The Rice Cultivator, Vol. 1, 2008, pp. 30-45

Usability of Election Technologies: Effects of Political Motivation and Instruction Use

Stephen N. Goggin[1]

Abstract. Election technology presents a unique challenge: voters of all backgrounds should be able to efficiently and effectively cast a ballot with ease. Unfortunately, with the recent adoption of new voting technologies, few studies have examined the usability of these machines. Those laboratory studies that have examined these technologies have overlooked two critical factors: the unique political motivation and identification with specific candidates voters experience on election day, and the instructions and training given to voters. In the current study, seventy-two adults were observed using Direct Recording Electronic (DRE) machines, optical-scan ballots, punch-card ballots, and lever machines. Although the results of the study reveal no impact of voter motivation and instruction use in a laboratory setting, further systematic study of their impact during real elections is needed. Furthermore, the results of the study provide evidence for more general usability problems in using election technology by all segments of the voting population.

Introduction

As a policy issue that traditionally remained out of public sight and scrutiny, the administration of elections in the United States quickly came to public attention after the 2000 presidential election. After the debacle of the butterfly ballot in certain Florida counties, concerns over the specific election technologies in use by counties across the United States have risen dramatically. In 2002 the Help America Vote Act (HAVA) provided vast funds to states to modernize their voting equipment in hopes of preventing future controversies. Even with this significant financial commitment from Congress, many security, accuracy, secrecy, auditability, and transparency concerns over election technologies remain.

Throughout the United States, a daunting patchwork of federal, state, and county laws govern election administration. The funds HAVA provided allowed many counties to modernize the technologies they used to administer elections, and many purchased Direct Recording Electronic (DRE) voting machines. Election officials largely cast aside previous voting technologies such as the punch-card, lever, and some types of paper ballots in favor of the new technologies (Wiltenburg, 2003). In fact, in the 2006 midterm elections, 33.7 percent of voters used DRE machines, while 48.2 percent used optical scan paper ballot systems, with punch-card and lever machines each being used by fewer than 4 percent of voters (Election Data Services, 2006).

With the quick rush to adopt new technologies, many states and counties failed to examine the new technologies in great detail. Because few federal regulatory laws regarding certification of the new machines existed,

[1] Stephen Goggin is a Rice University senior; goggin@rice.edu.

many states and counties were left with only the voting machine vendor's promises and demonstrations as evidence of the machine's quality. While many states require Independent Testing Agencies (ITAs) to certify voting equipment before it is purchased, ITAs are contracted directly by the machine manufacturers to obtain certification. Furthermore, contracts between vendors and purchasing states or counties are often subject to strict legal contracts regarding the use and testing of the equipment (Hall, 2007). These strict legal contracts largely prevented many counties from hiring outside parties, such as researchers, to investigate the machines, further complicating the study of election administration.

The activity of voting is itself a unique event. Because every citizen, with few exceptions, is guaranteed the right to vote, voting technologies must be designed for use by any person, including the elderly and disabled. In addition to simply being usable by voters, the technologies must still meet the required local and state laws regarding security, secrecy, accuracy, transparency, and auditability. While different fields of researchers study each of these characteristics in isolation, it is important to note that these criteria are all goals of election administration in the United States. Furthermore, while studied alone, they often have very clear impacts on each other; for example, as a machine becomes more secure against tampering by implementing cryptography, the interface may become too complex for many users.

Studying Usability

The usability of a technology cannot be captured by a simple, single scale; it encompasses several aspects of the human-machine interaction. In 2004 a National Institute of Standards and Technology special report highlighted three metrics of usability recommended for use in testing voting technologies: efficiency, effectiveness, and satisfaction (Laskowski et al., 2004). These three metrics have been commonly used in prior usability studies of voting technologies (Everett, 2007; Byrne et al., 2007). Two of the metrics are objective measures, which are easily operationally defined: efficiency and effectiveness, as the amount of time a voter takes to cast his or her ballot and errors in accurately casting the ballot, respectively. The third, voters' satisfaction with the technology, can be measured by a standardized subjective questionnaire such as the System Usability Scale (SUS) (Brooke, 1996). While studies such as Everett (2007) have noted high satisfaction with newer electronic technologies, satisfaction alone is not a proper measure of usability. Machines that appear simple to the users may in fact have many interface problems, causing errors or longer usage times.

While the metric of effectiveness is often thought to be most important to the study of elections, as it measures the accuracy of the capturing of a voter's intent, the other two metrics also hold great importance. As DRE technologies are often more expensive than their predecessors, counties usually have fewer machines. This highlights the importance of efficiency—if few machines are available relative to the number of voters, machine use should be efficient and quick so that as many voters as possible have the opportunity to vote without long waiting times. Finally, the metric of satisfaction is no less important—as elections are the source of legitimacy for any democracy, voters should feel satisfied that their vote was cast accurately, and they should wish to vote again.

While the investment of federal money after HAVA was an attempt to stem the problems of the 2000 presidential election, it did not eliminate errors. Actual problems such as those experienced with Voter Verified Paper Audit Trails (VVPAT) in Cuyahoga County, Ohio, (Election Science Institute, 2006) and those demonstrated with usability in previous laboratory studies (Everett, 2007; Byrne et al., 2007) have shown that our voting systems are far from perfect. Byrne et al.'s laboratory study of usability indicated that a significant number of ballots had at least one error due to ineffective use by a participant. Even studies outside the laboratory, such as archival studies of voting records from enormous samples of counties across multiple elections, reveal similar results. Ansolabehere and Stewart (2005) found that the types of voting technologies in use have an impact on under- or over-vote occurrences, known as a residual vote rate. Residual vote rates have long been studied by political scientists; however, only recently have researchers examined the effects of election technology on these rates.

As researchers have focused on usability problems with current voting technologies, they largely have overlooked many potentially strong influences on usability. While a wide variety of factors affect machine usability, generally they can be assigned to two groups: characteristics of the user and characteristics of the device. Much of the research regarding voting technologies stresses the differences among devices, but this paper examines characteristics that are common to nearly all of the voting technologies in use and characteristics variable across voters regardless of the technology used.

First, we examine the instructions and information given to voters on election day. The impact of instructions has been studied in many other arenas but has been largely passed over in the election literature. Next, we examine the impact of political motivation of voters on election day; few other human-machine interactions have a strong ideological, emotional, or rational importance tied to them, and it is unknown what effect this may have on their usability.

VOTING EQUIPMENT INSTRUCTIONS

One feature common to all technology is the instructions provided for its use. Because the instructions may guide the user through a series of novel and complex operations, how the instructions are presented and interpreted can have an enormous impact on the success of the user in reaching their goals. Because voting is a rare event, with national elections held every two years, voters often do not have the requisite experience to use complex voting technologies solely from memory. While users may use existing schemas and scripts for using other technologies to help them learn new technologies, instructions provide a bridge to learning to use a new technology effectively.

In attempts to make the newly purchased election equipment more usable by voters, voting machine vendors and bureaucratic entities such as county clerks have responded with the issuance of numerous instructions, videos, and training materials. The efficacy of these materials is largely unknown. Because jurisdictions are primarily autonomous in their control of elections, few regulations are in effect standardizing how voters are trained and shown how to vote. For example, the Election Assistance Commission (EAC), created by HAVA in 2002, published the 2005 Voluntary Voting System Guidelines (VVSG). While the VVSG outlines numerous standards regarding voting procedures and machine security and use, the guidelines are voluntary due to the lack of federal control in election administration. Although instructions of all types, including video tutorials, are provided to voters in the United States, the main focus of this study is on the impact of step-by-step text instructions presented during the act of voting. This type of instruction was chosen because of its availability for manipulation in the laboratory setting and its commonality in election scenarios.

There has been much applied research regarding instructional methods and use within the broader realm of psychology. Much of the relevant research has focused on the impact of cognitive aging and computer-based technologies (Van Gerven et al., 2006). This research is very specific and applied; for example, numerous studies regarding Automatic Teller Machine (ATM) usage and design have been conducted (Mead and Fisk, 1998; Mead et al., 1999; Jameson and Rogers, 2000). Fortunately, the task of using an ATM is actually quite similar to and is likely subject to many of the same theories as DRE use. Users must interact with a computerized interface designed for a highly specialized task, requiring little user input besides forced-choice selections among few alternatives. Van Gerven et al. (2006) apply two theories to the design of instructions for computer tasks for those with reduced cognitive ability: Cognitive Load Theory (CLT) and the Cognitive Theory of Multimedia Learning (CTML).

Both CLT and CTML are based on a theoretical notion that there is a limited capacity of working memory, and therefore novice learners choose carefully the information they gather from instructions. Therefore, CLT and CTML suggest that reducing extraneous cognitive load by optimizing the amount of relevant information in instructions is ideal. Furthermore, Morrell and Echt (1997) discuss how instructions should be presented in an explicit, familiar manner, such as in a concise list of the proper actions to take.

While much empirical evidence supports these theories, their focus is mainly on the procedural memory for a given task, not performing a novel task aided by instructions. Specifically, many of the ATM studies rely on training sessions with instructions before actual ATM usage, rather than establishing proper usage during the user's first interaction with the technology. Due to the infrequency of voting, a more thorough, explanatory step-by-step instructional approach might be more fitting, relating the proper procedure to known schemas or scripts. This study uses these theoretical constructs to design an "ideal" set of usable instructions for each voting technology, attempting to demonstrate that proper instructions can offset inherent usability flaws in election technology.

Political Motivation and Laboratory Studies

While tightly controlled usability studies regarding different technologies in laboratories can yield extraordinary insight into the processes and problems in human-technology interaction, these studies inherently lack the campaign effects and political motivation of a real election. While demand characteristics and the "strong situation" of a laboratory setting could be argued to suffice for the motivation to accurately and thoroughly cast one's ballot, the lack of true intrinsic or extrinsic motivation inherent in political participation is a major criticism of laboratory-based election usability studies.

Voting has been argued to be an economic event (Downs, 1957) in which voters apply a cost-benefit model to choosing a candidate and choosing whether or not to vote. Because of this, reconstructing an election in the laboratory may not apply the motivational effects that are present in actual use of the machines during elections. Downs offers the explanation of voting as a purely economic activity—a voter weighs the costs and benefits of voting in deciding to vote. Downs's model was revolutionary in that he was the first to propose that choosing not to vote could be a rational decision. In prior laboratory usability studies such as Everett (2007) and others, a voter has no external motivation besides experimental demand characteristics to be effective and accurate in casting his/her ballot. In a real voting situation, Downs's theory and others note that voters would perceive the benefits of policies in line with their ideology by electing a preferred candidate. Campaign effects, including any implicit or explicit motivation to vote for candidates, are entirely missing from laboratory studies of usability. Because of these missing effects, our voters in a laboratory setting may not experience the intrinsic political motivation to correctly cast their ballot and ensure its accuracy.

While we cannot replicate the motivation of costs and benefits in controlled studies, we hope we can strike a balance that is more reflective of the real world than in previous usability studies. While there is an important difference between implicit and explicit task motivation, we hypothesize that the presence of any "benefit" should improve voters' performance. By providing external motivation through offering a five-dollar cash payment for rendering an accurate vote, the motivation to vote with care and accuracy should be induced. While strong motivation to vote for a candidate in a real election cannot be equated with a small cash payment, any finding of a motivational effect is important because it suggests participants in laboratory studies could improve their performance by increased motivation. We expect the motivation for increased accuracy may indeed demonstrate a time/accuracy tradeoff, with those given the incentive taking longer to cast their ballots, due to their motivation to accurately do so.

We also assess the impact of this additional motivation by including another variable in the experimental design. The use of a voter's guide filled with fake candidate information was provided to some voters, while others are simply given a list of candidates to vote for. If motivation does indeed have an impact on the effectiveness of the participants, it should affect these conditions differentially due to the personal choice and association with certain candidates on the ballot by those participants allowed to choose their votes.

Methods

Participants

Seventy-two participants from the greater Houston, Texas, area were recruited through newspaper and online ads for participation in a mock election study. The participants consisted of thirty-two men and forty women, with an average age of 47 years old (SD = 15). All possessed normal or corrected-to-normal vision and were fluent in English. More demographic characteristics of the participants are shown below in Table 1.

Yearly Wages	**# of Participants**	**Percentage**
below $20,000	16	22.20%
$20,000 - $40,000	22	30.60%
$40,000 - $60,000	24	33.30%
above $60,000	10	13.90%
Education Level		
High school or less	7	9.70%
Some College	27	37.50%
4-year Bachelor's	26	36.10%
Advanced Degree	12	16.70%
Race/Ethnicity		
Non-Hispanic White	35	48.60%
Black	22	30.60%
Hispanic	6	8.40%
Other	9	12.50%

Table 1. Economic, educational, and race/ethnicity distribution of participants

Materials

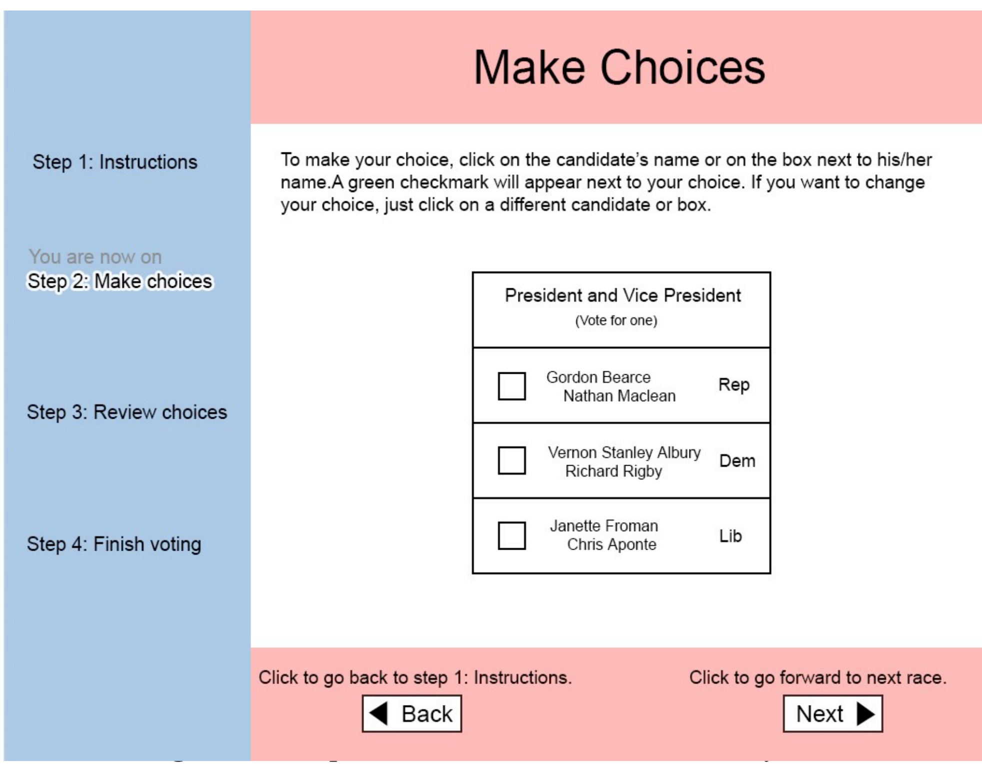
Because of the enormous differences in implementations of voting systems by states and counties across the United States, we used several prototypical election systems to study the hypothesized effects. The VoteBox DRE system (Figure 1), developed for usability studies (Everett, 2007; Sandler and Wallach, 2007; Everett et al., 2008) was used in the current study to examine participant performance on a prototypical DRE system. In addition, voters cast ballots on a traditional optical scan ballot (Figure 2), a lever machine (Figure 3), and a punch-card ballot system (Figure 4).

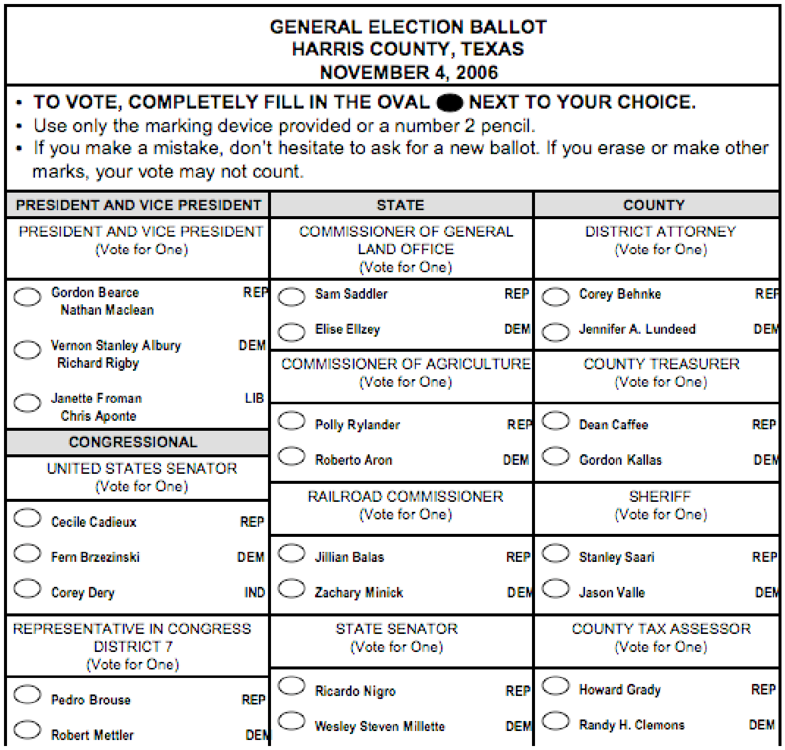
A fictional but realistic 27-race ballot, containing twenty-one political candidacies and six propositions, first used in Greene et al. (2006), was programmed into the four different voting systems. The fictional voter guide first developed for Greene et al. (2006) was given to half of the participants. This guide provided specific information about each candidate for office, in addition to arguments for and against the propositions on the ballot.

The "ideal" set of instructions given to participants in this study were developed from examinations of actual instructions present inside voting booths and on similar types of election technologies. A separate instructional walkthrough was created for each of the four technologies. See Appendix A for the full sets of instructions.

Design

Four between-subjects variables are manipulated in the current study. The first, the ideal instruction set, was given to half of the participants prior to using the selected voting technology, while the other half received no additional instructions beyond the very basic instructions printed on the ballot or the machine. The second, the extrinsic motivation, was manipulated in the form of a five-




GENERAL ELECTION BALLOT
HARRIS COUNTY, TEXAS
NOVEMBER 4, 2006

- **TO VOTE, COMPLETELY FILL IN THE OVAL ⬤ NEXT TO YOUR CHOICE.**
- Use only the marking device provided or a number 2 pencil.
- If you make a mistake, don't hesitate to ask for a new ballot. If you erase or make other marks, your vote may not count.

PRESIDENT AND VICE PRESIDENT	STATE	COUNTY
PRESIDENT AND VICE PRESIDENT (Vote for One)	COMMISSIONER OF GENERAL LAND OFFICE (Vote for One)	DISTRICT ATTORNEY (Vote for One)
Gordon Bearce / Nathan Maclean REP	Sam Saddler REP	Corey Behnke REP
Vernon Stanley Albury / Richard Rigby DEM	Elise Ellzey DEM	Jennifer A. Lundeed DEM
Janette Froman / Chris Aponte LIB	COMMISSIONER OF AGRICULTURE (Vote for One)	COUNTY TREASURER (Vote for One)
CONGRESSIONAL	Polly Rylander REP	Dean Caffee REP
UNITED STATES SENATOR (Vote for One)	Roberto Aron DEM	Gordon Kallas DEM
Cecile Cadieux REP	RAILROAD COMMISSIONER (Vote for One)	SHERIFF (Vote for One)
Fern Brzezinski DEM	Jillian Balas REP	Stanley Saari REP
Corey Dery IND	Zachary Minick DEM	Jason Valle DEM
REPRESENTATIVE IN CONGRESS DISTRICT 7 (Vote for One)	STATE SENATOR (Vote for One)	COUNTY TAX ASSESSOR (Vote for One)
Pedro Brouse REP	Ricardo Nigro REP	Howard Grady REP
Robert Mettler DEM	Wesley Steven Millette DEM	Randy H. Clemons DEM

Figure 2. Partial Optical Scan ballot

dollar "bonus" payment for correctly casting the ballot consistently for the same candidates; due to the nature of the experiment, we were not able to determine the accuracy of votes immediately, so all participants were paid the bonus at the conclusion of the experiment.

For the third variable, the voter's decision method, half of participants were given a fictional voter guide, while half were given a list of candidates for whom they were directed to cast their ballot. The last between-subjects variable was the older technology participants voted on. All participants voted on the VoteBox DRE system and on one other, older technology—either the lever machine, optical scan ballot, or punch-card system.

Participants were measured on two objective metrics—efficiency and effectiveness—while satisfaction was measured with a subjective scale. Effectiveness was measured in terms of error rates: under-votes, over-votes, and "switched" votes. All three of these occurrences were counted as separate types of errors. Efficiency was measured as time spent voting; participants were timed from when they were handed the instructions and told to begin voting until they completed the ballot. Finally, satisfaction was measured using the System Usability Scale previously mentioned—a ten question battery of questions about one's opinion of the technology. Finally, an extensive voting history and demographic information was collected from participants, including several manipulation checks.

Procedure

Participants were first given a set of written instructions explaining the purpose of the study, accompanied by either a slate of candidates they were directed to vote for or a fictional voter guide. If given the voter guide, the participants were given as much time as they wanted to familiarize themselves with the political candidates and propositions. Once the

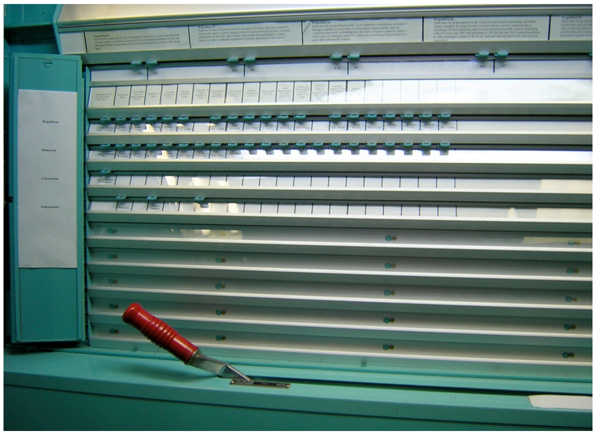


Figure 3. Front view of lever machine

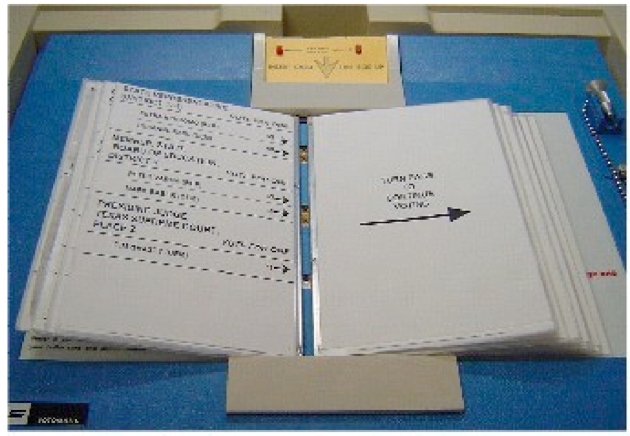


Figure 4. Punch-card ballot system

participant indicated he or she was ready to begin, several key instructions were reiterated verbally, including the presentation of the five-dollar extra incentive for accuracy in voting. Participants were instructed to vote for the same candidates on all ballots. If a participant was selected to receive the additional "ideal" instructions, they were given these immediately before approaching the voting equipment and beginning to cast their ballot.

The participant then voted using either the VoteBox system or punch-card, lever, or optical scan technology. After the first ballot was complete, the participant filled out a SUS for the technology they used. Next, the participant voted on a different technology, followed by another SUS; every participant voted on the VoteBox system and one other technology. Next, if the participant was in the voter guide condition, an exit interview was conducted, and the participant was asked verbally to reveal their vote choices. Finally, the participant filled out a longer questionnaire about their voting history and demographic information.

Results

Manipulation Checks

Prior to the data analysis, several questions were examined to determine whether the participants noticed the experimental manipulations and were affected by them. While these questions are subjective and rely on the participant's honesty, they provide some indication of the manipulation's success. Of the participants given additional instructions, 35 of the 36 participants said that they had read them prior to beginning, with all 35 noting that the instructions were helpful. Of those 35, 14 admitted to looking back at the instructions while voting.

Next, to check the incentive manipulation's effect on motivation, the participants were asked how important it was to them that their vote was accurately cast in the mock election. Additionally, participants were asked whether they double-checked their votes in the mock election, which would be an expected effect of the incentive. Surprisingly, there is no evidence that the subjective level of importance, rated on a 1–5 scale with 1 reflecting "most important," was different between the group that received the incentive (M = 1.44, SD = 0.65) and the group that did not receive the additional \$5 incentive (M = 1.34, SD = 0.54), $t(69) = -0.714$, $p = 0.48$. Additionally, there is no evidence that participants receiving the incentive were more likely to double-check their votes, $t(67.96) = 1.34$, $p = 0.19$. The degrees of freedom in this test were corrected

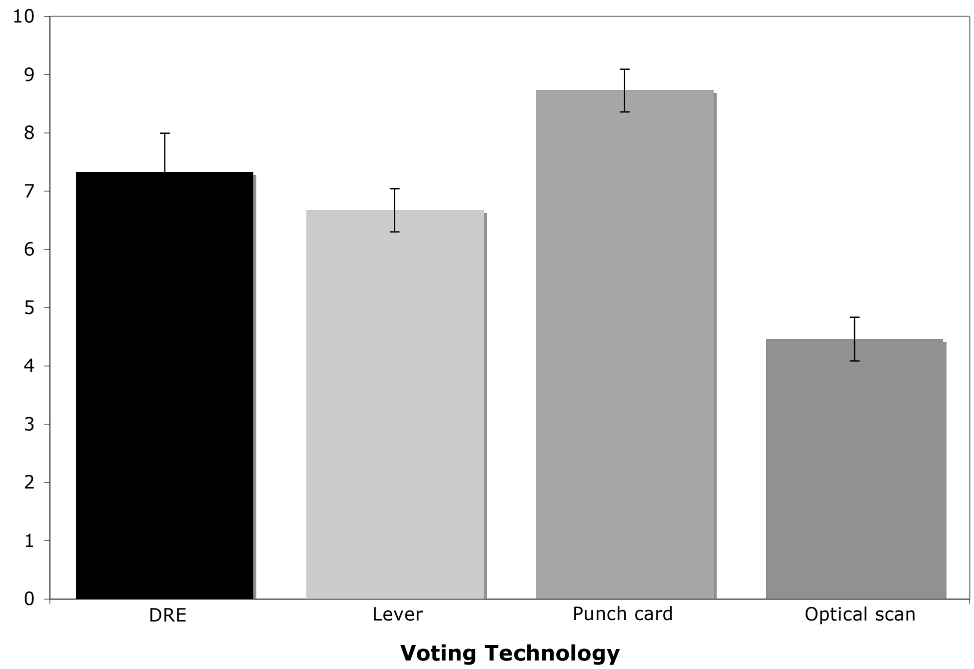


Figure 5. Ballot completion time by voting technology ± 1 SEM

for heteroscedasticity of variances. Because these measures were simple self-report questions of motivation, they do not necessarily predict the measurable effects on the metrics; however, they do indicate that there is no difference in the amount of conscious importance participants placed on the task of voting accurately during the study. Furthermore, the means for both groups were extremely close to the highest possible value for importance—meaning that regardless of the incentive, participants thought it extremely important to be accurate in the study.

Efficiency

One participant was excluded from these analyses due to extremely high completion times on both technologies. We find no reliable differences between the DRE and the other technologies on completion times, $F(1, 25) = 0.38$, $p = 0.55$ (See Figure 5).

We also find that participants given the monetary incentive take less time (M = 6.1 minutes, SD = 0.57) than those not given the incentive (M = 7.6, SD = 0.56). This effect is statistically reliable, $F(1, 25) = 4.91$, $p = 0.036$. Additionally, those choosing candidates from the voter guide take more time (M = 8.7, SD = 0.57) than those participants directed to vote for specific candidates (M =

Voting Technology	Percentage	St. Error	n
DRE	16.0%	4.0%	72
Lever	28.9%	8.4%	24
Punch card	31.2%	8.3%	24
Optical scan	12.5%	8.3%	24

Table 2. Percentage of ballots with at least one error

5.0, SD = 0.56); this effect is also statistically reliable, $F(1, 25) = 21.54$, $p < 0.001$. There is no evidence for an effect of instructions on participant efficiency, $F(1, 25) = 1.07$, $p = 0.31$.

Effectiveness

Error rates were calculated differently between the two decision conditions. For the voters that used the voter guide, an exit interview was conducted, and this was used as a third "vote." The vote inconsistent with the other two was deemed an error. For the directed, or "slate" condition, in which voters were told to vote for specific candidates, errors were calculated simply as incorrect selections.

If we measure error as a dichotomous variable at the ballot level—that is, whether a ballot contains at least one error or not, we find no statistically reliable differences between the four technologies. Additionally, there are no main effects of incentive, $F(1, 25) = 1.64$, $p = .21$, instructions, $F(1, 25) = 1.70$, $p = .21$, or voter's decision, $F(1, 25) = .96$, $p = .34$. We do find a significant interaction (See Figure 6) between the voter's decision and instructions, $F(1, 25) = 12.93$, $p = .001$.

To get a more precise measure of error, we measure at the level of each race, with the three different types of errors users could make. At this level, there are no significant effects of the technology or any of the independent variables. The mean error rates are shown below in Table 3.

Satisfaction

The SUS scores are computed on a 0–100 scale, with 100 indicating an ideal technology in terms of subjective satisfaction. The SUS scores reveal a clear subjective preference for DRE's (M = 91.3, SD = 14.5) over the other technologies (M = 69.3, SD = 23.7) (See Figure 7). This difference is statistically significant, $F(1, 25) = 32.91$, $p < .001$. We find that the SUS scores for the lever machine (M = 62.4, SD = 4.7), the punch-card ballots (M =

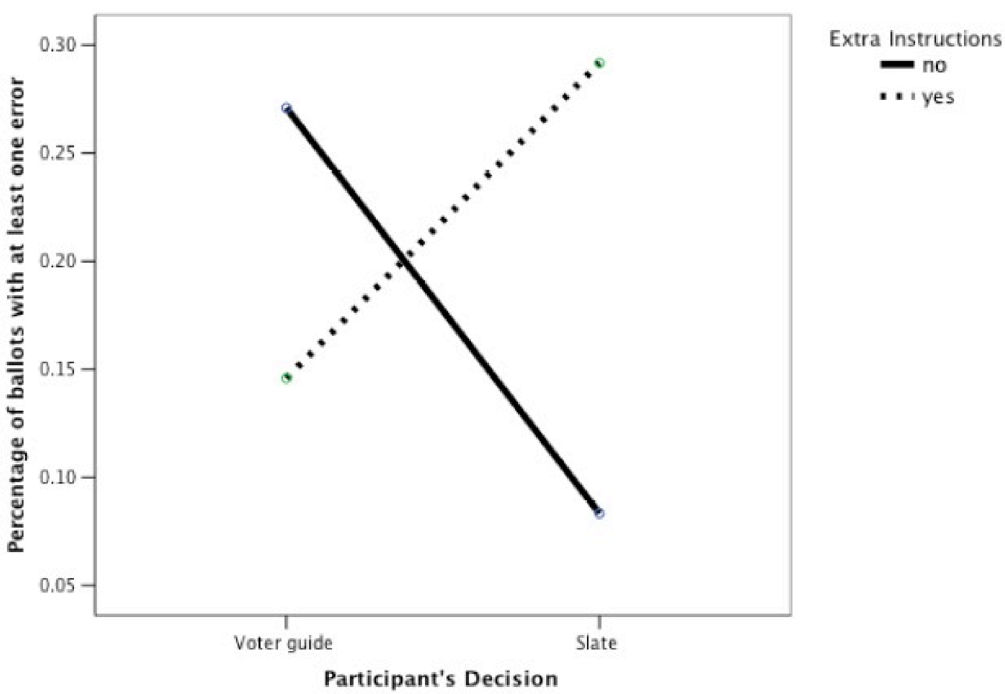


Figure 6. Interaction of participant's decision and instructions on error rate

Technology	Overall Rate	Overvote Rate	Undervote Rate	Wrong-vote Rate
DRE	1.1%	0.0%	1.2%	2.0%
Lever	0.4%	0.0%	0.2%	0.8%
Punch card	2.2%	0.5%	4.3%	2.0%
Optical Scan	1.6%	0.0%	3.1%	1.7%

Table 3. Mean error rates by technology and type of error

67.3, SD = 4.7), and the optical-scan ballots (M = 78.3, SD = 4.7), are not reliably different from one another, $F(2, 69) = 3.01$, $p = .056$.

While the differences between the technologies appear in the SUS scores, no other effects approached statistical significance. Neither prior use of a technology nor subjective level of computer expertise affected the SUS scores. There was no main effect of the incentive to increase motivation, $F(1, 25) = 0.018$, $p = 0.894$, the additional instructions, $F(1, 25) = 2.73$, $p = 0.11$, or the decision method on the SUS scores, $F(1, 25) = 3.37$, $p = 0.078$.

Subjective Responses / Voter Background

Due to the failure of the additional incentive to produce an effect on any of the three metrics above, we sought to examine whether a disparity in motivation actually existed between voting in the laboratory setting and during an actual election. Participants were asked for a 1–5 rating of the personal importance that their ballot was cast correctly during

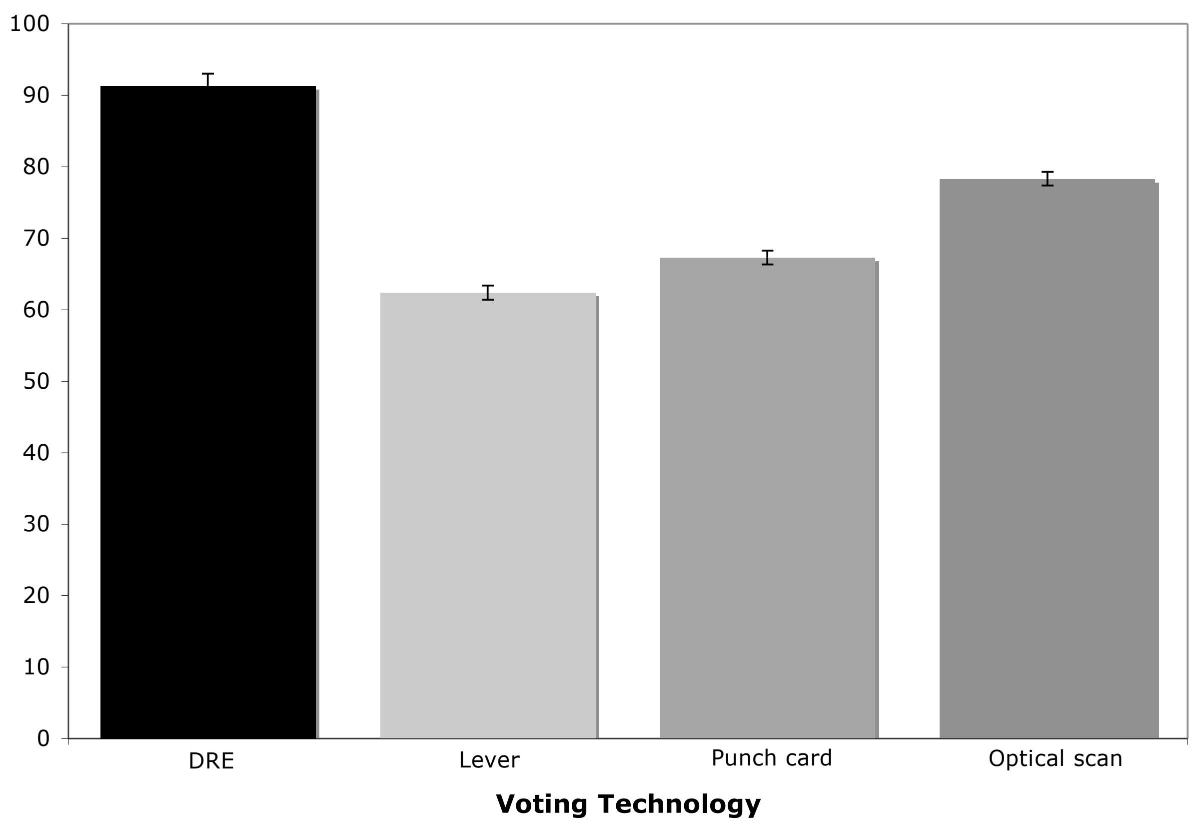


Figure 7. SUS scores by voting technology ± 1 SEM

a typical, real election, with 1 being "most important," in addition to an analogous question asking about the importance of their ballots being cast correctly during the experiment. We find that participants rated the importance of accuracy in a real election (M = 1.06, SD = 0.24) significantly higher than the importance of accuracy in the experiment (M = 1.38, SD = 0.6), $t(68) = 4.56$, $p < 0.001$. Although this difference in subjective importance is found, there was no evidence for a difference between rates of double-checking ballots for accuracy in real elections than mock elections, $t(69) = -0.94$, $p = 0.35$.

Although our sample is not meant to be representative of the electoral population, we collected some data regarding instruction use in real elections. 23.6 percent of our participants reported seeking instructions for voting equipment prior to election day in a previous election, while 19.4 percent have watched an online video tutorial for the equipment. 79.2 percent reported using instructions provided in voting booths to help them use the technology.

DISCUSSION

The manipulation checks conducted provided clear evidence for problems with the manipulations of the independent variables. Although the manipulation checks were simply subjective questions about the participant's use of the voting technologies, it is clear that the incentive did not increase voter motivation for accuracy. In addition, the additional "walkthrough" instructions provided to voters were only utilized during the process of voting by 14 of the 36 that were given the instructions.

Although the study produces few findings regarding the impact of voter motivation and instruction use on the voting technologies, it is still clear future research in this area is needed. Participants still rated the personal importance of accuracy in their ballot higher during real elections than in the laboratory experiment. This finding suggests that the laboratory study of voting machine usability may lack important ecological validity in terms of voter motivation. Because voters noted that they were consciously more motivated in real elections than in the current study to cast an accurate ballot, this could present a problem in generalizing the results of this study. It is extremely important to note, however, that there is a major difference between what a user wants to do and what a user is capable of doing in interacting with a technology. Even if voters are more motivated during an election, this does not necessarily mean they will have improved accuracy. Regardless, voters experience large amounts of pressure from political campaigns and their own decision-making processes in choosing to vote, and their votes in real elections carry actual impact, while they see no impact of their vote in a laboratory usability testing scenario. This poses problems for the ecological validity of our laboratory studies.

Even without conclusive findings regarding the two independent variables of interest, this study still provides clear data regarding the impact of the different technologies on usability. Although no differences between the technologies were found in terms of per-ballot error rates and per-race error rates, the mere finding that these error rates are nonzero is cause for concern. These findings of high, double-digit per-ballot error rates replicate previous work on older, non-DRE voting machines (Everett, 2007; Byrne, Greene, and Everett, 2007). While any level of error rate is cause for concern, the relative rareness of errors in voting studies such as this make it extremely hard to examine the causes of these errors. Because errors are a relatively rare event, between 0.4% and 2.2% per-race in this study, it is extremely hard to find reliable differences between experimental conditions, due to the extreme positive skew of the data and a modal error rate of zero. Because of this, the lack of evidence for an effect of increased motivation through incentives and better performance through instruction use is somewhat expected. The interaction between voters' decision method and whether they received additional instructions on per-ballot error rates is not readily explainable. Voters with the voter guide had higher error rates when not given instructions, while voters given a specific slate of candidates to vote for had higher error rates when they were given instructions.

With regard to the subjective satisfaction of the technologies, the prototypical DRE system, VoteBox, is clearly well liked by participants across the board. This finding of extremely high satisfaction ratings for the newer, DRE systems replicates earlier studies (Everett et al., 2008). Of interest, while DREs outperform the other technologies clearly in satisfaction, there is not evidence for a difference in efficiency or effectiveness. This finding, however, is somewhat counterintuitive. With the mounting news coverage over the potential flaws and problems with newer electronic systems, one would expect voters to not be as satisfied with the DRE systems. However, because the SUS scale measures the voter's subjective usability of the system, we see that voters have high ratings of the usability of the system, even though voters may have doubts over the integrity and security of the electronic voting system.

From the efficiency data, we see that all of the technologies do not differ significantly in the amount of time that voters took to complete a ballot. Of note, the completion times observed were on 27-race ballots; many states and jurisdictions have much longer ballots, including some ballots with over 100 races, which would dramatically increase the amount of time a voter needs to cast a ballot. In the study, voters that received the additional monetary incentive for correctly casting a ballot actually took less time than those that did not receive the incentive. This finding is quite counterintuitive, as we would expect those risking an incentive to check their votes to ensure their accuracy and therefore take more time.

While this study faced clear limitations, the impact of instructions and political motivation are still valid factors for study in future election usability studies. Furthermore, this study replicates several key findings of previous research, including the high subjective ratings of newer, electronic systems, accompanied by a relative lack of improvement in efficiency and effectiveness. The use of instructions and training of voters is also a key component of making election technologies more usable. Because elections are unusually rare events, traditional models of training and habitual learning do not apply—voters must be helped to perform a relatively novel task, while still maintaining their privacy. Tightly controlled laboratory studies of usability have their benefits, but elections are a unique event. Further studies of elections with high ecological validity and the presence of campaign effects from actual elections are needed.

Acknowledgements

This research was supported by the Rice Undergraduate Scholars Program (RUSP) and by the National Science Foundation under grant #CNS-0524211 (the ACCURATE center). The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of the NSF, the U.S. Government, or any other organization. I would like to thank my advisor, Dr. Michael Byrne, and Bryan

Campbell and Kristen Greene for their assistance with this research.

REFERENCES CITED

Ansolabehere, S. & C. Stewart III. 2005. Residual votes attributable to technology. *The Journal of Politics*. 67:2, 365–389.

Brooke, J., Jordan, P. W., Thomas, B., Weerdmeester, B. A., & McClelland, A. L. 1996. SUS: A "quick and dirty" usability scale. *Usability Evaluation in Industry*: 189–194. London: Taylor and Francis.

Byrne, M. D., K. K. Greene, & S. P. Everett. 2007. Usability of voting systems: Baseline data for paper, punch cards, and lever machines. *Human Factors in Computing Systems: Proceedings of CHI 2007*: 171–180. New York: ACM.

Downs, A. 1957. *An Economic Theory of Democracy.* New York: Harper & Brothers.

Election Assistance Commission. 2005. Voluntary Voting System Guidelines. <http://vote.nist.gov/VVSG2005Vol1.pdf> <http://vote.nist.gov/VVSG2005Vol2.pdf>

Election Data Services. 6 February 2006. 69 Million Voters will use Optical Scan Ballots in 2006; 66 Million Voters will use Electronic Equipment. Election Data Services Press Release. <http://www.electiondataservices.com/EDSInc_VEStudy2006.pdf>

Election Science Institute. 2006. DRE Analysis for May 2006 Primary Cuyahoga County, Ohio. San Fransisco, CA.

Everett, S. P. 2007. *The usability of electronic voting machines and how votes can be changed without detection*. Unpublished doctoral dissertation, Rice University, Houston, TX.

Everett, S. P., K. K. Greene, M. D. Byrne, D. S. Wallach, K. Derr, D. Sandler, and T. Torous. 2008. Electronic voting machines versus traditional methods: Improved preference, similar performance. *Human Factors in Computing Systems: Proceedings of CHI 2008*. New York: ACM.

Greene, K. K., M. D. Byrne, and S. P. Everett. 2006. A comparison of usability between voting methods. *Proceedings of the 2006 USENIX/ACCURATE Electronic Voting Technology Workshop*. Vancouver, BC, Canada.

Hall, J. L. 2007. Contractual barriers to transparency in electronic voting. *Proceedings of the 2007 USENIX/ACCURATE Electronic Voting Technology Workshop.* Boston, MA.

Herrnson, P. S., B. B. Bederson, R. G. Niemi, F. Conrad, M. J. Hanmer, M. Traugott, & P. L. Francia. 2005. 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.

Jamieson, B. A., & W. A. Rogers. 2000. Age-related effects of blocked and random practice schedules on learning a new technology. *Journal of Gerontology: Psychological Sciences*, *55B*, 343–353.

Laskowski, S. J., M. Autry, J. Cugini, W. Killam, and J. Yen. 2004. *Improving the usability and accessibility of voting systems and products* (Publication No. 500–256). NIST.

Mead, S., & A. D. Fisk. 1998. Measuring skill acquisition and retention with an ATM simulator: The need for age-specific training. *Human Factors 40*, 516–523.

Mead, S. E., P. Batsakes, A. D. Fisk, and A. Mykityshyn. 1999. Application of cognitive theory to training and design solutions for age-related computer use. *International Journal of Behavioral Development*, *23*, 553–573.

Morrell, R. W., K. V. Echt, A. D. Fisk, and W. A. Rogers. 1997. Designing written instructions for older adults: Learning to use computers. *Handbook of Human Factors and the Older Adult*: 335–361. San Diego, CA: Academic Press.

Sandler, D., and D. S. Wallach. 2007. Casting votes in the auditorium. *Proceedings of the USENIX/ACCURATE Electronic Voting Technology 2007 Workshop*. Boston, MA.

Van Gerven, P. W., F. Paas, and H. K. Tabbers. 2006. Cognitive aging and computer-based instructional design: Where do we go from here? *Educational Psychology Review 18*: 141–157.

Wiltenburg, M. 3 November 2003. A better ballot. *Christian Science Monitor.*

Appendix. Additional instructional walkthroughs provided to voters

Electronic Voting Machine (VoteBox) Instructions

Step 1: To go to first race, click the "next page" button in the bottom right of the screen with the computer mouse.
Step 2: To select your vote choices, click on the checkbox next to the choice you prefer.
Step 3: Click the next button in the bottom right of the screen to advance to the next screen.
Step 4: After making all your choices, review the summary screen to ensure your choices are correct, and then select "next page".
Step 5: Click the "Record Vote" button in the bottom right of the final page to cast your ballot.

Optical Scan (Bubble Ballot) Voting Instructions

Step 1: Using the #2 pencil provided to you, completely fill in the oval next to your preferred choice in each race.
Step 2: Vote for the races in all three columns on both the front and the back of the ballot.
Step 3: After you have selected all your choices, notify the experimenter.

Punch-Card Voting Instructions

Step 1: Carefully insert the punch-card into the slot at the top of the machine. Make sure the correct side of the card is facing up.
Step 2: Push the card far enough down so that the holes in the card come to rest over the posts at the top of the machine.
Step 3: Turn the book of candidates to the first page, if it is not already there.
Step 3: Using ONLY the punching apparatus attached to the voting booth, vote for your choices by pressing the needle straight down through the hole next to your choice in each race.
Step 4: Turn the pages of the book and vote in every race.
Step 5: When finished voting, remove the punch-card from the machine by pulling it out the slot at the top of the machine and return it to the experimenter.

Lever Machine Voting Instructions

Step 1: Pull the large red handle to the right, and then do not move it.
Step 2: Each race is held in a separate vertical column. To indicate your votes, pull the small lever above the candidate or proposition you support downwards. If you make a mistake, you may simply push the small lever back up.
Step 3: After you have selected all your choices, notify the experimenter.
Step 4: DO NOT pull the red handle back to the left until you are told to do so.