

Usability Review of the Diebold DRE System for Four Counties in the State of Maryland



Benjamin B. Bederson
Director, Human-Computer Interaction Lab
Computer Science Dept.
University of Maryland
bederson@cs.umd.edu
(301) 405-2764

Paul S. Herrnson
Director, Center for American Politics and Citizenship
Dept. of Government and Politics
University of Maryland
pherrnson@capc.umd.edu
(301) 405-4123

Assisted by: Bongshin Lee, Scott Richman and Bongwon Suh

Executive Summary

We assisted Allegany, Dorchester, Montgomery, and Prince George's counties with acceptance testing of the Diebold AccuVote-TS voting system using three commonly used techniques: expert review, close-up observation, and field testing. Our findings show that the system has both strengths and weaknesses, some of which election officials may wish to have addressed.

Our major findings are as follows:

- One of the two machines we tested exhibited catastrophic failure.
- The system presents inconsistent terminology, which could confuse voters.
- The help instructions are unclear and unnecessarily long.
- There is no help button that can be used while voting.
- The layout could use improvement, especially because the review screen is organized differently than the voting screens.
- No warning is given for overvoting.
- The screen graphics could use improvement.
- Potentially distracting information related to election management and system operations appears on the screen.
- The system's legs are wobbly and flimsy.
- The audio-only (disabled) system has shortcomings: it is hard to navigate, it is difficult to have questions repeated, the audio quality is poor, the buttons provide no feedback, there is no opportunity to review the ballot.
- Most of the voters who tested the system responded favorably to it. However, the responses of a significant number indicated that improvements could be made to strengthen voters' overall impressions of the system, their comfort in using it, their ability to read and understanding the terminology presented on the screen, their ability to correct mistakes, and their trust in the system's ability to accurately record their votes.

Introduction

At the request of election officials from Allegany, Dorchester, Montgomery, and Prince George's counties, we evaluated the Diebold AccuVote-TS voting system using three commonly used techniques: expert review, close-up observation, and field testing. The result of each technique suggests the voting has some shortcomings that election officials may wish to have addressed.

Expert Review

The first methodology, expert review, enabled us to analyze the systems in detail, comparing it to the state-of-the-art in touch-screen user interfaces. Expert review consists of having several individuals with significant experience in user interface design walk through the system in detail, perform representative tasks, and record where weaknesses occur. This approach relies on the expertise of reviewers who are very familiar with the state-of-the-art in user interface design, both in terms of other existing commercial systems as well as scientific research about what works and doesn't. Another benefit of expert review is that based on their knowledge, the experts can make concrete suggestions about how to improve problematic aspects of the design. The downside of expert reviews is that even experts can't predict how every person will actually respond to a new situation – and so while they are likely to identify general problem areas and specific things to look for, they are likely to miss some things that a broad population will run into.

We performed our expert review with five faculty and staff at the Human-Computer Interaction Laboratory at the University of Maryland. Each person spent approximately one hour using the DRE system and independently reported their concerns and suggested solutions. The standard and audio-only (disabled) systems were evaluated independently.

Expert Review Results

Each problem area is listed with the number of experts that reported that problem and suggestions for fixing the problem.

Traditional Visual System

1. Inconsistent Terminology/Labeling (# reviewers – 5). The text used throughout the interface is crucial, and must be written very carefully since this is the primary way that voters understand what to do with the interface. Several specific issues are listed here:
 - "Review Ballot" is confusing because this navigational button is presented while actively reviewing the ballot. It has same functionality as Previous button.
 - In review mode, "office" is the wrong word since it doesn't represent referendums.
 - In the write-in screen, an instruction refers to "backspace" but there is no key with such label.
 - The instructions are not consistent with the actual interface. Example: Instructions include "CAST BALLOT, NEXT" and the actual buttons are "Cast Ballot, Next".
 - For question 1, YES/NO is better than 'change/don't change the amendment'.
 - The help screen only mentions candidates (not answers to questions).

Suggestions:

Rename “Review Ballot” to "Back to full ballot"
Carefully choose terms and use them consistently.

2. Color usage (# reviewers – 4)
 - The “X” on red background is too low contrast.
 - The red X indicating selection doesn’t always look red.
3. Inserting/Removing card (# reviewers – 4)
 - Difficult to insert card, and to know where to insert card
 - There is a little delay between inserting the card and having the machine react

Suggestions:

Should have physical label by reader in addition to onscreen arrow
The message on the screen should describe a 2-step operation: “Insert the card and push it in lightly until it is fully inserted”
Provide feedback at the point when the card is just inserted and the machine has not yet responded.
Add a reminder sound when the card has been ejected, but not yet removed as ATM machines do.

4. Help / Instructions (# reviewers – 4)
 - Instructions are long and unclear
 - No overview of the process
 - No help button during voting

Suggestions:

Instructions should have interactive example (a working button rather than just a picture of one), and abstracted views of voting screen rather textual descriptions (e.g., "bottom right")
Should always have access to help, i.e.:



5. Layout (# reviewers – 4)
 - There is concern about how review be shown when it is larger than one screen.
 - There is concern about what the ballot will look like when the list of candidates is more than a column long or when names are exceptionally long?
 - The review screen is organized differently than voting screens which may cause confusion.
 - The three lines of the instructions in the review page should be left justified and similar in format to the instruction page.
6. System information shown (# reviewers – 4)
 - Information about the system was shown (including “Unit” and “Version”).

- Information about the election was shown (including "Election Total", "Election", and "Precinct").

Suggestions:

Don't show election management information

Or if necessary, move that information to a different screen and let it be accessed by a special poll worker card only.

7. Glare on screen (# reviewers – 3)

Suggestions:

Light must be positioned to minimize glare on screen

8. Feedback / Warning (# reviewers – 2)

- If you try to overvote, no warning is given - instead, nothing happens at all.
- There isn't any audio feedback on the next/prev buttons - only when casting votes.

Suggestions:

Give a warning when user tries to overvote.

Add audio feedback to all interactions

9. Poor graphic quality (# reviewers – 2)

- Start screen image is too low resolution.
- The image of the card on the start screen doesn't match the actual physical voter card.
- There are too many colors and font styles in help and review.
- The "X" in the button that shows a vote is ugly - it goes outside the box.

Solutions:

Hire a graphic designer.

10. Privacy (# reviewers – 1)

- Not private enough - others might be able to see my vote

Suggestions:

Polling place will have to construct larger privacy shields

11. Specific to Review Screen (# reviewers – 1)

- The message at the top should be context dependent.
- If all races or referendum questions have been answered, say so.
- Undervotes aren't highlighted.
- At the same time, make it clear that it's ok to cast incomplete ballots.
- Red may imply to some users that they did something wrong, not that their ballot is incomplete.
- When you select a particular item, you are taken to the original page the item was in - but there is no indication which item you asked to modify.
- The "Cast Ballot" text should be centered within its button.
- The "Cast Ballot" might be confusing to non native English speaker.

- No chance to cancel “Cast Ballot” if pressed by mistake.

Suggestions:

Highlight undervotes in red.

Highlight the item you are brought to, perhaps with an arrow and some text. Or, instead go to a special screen to modify just that item. Otherwise, there could be navigational problems for large ballots.

Have instructions above the buttons:

To make changes in your vote:

PREVIOUS

To confirm your vote

Cast ballot and END

12. Miscellaneous (# reviewers – 1)

- The hourglass won't be understandable to all users.
- The bottom middle button has some missing pixels in the border.
- The legs on the device are somewhat weak and wobbly.
- Will wheelchair fit between legs of ballot machine?
- There is no cancel and don't vote button.

Suggestions:

Replace hourglass with a clear text message (e.g., "Please wait a moment")

General “cancel” button may not be necessary for voters, but it may be for the election official if a voter leaves in the middle of the vote.

Audio-only (disabled) System

1. Inappropriate Keypad Mapping (# reviewers – 5)

- It is very strange. It's hard to remember which number assigned to which function.
- Hard to navigate
- Having the question repeated is very difficult.

Suggestions:

Relayout touchtone keyboard functionality:

1: previous race or question	2	3: next race or question
4: previous candidate	5: select	6: next candidate
7	8	9: cast ballot
	0: repeat questions, instructions	

Give users the option to practice with the keypad and the mapping of the keys before starting.

2. Audio Quality (# reviewers – 5)

- The audio quality is full of static and very difficult to understand.
- There are delays between audio segments which make it harder to understand.

- Different voices are distracting.
- Annoying clicks between sentences.

Suggestions:

The segments could be woven together more smoothly, especially eliminating the click sound and delay between segments.

3. Review Ballot (# reviewers – 3)
 - There is no review of the ballot before casting it.

Suggestions:

Give the option to have the whole ballot summary read to the users in a concise way. Make it possible to stop the review at any time to return to the voting procedure or cast the ballot.

4. Feedback (# reviewers – 2)
 - Buttons don't have any audio feedback when pressed.
 - If you try to overvote, nothing is said at all, so it is impossible to know that you didn't vote.
 - Undervote noted but not specific.
5. "Write in candidate" (# reviewers – 2)
 - It is loud, in a different voice, and unimplemented.
6. Cast Ballot (# reviewers – 2)
 - After voting, the users are forced to go through the rest of the candidates. This is slow and frustrating.
7. Script (# reviewers – 1)
 - The system starts with demeaning phrase "Visual Impaired Ballot System."
 - Highly repetitious
8. The volume control doesn't indicate which way is loud or soft (# reviewers – 1).

Close-up Observation

Close-up observation consists of our monitoring non-expert individuals interacting with the voting machines. We observed and videotaped individuals responding to different aspects of the voting process, including inserting the ballot card, selecting the candidates, and casting their ballot. We also asked the voters to "think out loud while they were voting" so we could learn about their immediate responses to different aspects of the process. Finally, we had the voters fill out a questionnaire describing their reactions to the new voting system. For each participant, we measured how long it took them to vote the entire election from the time they walked up to the machine to the time they left. We also counted how many errors they made.

One of the strengths of this methodology is that it enabled us to focus on the voters' interactions with the machines and their thought processes in a controlled environment that was free from distractions. This enabled us to pinpoint areas where they encountered shortcomings in the system. A second strength was that we could hear the voters describe aspects of the voting machines that met their approval or made them uncomfortable. The major weakness of this approach is it lacks in realism and can only be used to record the assessments of a small set of voters. In this case, the group consisted of 47 University of Maryland members primarily including students, but also including faculty and staff — a highly educated population group that utilizes computer technology on a regular basis. The election that was tested included five races and one question that was split between two screens.

Close-up Observation Results

The average time to complete the ballot was 2 minutes and 10 seconds. All participants except one completed their vote successfully. The one problem occurred when the participant was unable to figure how to write-in a candidate.

The participants generally liked the DREs, rating their overall comfort 7.7 out of 9 (on a 1-9 scale where 9 represented highest level of comfort) in terms of overall comfort. They found the screen layouts and color more problematic (6.9 out of 9).

Some representative comments from participants include:

Positive comments

- Easy to use, straightforward
- Excellent idea

Negative comments

- Inserting card was very confusing.
- Concerns about reliability
- Colors are not well chosen.
- Font size could be bigger.
- Layout of the ballot was confusing.

One subject pointed out that training is very important. It would be helpful if the county provide a homepage where voters can practice to vote.

We then watched the videotapes again, and compiled our analysis of this observations. We summarize our findings here:

System Failure

One of the main concerns of the electronic voting system is reliability. We didn't expect we could measure the robustness of this system because this simulated election was so simple and the number of subjects was not large enough to test it. However, at the very start of the experiment, one of the two machines malfunctioned and was unusable (it would not return the voter card). This raises serious concerns.

Card Insertion

It is necessary to insert an access card into the machine for voters to start voting, and many participants showed difficulties in inserting card. They expected the machine to accept the card as ATMs do. So, they just put the card in the slot gently and waited until the machine took it in. This situation might be improved by providing better instruction with some animation.

Help/Instructions

Most of the subjects spent a fair amount of time reading the instructions on the screen which reminds us of the importance of training. The county should provide a chance for voters to use this system (or a simulated one on the internet) in advance of all elections. In addition, they should provide printed instruction before the election. The online instructions should not be removed, but it is important to give as much advance instruction as possible.

Layout

While only a small number of subjects were concerned with the layout of the ballot, we should point out that test ballot was very simple, and does not represent the complexities of many actual ballots. All the candidates of each race fit in one column, and the entire election was only two pages. We caution that the importance of the layout of the ballot should not be underestimated.

Language Selection

There were two language options, English and Spanish, and English was selected by default. The shape and lay out of the buttons were not clear. So, most of the subjects touched the “English” button and then waited for the next screen. It often took several seconds for voters to recognize they also have to press “Start” button.

Undervoting

The system offers a summary page to the voter once the voter has sequenced through the entire ballot. The summary page indicates via a distinct color which races haven’t been voted at all. However, if a race was undervoted (i.e., the full number of candidates haven’t been voted for), then that race is not highlighted and is reported on the summary page in the same way that a fully voted for race is.

Field Tests

The third aspect of the study — the field test — complements the other two approaches because it is designed to involve a larger, more heterogeneous, and more representative group of individuals in a more natural setting. There were three important components to the field test we designed for implementation by the election officials: 1) the observation and recording of information about individuals’ interactions with the new voting systems, 2) the administration of a questionnaire to record the voters’ assessments of the systems, and 3) the administration of parts 1 and 2 to a large heterogeneous group of voters, including some Spanish-speaking individuals who were to receive a Spanish language ballot and questionnaire. The importance of the first two points—observation and recording voters’ interactions and administering the questionnaire—have been discussed above in connection with close-up observation.

The third component, involving a large heterogeneous group of voters is also important. Ideally, the participants would have included voters who differed on issues such as computer ownership, computer usage, Internet usage, age, education, sex, native language, race, and a host of other relative background variables. This would have enabled us to learn how individuals from a set of backgrounds that is almost as diverse as Maryland voters assessed the new voting system. We would have used this information to discern with a high degree of accuracy: 1) the level of challenges that the new voting machines would pose for Maryland voters, and 2) whether certain population groups would face more hurdles in acclimating to the new machines than others. Both sets of information would have been useful in terms of identifying general shortcomings in the usability of the voting system and which population groups should be targeted for an educational campaign about using the system.

Caveats

Unfortunately, the field study had two major shortcomings. The election officials did not record information about individuals' interactions with the voting machines, and they did not involve a large heterogeneous population in the assessment (this was mainly due to the fact that the vast majority of participants – 365 – came from Montgomery County). Thus we have no record of voters' interactions with the voting machines, and we only have responses from a very narrow slice of the population of Maryland voters. The voters who participated in the study consist of individuals from a relatively affluent retirement community, four libraries, a shopping mall, and the lobby of the Prince George's County Administration Building. Because they are mostly an economically and socially elite population group, whose levels of educational attainment, computer usage, and Internet usage are higher than the general population of Maryland voters, the experiences these citizens had with the new voting system are not representative of those of Maryland voters in general. Virtually absent from the field test are the experiences of individuals in rural or farming communities, individuals 34 years of age or younger (more than 60 percent of the participants are over 65 years of age), individuals who have not earned a high school diploma (over half have a degree from a four-year college and 32 percent have done some post-graduate work), members of most minority populations (Latinos, Asian Americans, Native Americans, and multiracial citizens each comprised less than 3 percent of the participants and African Americans accounted for only 8 percent), and individuals born outside the United States or whose native language is not English. Thus the results we report below paint an incomplete and probably overly favorable assessment of how Marylanders can be expected to respond to the new voting systems absent a major educational campaign. We expect that if a less elite and more diverse group of citizens been involved in the field test, we would have reported that a substantially larger number of voters found using the voting system challenging.

A final caveat concerns the nature of the ballot used in testing the voting machines. In all three cases, a very short and simple ballot was used. The simplicity of the ballot is important because the challenges that participants faced is lower than the challenges that voters are likely to encounter on election day.

Field Test Results

The field study produced the following results:

1) When asked to report their **overall impressions** about using the system on a scale of 1 through 9 (where 1 records the highest level of difficulty and 9 represents the highest level of ease), 80 percent of the respondents reported the system was easy to use (rated 8 or 9), 10 percent reported it was moderately easy to use (rated 7), and the responses of the remaining 10 percent indicate it was anywhere from difficult to somewhat challenging for them to use (rated from 1 to 6). Although 10 percent seems a small portion, it is important to recall that this is 10 percent of a group of social and economic elites and 10 percent of Maryland's voting age population equals roughly 383,000 voters, according to 1998 estimates. Had the field test been conducted on a more representative and diverse group of voters, we believe that both the percentage and number of voters who reported they found the new voting system somewhat challenging would have been substantially larger.

Despite the overall homogeneity of the sample, there was some variation of opinion among the respondents. Individuals who own a personal computer, use computers frequently, or live in a city or suburban area had more favorable overall impressions of the new voting system than did others. Women had more favorable impressions than did men.

2) When asked to report whether they **felt comfortable using** the system on a scale of 1 through 9 (where 1 records the lowest level of comfort and 9 represents the highest level) 86 percent of the respondents reported they were comfortable using the system (rated 8 or 9), 7 percent reported they were moderately comfortable (rated 7), and the responses of the remaining 7 percent indicate they were anywhere from uncomfortable to somewhat comfortable using the system (rated from 1 to 6).

Once again, individuals who own personal computers, use computers frequently, or live in a city or suburban area felt more favorable using the new voting system than did others. Women were more comfortable than were men. (Note: all of the comparisons reported here and below are statistically significant.)

3) When asked how easy it was to **read the characters on the screen** using a scale of 1 through 9 (where 1 records the greatest difficulty of reading and 9 represents its being easy to read) 86 percent of the respondents reported they it was easy to read the screen (rated 8 or 9), 8 percent reported they found the screen moderately easy to read (rated 7), and the responses of the remaining 6 percent indicate they found the screen anywhere from hard to read to somewhat easy to (rated from 1 to 6).

Individuals who are older or have higher levels of educational attainment had more difficulty reading the characters on the screen than did others. This finding conforms to what is generally known about variations in the eyesight of these population groups.

4) When asked to assess the **terminology on the voting system's screen** using a scale of 1 through 9 (where 1 records it was very ambiguous and 9 represents its being very precise) 83 percent respondents reported they the terminology was precise (rated 8 or 9), 10 percent reported

they found the screen moderately precise (rated 7), and the responses of the remaining 7 percent indicate they found the screen anywhere from hard to read to somewhat easy to (rated from 1 to 6).

Individuals who use personal computers less frequently found the terminology on the system's screen more ambiguous than did others.

5) When asked to report whether **correcting mistakes was easy** using a scale of 1 through 9 (where 1 records the highest level of difficulty and 9 represents the highest level of ease), 81 percent of the respondents reported the system was easy to use (rated 8 or 9), 11 percent reported it was moderately easy to use (rated 7), and the responses of the remaining 8 percent indicate it was anywhere from difficult to somewhat challenging to use (rated from 1 to 6).

Individuals who use computers frequently found it easier to correct mistakes than did others.

6) When asked to report whether they **trusted that the system recorded the votes they intended to cast** using a scale of 1 through 9 (where 1 records the voter did not trust the system and 9 records the voter trusted the system), 85 percent of the respondents reported they trusted the system (rated 8 or 9), 7 percent reported they trusted the system moderately (rated 7), and the responses of the remaining 8 percent indicate the voters did not trust the system or they only trusted it somewhat (rated from 1 to 6).

Individuals who use computers frequently reported having less trust in the new voting systems than did others. This result probably stems from frequent computers users' greater awareness of the limitations of computer technology, exposure to computer "crashes," familiarity with viruses, and other challenges facing the computer industry.

Summary

The results of the expert review, close-up observation, and field test suggest that the Diebold AccuVote-TS voting system has both strengths and weaknesses, some of which election officials may wish to have addressed.