American Politics Research
Volume 36 Number 4
July 2008 580-611

10.1177/1532673X08316667
http://apr.sagepub.com
hosted at
http://online.sagepub.com

Voters' Evaluations of Electronic Voting Systems

Results From a Usability Field Study

Paul S. Herrnson
University of Maryland
Richard G. Niemi
University of Rochester
Michael J. Hanmer
University of Maryland
Peter L. Francia
East Carolina University
Benjamin B. Bederson
University of Maryland
Frederick G. Conrad
University of Michigan, University of Maryland
Michael W. Traugott
University of Michigan

Electronic voting systems were developed, in part, to make voting easier and to boost voters' confidence in the election process. Using three new approaches to studying electronic voting systems—focusing on a large-scale field study of the usability of a representative set of systems—we demonstrate that voters view these systems favorably but that design differences have a substantial impact on voters' satisfaction with the voting process and on the need to request help. Factors associated with the digital divide played only a small role with respect to overall satisfaction but they were strongly associated with feeling the need for help. Results suggest numerous possible improvements in electronic voting systems as well as the need for continued analysis that assesses specific characteristics of both optical scan and direct recording electronic systems.

Keywords: *election reform; voting technology; public opinion; usability; voting machines*

Political scientists' interest in voting systems and ballots may seem relatively recent but research on these topics goes back to the beginning of the profession and includes a long line of work on ballot content (Allen, 1906; Beard, 1909), ballot structure (Bain and Hecock 1957; Campbell, Converse, Miller, & Stokes, 1960; Walker 1966), and ballot order effects (Krosnick, Miller, & Tichy, 2003; Mueller, 1970). The 2000 presidential election and the subsequent passage of the Help America Vote Act (HAVA) rekindled interest in voting systems per se. That election was a reminder that voting technology and ballot design affect not only election outcomes (Wand et al., 2001) but also the fortunes of political parties (Campbell & Miller, 1957; Rusk, 1970), voters' ability to exercise their right to vote (Bensel, 2004; Keyssar, 2000, pp. 142-144), and voters' willingness to accept the legitimacy of an election (Saltman, 2006).

Computer-based direct recording electronic (DRE) voting systems, which are the focus of most new inquiries, offer the promise of faster and more accurate voting. They raise a variety of other issues as well, including voters' confidence that their votes were accurately counted and recorded, their ability to cast their votes independently and without help, their overall satisfaction with the voting process, and whether these concerns weigh more heavily for certain types of individuals.[1] The use of electronic equipment for voting purposes presents new challenges in that it must be usable by nearly every citizen 18 and older—including the elderly and disabled, those with little formal education, and those who have opted out of using computerized equipment. Moreover, voting systems are unique to the electoral process, reducing the transfer of knowledge from other electronic systems. And because voting typically occurs in a public venue, it can be accompanied by considerable social pressure, discouraging voters from asking for assistance. Given the characteristics of computer-based voting systems, reports of voters being challenged by them, and studies suggesting that the systems themselves may be a source of voter errors (e.g., Caltech/MIT 2001), we apply approaches to the study of voting systems that draw from the field of usability.

Usability studies typically include assessments of learnability, memorability, efficiency, errors, accuracy, speed, and user satisfaction (Nielsen, 1994, 2003). In this article, we report on a study of voting systems and ballots that draws from three forms of usability research: expert reviews, a laboratory experiment, and especially a large-scale field test. We investigated the

Authors' Note: This work was supported in part by National Science Foundation Grants EAI-0201634 and IIS-0306698 and Carnegie Corporation of New York Grant D05008. We thank the editor, guest editor Mike Alvarez, and the anonymous reviewers for their comments on the article. We also thank the voting system vendors, the participants in our study, the staff at the Center for American Politics and Citizenship of the University of Maryland, and the numerous undergraduate and graduate research assistants at the University of Maryland, University of Rochester, and University of Michigan.

usability of six voting systems, both paper based and electronic, representing a variety of interfaces and other attributes. Because ballot design is such an integral part of the voting interface, each system was tested with two styles of ballot. Our goal is to assess the impact of voting systems and ballots on voter satisfaction and the ability to vote independently, without help.

Understanding how voters react to new voting technologies and ballot designs is crucial. Research in this area could help explain the patterns of errors and spoiled ballots reported in other studies, uncover potential problems arising from voter–voting system interactions, identify the impact of different voter characteristics on their voting experience, and suggest improvements in the design of voting interfaces and election administration. Such research also has broader implications for the legitimacy of the electoral process. If those seeking to communicate their political choices through voting—the one political act designed to weigh each citizen's wishes equally—lack confidence in the voting process, this can result in challenges to the legitimacy of elections, such as those raised during the 2000 presidential election, and perhaps lower turnout in the future (see Alvarez, Hall, & Llewellyn, in press).

We approach these issues from the framework of usability research, which includes a variety of approaches and techniques. We study the interactions between voters and a representative set of voting systems via three usability approaches: review by experts, observation of subjects in a laboratory setting, and field tests. The results we report here are derived mostly from the field tests and examine voters' opinions about the systems.[2] We also investigated how voters interact with the systems when different ballot formats were used.

Literature, Theory, and Expectations

Recent research related to voting systems has consisted mainly of aggregate studies of residual votes (which combine overvotes, undervotes, and spoiled ballots into a single measure). Initial studies have provided important foundational work by showing that voter errors vary by location according to whether traditional paper ballotoptical scan voting systems, mechanical lever systems, or DREs were used (e.g., Caltech/MIT, 2001; Ansolabehere & Stewart, 2005). More recent studies relying on the residual vote have explored the effects of precinct-based versus central optical scan systems and various aspects of ballot formats (Bullock & Hood, 2002; Kimball & Kropf, 2005, 2006). Another focus has been on the effects of education, race, and other demographic factors on voter errors. Most of the research reports that precincts comprising members of traditionally underrepresented groups tend to produce more residual votes than do others (Ansolabehere & Stewart, 2005; Brady, Buchler, Jarvis, & McNulty, 2001; Kimball & Kropf, 2005; Tomz & Van Houweling, 2003), although Herron and Sekhon (2005) find that some undervoting on the part of Blacks may be strategic and related to the absence of an African American candidate on the ballot.

These studies have made significant contributions to the understanding of election technology, but they are limited by their reliance on a common methodology: They are all based on the residual vote (or some portion of that measure), which captures only a portion of the kinds of problems voters can have. Most gloss over significant differences among voting interfaces that affect how citizens actually cast their ballots, and because most rely on aggregate data, they cannot measure some aspects of individual behavior that could be the source of the voting errors they report. Moreover, studies using aggregate residual vote data must take special care to separate polling place votes from absentee votes as the voting systems used for these two methods of voting might differ (see Traugott et al., 2008).

An important next step is to move from election forensics (Mebane, 2006) based on aggregate data to direct observation of the interactions between voters and voting systems. This study is an attempt to build on the foundation provided by prior work using a new methodology and data set that directly compare individual voters' experiences when using different voting systems and ballots. Measures recording certain aspects of the voting experience, such as ease of use and trust in the system and whether voters feel the need for help while voting, are important because they can provide insight into what leads to satisfaction and accuracy in voting or, conversely, what causes voter errors and displeasure.

The first step in our research was to assemble a group of human-computer interaction (HCI) experts to conduct a review to assist us with developing measures and hypotheses to guide the field study. HCI draws expertise from psychology, design engineering, and computer science. HCI researchers have developed standard criteria that they apply to computerized hardware and software to assess how individuals can be expected to interact with them and to recommend improvements. One of their major foci is the impact that the complexity of an interface has on the quality of the user experience, including individuals' abilities to accomplish the desired task and to do so with confidence and with little or no need for assistance. HCI assessments typically rely on a set of core heuristics that are modified to meet the needs of specific computer applications. Once these criteria are formalized, the HCI experts apply them in a rigorous and systematic fashion (Nielsen, 1994, 2003).

Working with a team of 12 internationally recognized HCI experts, we developed eight criteria for assessing voting systems: confidence in the system's ability to accurately record votes, ease of use, voter comfort using the system, readability of the ballot and other system characters, ability to understand the ballot, ease of correcting mistakes, ease of changing a vote, and ease of casting a write-in vote (Bederson, Conrad, Herrnson, Niemi, & Traugott, 2005). These criteria served as major dependent variables in our study. We examine them independently and then in the form of an additive index. The other dependent variable on which we focus is whether the voter felt the need for help while voting, as opposed to voting independently.

Our first set of hypotheses draws on the literature on the digital divide that identifies the highly educated, the young, those well off financially, males, and Whites as the most likely to have experience with computers and other electronic devices (Alvarez & Hall, 2004; U.S. Census Bureau, 2005). We expect that voters whose reported computer usage and other background characteristics place them at higher levels of the divide will provide more positive evaluations of voting systems that exhibit the most computerization and complexity. Voters with these characteristics also should be more likely to use the various systems without asking for help.[3]

The demographic characteristics associated with the digital divide also capture other important considerations. For example, age and education are associated with the ability to use a computer efficiently and effectively. Older adults have been shown to experience more difficulty performing tasks with a computer (Riviere & Thakor, 1996), and they consistently perform less well both with respect to time required and number of errors committed (Kubeck, Delp, Haslett, & McDaniel, 1996). We include income, race, and sex because of their relationships to past practices and persistent norms such as the availability of technical training, subject-matter biases in education, the overall use of machinery, and as a reflection of the way political jurisdictions are drawn.[4] These forces all pull in the same direction as those associated with the digital divide. Another group that might experience problems with the new voting systems consists of those for whom English is not the primary language.

A second set of hypotheses also draws on the field of usability and concerns the effects of prior experience. Owing to familiarity with the process and experiential learning, we expect individuals who have voted before to provide more positive appraisals of the voting systems and to be less likely to feel the need for help (Grissom & Perlman, 1995; Nielsen, 1994). This is logically extended to suggest that specific experiences with voting systems should also matter; voters who have previously used systems similar to those studied may provide more positive appraisals of voting systems with those features and should be less likely to feel the need for help when using them.

Our third hypothesis is that the type of ballot the participants encounter will influence their voting experience. The political science literature (e.g., Kimball & Kropf, 2005; Niemi & Herrnson, 2003) and the HCI experts have identified ballot design as an important factor in voting. Here we focus on one specific feature of ballot design. We hypothesize that individuals using a ballot with a straight-party option will provide lower evaluations of the systems and will be more likely to feel the need for help than will those using a standard office-bloc ballot because the former adds another layer of complexity to the voting process.[5]

Following the expert review, we used a second method to evaluate the voting systems: observation in a usability laboratory. The lab experiment enabled us to record and analyze the voting experience of 42 voters with diverse background characteristics. The findings from the lab study lent some support to our hypotheses about the relationships between voting system attributes, voter evaluations of the systems, and voter characteristics. Because the lab study was videotaped, we were able to code the sequence of actions voters took and assess the amount of time it took them to vote. The findings from the expert review along with the lab data were especially useful in pinpointing voting system characteristics that impacted voters' evaluations of the systems and their need for help when using them (Conrad et al., 2006).

Research Design

Voting Systems and Ballots

We began by reviewing the dozens of paper ballot/optical scan, and DRE systems that are available and selecting six that incorporate the major design principles used in contemporary voting systems. Two considerations limited the number tested. First was the number of systems we could expect subjects to test. Athough we paid subjects $5 or $10 (varying by location) for their participation, asking them to evaluate more than six systems in a given setting seemed infeasible. Second was the willingness of voting system manufacturers to participate. Fortunately, the manufacturers of five voting systems having characteristics we deemed important to test agreed to loan us voting machines and to provide programming and other assistance. To these systems we added a prototype that incorporates design features that rate highly on usability tests but have yet to be incorporated into voting systems (Bederson, 2001).

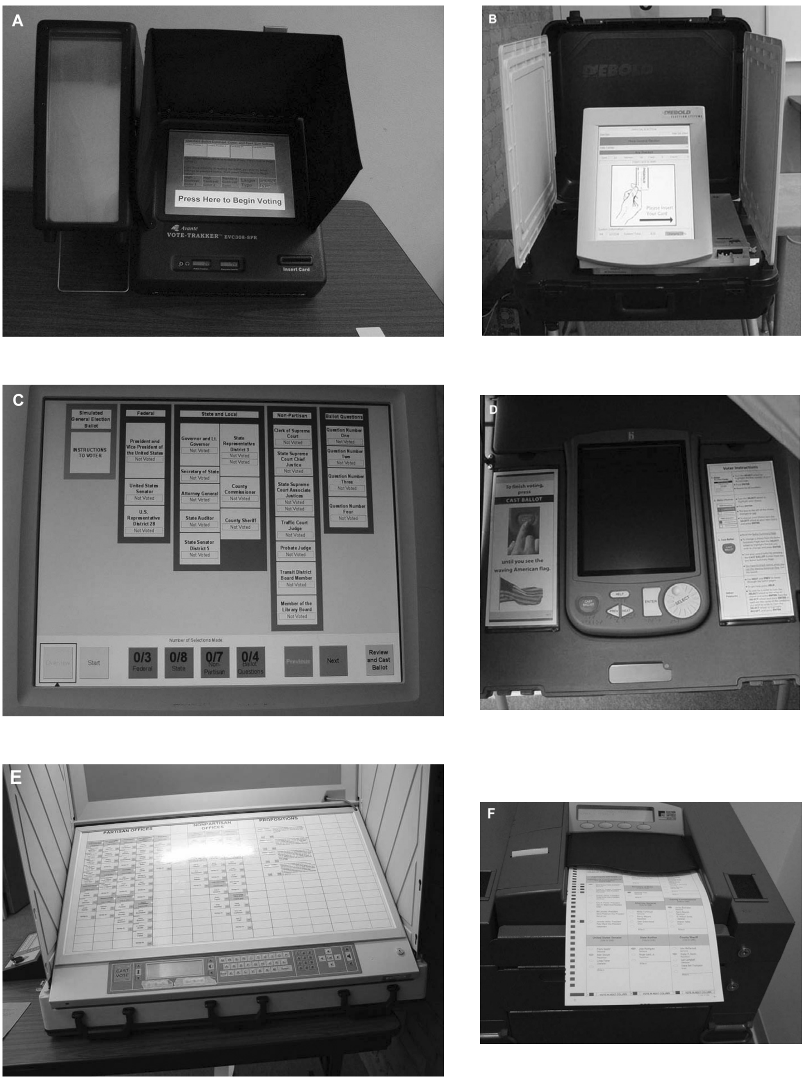
The systems included in the study incorporate a representative array of design principles.[6] Foremost among the differences is the voting interface. Three are touch screen (Avante Vote-Trakker, see Figure 1a; Diebold AccuVote-TS, see Figure 1b; and Zoomable protype, see Figure 1c). Another (Hart Intercivic eSlate, see Figure 1d) uses a wheel-and-button system to navigate and vote. One (Nedap LibertyVote, see Figure 1e) presents the entire ballot at once and requires voters to press "membrane buttons" (smooth buttons located behind the ballot) to vote. The final system (ES&S Model 100, see Figure 1f) uses a paper ballot and an electronic optical scan system for checking the ballot before it is cast. Another major difference was whether the system included a so-called voter-verifiable paper trail. This feature was on one of the systems tested (and is inherent in a paper ballot). Other differences include whether the ballot advances automatically after a vote is recorded and the type of online help it offers. (More full descriptions are provided in Herrnson et al., 2008.)

The same ballot—of our own design—was used on all voting systems. It was about the length of ballots used recently in many states, containing 18 offices and four ballot questions. The ballot contained most of the features that occur in ballots across the states: partisan and nonpartisan sections, federal, state, and local offices, multiple parties, an uncontested race, and races with more than one candidate to be elected. There were two versions: a standard office bloc design and another that differed only in that it included a straight-party option for partisan offices (a party-column 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 voting systems offered different programming options for ballot presentation or other features, we instructed the manufacturers to use the configurations most frequently requested by election officials.

The Field Study

The goals of the field study, our third method for assessing voting systems, made it important to recruit a diverse set of participants. We made special efforts to involve individuals with a wide range of relevant experience. We included frequent computer users and individuals who had not used computers at all. Likewise, we recruited especially among the young (who had little voting experience of any sort) and the elderly (who might have physical problems that would make voting difficult and who might be unfamiliar with computers). The 1,540 participants were recruited from

Figure 1
Voting Systems: (a) Avante Vote-Trakker, (b) Diebold AccuVote-TS, (c) Zoomable Protype, (d) Hart Intercivic eSlate, (e) Nedap LibertyVote, and (f) ES&S Model 100



university student bodies, faculty, and professional and maintenance staffs; upscale and downscale shopping malls; senior citizens' facilities; community centers; and other places. The study was conducted in a variety of urban and suburban areas in Maryland, Michigan, and New York. These states rely on different voting technologies, ensuring that participants had exposure to a variety of voting systems and ballot designs.[7] The result of our recruitment efforts is a sample of participants with significant variation on key background variables (see Web appendix).[8]

As part of a brief orientation to the project, participants were asked to read a voter guide (or information booklet). The voter guide resembles the nonpartisan sample ballots mailed to registered voters by election officials in some localities and published in some local newspapers. Voters frequently circle their choices on these samples and take them into the voting booth to aid them in voting. By recording their voting decisions ahead of time, voters are able to cast their ballots without relying entirely on memory. For each race, the voters were asked either to make their own choices or to vote a particular way. We asked voters to make some selections on their own, in response to information about specific candidates, to get and keep them engaged in the voting process. The primary reason for asking voters to vote a particular way in some contests was to have them do the things that voters frequently do in the polling place: change a vote, cast a write-in vote, and omit a vote for one office. Participants were directed to vote on one of the systems and to complete a questionnaire when finished. Some 788—just over half—of the participants made open-ended comments on one or more of the systems, which indicates the seriousness with which they undertook the exercise.[9] This procedure was repeated until each had voted on all six systems. The order in which participants voted on various systems was determined using a 6×6 randomized Latin square design to mitigate the impact of learning effects (or a 4×4 design in which individuals were tested on only four systems). Participants concluded the experiment by completing a demographic questionnaire. (The instruction booklet and questionnaires are provided in Herrnson et al., 2008, and the web appendix).

Results

Assessments of Voting System Characteristics

Given all the publicity about problems with voting procedures, our initial, most general finding is how favorably the various systems were rated. An

Table 1
Evaluations of Voting Systems on Various Usability Criteria

Criterion	ES&S Model 100 (Paper Ballot/ Optical Scan)	Diebold AccuVote-TS (Manual Advance TS)	Avante Vote-Trakker (Auto Advance TS With Paper)	Zoomable TS (Prototype)	Hart InterCivic eSlate (Dial and Buttons)	Nedap LibertyVote (Full Face With Buttons)
Confident vote recorded accurately	5.31	5.79	5.59	5.80	5.15	5.15
Ease of use	5.60	5.92	5.21	5.77	4.10	5.04
Comfort	5.56	5.89	5.14	5.71	4.18	4.95
Characters easy to read	5.81	6.12	5.71	6.03	5.44	5.20
Ballot easy to understand	5.78	6.10	5.63	5.98	5.29	5.31
Correcting mistakes easy	4.88	5.76	4.72	6.06	4.61	4.96
Changing vote easy	4.83	5.83	4.74	6.12	4.70	5.00
Casting write-in vote easy	6.09	5.99	5.41	6.10	4.15	5.10
Average of evaluations	5.48	5.92	5.27	5.94	4.70	5.08
Approximate *n*	1,503	1,490	1,490	1,492	1,505	1,492

Note: Entries are mean scores on 7-point scales; 1 = *strongly disagree*; 7 = *strongly agree*. TS = touch screen.

overall evaluation of the systems, created by averaging the responses for each system on the eight usability characteristics, where evaluations range from 1 (lowest) to 7 (highest), is presented in the bottom row of Table 1. Five of the systems received an average score above 5.0, and two of the touch screen systems scored close to 6.0.

Nevertheless, participants varied in their responses on some criteria and viewed some systems more favorably than others. Most important, perhaps, are the relatively high levels of voter confidence, especially in the DRE systems. When asked how much they disagreed or agreed with the statement "I am confident that my vote was accurately recorded," the three touch screen systems were rated more highly than the paper ballot/optical scan system (for each of the three comparisons with the ES&S, $p < .001$, two-tailed). Far from expressing concern over the "black box" of electronic gadgetry, voters judged the touch screen systems as highly reliable.

Most of the lower ratings involve changing votes or simply correcting mistakes. Even on the systems most highly rated on this feature, Diebold and Zoomable, this was typically the lowest rated feature. Based on our observations, and the analyses from the lab experiment and review by HCI experts (Bederson et al., 2005; Conrad et al., 2006), it is clear that some voters were confused by the need to deselect a candidate before voting for another one on these systems.

We can better explain the usability ratings by linking them, insofar as possible, to the features of the individual systems. We organize the discussion by the type of voting interface, as that is the most significant difference between systems. In making connections between features and rankings, we rely heavily on the judgments the HCI experts expressed during the expert review, the almost 800 comments of those participating in the field study and observations we made during that study, and by the laboratory experiment.[10] Following this discussion of the various system characteristics, we present the results from our models of the effect of individual characteristics on overall satisfaction and the need for help.

The paper ballot with optical scan. A priori, one might have thought that paper ballots would set the standard against which all other methods of voting would be judged. After all, a paper ballot is familiar to nearly everyone, and a mark on a piece of paper is tangible and remains visible throughout the voting process. Indeed, much of the discussion about DREs has been about the absence of a paper trail, which many activists argue is the ultimate backup. Despite all this, the paper ballot system was not rated especially well in terms of the confidence voters expressed that their vote would be accurately recorded.

The explanation for the middling confidence expressed in the paper system may lie in a number of features of the ballot and of the voting process. First, everything on the ballot is visible at once (technically, with a flip of the page): general instructions, reminders to "vote in next column," the 18 offices to be voted on, and the four ballot questions. For some, this amount of information may have been overwhelming, leading them to wonder whether they correctly filled out the ballot. At least two other features led to the low confidence rating. One was the difficulty of making changes. If voters made a mistake, they were supposed to obtain a new ballot. If they did trade ballots, they may have felt that filling out the ballot a second time created more opportunities for mistakes. If they erased their mark and revoted on the same ballot (as many did), they may have felt that the erasure could create a problem. Indeed, of the 382 individuals commenting negatively on the paper ballot system, nearly a quarter—considerably more than for any other category—wrote about changing votes: "Changing anything was time consuming," "having to get a new ballot with mistake was a pain," and it was "time consuming if you make a mistake."

The second shortcoming was that the ES&S system required two steps, yet there was little meaningful feedback. Most voters seemed oblivious to the fact that their ballots were being scanned for overvotes when they put them into the ballot feeder attached to the ballot box.[11] Apart from feedback related to overvoting, there was no indication that the ballot was filled out properly or completely. Lacking a review feature, voters could not be sure their votes were cast as they intended.

On criteria such as ease of use, comfort, readability, and ease of understanding, the paper ballot was judged more favorably, although still not at the level of the highest-rated DREs.[12] Some of the same factors—a large volume of information presented at one time and shortcomings associated with changing a vote—may have detracted from the voting experience.

Casting a write-in vote was judged to be very easy. This is not surprising because there was no keyboard to contend with; on a paper ballot, one truly writes in the name. However, simplicity does not necessarily mean that individuals voted correctly. Indeed, an analysis of the ballots shows that 26% of the voters failed to fill in the bubble beside the write-in, which is how the computer indicates that a ballot needs to be counted by hand (Herrnson et al., 2008). In this regard, paper ballots may be deceptively easy.

Touch screens. Touch screens have been hailed as the modern way to vote, similar to ATMs at banks, check-in options at airports, and fuel pumps at gasoline stations. They also have been maligned by some activists as unfair to those with little computer experience and as problematic for their reliance on fallible electronics, their lack of transparency, and their questionable security. Our field studies demonstrate that the general public does not perceive these to be serious problems. On the crucial rating of confidence, all three were rated higher than the alternative systems ($p < .001$ for all comparisons, two-tailed), and two—the Diebold and the Zoomable systems—were given especially high ratings—a mean of 5.8, with 50% expressing the strongest level of agreement (a score of 7) with the statement that their votes would be accurately recorded and more than 70% giving them a 6 or 7.

The high ratings on the confidence measure for the Diebold and Zoomable systems indicate that voters do not necessarily need a paper ballot or paper trail to trust that their votes are recorded accurately. The fact that the Avante system, which has a paper trail, was rated highly on confidence relative to how it was rated on other measures, suggests some potential benefit of having a paper record. Nevertheless, most of the evidence is to the contrary. We observed that few of the participants took the time to thoroughly examine the paper record, despite our calling it to their attention when introducing them to the system; video from the lab study corroborates these observations from the field (Conrad et al., 2006).

The Diebold and Zoomable systems were rated highly across the board. Few voters had problems with the Diebold, and most of those had to do with getting started—shortcomings easily remedied without violating the privacy of the vote. The Zoomable system was rated only slightly lower on ease and comfort. Of the 329 subjects who made negative comments on this system, about 10% found its unique zooming feature distracting or confusing.

The Diebold and Zoomable systems also were rated favorably on ease of correcting mistakes, making changes, and casting write-in votes. On the Diebold system, some subjects were unaware of the need to deselect before making another selection, leading to somewhat reduced ratings on correcting mistakes and changing a vote. As for write-in votes, the Zoomable system was rated best (although not statistically better than the paper ballot), evidently because a keyboard popped up front and center in the middle of the screen.

Both systems suffered slightly from problems related to casting a vote. On the Diebold system, the reaction to the instruction "CAST BALLOT" was too slow for some voters, causing them to push it extra times rather than waiting for the system to complete the task. With the Zoomable system, the screen target area labeled "review and cast ballot" was visible all of the time, leading some voters to touch it after selecting candidates for every office. This created problems in navigating to the next office. In addition, after one person completed voting, the "begin vote" for the next voter appeared, leading some individuals to write comments indicating that they were unsure whether they had actually finished voting.

Voters provided less positive feedback on the Avante system, rating it lower on all measures than the other touch-screen systems. There was also greater differentiation in its ratings. As noted, it was given a relatively high score on confidence that one's vote would be recorded accurately. It was also judged relatively easy to read and understand (but significantly lower than the other touch screen systems, $p < .002$, two-tailed). However, voters generally rated the Avante system lower on comfort and ease of use.[13] Voters rated it especially low for correcting mistakes and changing votes, rating it a point lower than the Diebold and Zoomable systems (for each measure and machine comparison, $p < .001$, two-tailed). Among the 446 voters who commented on the Avante system, a quarter noted this factor.

The lack of comfort and the low ratings on changing votes can probably be traced to the automatic advance mechanism, which distinguished this system from all others. The HCI experts were quick to note that this feature added to the complexity of voting and reduced one's ability to control navigation through the ballot. Ten percent of the voters who commented about the system referred to this. Casting a write-in vote was also a problem for many voters on the Avante system. One needed to enter the first name, tab, and then enter the last name. This feature often went unnoticed by voters, who typed in both first and last names at once and then had to delete the last name and re-enter it. Finally, the comments of some voters as well as our viewing of the videotapes from the laboratory experiment indicate they found the review and voting process somewhat confusing. After a voter pressed the screen to vote, the system printed the paper record and asked if the voter wished to accept the vote as shown on the record. However, this only allowed them to verify their vote, not change it. A few voters made comments indicating that they did not quite understand the final step. The response of one voter was typical: "At the end I thought I had cast my vote, but I had to cast it again."

Wheels and buttons. The most unique system had a mechanical interface for navigating and selecting and entering one's choices. The HCI experts warned that the wheel could be difficult to learn and would lead to confusion at various stages of the voting process. Compared to the touch-screen systems, voters reported the wheel and buttons as less comfortable and the ballot not as easy to understand (for each measure and machine comparison $p < .001$, two-tailed). The wheel and button mechanisms posed challenges to voters when they activated the system, entered candidate selections, made corrections, and changed a vote. Many commented that the mechanisms were slow and cumbersome, reflecting the fact that it took more steps to perform most actions, including voting for candidates, than did it on the touch-screen systems. These issues showed up clearly in the analysis of the videotapes from the laboratory experiment (Conrad et al., 2006).

Another reason for the low ratings of this system was related to navigation across the ballot. The wheel does not provide one-to-one tracking—that is, movement of the wheel does not correspond to movement across the screen. This added to the problems some voters had in locating 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 would appear. When they figured this out (perhaps after assistance), they would turn the dial in the other direction, often moving too far 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. Of the 653 people who commented on the system, 40% referred to its navigation features.

The HCI experts also noted that the process for casting the ballot could cause confusion. When it came time to cast a ballot, voters pushed a cast-ballot button, which took them to a review screen. Often, after looking it over, they would push the vote button again, thinking they were done. What they got instead was a second review screen, which listed the remaining offices on the ballot. Many voters reported this confused and frustrated them. The low confidence rating voters assigned to this system may reflect the accumulated frustration many voiced about it.[14]

Membrane button, full-screen system. The full-face ballot system we tested was rated relatively low across the board. Such variation as did occur found the highest rating for ease of understanding—probably owing to the fact that the entire ballot was laid before the voter. The low ratings were probably due to multiple challenges faced by voters. First, because the system had no shielding to prevent glare, it was difficult to see the screen in a well-lit room.

Second, the membrane buttons may have contributed to the low ratings on comfort and ease of use.[15] Although the button was covered by the ballot, so one did not actually see it, one had to push directly on it using some force. Thus, some voters pushed on the ballot very close to the button but without any effect. One third of the 583 subjects commented negatively on the buttons.

Correcting mistakes required deselecting one candidate before selecting another. This may have combined with the workings of the membrane buttons to further contribute to the low ratings for those processes. Casting a write-in vote was given a low rating as well. The text window was small, making it hard for some to see the results of their actions, and there was also no way to check one's write-in vote just prior to casting the ballot.[16]

There were additional problems when casting the ballot. When some offices were left blank, the system informed voters they had undervoted, but the text window was so small that some did not notice the message. When they did understand the problem, they sometimes voted for one office they had left blank but failed to notice other such offices, leading 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." All of these difficulties undoubtedly contributed to the low rating given this machine on confidence that one's vote would be recorded accurately.[17]

The Need for Help

Although the overall ratings of the voting systems were quite high, many voters needed assistance in using them. This is cause for concern, as not everyone who feels the need for help asks for it, which potentially results in some voters taking more time than is necessary, being less happy with the experience, and even making mistakes. Moreover, when election officials provide help, it typically results in some sacrifice of privacy. Roughly 18 to 24% reported that they needed help with the paper ballot/optical scan system and with the two most highly rated DREs (see Table 2, bottom row).[18] Voters needed considerably more help with some of the other systems.

The percentages asking for help indicate that some aspects of voting are not intuitive and readily performed by all. Computer-savvy, nimble-fingered voters have little difficulty with any of these systems (as our multivariate analysis below demonstrates). Navigating backward and forward, using review screens, deselecting before reselecting, typing in names, and touching the screen are second nature to these individuals, and they can perform them dexterously. But voters who are unused to computers or whose fingers and joints are not so agile find at least some steps problematic, especially when asked to perform them on unfamiliar equipment.

For the paper ballot/optical scan system, requests for help were mostly related to two steps: changing a vote and inserting the ballot into the optical-scan vote checker. In our field study, it was clear that voters did not understand the concept of an overvote. They often needed instruction to correct their ballot or to override the system if they decided to cast the flawed ballot.

Table 2
Perceived Need for Help in Using Different Voting Systems and Different Ballot Formats

Ballot Format	ES&S Model 100 (Paper Ballot/ Optical Scan)	Diebold AccuVote-TS (Manual Advance TS)	Avante Vote-Trakker (Auto Advance TS With Paper)	Zoomable TS (Prototype)	Hart InterCivic eSlate (Dial and Buttons)	Nedap LibertyVote (Full Face With Buttons)
Office bloc	16.0	17.8	27.5	20.9	33.0	37.6
Straight party	29.6***	19.1	31.3*	26.4***	39.2***	50.5***
Total	22.5	18.4	29.3	23.5	35.9	43.7

Note: Entries are percentages. The LibertyVote system was tested using a party-column ballot, as it could not accommodate a straight-party option.
TS = touch screen
*$p \leq .10$. ***$p \leq .01$ (for the difference between ballot formats).

Changing votes and correcting mistakes also led to many of the requests for help on the other systems, although in the case of the DREs the need to deselect before reselecting (rather than the question of erasures) was at issue. Similarly, casting the vote resulted in some confusion, although with the touch screens, the questions involved the need to push “vote” more than once. In the case of the wheel-and-buttons system, navigation problems were also at the root of many requests for help.

The findings also demonstrate the importance of ballot format. Approximately two thirds of the participants in our study had not previously encountered a straight-party option, and those who voted with one asked for help more often than those did who voted using an office-bloc ballot. The largest difference was for the ES&S system and the smallest (and statistically insignificant) was for the Diebold system. The likely explanation is that voters are more likely to become confused when confronted with all of the choices at one time, including the possibility of casting a straight-party vote, than when a system gives them the opportunity to vote straight party and then automatically advances them down the ballot. Although the Nedap system could not be programmed with a straight-party option, the results for it also demonstrate that the interaction of ballot style and voting system matters. Voters who encountered the party-column ballot, which is the system standard, felt the need to ask for help more often than did those who were presented with an office-bloc ballot.

Multivariate Analysis

Contrary to our initial expectations, voters’ ratings on the various system characteristics were highly correlated, appearing to tap into an overall assessment of each voting system’s perceived usability (Cronbach’s alpha was .90 or higher for each system). As a result, we created an index defined as the average rating across the eight usability questions.[19]

To test the hypotheses set out earlier, we regressed this index on a variety of individual-level characteristics as well as the type of ballot voters used and, as controls, dummy variables for two of the three locations in which the experiments were carried out and a variable that records the order in which each individual voted on each voting system.[20] Table 3 reports the results. The most striking finding is the predominance of substantively modest results.[21] Although a number of studies (e.g., Bullock & Hood, 2002; Herron & Sekhon, 2003; Knack & Kropf, 2003; Tomz & Van Houweling, 2003) have found a strong connection between individual characteristics and residual voting rates, we do not find a similar relationship with respect to satisfaction with the voting systems.

Table 3
The Impact of the Digital Divide, Previous Voting Experience, and Ballot Design on Voter Satisfaction (*b*, *SE* in parentheses)

	ES&S Model 100	Diebold AccuVote-TS	Avante Vote-Trakker	Zoomable Prototype	Hart InterCivic eSlate	Nedap LibertyVote
Digital divide						
Computer use	–0.082** (0.03)	0.075** (0.03)	–0.027 (0.03)	0.109*** (0.03)	0.002 (0.04)	–0.019 (0.04)
Income	0.050** (0.02)	0.059** (0.02)	0.056** (0.03)	0.056** (0.02)	0.023 (0.03)	0.059** (0.03)
Education	0.013 (0.03)	–0.023 (0.03)	–0.062** (0.03)	–0.027 (0.03)	–0.098*** (0.03)	–0.044 (0.03)
Age	0.042 (0.03)	0.006 (0.03)	–0.058* (0.03)	–0.098*** (0.03)	–0.084** (0.04)	0.037 (0.04)
English	0.283 (0.25)	0.166 (0.23)	0.273 (0.27)	0.258 (0.24)	–0.343 (0.28)	0.346 (0.28)
Male	–0.061 (0.08)	–0.059 (0.08)	–0.056 (0.08)	0.030 (0.08)	0.165* (0.09)	0.029 (0.09)
Black non-Hispanic	0.357*** (0.12)	0.203* (0.11)	0.360*** (0.13)	0.140 (0.11)	0.431*** (0.13)	–0.080 (0.14)
Voting experience						
Previously voted	–0.402*** (0.12)	–0.219** (0.11)	–0.223* (0.12)	–0.073 (0.11)	–0.345** (0.14)	–0.286** (0.14)
Used similar system	0.238** (0.09)	0.120 (0.10)	–0.075 (0.12)	–0.063 (0.10)	0.293*** (0.11)	0.128 (0.11)
Ballot						
No straight party	0.142* (0.09)	–0.021 (0.08)	0.137 (0.09)	0.164** (0.08)	0.363*** (0.10)	0.022 (0.10)
Research factors						
Michigan	0.197** (0.10)	0.067 (0.09)	0.129 (0.10)	–0.054 (0.09)	0.177* (0.10)	–0.170* (0.10)
New York	0.183 (0.12)	0.221* (0.12)	0.121 (0.13)	0.162 (0.12)	0.170 (0.13)	0.045 (0.13)
Order	–0.036 (0.02)	–0.039* (0.02)	–0.022 (0.02)	0.040* (0.02)	–0.101*** (0.03)	–0.003 (0.03)
Constant	5.420*** (0.33)	5.471*** (0.31)	5.459*** (0.35)	5.127*** (0.31)	5.600*** (0.38)	4.908*** (0.37)
Adj. R^2	0.03	0.01	0.02	0.04	0.05	0.01
N	1,233	1,222	1,224	1,224	1,235	1,229

$^*p \le .10$. $^{**}p \le .05$. $^{***}p \le .01$ (two-tailed).

With respect to the hypotheses relating to the individual characteristics associated with the digital divide, our most direct measure, computer usage, is positive and statistically significant for two of the most highly computerized systems: Diebold and Zoomable. However, the effect was substantively small.[22] For example, voters who reported using a computer 5 days a week rated the Zoomable system on average only one half of a point higher than did voters who never used a computer. For the ES&S optical-scan system, although the effect was small it indicates that frequent computer users were less satisfied with this largely paper-based system.

Even when the other individual characteristics had effects that were in the expected direction, the effect sizes were modest. For example, whereas the results for income indicate that wealthier voters rated all but the Hart system more highly than did less affluent voters, the difference, even between those with the highest and lowest incomes was substantively small on average, amounting to only about one third of a point on the 7-point scale. Though, as expected, older voters were less satisfied than younger voters were on systems such as the Hart (also Avante and Zoomable), which as the experts noted had a number of confusing features, again in terms of substantive significance the differences were quite small. One other characteristic deserves attention—that is, race. Contrary to what one might expect from the literature on the digital divide, our results suggest that African American voters were slightly more satisfied with the new systems than were White voters and those of other races, although the effect was never greater than one half of a point.

Whereas the variables for experience with similar technologies tended to have a positive effect, as was the case with the other individual characteristics, this type of experience did not meaningfully affect the ratings.[23] The results for previous voting experience stand out: Those who had previously voted consistently provided slightly lower ratings than did those who had never voted. Perhaps it reflects the experienced voters' discomfort with change or the lower expectations among nonvoters.

Individuals using the standard office-bloc ballot generally responded more favorably than did those using the office-bloc ballot with the straight-party option (or the party-column ballot in the case of the Nedap system). However, this effect was rather small and statistically significant for only half of the systems (ES&S, Zoomable, and Hart). Of course, voters who encountered the straight-party feature had to deal with more complexity in that they had to decide whether to use it and what to do if they did. Strictly speaking, of course, this is a problem caused by the ballot, not by the voting system per se. Nevertheless, this added complexity spilled over into participants' ratings of the voting systems even when additional factors were taken into account.

In contrast to the results for satisfaction, the individual characteristics associated with the digital divide and political cleavages had a much greater impact on voters' abilities to cast their ballots independently and without feeling the need to request any assistance (Table 4).[24] In general, those with little computer experience, the lowest incomes, older participants, and women were the most likely to report that they felt the need for help. Although the results were statistically significant on just one system (Avante), those who regularly speak a language other than English were also more likely to feel the need for help. The differences for African Americans suggested that they also were more likely to feel the need for help, but the effect was statistically significant only for the Zoomable system.

To provide a clearer picture of what these results mean, we calculated the marginal effects for the variables of greatest interest (Table 5). The effect of computer usage varied across the systems. Whereas it barely mattered for the ES&S and Avante systems, a small increase in computer usage led to a 3- to 4-percentage-point drop in the likelihood of feeling the need for help on two of the most highly computerized systems (Diebold and Zoomable) and, somewhat surprisingly, to a 6-point decrease on the Hart system.[25] Education and income had little substantive impact across all of the systems, but age and gender had larger effects. The impact of age reached as high as 12 percentage points on the Hart and Avante systems. Age mattered less for the other systems. Not surprisingly, it mattered least—only about 4 percentage points—for the paper ballot/optical scan system. With respect to gender, men were 10 points less likely to report feeling the need for help on the Hart system and 2 to 6 points less likely to report feeling the need for help on each of the other systems. Although the effects of language were statistically significant only for the Avante Vote-Trakker, they were huge: Those who do not speak English regularly were 29% more likely to feel the need for help on that system.[26]

Previous voting experience was generally associated with a lower likelihood of feeling the need for assistance when voting. On all but the ES&S and Hart systems, the effect was both substantively and statistically significant. Having previously voted led to a 7- to 13-point reduction in the probability of feeling the need for help on the other four systems. Those who had previously voted on touch-screen, mechanical, or paper-based voting systems were not much less likely to feel the need for help on systems that used similar interfaces.

Differences across ballot types had a noticeable impact on citizens' experiences with the voting systems. Participants who used standard office-bloc ballots were less likely to report feeling the need for help than were

Table 4
The Impact of the Digital Divide, Previous Voting Experience, and Ballot Design on the Need for Help (*b*, *SE* in parentheses)

	ES&S Model 100	Diebold AccuVote-TS	Avante Vote-Trakker	Zoomable Prototype	Hart InterCivic eSlate	Nedap Liberty Vote
Digital divide						
Computer use	–0.031 (0.06)	–0.210*** (0.06)	–0.042 (0.05)	–0.240*** (0.05)	–0.240*** (0.05)	–0.097* (0.05)
Income	–0.117** (0.05)	–0.119** (0.05)	–0.042 (0.05)	–0.059 (0.05)	–0.092** (0.04)	–0.075* (0.04)
Education	0.110** (0.06)	–0.075 (0.06)	–0.091* (0.05)	–0.058 (0.06)	–0.030 (0.05)	–0.037 (0.05)
Age	0.290*** (0.06)	0.505*** (0.07)	0.568*** (0.06)	0.503*** (0.06)	0.557*** (0.06)	0.282*** (0.05)
English	–0.411 (0.43)	–0.495 (0.46)	–1.230*** (0.39)	–0.667 (0.44)	–0.294 (0.41)	–0.464 (0.38)
Male	–0.174 (0.15)	–0.315* (0.17)	–0.321** (0.15)	–0.405** (0.17)	–0.468*** (0.14)	–0.190 (0.13)
Black non-Hispanic	0.021 (0.23)	0.273 (0.25)	0.156 (0.21)	0.444* (0.23)	0.063 (0.20)	0.123 (0.19)
Voting experience						
Previously voted	–0.255 (0.24)	–0.452* (0.26)	–0.436** (0.22)	–0.686*** (0.24)	–0.121 (0.22)	–0.465** (0.19)
Used similar system	–0.285 (0.18)	–0.370 (0.23)	0.003 (0.19)	0.047 (0.21)	–0.119 (0.17)	0.172 (0.15)
Ballot						
No straight party	–0.859*** (0.16)	–0.149 (0.18)	–0.367** (0.15)	–0.443*** (0.17)	–0.502*** (0.15)	–0.523*** (0.13)
Research factors						
Michigan	–0.153 (0.19)	–0.022 (0.21)	–0.419** (0.17)	–0.426** (0.20)	–0.147 (0.16)	0.053 (0.14)
New York	–0.073 (0.20)	–0.469* (0.25)	–0.797*** (0.22)	–0.456* (0.23)	–0.577*** (0.20)	–0.077 (0.18)
Order	–0.228*** (0.05)	–0.104** (0.05)	0.026 (0.04)	–0.234*** (0.05)	–0.093** (0.04)	–0.148*** (0.04)
Constant	–0.060 (0.57)	0.107 (0.62)	0.110 (0.53)	1.266** (0.59)	0.788 (0.56)	1.367*** (0.51)
Log likelihood	–571.4	–478.0	–621.8	–514.9	–652.8	–775.6
N	1,220	1,218	1,215	1,221	1,224	1,223

*$p \le .10$. **$p \le .05$. ***$p \le .01$ (two-tailed).

Table 5
Marginal Effect on Feeling the Need for Voting Assistance, Selected Characteristics (percentage points)

	ES&S Model 100 (Paper Ballot/ Optical Scan)	Diebold AccuVote-TS (Manual Advance TS)	Avante Vote-Trakker (Auto Advance TS With Paper)	Zoomable TS (Prototype)	Hart InterCivic eSlate (Dial and Buttons)	Nedap LibertyVote (Full Face With Buttons)
Digital divide						
Computer use	0	–3	–1	–4	–6	–2
Income	–2	–2	–1	–1	–2	–2
Education	2	–1	–2	–1	–1	–1
Age	4	7	12	8	13	7
English	–7	–8	–29	–13	–7	–11
Male	–2	–4	–6	–6	–10	–4
Black non-Hispanic	0	4	3	8	1	3
Voting experience						
Previously voted	–4	–7	–10	–13	–3	–11
Used similar system	–4	–5	0	1	–3	4
Ballot						
No straight party	–16	–2	–8	–8	–12	–13

Note. Results are based on the results in Table 4. When computing the effect of each variable, other variables are held at their mean or modal (for dummy variables) values. TS = touch screen.

those who used a straight-party ballot (or party-column arrangement on the Nedap system). The effect was statistically significant for all of the systems except Diebold. The size of the ballot effects ranged from an 8-percentage-point decrease on the Zoomable and Avante systems to 16-percentage-point decrease on the ES&S. These results suggest that debates over the inclusion of a straight-party device should be extended beyond the usual arguments about partisan advantage to include issues concerning the challenges voters face when using these ballots.

Conclusion

The results of our field study demonstrate that all six voting systems we tested were judged quite favorably. Despite the absence of training preceding the tests, voters were able to negotiate their way through the systems. At the same time, voters found some design features annoying, perplexing, or disconcerting; they expressed varying levels of confidence that their votes would be accurately recorded; and they often felt the need to ask for help in completing the voting process.

Especially interesting was that voters were more confident that their votes would be recorded accurately by the paperless touch-screen systems than by other systems, including the system using a paper ballot. The favorable responses that voters had across the board to the Diebold and Zoomable systems suggest that critics of touch-screen voting equipment may be somewhat premature and harsh in their judgments. The fact that the Avante system was judged less favorably in terms of correcting mistakes, changing votes, and the ability to vote without asking for help suggests that voters were not entirely comfortable with its automatic advance mechanism but that they prefer systems that allow them to exercise more control over the voting process. The DRE systems with the lowest levels of visible computerization—the Hart and the Nedap—were not evaluated as favorably as were the touch-screen systems on most criteria. Together with some results showing that the Hart system is prone to produce higher levels of residual votes and voter errors (Brady & Hui, 2006; Herrnson et al., 2008, Chapter 4; Kimball, 2004), our findings suggest a possible link between attitudes and performance. Research investigating the possibility of this connection would represent a significant advance.

We found little support for our hypotheses regarding which individual-level characteristics explain satisfaction with the systems. The results for voters' perceived need for help provided substantially more support for our expectations. Voters with little computer experience, senior citizens, and individuals whose predominant language is not English had a greater need to ask for help on most of the systems. Having previously voted also reduced the probability that voters would feel the need to request help. Collectively, the findings for satisfaction and need for help lead to complementary implications. The former suggests that voting manufacturers need to be more careful about their designs and more diligent about testing them for usability. The latter suggests that, in addition to improvements to the systems by the manufacturers, election officials need to consider the needs of certain types of voters, perhaps deploying more poll workers in precincts with high concentrations of the elderly and others who, our results indicate, are more likely to need help.

The results also establish support for our hypothesis regarding ballot design. This was a significant factor in conditioning voters' requests for help on all of the voting systems except the Diebold. The consistency of the results suggests that ballot types can complicate the voting experience. This may be especially so when state laws permit "straight-party voting with exception," with its complicated instructions (Niemi & Herrnson, 2003). Based on these results, as states and localities transition to new voting systems they should take seriously the opportunity to program those systems with ballots that are more user-friendly.[27]

Our findings have implications for scholars, voting system manufacturers, ballot designers, and election officials. First, they introduce three new approaches to testing voting systems and ballots: expert reviews, laboratory experiments, and field studies. Second, they identify features that contribute to or detract from the usability of voting systems and ballots, thereby suggesting potential improvements. Control over navigation across the ballot is a key case in point. Third, the findings demonstrate that substantial numbers of voters feel the need for assistance and, fourth, that certain groups of voters are especially likely to need extra help. The latter results, connected as they are with the digital divide, show that inequalities that exist in other parts of the electoral process carry over to the need for help when using the voting systems. Combined with long lines at the polls, less contact from campaigns, greater problems with registration, and so on, the impediments and inequalities posed by new voting systems might further skew satisfaction with the election system.

Not surprisingly, our analysis leaves some questions unanswered, including how many times voters must use a new voting system before they become accustomed to it and more confident in the system's performance. Moreover, we have yet to assess the impact of the voting systems and ballot designs on the amount of time it takes to vote. More research on the interactions between voting systems and the broader context in which citizens vote, including how elections are administered, is also needed. Whereas we find that satisfaction, including confidence, is similar for Black and White voters, Alvarez et al. (in press) find that although confidence remained steady for White voters from 2000 to 2004 it dropped markedly among Black voters. Our finding that interaction with the voting systems does not reveal racial differences lends weight to their suspicion that issues related to the politics and administration of the 2000 and 2004 elections, besides voting interfaces, explain lower confidence levels among Black voters. Additional research also should address the influence of poll workers on the voting experience. Atkeson and Saunders (2007) show that voters' opinions about elections are influenced by how poll workers perform their jobs. Additional studies of actual elections should capture voter reactions to the type of voting system and ballots used as well as the number and nature of voter–poll worker interactions. Finally, we have tested only six voting systems and ballots in a limited number of test sites. Further research is needed to broaden the investigation to include more voting systems, ballots, and localities. Regardless, we have demonstrated that the design of voting systems and ballots influences voter satisfaction and need for help. Given the centrality of voting to the democratic process, these are important concerns.

Notes

1. Other concerns include voter accuracy (the ability to cast votes as intended) (Herrnson et al., 2008) and reliability and security (Feldman, Halderman, & Felten, 2007; Rubin, 2006). These topics are beyond the scope of this article.

2. Our field tests involved the recruitment of a diverse set of participants in shopping malls, offices, and other settings to cast votes in a simulated election on each of the systems. The field tests differ from field experiments (e.g., see Green & Gerber, 2004) in that we did not randomly select participants and did not have a definitive control group; however, we did randomly assign the order in which the participants voted on the systems.

3. Because the voting systems vary with respect to computerization, we do not expect these characteristics to have a uniform influence across systems. For example, computerization is limited on optical scan systems, suggesting that whereas those with high levels of computer experience might find them less satisfying than direct recording electronics (DREs), older voters might find the optical scan systems more satisfying. We incorporate these expectations into our discussion below.

4. Sex may, in fact, yield small or nil differences. Women are only a few percentage points less likely to live in a home without a computer, and in those homes slightly more women than men report using the computer. More women also report using a computer at work (U.S. Census Bureau, 2005). African Americans are much less likely than are Whites or Asians to live in a home with a computer, but among those who have computers the difference in usage is less than

5%. Race is an important factor to consider given the role that it plays in the composition of U.S. political jurisdictions and the findings in the literature that behavioral outcomes differ by race. Studies relying on aggregate data to assess the impact of income, education, or the percentage of newly registered voters have fairly consistent results; findings about the influence of race are much less consistent. For recent work, see Brady, Buchler, Jarvis, and McNulty (2001); Bullock and Hood (2002); Knack and Kropf (2003); Tomz and Van Houweling (2003); Herron and Sekhon (2005); Kropf and Knack (2004); Alvarez, Sinclair, and Wilson (2004); and Kimball and Kropf (2006).

5. Other attributes of ballots, such as excessive length and listing candidates for one office on multiple pages, also can complicate the voting task (Herron & Sekhon, 2005; Jewett, 2001; Kimball & Kropf, 2005; Kimball, Owens, & Keeney, 2004; Lausen, 2007; Wattenberg, McAllister, & Salvanto, 2000). We cannot investigate all such factors here but because we use only two ballots (differing by a single characteristic) they cannot be responsible for variations in results across our ballots.

6. For a comprehensive listing of voting systems, see (Herrnson et al., 2003). As of November 2006, no single machine was used by 12% of all registered voters. The ES&S Model 100 was used by 10.1%, the Diebold AccuVote-TS by 11.6%, and the Hart eSlate by 1.6%. All together, optical scan machines were used by 49% and DREs by 38%. Full-face machines, such as the Nedap LibertyVote, will be used in New York and perhaps one or two other states. See Election Data Services (2006).

7. For the most part, Maryland used lever or punch card systems before switching to DREs in 2004; Michigan used a variety of systems but most voters voted on optical scan ballots; New York used lever machines. Michigan is the only state that offered voters a straight-party option.

8. The Web appendix can be found at http://www.capc.umd.edu/rpts/votingtech_par.html.

9. Moreover, our observations in the field as well as the finding that only a handful of respondents used the write-in option to make entries that were obviously not going to count as correct (e.g., Bill Clinton, who was not a valid write-in candidate) attest further to the seriousness with which the participants took the task.

10. One might think to regress voter ratings of the systems (as the dependent variable) on the various system characteristics and then use the regression coefficients to determine the apparent impact of each feature on those ratings. Unfortunately, high (even complete) multicollinearity among the characteristics—owing to fact that there are a multitude of features and only six machines—makes this approach unworkable.

11. In some jurisdictions, voters simply drop their ballots into a box; ballots are scanned centrally at the end of the day. This simplifies the process but it means that voters do not even receive feedback about overvotes.

12. On each of these measures, the Diebold and Zoomable systems were the highest rated. Individual t tests (two-tailed) for each of these measures reveal that the differences between the ES&S system and Diebold are significant at $p < .001$ and differences between the ES&S system and the Zoomable prototype are significant at $p < .02$.

13. The ratings on comfort and ease of use for the Avante system were significantly lower ($p < .002$, two-tailed) than ratings for the ES&S, Diebold, and Zoomable systems and significantly higher ($p < .01$, two-tailed) than ratings for the other two systems.

14. The Hart system was rated significantly lower than ES&S ($p = .022$), Diebold ($p < .002$), Avante ($p < .002$), and Zoomable ($p < .002$), all two-tailed.

15. Ratings on comfort and ease of use for the Nedap system were lower than ratings for ES&S ($p < .002$), Diebold ($p < .002$), Avante ($p < .01$), and Zoomable ($p < .002$) and higher than Hart ($p < .002$), all two-tailed.

16. A technical problem with the write-in vote meant that a number of people could not enter the name and the message in the window was uninterpretable. To get around the problem, we had them start the write-in process again.

17. The Nedap system was rated significantly lower than ES&S ($p = .026$), Diebold ($p < .002$), Avante ($p < .002$), and Zoomable ($p < 0.002$), all two-tailed.

18. At one test site, we recorded whether individuals actually received help using the voting systems. For the six systems, the average correlation between reporting the need for help and receiving help was $r = .82$ ($p < .001$).

19. The index was created for respondents who answered at least six of the eight usability items and has a theoretical range from 1 to 7. We decided to use OLS rather than ordered probit given the large number of values the variable takes on and a preference for not having to make somewhat arbitrary choices about how to collapse the values into a smaller set of categories.

20. The location variables are, in a sense, catch-alls that pick up differences in the populations across the three sites and any differences in how the field tests were administered. The order variable is to control for learning and fatigue effects. It is coded from 1 (*the first machine on which the participant voted*) to 6 (*the sixth machine on which the participant voted*). Negative coefficients indicate that the later the voting system was used in the sequence the less satisfied the voter was. Note that the drop in sample size from the earlier tables is largely because of item nonresponse on the income question. Although we do not have strong theoretical expectations with respect to nonlinear effects, we did check for this by examining the relevant bivariate relationships, finding little, if any, reason to reject the linear specification of these variables.

21. To see whether the absence of statistical significance is because of multicollinearity, we reran the regression models and computed the variance inflation factor. The rule of thumb is that variance inflation factors larger than 10 suggest problems. In our case, the variance inflation factor averaged under 2 and was never larger than 5. In addition, we experimented with deleting variables that we thought might be correlated with others in the model, but the results did not change materially.

22. The auto-advance mechanism on the Avante might explain why frequent computer users were not more satisfied with the system. That is, this feature might have been especially annoying for the computer savvy, who are used to having more control when using computerized devices.

23. The negative (though insignificant) coefficients for previous experience with similar systems for the Avante and Zoomable systems likely stem from lack of experience with the unique autoadvance and zooming mechanisms, respectively.

24. As noted above (Note 17), there was a strong correlation between reporting the need for help and receiving help. The results reported in the text and tables are for the perceived need for help. Because the dependent variable was binary (coded 1 if an individual reported feeling the need for help and 0 otherwise), we used a logit model. In this model, negative coefficients on the order variable indicate that the later the voting system was used in the sequence the less likely it was that the voter felt the need for help.

25. With respect to the Avante system, although its high level of computerization would lead one to expect frequent computer users to be less likely to need help, it is possible that using a computer frequently does not translate into greater ability to figure out how to gain more control over the autoadvance mechanism.

26. The lack of statistical significance for language on the other systems may, in part, be because of the small number of voters in our study who do not speak English regularly.

Although all of the substantive effects were large, it should be pointed out that the machines we tested were not set up to operate in other languages, and voters who spoke only another language did not participate.

27. The experience of the state of Maryland provides an example. Prior to the statewide adoption of the Diebold system in 2004, some counties in Maryland used party-column ballots, but with the transition to the new system all counties moved to an office-bloc ballot.

References

Allen, P. L. (1906). Ballot laws and their workings. *Political Science Quarterly*, *21*, 38-58.

Alvarez, R. M., Sinclair, D. E., & Wilson, C. (2004). Counting ballots and the 2000 election: What went wrong? In A. Crigler, M. Just, & E. McCaffery (Eds.), *Rethinking the vote* (pp. 34-50). New York: Oxford University Press.

Alvarez, R. M., & Hall, T. E. (2004). *Point, click, and vote: The future of Internet voting.* Washington, DC: Brookings.

Alvarez, R. M., Hall, T. E., & Llewellyn, M. (in press). Are Americans confident their ballots are counted? *Journal of Politics*.

Ansolabehere, S., & Stewart, C., III. (2005). Residual votes attributable to technology. *Journal of Politics*, *67*, 365-389.

Atkeson, L. R., & Saunders, K. L. (2007). The effect of election administration on voter confidence: A local matter. *Political Science and Politics*, *60*, 655-660.

Bain, H. M., & Hecock, D. S. (1957). *Ballot position and voters choice.* Detroit, MI: Wayne State University Press.

Beard, C. A. (1909). The ballot's burden. *Political Science Quarterly*, *24*, 589-614.

Bederson, B. B. (2001). PhotoMesa: A zoomable image browser using quantum treemaps and bubblemaps. UIST 2001, ACM Symposium on User Interface Software and Technology. *CHI Letters*, *3*, 71-80.

Bederson, B. B., Conrad, F. G., Herrnson, P. S., Niemi, R. G., & Traugott, M. W. (2005, February). *Designing interfaces for voting machines.* Paper presented at the annual meetings of the National Association of Election Directors and American Association for the Advancement of Science, Washington, DC.

Bensel, R. F. (2004). *The American ballot box in the mid-nineteenth century.* Cambridge, UK: Cambridge University Press.

Brady, H. E., Buchler, J., Jarvis, M., & McNulty, J. (2001). *Counting all the votes: The performance of voting technology in the United States.* Survey Research Center and Institute of Governmental Studies, University of California–Berkeley. Retrieved January, 12, 2006, from http://ucdata.berkeley.edu/new_web/countingallthevotes.pdf

Brady, H. E., & Hui, I. (2006, June). *Let's not forget about accuracy.* Paper presented at the workshop on the Usability and Security of Electronic Voting Systems, College Park, MD.

Bullock, C. S., III, & Hood, M. V., III. (2002). One person–no vote, one vote, two votes: Voting methods, ballot types, and undervote frequency in the 2000 presidential election. *Social Science Quarterly*, *83*, 981-993.

Caltech/MIT. (2001). *Residual votes attributable to technology: An assessment of the reliability of existing voting equipment*. Retrieved February 2, 2007, from http://www.vote.caltech.edu/Reports

Campbell, A, Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American voter*. New York: Wiley.

Campbell, A., and W. E. Miller. (1957). The motivational basis of straight and split ticket voting. *American Political Science Review*, *51*, 293-312.

Conrad, F. G., Lewis, B., Peytcheva, E., Traugott, M. W., Hanmer, M. J., Herrnson, et al. (2006, April). *The usability of electronic voting systems: Results from a laboratory study*. Paper presented at the annual meeting of the Midwest Political Science Association, Chicago.

Election Data Services. (2006). *About 55 million, or one-third of the nation's voters, will face new voting equipment in 2006 election*. Retrieved December 1, 2007, from http://www.edssurvey.com/images/File/ve2006_nrpt.pdf

Feldman, A. J., Halderman, J. A., & Felten, E. W. (2007). *Security analysis of the Diebold AccuVote TS voting machine*. Princeton, NJ: Center for Information Technology Policy, Princeton University. Retrieved February 2, 2007, from http://itpolicy.princeton.edu/voting

Green, D. P., & Gerber, A. S. (2004). *Get out the vote! How to increase voter turnout*. Washington, DC: Brookings.

Grissom, S. B., & Perlman, G. (1995). StEP(3D): A standardized evaluation plan for three-dimensional interaction techniques. *International Journal of Human-Computer Studies*, *43*, 15-41.

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 task of casting a ballot.* Washington, DC: Brookings.

Herrnson, P. S., Niemi, R. G., Horowitz, J. M., & Richman, S. (2003). *Characteristics of contemporary voting machines*. College Park: Center for American Politics and Citizenship, University of Maryland. Retrieved February 2, 2007, from http://www.capc.umd.edu/rpts/md_evote_ContempVotingMach.pdf

Herron, M. C., & Sekhon, J. S. (2003). Overvoting and representation: An examination of overvoted presidential ballots in Broward and Miami-Dade Counties. *Electoral Studies*, *22*, 21-47.

Herron, M. C., & Sekhon, J. S. (2005). Black candidates and Black voters: Assessing the impact of candidate race on uncounted vote rates. *Journal of Politics*, *67*, 154-177.

Jewett, A. (2001, August). *Explaining variation in ballot invalidation among Florida counties in the 2000 election.* Presented at the annual meeting of the American Political Science Association, San Francisco.

Keyssar, A. (2000). *The right to vote.* New York: Basic Books.

Kimball, D. C., (2004). Assessing voting methods in 2002. Unpublished manuscript, University of Missouri–St. Louis.

Kimball, D. C., & Kropf, M. (2005). Ballot design and unrecorded votes on paper-based ballots. *Public Opinion Quarterly*, *69*, 508-529.

Kimball, D. C., & Kropf, M. (2006, January). *Ballot initiatives and residual ballots in the 2004 presidential election*. Paper presented at the annual meeting of the Southern Political Science Association, Atlanta, GA.

Kimball, D. C., Owens, C. T., & Keeney, K. M. (2004). Residual votes and political representation. In R. P. Watson (Ed.), *Counting votes: Lessons from the 2000 presidential election in Florida.* Gainesville: University of Florida Press.

Knack, S., & Kropf, M. (2003). Voided ballots in the 1996 presidential election: A county-level analysis. *Journal of Politics*, *65*, 881-897.

Kropf, M. E., & Knack, S. (2004). Balancing competing interests: Voting equipment in presidential elections. In R. P. Watson (Ed.), *Counting votes: Lessons from the 2000 presidential election in Florida.* Gainesville: University of Florida Press.

Krosnick, J. A., Miller, J. M., & Tichy, M. P. (2003). An unrecognized need for ballot reform: The effects of candidate name order on election outcomes. In A. N. Crigler, M. R. Just, & E. J. McCaffery (Eds.), *Rethinking the vote: The politics and prospects of American election reform.* New York: Oxford University Press.

Kubeck, J. E., Delp, N. D., Haslett, T. K., & McDaniel, M. A. (1996). Does job-related training performance decline with age. *Psychology and Aging*, *11*, 92-107.

Lausen, M. (2007). *Design for democracy.* Chicago: University of Chicago Press.

Mebane, W. R., Jr. (2006, July). *Election forensics: Vote counts and Benford's Law.* Presented at the annual meeting of the Political Methodology Society, University of California–Davis.

Mueller, J. E. (1970). Choosing among 133 candidates. *Public Opinion Quarterly*, *34*, 395-402.

Nielsen, J. (1994). *Usability engineering.* Morgan Kaufmann: San Francisco.

Nielsen, J. (2003). *Usability 101: Introduction to usability*. Retrieved February 2, 2007, from http://www.useit.com/alertbox/20030825.html

Niemi, R. G., & Herrnson, P. S. (2003). Beyond the butterfly: The complexity of U.S. ballots. *Perspectives on Politics*, *1*, 317-326.

Riviere, C. N., & Thakor, N. V. (1996). Effects of age and disability on tracking tasks with a computer mouse: Accuracy and linearity. *Journal of Rehabilitation Research and Development*, *33*, 6-16.

Rubin, A. D. (2006). *Brave new ballot.* New York: Random House.

Rusk, J. G. (1970). The effect of the Australian ballot reform on split ticket voting. *American Political Science Review*, *64*, 1220-1238.

Saltman, R. G. (2006). *The history and politics of voting technology.* New York: Palgrave Macmillan.

Tomz, M., & Van Houweling, R. P. (2003). How does voting equipment affect the racial gap in voided ballots? *American Journal of Political Science*, *47*, 46-60.

Traugott, M. W., Hanmer, M. J., Park, W., Niemi, R. G., Herrnson, P. S., Conrad, F. G., et al. (2008). *Losing fewer votes: The impact of changing voting systems on residual votes*. Unpublished manuscript.

U.S. Census Bureau. (2005). *Computer and Internet use in the United States: 2003*. Retrieved November 2, 2005, from http://www.census.gov/prod/2005pubs/p23-208.pdf

Walker, J. L. (1966). Ballot forms and voter fatigue: An analysis of the office block and party column ballots. *Midwest Journal of Political Science*, *10*, 448-463.

Wand, J. N., Shotts, K. W., Sekhon, J. S., Mebane, W. R., Jr., 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*, 793-810.

Wattenberg, M. P., McAllister, I., & Salvanto, A. (2000). How voting is like taking an SAT test: An analysis of American voter rolloff. *American Politics Quarterly*, *28*, 234-250.

Paul S. Herrnson is director of the Center for American Politics and Citizenship and professor of Government and Politics at the University of Maryland. His research focuses on money and politics, political parties and elections, and voting technology. He is author of *Congressional Elections: Campaigning at Home and in Washington*.

Richard G. Niemi is Don Alonzo Watson Professor of Political Science at the University of Rochester. He has published widely on civic education, term limits, public opinion, and voting technology and coedits the long-running *Vital Statistics on American Politics*.

Michael J. Hanmer is an assistant professor in the Department of Government and Politics at the University of Maryland. He has published on election-day registration, vote by mail, vote overreporting, and voting technology.

Peter L. Francia is associate professor of political science at East Carolina University. He is the author of numerous publications on American elections and politics, including his most recent book, *Conventional Wisdom and American Elections: Exploding Myths, Exploring Misconceptions*.

Benjamin B. Bederson is associate professor in the Department of Computer Science and past director of the Human-Computer Interaction Lab, Institute for Advanced Computer Studies, University of Maryland. He is a cofounder and chief scientist for Zumobi, Inc.

Frederick G. Conrad is a research associate professor in the Institute for Social Research at the University of Michigan. He works at the interface of cognitive psychology, survey methodology, and human-computer interaction. He is on the editorial boards of *Public Opinion Quarterly* and *The Journal of Official Statistics*.

Michael W. Traugott is professor of communication studies and political science at the University of Michigan. He has published extensively on campaigns and elections, political communication, and election administration.