., Singer, J., & Wallach, D. S. (2004). Hack-a-vote: Security issues with electronic voting systems. *IEEE Security & Privacy, 2*(1), 32-37.

Bernstein, A. (2008) Critics dislike eSlate voting, but officials praise it: County clerk sys the system has far fewer problems than "touch-screen." *Houston Chronicle*. Published October, 29, 2008.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.

Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive science, 21*(1), 31-61.

Byrne, M. D., Greene, K. K., & Everett, S. P. (2007). Usability of voting systems: Baseline data for paper, punch cards, and lever machines. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 171-180). ACM.

Epstein, J. (2015). Weakness in Depth: A Voting Machine's Demise. *IEEE Security & Privacy, 13*(3), 55-58.

Everett, S. P., Greene, K. K., Byrne, M. D., Wallach, D. S., Derr, K., Sandler, D., & Torous, T. (2008, April). Electronic voting machines versus traditional methods: Improved preference, similar performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 883-892). ACM.

Fischer, E. A. (2003, November). *Election reform and electronic voting systems (DREs): Analysis of security issues.* Congressional Research Service, the Library of Congress.

Hart InterCivic. (2005). *What Makes eSlate The Right Choice? Twelve Reasons to Choose the eSlate Electronic Voting System from Hart InterCivic* http://www.hartintercivic.com/files/eSlate_Top_12_Hart_Firsts.pdf. Accessed on March 3, 2017.

Herrnson, P. S., Niemi, R. G., Hanmer, M. J., Francia, P. L., Bederson, B. B., Conrad, F. G., & Traugott, M. W. (2008). Voters' evaluations of electronic voting systems: Results from a usability field study. *American Politics Research, 36*(4), 580-611.

Herrnson, P. S., Niemi, R. G., Hanmer, M. J., & Bederson, B. B. (2009). *Voting technology: The not-so-simple act of casting a ballot*. Brookings Institution Press.

Inguva, S., Rescorla, E., Shacham, H., & Wallach, D. S. (2007). *Source code review of the Hart InterCivic voting system.* University of California, Berkeley under contract to the California Secretary of State.

ISO/IEC. (1998). *9241-11 Ergonomic Requirements for Office Work with Visual Display Terminals (VDT)s-Part II Guidance on Usability*. International Organization for Standardization.

Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the System Usability Scale. *International Journal of Human-Computer Interaction, 29*(2), 67-76.

Loy, B. (2006). Just How Hard Are Electronic Voting Machines to Use? *The Irish Trojan's Blog* at http://www.brendanloy.com/wp/2006/11/just-how-hard-to-use-are-electronic-voting-machines.html.

Proebstel, E., Riddle, S., Hsu, F., Cummins, J., Oakley, F., Stanionis, T., & Bishop, M. (2007, August). *An Analysis of the Hart Intercivic DAU eSlate*. In EVT.

Raita, E., & Oulasvirta, A. (2011). Too good to be bad: Favorable product expectations boost subjective usability ratings. *Interacting with Computers, 23*(4), 363-371.

San Mateo County. (*No date*). ESlate Voting FAQ. *Shape the Future. Vote! Official Election Site of San Mateo County.* Retrieved from https://www.shapethefuture.org/votingoptions/eSlate/faq/. Accessed on March 1, 2017.

Sauro, J. (2011). *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC.