THE IMPORTANCE OF USABILITY TESTING OF VOTING SYSTEMS

Paul S. Herrnson, *University of Maryland*, Richard G. Niemi, University of Rochester, Michael J. Hanmer, *Georgetown University*, Benjamin B. Bederson, *University of Maryland*, Frederick G. Conrad, *University of Michigan*, Michael Traugott, *University of Michigan*

Abstract
Expert reviews, laboratory tests, and a large-scale field study of one paper/optical scan and five electronic voting systems suggested numerous possible improvements. Changes could be made in all aspects of the process—signing-on, navigating across the ballot, checking and changing votes, casting write-in votes, and reviewing and casting the ballot. A paper trail was largely ignored by voters. Voters generally cast votes as intended, but complexities, such as changing votes and using a ballot with a straight-party feature, reduced voting accuracy. We call for additional usability research to examine new and altered systems, especially considering add-ons such as voter verifiable paper trails.

Introduction

The 2000 presidential election and the subsequent passage of the Help America Vote Act (HAVA) increased interest in voting systems, particularly those that rely on electronic technology. The election was a reminder that voting technology and ballot design can affect election outcomes [10], voters' ability to exercise their right to vote [3], and voters' willingness to accept the legitimacy of an election [8]. It also made evident in stark terms the systematic variations in the election equipment used by different racial and income groups, which raised civil rights issues [9]. More fundamentally, the election signified the importance of testing the voting machine interfaces and features.

Electronic voting systems offer the promise of faster and more accurate voting and ballot tabulation, but they present new challenges. Evaluation of electronic voting systems has begun only recently [8] and has been bifurcated in its approach. Computer programming experts have focused primarily on election security, calling for investigation into the possibility that voting systems could be maliciously compromised and elections rigged. Political scientists and usability researchers have focused on the interaction between citizens and voting systems, pointing out the potential for electronic voting systems to reduce voter errors at the polls. Most of this research has focused on the so-called residual vote, a measure that includes undervotes, overvotes, and spoiled ballots [1].

However, research on citizens' abilities to use voting systems needs to extend beyond lost or uncast votes to a broader set of considerations. These include the ability of voters to have a satisfactory experience at the polls, including being able to use the systems with relative ease, require minimal assistance, and trust the overall process. It also should ascertain the impact of voting system designs on individuals' abilities to cast their votes as intended. Because voting equipment must be usable by nearly every citizen 18 and older—including the elderly and disabled, those with little formal education, those who do not speak English regularly, and those who have opted out of using computerized technology--research on voting systems must include subjects with a wide array of background characteristics.

In this position paper, we summarize a first round of studies that investigated several voting systems currently available on the market and a new prototype voting system [2, 5, 6, 7] and we advocate further usability research. The studies we have conducted demonstrate that current voting systems typically perform well, but that many have features that violate the criteria set out above by making the voting experience less pleasant and more difficult than necessary while causing voters to ask questions about what should be a simple process and to doubt the validity of the outcome.

The studies include six voting systems selected to represent an array of design principles. Foremost among the differences is the voter interface. One system (ES&S Model 100) uses a paper ballot and an optical scanner to checking the ballot before it is cast. Three systems use touch screens (Avante Vote-Trakker, Diebold AccuVote-TS, and a Zoomable prototype developed specifically for the study representing a unique interface that holds promise for voting): voters touch the screen to navigate through the ballot and to record their votes. A fifth system (Hart Intercivc eSlate) uses a dial and buttons to move through the ballot and vote. The final system (Nedap LibertyVote), presents the entire ballot at once and requires voters to press buttons located behind the ballot to vote. Another major difference concerns the so-called voter-verifiable paper trail. The Avante system has this feature and it is inherent in a paper ballot. Other differences include whether or not the ballot advances automatically after a vote for a particular candidate is recorded and the type of help the system offers.

All of the systems were tested using the same set

of elections. The ballot was typical in its length and features. We tested two versions: an office bloc design and one that included a straight-party device for partisan offices (a party-row design on the LibertyVote). The ballots were programmed onto the systems by their manufacturers (or with their assistance) to ensure that voters were presented with the best possible voting interface. When options were available, the configurations were set to those most frequently requested by election officials.

Our studies relied on four research methodologies: review by computer-human interaction experts, a laboratory experiment, a large scale field experiment (N=1,540), and natural experiments conducted in Florida and Michigan. The results of the studies were shared with voting system manufacturers and disseminated at conferences attended by the manufacturers, state secretaries of state, directors of state boards of elections, academics, and others interested in the voting process. We summarize here some of the types of usability problems encountered in the first round of testing. (Fuller descriptions of the voting systems, pictures, descriptions of the research methodologies, the study results and discussions of other usability issues are available at http://www.capc.umd.edu/rpts/VotingTech_par.html.. We begin at the start of the process—how one begins voting—and then discuss varies aspects of the voting process.

Beginning the Voting Process

Sometimes voters simply did not know what to do when they first approached a voting system. Filling in the circles on the paper ballot for the optical scan system posed few challenges, even to first time voters. However, some were unfamiliar with the scanning system and needed assistance with feeding their ballot into it. Two of the touch screen systems posed more challenges. The need to insert a card into a slot might be widely understood after a few elections with the same voting system and can perhaps be ignored. But that process on the Diebold system could be made simpler and more reliable, as the card often did not work as easily as, for example, slots on ATM machines. The Hart system requires voters to enter a four-digit identification number. While conceptually simple, this is the voters' first introduction to its mechanical navigation system, which as noted below is quite cumbersome. Both the Nedap and Zoomable systems place the onus for activating the system on election officials, leaving voters to begin selecting candidates free of preliminary efforts.

Navigating the Ballot

Navigating the ES&S system's paper ballot is simple on the plain office-bloc ballot (but rose dramatically in complexity when the ballot with the straight-party option was used.) Voters fill in the circles associated with their preferred candidates in any order they wish, flipping the ballot in order to complete both sides.

Two of the touch screen systems—the Diebold and Zoomable systems—use navigation systems requiring voters to indicate when they want to move through the ballot. On the Diebold one votes for one office and then advances to the next by touching the appropriate target area on the screen. After completing the first page, voters manually direct the system to move them to the next. The Zoomable system provides a visual overview of the full ballot and allows the voter to navigate freely between an overview of the entire ballot and the details of a specific race. For example, if the voter touches the box on the screen titled U.S. Senate, then the screen that lists the candidates for U.S. Senate will zoom into view, allowing the individual to vote for one of them. Then, if the voter reselects the overview, the screen shrinks back revealing the entire ballot. Most voters found both systems easy to use, but some described the amount of information on the Zoomable's screen as overwhelming and described its zooming transitions as jarring.

The Avante system, also a touch screen, offers voters less control in navigating the ballot. After a candidate for office is selected, the software automatically moves the voter to the election for the next office until the ballot is completed. The speed at which this occurs, and the loss of control over the voting process, led many to rate the Avante system lower than the other two touch screen systems.

A dial is central to the Hart system's navigation. Voters use it, or the triangle-shaped keys labeled prev and next, to move to the next screen after a selection is made. To vote, the user presses an enter key. The wheel posed challenges to some voters, leading them to rate it lower than the other systems in terms of comfort and ease of use. At high rates of rotation the dial does not provide one-to-one tracking. That is, faster movement of the dial does not correspond to faster movement through the ballot. This caused some voters to become confused as to where they were on the ballot. Many who asked for help on this system were looking for particular candidates and did not realize that they had moved several offices beyond the bloc in which the candidates' names appeared. After figuring this out (perhaps after assistance), they would turn the dial in the other direction, sometimes moving back beyond offices

they had already voted on. Even when they had the correct office in front of them, some voters found it difficult to stop the wheel on their chosen candidate. Forty percent of the people who commented on the system were critical of its navigation features.

The Nedap system is similar to that of the ES&S's paper ballot in that all of the information—offices, candidates, etc.—is presented at once. However, this system was rated significantly lower across the board due to the challenges it poses for voters. First, glare on the ballot surface combined with the small blue lights that glow when voters make a selection (rather than a large X or lighting up the entire box containing the candidate's name) make it difficult to see how one has voted if the room is brightly lit. Second, the membrane buttons a voter must push to make a selection are covered by the ballot so one does not actually see them, and they must be pushed directly using some force. A third of the subjects who commented on this system criticized the buttons.

Correcting or Changing a Vote

Several of the systems pose challenges to what should be the simple process associated with correcting or changing a vote. The ES&S system is one of them. The first challenge is whether to follow the manufacturer's suggestion and obtain a new ballot and start over or to try to erase the error and continue voting. Most study participants did the latter. This might have saved them time or prevented them from creating more errors, but it also may have lowered confidence that their vote would be correctly recorded. Nearly one-fourth of the voters who commented on the system were critical of the procedure used for changing votes.

Changing votes varies substantially across the touch screen systems. Most participants did not report changing votes on the Diebold system to be problematic, although it took some participants time to realize they must first deselect the initially chosen candidate before picking a new one. Changing votes on the Zoomable system is easier. It requires touching the area for the new candidate but not deselecting an earlier choice. The Avante system is perhaps the most challenging of the three systems. Because of the automatic navigation system, voters have to wait until the review stage of the voting process to make changes—if they remembered to make them. This caused voters to rate the system considerably lower than the other touch screen systems on comfort and ease of use. A quarter of those who commented on the system, disapproved of this aspect of their experience.

Changing votes on the final two systems was not as trying as on the Avante system, but not without challenges. On the Hart system, the main difficulty, again, was rotating the dial to the new candidate. The Nedap system was similar to the Diebold in that it required voters to deselect a candidate before selecting a new one. However, voters appeared to find this system more taxing because of the need to locate the membrane buttons behind the ballot.

Write-in Votes

Write-in candidates present a major challenge for voters. For a paper ballot/optical scan system, the problem is that they often forget to fill in the oval or complete the arrow that signals to the optical scanner that a write-in vote has been cast, a failure that could result in these ballots not being counted at all. With electronic systems, write-ins took more time than with the paper ballot, but other problems existed as well. On the Nedap LibertyVote, the size of the window was very small and it was below the large ballot. On the Avante system, the need to enter the first name, then tab to a second field to enter the last name resulted in considerable confusion. On the Hart system, the general difficulty in navigation extended to selecting letters for the name to be written in.

Reviewing and Casting the Ballot

According to HAVA, voters must have an opportunity to review their ballot prior to casting it. Voters have at least two opportunities to review their ballot on the ES&S system. They can read it prior to inserting it into the optical scanner or they can check it after if the scanner alerts them to an error, such as an overvote. If one or more errors are detected, voters have the option of removing and correcting their ballot, replacing it with a new one, or pressing a button that activates the computer to accept it as is, complete with the mistake(s). If they correct the ballot and repeat the process, they are alerted to any additional errors. This aspect of the system did not instill confidence in voters. Some were unaware of the message on the screen and wondered why their ballot had not been accepted. Others felt that not enough meaningful feedback was provided at the review stage.

Voters found reviewing ballots on the touch screen and mechanical systems significantly simpler. It is impossible to overvote on these systems. On the Diebold, voters are given the opportunity to review selections before pushing the target area on the screen to cast their ballot. With a ballot of typical length, not all of the selections are visible on one page, which can result in some voters forgetting to review all of them. The Zoomable system is similar

to the Diebold, but it lists all of the selections on one screen before pressing the target area to cast the ballot. The Avante system was similar to the Zoomable. (However, as noted below, the voting process has one more step.) The Hart machine's review system also displayed the votes on one screen, and voters cast their ballot by pressing a separate button.

Reviewing a ballot on the Nedap system is simple, and it is impossible to overvote on it. Still, more errors are likely on it than on the previous four systems. As was the case with the paper ballot, voters review their selections which are displayed throughout the entire voting process. Voters are warned of undervotes on a small window at the bottom of the system and given the option of filling in missing votes or casting the ballot as is. However, the text window is so small that some participants did not notice the message. Others did not understand the problem; some voted for one office they had left blank, failing to notice other blank offices, which led to a series of undervote messages. In addition, when the screen said Ballot Complete, voters often failed to realize that they still had to press Cast Ballot. These difficulties contributed to the low rating given this machine on confidence that one's vote would be recorded accurately.

Voter Verifiable Paper Trail

The ES&S and Avante systems both have voter verifiable paper trails. The ES&S system's trail is the paper ballot itself. The Avante system prints a paper record behind a plastic screen after an individual presses the screen to vote and asks if the voters if they wish to accept the vote as shown. Despite all the publicity given to the paper trail issue, voters seemed largely to ignore the paper record. Some of those who did not ignore it became confused, perhaps because it only allows voters to verify their votes, not change them. A lack of familiarity with the concept of checking a paper record against what one entered into a computer and a lack of clear instructions may have contributed to this.

Regardless of the difference in the design of the ES&S and Avante systems, our findings indicate that their use of paper did not add much of the voting process. To our surprise, voters had less confidence that their vote was recorded accurately on these systems than on the Diebold or Zoomable systems. The implications of our findings are that the design of current paper-based systems and attempts to retrofit existing systems with paper trails increase the difficulties faced by voters while adding nothing to their experience at the polls.

The Need for Help

A question of substantial importance to election officials is: Can voters cast their ballots unassisted? The answer is that most but not all can. The Diebold, ES&S, and Zoomable systems performed the best in this regard, with 18, 24, and 22 percent, respectively, reporting the need for assistance. The Avante system came next, with 29 percent stating they required some help. Finally, 36 percent stated they felt the need for assistance using the Hart system and 44 percent gave the same response regarding the Nedap system. (In our test, more than the usual number requested help because of the tasks we imposed.)

Voter Accuracy

One of the most important questions one can ask about voting systems is: Do they enable individuals to vote as intended? The answer is mostly, but not always. Participants in our study were able to cast their ballots with 97-98 percent accuracy when all they did was select a candidate. While a 2-3 percent error rate is not huge, it should be cause for concern, especially in a close election. Of importance is that the most frequent errors were proximity errors, i.e., voting for candidate just above or below the candidate they intended to vote for. This means that their preferred candidate probably suffered a double whammy—they did not receive a vote that was meant for them and that vote was received by an opponent. Voting for some other candidate or failing to vote at all each accounted for one percent or fewer voter errors.

When voters were asked to do less straightforward voting tasks their voting accuracy declined. For example, when asked to vote for two candidates for one office (as is often done in judicial elections), 5-6 percent committed errors using the Diebold and Zoomable systems, 6-7 percent made errors on the Avante and ES&S systems, almost 10 percent did likewise on the Nedap system, and more than 16 percent made errors on the Hart system. Higher levels of voter errors occurred when participants tried to change a vote.

The Impact of Ballot Design

Ballot design had a significant impact on every aspect of the voting process, with the exception of getting started. Individuals who voted on an office-bloc ballot provided more favorable assessments of all six voting systems, required less assistance, and made fewer errors than those using a ballot with the straight-party option (party-row ballot on the Nedap system).

Some voters simply do not understand what it means to vote a straight party or that a straight-party vote applies only to partisan elections. The combination of partisan and nonpartisan elections on the same ballot thus requires special thought about how to program electronic systems so that voters understand how many buttons they need to touch to complete a ballot. For paper ballots, one might best program a precinct ballot checker to remind voters, if necessary, that they have failed to vote for any nonpartisan offices or for ballot questions.

Variations in Voter Responses

Just as not all voting systems are the same, there also are substantial differences among voters. Our findings demonstrate that the voters' overall satisfaction, need for help, and ability, to cast their votes accurately are influenced by factors related to the digital divide. That is, frequent computer users, the more educated, younger and middle-aged individuals, voters who primarily speak English, men, and non-Hispanic whites tend to have more positive reactions to the voting systems, need less assistance voting, and they make fewer errors. This suggests that election officials should be mindful of the characteristics of voters in their communities, and perhaps make extra staff and voting systems available in areas inhabited by large numbers of voters not sharing these traits.

A Call for Further Research

We believe our studies make significant contributions to the study of voting technology and ballot design, but additional research is warranted. First, we studied only six voting systems; others are commercially available and currently in use. In addition, manufacturers have created new voting systems and modified old ones (sometimes in response to our research). It would be useful to know whether these alterations have had a positive impact on the voting systems. If modifications consist mainly of retrofits (e.g., adding a printer to an existing system), there is ample reason to be concerned about usability. Further studies could measure the time it takes to vote on different systems and the efforts election officials must undertake to set up the systems, keep them operational, and transmit the results from the polling place to a central counting facility. Some progress has been made in expanding usability studies [4], as our research has been replicated in Utah and has informed research conducted by the National Institute of Standards and Technology. Nevertheless, given the importance of voting in democracies, the possibility of another flawed election such as occurred in 2000, and concerns raised by politicians, election officials, advocates, and others, more research into the usability of voting systems is needed.

References

[1] S. Ansolabehere and C. Stewart III. Residual votes attributable to technology. *Journal of Politics* 67:365-89, 2005

[2] B. Bederson, B. Lee, R. Sherman, P. Herrnson, and R. Niemi. Electronic voting system usability issues, ACM Conference on Human Factors in Computing Systems, *CHI Letters*, vol. 5, 2003.

[3] R. Bensel. *The American ballot box in the mid-nineteenth century.* Cambridge, UK: Cambridge University Press, 2004.

[4] S. Cohen. Auditing technology for electronic voting machines. Undergraduate thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2005.

[5] F. Conrad, E. Peytcheva, M. Traugott, M. Hanmer, P. Herrnson, B. Bederson, and R. Niemi. Voter intent, voting technology and measurement error. Annual meeting of the American Association for Public Opinion Research, Miami Beach, FL, May, 2005.

[6] P. Herrnson, R. Niemi, M. Hanmer, B. Bederson, F. Conrad, and M. Traugott. The not so simple act of voting: an examination of voter errors with electronic voting, Annual meeting of the Southern Political Science Association, Washington, DC, January 4-7, 2006.

[7] P. Herrnson, R. Niemi, M. Hanmer, P. Francia, B. Bederson, F. Conrad, and M. Traugott. The promise and pitfalls of electronic voting: results from a usability field test. Paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL, April 7-10, 2005.

[8] R. Saltman. *The history and politics of voting technology*. New York: Palgrave Macmillan, 2006.

[9] USCCR. Voting irregularities in Florida during the 2000 presidential election, 2001. http://www.usccr.gov.

[10] J. Wand, K. Shotts, J. Sekhon, W. Mebane, Jr., M. Herron, and H. Brady. The butterfly did it: the aberrant vote for Buchanan in Palm Beach County,

Florida. *American Political Science Review* 95:793-810, 2001.