# Voter Review and Verification of Ballots: Review of the Literature and Research Approaches

Whitney Quesenbery
Suzanne Chapman
Christopher Patten
Center for Civic Design

Roberto Spreggiaro
University of Milan

Sharon J. Laskowski
National Institute of Standards and Technology

**Disclaimer**

Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

**Note**

This work was performed by the Center for Civic Design in collaboration with NIST under prime contract GS06F0942Z, subcontract DCI-18-00019 from the U.S. Departmet of Commerce, National Institute of Standards and Technology.

# Contents

# Context for this work

The last comprehensive review of the current landscape and standards for improving the usability and accessibility of voting systems was published 15 years ago (Laskowski et al., 2004). This National Institute of Standards and Technology (NIST) special publication reviewed the research on human factors engineering, usability and accessibility including a small number of reports written between 2001 and 2003 in the aftermath to the 2000 U.S. Presidential election. Many of the recommendations in that report have been incorporated into the U.S. Election Assistance Commission (EAC) Voluntary Voting System Guidelines 1.0, 1.1, and 2.0 (EAC 2020).

Since then, the focus of work on usability and accessibility of voting systems has shifted several times: from computerized systems, called "direct recording electronic" or DRE, to paper ballots marked by hand and cast in a ballot scanner to hybrid systems called ballot marking devices (BMDs).

BMDs have been in use for many years as an option for the accessible voting system required by the Help America Vote Act of 2002 (HAVA). They provide a way for people with disabilities to mark their ballot independently and privately, then print and cast it with other ballots and are sometimes used for alternative language access (required by the Voting Rights Act). Attention to BMDs has increased in recent years as elections departments have considered them for more widespread use in vote centers and polling places as offering the advantages of both an electronic marking interface and the security of a paper ballot.

One of the major issues for voting systems today is whether they provide voters with a meaningful opportunity to verify their ballot before casting it. This opportunity is important in helping them vote their intent by catching errors or omissions made during marking, such as skipping a contest or mis-marking a selection. It is also an opportunity to confirm that whether the ballot is marked by hand or printed by a BMD, it accurately reflects their choices – including finding any malicious changes (or software errors) made by the voting system between the voter's review and the printing of the ballot.

This review of the literature was done in preparation for research to understand how voters approach the task of marking, reviewing, verifying, and casting a ballot. We included not only research on voting systems and voting, but on related issues of trust, privacy, and mental models.

One of the challenges in reviewing previous research is that voting systems and issues ripe for investigation have changed over the years. For example, can studies of electronic review screens be compared to studies of paper ballots and verification on BMDs? It can also be hard to assess the impact of using prototypes rather than real voting systems and all the other variables of setting up a research session.

In addition, we were interested in what is known about how to conduct this research. Voting is an activity that is normally conducted privately. Much as we might want to follow voters into the polling place and interview them as they vote, this is not ethically or legally possible. We looked at the methodologies of other researchers and used that information to help design our own research.

## The journey of casting a ballot

The act of casting a ballot is the central step in the larger voter journey (Laskowski et al., 2015; Chisnell, 2017), which includes registering to vote, learning about the election, and learning the results.

Similarly, the act of voting itself is a micro-journey, each with a different task. This means that during marking, reviewing, verifying, and casting a ballot, different voter mental models and therefore different design requirements are in play:

- **Marking.** The voter emphasis is on making selections. The ballot and/or voting system assists with an emphasis on the rules and choices, helping the voter focus on each contest, one at a time.
- **Reviewing.** On an electronic voting interface, the review is a separate stage when the voter can check that they have completed the ballot as they planned as well as confirm their choices.
- **Verifying.** The printed ballot is a confirmation, allowing the voter to verify all of the contests and selections (and undervotes) before casting the ballot.

The way each voting system designs these stages will be specific to the architecture and how voters will interact with it. For example, a ballot marking device that expects voters to come to the polling place with a list of choices, transferring their selections to a system at the polling place such as the experimental *Balloting* (Lola et al., 2013) or the pre-marking option in the Voting System for All People (VSAP), the new voting system for Los Angeles County (described in a video at https://vsap.lavote.net/) would place greater emphasis on the review stage than the initial marking of selections.

## Detecting voting contrary to intent

Because voting is a private process, it is hard to determine whether a problem detected on a ballot is intentional or not. The CalTech/MIT Voting Technology Project (2001) created the term "residual vote," combining the number of overvotes, undervotes, and spoiled ballots into a single aggregate-level measure. This measure is useful, but cannot tell whether an undervote is intentional or not, and cannot identify votes that are mistakenly cast for a candidate other than the one the voter intended. The terms are described in Norden (2006) as:

- An **overvote** occurs when the number of selections made is more than the maximum number allowed. An overvote is always considered an error because it means that no vote is counted for that contest.
- An **undervote** occurs when the number of voter selections in a contest is less than the maximum number allowed for that contest or when no selection is made. The number of undervotes is equal to the number of votes lost, for example, if no selection is made in a vote for two contest the number of votes lost is two.
- The **residual vote rate** is the difference between the number of ballots cast and the number of valid votes cast in a particular contest.

Looking across the interactions of marking, reviewing, and verifying, there are several root causes why a ballot might be cast that did not reflect the voter's intent. The terms we have used in this document are:

- **Errors** include overvotes (if the system allows it), selection mistakes (a candidate other than the one they want to vote for), and unintentional undervotes (skipping a contest they intended to vote in or under-voting a contest because they did not know they could choose more than one candidate).
- **Malicious changes** include any changed selection made by the system, whether as a result of a bug in the voting system, a mistake in setting up the election, or a deliberate attack on the election.

On a BMD, the marking interface and review screen are typically optimized to catch and correct voter errors. The verification of the printed ballot is primarily intended to detect malicious changes, though it also serves as a last chance for the voter to catch their errors.

# Cybersecurity concerns and voter-verified paper ballots

Cybersecurity is one of the foremost concerns in elections administration[1]. The general focus on elections security has led to wide-spread interest in audits, known as post-election audits and most recently risk limiting audits (RLA), with a growing list of states conducting pilots, and allowing or requiring them (NSCL, 2019). RLAs are statistical audits designed to limit the risk that the elections results are correct and ballots have been tabulated accurately (Morrell, 2019).

Because RLA rely on paper ballots that have been verified by voters and which can be compared to the reported results either individually or through a comparison of batches, the usability and accessibility of the verification process is critical. There are, broadly, two styles of paper ballots that voters can verify: target-area ballots and summary ballots. We compare their usage below. [2]

## Comparison of ballot types

| | **Target-area style ballots** | **Summary style ballots** |
|---|---|---|
| How they are produced | Hand marking or printing by a BMD | Printing by a BMD only |
| How selections are represented | Marking targets (in the shape of ovals, squares or arrows) are filled in on a ballot card. The candidate or voting choice they represent is indicated by proximity. | A printed list of contest titles and the voter's selection or that they made no selection.<br>Summary ballots may also contain a coded representation (typically a bar code or QR code)[3] |

[1] Note that issues such as protecting voter registration databases and network-connected systems like electronic pollbooks from attacks are outside of the scope of this report.

[2] A related topic is the question of how the paper ballots are counted, but this is not discussed here.

[3] There is a long-running debate about the use of barcodes and QR codes on ballots that is not addressed in this report.

| | Target-area style ballots | Summary style ballots |
|---|---|---|
| How they are counted | An optical or digital scanner which reads selection marks and tallies the election by comparing the position of each mark to a definition of its meaning in the election management system. | Some systems count summary ballots using optical character recognition (OCR).<br>Others read the coded representation and match the internal coding to its meaning in the election management system. |
| How they are audits | Auditors compare the location of the marks to the candidate name to determine voter intent. | Auditors read the human-readable list of contest selections to determine voter intent. |
| How they are verified | Voters compare the location of the marks to the candidate name or voting choice. | Voters read the list of contests and their choices. |
| Can they be verified accessibly | Some voting systems can read the ballot and display its selections for review.<br>There is no personal assistive technology that reads bubble-style ballots. | Some voting systems can read the ballot and display its selections for review.<br>Personal assistive technology can read the list of selections, though success depends on the visual layout.<br>If a codebook is made available, an independent application can read the coded representation and report the selections. |

# Mental models and the voter journey

To start the review of the literature, we looked at research on voting and elections topics for discussions of the mental models of voting or other factors that might affect verification. We included related research on privacy and trust, because both of those issues are closely related to the goals of verification. They share common challenges including gaps between what people say and how they actually behave and a complex interaction between awareness or understanding of the problem, clear information about how to address the problem, and interactions that are either too inefficient or too difficult to complete effectively with the resources voters have available.

Other influences on voters' mental models of voting and how well-prepared they are to successfully verify their ballot include previous voting experiences, strength of political opinions, personality traits, reading ability, memory, knowledge of how elections work, and general attitudes about elections and government – all of which can affect both expectations and how voters react to the process and any problems they encounter.

## The paradox of attitudes versus behavior

One of the most interesting analogues for the challenge of understanding why, when, and how voters verify their ballot came from a case study of Facebook to understand why individuals who claim to be concerned about privacy on social media behave in ways that expose their information seemingly contrary to their belief. The authors (Garg, Benton & Camp, 2004) examined theories about this privacy paradox: that people do not care about privacy, that they do not know or understand how their actions affect their privacy, and that the controls are not usable. Through survey responses, they explored a large set of hypotheses including risk aversion, satisficing behaviors, attitudes, and demographic characteristics.

They found that the strongest factor in understanding use of privacy controls is that people don't know how to protect their privacy. The second factor was demographic influences of gender and education. The third factor was usability, or how easy it was for participants to use the privacy controls. Whether their survey respondents cared about privacy had only negligible impact.

Other privacy research from the Carnegie Mellon University Usable Privacy and Security Laboratory (Balebako et al., 2015) adds an element of timing, reporting that the timing of privacy notices and other information makes a difference, specifically that when they are show in context, users had better recall of the notice and its contents.

We can apply this model to verification: First, voters must know how and why they need to verify their ballots. Second, voter education needs to consider reading literacy, which affects the difficulty of marking, reviewing, and verifying the ballot. It can also affect general civic literacy, civic trust and engagement, and understanding of how elections are conducted and ballots counted. Finally, both the ballot to be verified and the required actions must be usable enough to make the verification practically possible.

## Personality traits and propensity for verification

There is little research on what personal characteristics might play a part in how carefully a voter checks for errors or verifies their ballot accurately. Everett's 2007 doctoral dissertation (Everett, 2007), examined data from studies of review and verification to see if there was any correlation demographics or personality traits.

Of all the traits considered, only one personality trait significantly predicted whether participants would notice a change on their ballot: openness to experience. People with high scores for this trait tend to be "intellectually curious, untraditional, creative, and imaginative." In other words, just the sort of voters who would examine their ballot closely and could imagine possible problems. The other predictor of careful verification was whether people reported that they followed news about election security and related issues.

## Perceptions of usability

Even beyond the question of the overall usability of a voting system, how the design manages the trade-offs between the critical qualities of security, accessibility, and usability will affect perceptions of trust in the voting system and the election, as will the details of the context and interactions (Evans & Paul, 2004, Little et al., 2008).

Almost everything about the voting experience can affect how successful voters are at casting a ballot. Even the first perception of a polling place makes a difference. In a study in which people were shown photographs of a polling place, the way the voting stations were arranged and elements such as space between the systems or the use of privacy screens changed the subjective ratings of the quality of the experience (Acemyan & Kortum, 2015).

General perceptions of what kinds of voting systems are the most usable and trustworthy have changed over time. The MIT Election Lab (Stewart, 2018) reports that between 2012 and 2018, respondents in their surveys have shifted from preferring Direct

Recording Electronic (DREs) to paper ballots and optical scanners to the opposite, although voters see both types of systems as safe from tampering.

This perception contrasts with the election security view that elections need to be designed around systems that produce evidence that the ballots cast reflect voter intent, as discussed in Stark and Wagner (2012). However, others point out that paper, too, has weaknesses including known methods of tampering with an election (Willemson, 2018) and flawed designs that can lead voters to make errors (Norden et al., 2008, 2012).

Early electronic voting systems (the DREs) added a final verification by printing a voter verified paper audit trail (VVPAT). These records were typically printed on a small roll of paper, shown to the voter behind glass, and then stored on a spool for use in an audit or recount. Although there are still a small number of current voting systems that use this method of creating a verification record, it has fallen out of favor because of the challenges of using the spooled paper in an election audit and the difficulty of reading and verifying the VVPAT through glass (Appel, 2018) as well as its inaccessibility to some voters with disabilities.

## Designing systems for trust

The act of participating in an election is, in part, an act of faith that the election will be run fairly and that all votes will be counted accurately. There are many documented examples of design and procedural flaws affecting elections from voters being given the wrong ballot (Tepfer, 2018) to reports of "vote flipping" caused by either poor calibration (Gilbert et al., 2012) or software bugs (Wallach, 2008; Garber, 2008). These are just a few examples of real issues of how the quality of the voting system and election administration can affect overall trust in elections.

A study of new e-government systems showed that trust cannot be built through technology alone. Any government technology interaction is colored by our perception of the service. An important factor is an individual's general perception of their ability to influence government and how the government interacts with them. The researchers concluded that "distrustful, low self-efficacy individuals will not increase their trust, irrespective of the medium of interaction. The quality of interaction, while important, is nonetheless secondary" (Parent et al., 2005).

Trust is also important in considering verification—including detecting errors and malicious changes to a ballot—because of the complex issues about whether a voter can or should trust their own actions or the actions of a computer. Research with focus groups that included discussions of futuristic scenarios of e-voting on a smartphone

showed how questions of trust and privacy are multi-dimensional and change with context (Little et al., 2008).

We wanted to know more about how trust is created and what aspects of a design or transaction can make a difference in how voters trust a voting system.

Themes included the value of civic participation in promoting trust, concepts for building trust in transactions, external factors that impact trust, and how design can influence trust.

We dipped into a body of political science literature on trust and civic participation, starting with the classic *Bowling Alone,* (Putnam, 2005), which connected an increase in civic engagement with an increased sense of trust. Other articles looked the relationship between a decline in trust and voting for more extreme candidates (Hooghe et al., 2011), resentment of immigrants (Udani & Kimball, 2017), and racial issues (Howell & Fagan, 1988).

More direct influences on trust in elections in general range from social media, candidate appearance and personality, the location of the polling place, the weather, a voter's emotional state, whether or not the voter is a parent or mother, political scandals, political campaigns, and even how well the economy is performing. (Sherman et al., 2012; Jarrett 2015).

Looking more closely at elections, two articles examined how well security measures such as end-to-end cryptography promoted trust in elections. Evans and Paul (2004) outlined different approaches to security from voter verified paper audit trails to cryptography and suggested that elements that make an election *seem* trustworthy can be important, even if they don't actually impact the security of the election. Similarly, Everett (2007) reported that 85% of participants in her research said that having an opportunity to review their selections on-screen made them feel confident that their vote would be counted accurately on a DRE voting system.

Another study looked at the relationship between trust and usability and concluded that as voting systems become more usable, trust in them increases, especially when voters have no choice. Acemyan and Kortum (2002) studied both trust in consumer products and voting systems. They found that more usable systems appeared more trustworthy to users, because the experience caused less need to think about what was going on inside the system. In voting systems, where there is no choice of which technological system to use, they found the connection between usability and trust is very strong.

When they have options, the same authors found that design factors like branding have more of an impact. (Acemyan & Kortum, 2016). This finding agrees with broader studies

of usability factors in trust in ecommerce transactions and the impact of navigational structure (Egger, 2003; Dit-Cosaque, et al., 2014). Other studies emphasized the need to accurately understand readers of online text-based communications (Feng, Lazar, & Preece, 2004).

Finally, we looked for ideas about how to measure trust. Most of the articles warned that this is a difficult concept to pin down into quantitative measures. One suggests that it is important to both observe the behavior and to ask follow-up questions, because attitudes and behavior do not always align (Glaeser et al., 2000).

Other articles suggested that looking for factors in making a trust decision might be a useful approach. Beyond looking at the design of the system, it is important to consider the personal and contextual properties that enable users to make decisions about whether to trust the system or not, and to distinguish between subconscious mental decisions (cognitive trust) and reactions to an external influence (affective trust) in understanding how trust is formed (McCarthy, Sasse & Riegelsberger, 2003; Reigeleberger, Sasse & McCarthy, 2005)

## Changing mental models

Usability testing innovative concepts for elections demonstrate how hard it can be to change mental models for the actions of voting. This seems to be especially true when a basic element of voting changes or when a new design alters long-established features. This is shown clearly in the research on systems based on end-to-end cryptography, as discussed in a later section in this report. A voter's existing mental model can affect usability of a process or design change both positively and negatively.

A study that tested an alternative contest navigation approach where voters started at the review screen, navigating from there to each contest, found that participants unintentionally undervoted and chose to vote in fewer contests. The authors observed that their work showed that seemingly small changes can have large effects on the effectiveness and general usability of a voting system (Greene et al., 2013). As they worked on early concepts for their integrated ballot box, the Los Angeles VSAP team found that the new method of casting a vote caused problems for 61% of the participants versus 11% who used a more conventional central ballot box, even though twice as many participants found it easier to use. IDEO, 2015b)

On the other hand, the concepts for a "pre-marking system" worked well in both a lab study (Lola et al., 2013) and in the development testing for the VSAP voting system. This

feature may have worked better because it builds on existing models of bringing a slate sheet or list of ballot selections to the polling place and using it as an aid to voting.

Certainly, new ways of voting can be established and mental models can change. However, voters' ideas about how elections work can be strongly ingrained, so new designs must be tested thoroughly and introduced carefully and cannot rely on voters' intuition.

# Studies of usable and accessible ballot marking, review and verification

One of the challenges of looking across the relevant research of the last fifteen years is that not only have voting systems changed, but so have the concerns about how those systems work to create a usable and accessible voting experience. At the same time, ideas about secure and trustworthy ballots have also changed, influencing the research being done.

This section looks at studies that focus on four different aspects of a voting system.

- **Ballot marking interface design** studies look at the voter interface for marking a ballot. We have focused on how well those ballots help voters understand the voting rules and accurately make their selections.
- **Review and verification effectiveness** studies include both electronic review screens with their support for detecting errors and undervotes and verification of the paper ballot.
- **End-to-end cryptographic design** studies are included because they show some of the challenges of introducing people to new ways of voting. They illustrate the difficulty of conveying the mechanics of voting methods and the importance for voters to understand the reasons for each action.
- **Designing a ballot for verification** looks at the challenges voters face in reading different styles of paper ballots, with a particular focus on accessibility.

## Ballot marking interface design

Although there has not been much new research on ballot marking interfaces, newer voting systems make use of general best practices in ballot design and have improved many aspects of voting system interface design. The EAC's Effective Designs for the Administration of Federal Elections (EAC, 2007) were a basis for several projects in their Accessible Voting Technology Initiative, including the "Anywhere Ballot" reference design (Summers et al., 2014).

Commercial voting systems have also been improved through requirements for usability and accessibility in the EAC Voluntary Voting System Guidelines 1.0, 1.1, and 2.0 (EAC, 2020). A review of the voting systems certified in Pennsylvania (PA DOS, 2019) shows that new systems have more consistent interfaces that can also improve voters' ability to use a voting system to accurately mark a ballot according to their intent, despite the problems in both usability and accessibility listed in the certification reports.

The question in this research is whether a well-designed voting experience can increase the accuracy of the process so that we can have greater confidence that ballots are cast as voters intend.

There is a rich body of information on accuracy of ballot marking, both in papers that studied this directly as well as data from studies with a broader focus. One finding across several projects is that the research shows that participants in a test environment mark their ballots with similar error rates using electronic interfaces, hand marked paper, and even punch cards or lever systems (de Jong et.al, 2007; Everett, 2007; Herrnson, 2012). For example:

- A comparison of three selection styles (bubble, arrow, open response) on hand-marked ballots found that all three performed similarly but also that over 11% of ballots contained at least one error (Everett et al., 2006)
- There is evidence that systems that use a full-face ballot design (all contests and candidates shown in a grid on one large screen) have higher residual vote rates than a scrolling interface, especially with lower income voters (Norden, et al., 2006).
- Where used, straight-party voting also provides some challenges. A study comparing paper ballots with straight-party to ballots without this found that the participants who used the straight-party option had fewer unintentional undervotes for partisan offices but many more errors overall (Herrnson et al., 2012).

This suggests that there is always some rate of residual votes in any election and that the normal human variation in personal habits, such as attention to detail or levels of interest, brought to voting mean that there will always be some error. However, we can hope to do better than the overvote rates reported in the California 2016 Primary, which reached 4.7% when the Senate race had an unexpectedly large number of candidates (Kimball & Kropf, 2017). Extensive reporting by the Brennan Center in 2008 and 2012 also relied on their analysis to identify elections with high rates of residual votes that could be explained by the ballot design (Norden et al., 2018, 2012).

There is some established good practice for effective design to support voters in the marking phase of voting, including techniques for confirming a selection in an electronic interface by changing the display of the selected candidate. For example, ballots with a clearly visible, large marking target are more effective in helping voters confirm their selection.

In a study where participants were presented with candidate names either with a visual checkbox included in the touch area or without, more voters touched the checkbox when it was present (Gilbert, et al., 2012). And in a small sample of older adults in a residence facility, many experienced problems trying to use their fingers to select candidates with a

small click target on an iPad (McGrew, 2013). Looking at usability trade-offs between competing design principles, a comparison of ballots with the marking target to the left or right of the candidate names found that proximity to the text was more important than alignment to a grid (Swierenga & Soucy, 2008).

A project to study of ranked choice voting ballot designs found that using a best practice design reduced the effect of different layouts for this novel voting method. A final test with 112 voters comparing 3 ballots was completed with no marking errors at all. (Center for Civic Design, 2017; Ramchandrani & MacArthur, 2017).

Low literacy is more widespread than generally understood, with 43% of literate adults reading at "basic" or "below basic" levels. (NCES, 2003). Information and instructions on ballots are often written in ways that are difficult or misleading for voters who do not read well, making it hard for them to understand the implications of their actions on how their ballot is cast (Summers et al., 2017; Redish et al., 2008).

Two more important areas of research into voting system usability and accessibility are studies of how well review and verification features work and the challenges of introducing new concepts as seen in studies on end-to-end systems.

## Review and verification effectiveness

A primary research concern is the effectiveness of review screens is accuracy of error detection and how well systems make both accidental and malicious errors salient to voters. There are some research studies that focused specifically on the accuracy of the review process as well as mentions issues on the review section of the voting experience as part of a larger study.

Research on the effectiveness of review screens and associated messages found specific usability issues:

- One study tested an alternative approach of in-line review where voters are prompted during the marking phase to confirm their selections following each contest. The study found that end-of-ballot review was more efficient and higher in voter satisfaction than in-line and that both had comparable efficacy (Holmes, 2015).
- A study of different types of messages aimed at motivating voters to verify in the context of Internet voting found that all significantly influenced voters' intentions, whether they acted on that intention or not (Olembo, 2014).
- Norden (2006) reported that the review screen is more successful when it looks different from contests and when it makes you see all selections.

- IDEO, working on the Los Angeles County Voting System Assessment Project (VSAP) discovered that participants needed explicit instructions or interactive controls to know to scroll through the long review screen. Some participants didn't realize there were more "below the fold" and didn't know how to use gestures to scroll (IDEO, 2015).
- Two different studies discovered that navigation from the review screen to a contest to make a change and back affected voters' willingness to correct a ballot "After altering or adding a vote for a single contest, however, most were frustrated by having to advance through the entire list of contests before arriving, again, at the review screen" (IDEO 2015; Summers et al. 2014).

The most recent study, conducted at the University of Michigan (Bernhard et al., 2020) studied whether voters can detect malicious changes in the machine-printed ballots produced by ballot marking devices. The researchers set up a realistic mock election polling place, used electronic marking interfaces, ballots based on the voting system in the local area, and a real scanner. The ballot marking device was programmed to change one selection on each ballot. In their first tests, less than half of the 241 participants verified their ballots at all, and only 6.6% found and mentioned the error to poll workers. Then, they tested 6 different interventions to promote verification, including signage, verbal instructions, and verbal instructions with a slate to simulate voter intent. The best performance came with a slate with one of the instruction scripts. In this experiment, 95% of the 21 participants were observed examining the ballot, and 86% reported the error to a poll worker.

In September 2019, researchers from Rice University (Byrne, 2020) observed a mock election of the Los Angeles County VSAP (Voting Solutions for all People) voting system before its first use in a live election. At 6 polling sites, they observed and timed many elements of the voting experience, including the time to vote and whether the voter showed evidence of verifying their printed ballot. They found that over half did take time to check their ballot, taking (on average) two minutes longer to vote than those who did not, suggesting a careful review.

One group of studies is of special importance for this review because of the consistency in the methodology. Four studies between 2007 and 2018 manipulated the results of a ballot to see if voters would find the anomalies during verification of the review screen or printed ballot.

Three studies from the same research group at Rice University (Everett, 2007, Campbell & Byrne, 2009; Acemyan et al., 2013) and a fourth by Gilbert et al. (2013) used the same

basic approach to study the effectiveness of review screens, varying details of the review process and design as they replicated the tests. In the basic session plan:

- Participants voted on an electronic interface using a prototype voting system
- The system manipulated the display of the review screen, introducing either no changes (for a control group), and 1, 2 or 8 anomalies mimicking a malicious hack.
- The changes either remove votes from a contest, creating an undervote or changed the votes, swapping candidate names.
- Voters were either directed in how to vote, voted in a self-directed way, or were given minimal directions.
- The ballot included 20-23 contests and ballot questions

The results of these studies show that a combination of the design and the degree to which participants have a strong intent does make a difference in how well voters are able to review accurately.

- Placing no emphasis on the review process produced the lowest detection of errors: 25% and 32%
- Adding instructions to review carefully and highlighting undervotes increased detection to 50%.
- Giving participants instructions for how to vote raised detection to 52% and 64%.
- Forcing slower navigation of the review screen raised the rate of detection to 74%
- Adding a second change at verification on a paper ballot produced the highest overall success at 90%.

The type of anomaly mattered. Campbell and Byrne (2009) found that the type of error made a difference; 69% of participants who had votes changed to undervotes noticed the change in the enhanced review screen, but only 39% noticed when one candidate name was swapped for different name. The contents of the ballot seemed to make little difference.

The lowest rate of detection was in a study with students being given credit for their participation. Acemyan and colleagues (2013) found that fictional contests based on real people and events on a university campus did not increase detection, suggesting that the type of interest participants have in a study is an important methodological question.

Giving participants directions about who to vote for in a list similar to someone bringing a voter guide or sample ballot with their voting plan to the polls affected detection. In Everett's original research in 2007, 52% of those who had a plan for voting noticed the errors compared to 12% of those who were not instructed.

The study in this group with the highest rate of accurate detection of errors was also the one that placed the most demands on the participants with a slower on-screen review process and adding the step of printing a paper ballot. This confirms a comment in earlier research on DRE systems that "...design elements that decrease efficiency or voter confidence may actually increase the accuracy of voters' selections." (Norden et al., 2006)

Another study of hand-marked paper ballots (DeMillo & Marks, 2018) took a different approach, manipulating the contests on a blank ballot showed to participants, not vote selections. This study showed that voters had difficulty identifying whether a ballot was the same as the one they just voted on, relying only on their memory. A companion study observed voters in a primary election to see how long they spent verifying a bubble-style ballot if at all.

The recommendations from Bernhard and colleagues (Bernard et al., 2020) show that the entire voting experience must be designed to encourage effective verification.

- Design polling places for verification, including a place to privately verify the ballot after it is printed by the ballot marking device.
- Provide poll workers with a script to give voters verbal instructions to verify. The script in the most successful experiment was simple: "Have you carefully reviewed each selection on your printed ballot?"
- Encourage voters to plan their voting in advance and bring a personalized list of their intended selections.
- Educate voters that the paper ballot is the official ballot that will be counted.
- Design printed ballots for legibility for voters and to make it possible for voters with impaired vision to use assistive technology to verify their ballots.
- Have procedures to help voters correct errors and carefully track problems. Take problems seriously, even if voters believe themselves to be at fault. Be sure poll workers have a way to report problems to a central office so that issues that affect multiple systems or locations can be identified quickly.

### Summary of the error-detection studies and results

| **Study**<br>Author and year | **Methodology**<br>Variations in the basic plan | **Participants**<br># and type shown errors | **Results**<br>% detecting changes |
|---|---|---|---|
| Everett 2007 | • System: VoteBox<br>• Ballot: Fictitious names | 53 general population from Houston | **32%** noticed the manipulation of the review screen:<br>• **52%** of those directed |

| **Study**<br>Author and year | **Methodology**<br>Variations in the basic plan | **Participants**<br># and type shown errors | **Results**<br>% detecting changes |
|---|---|---|---|
| | • Voting instructions: half directed, half self-directed | | • 12% of those not directed |
| Campbell & Byrne 2009 | • System: VoteBox<br>• Ballot: Fictitious names<br>• Voting instructions: Mix of directed, self-directed, and minimally directed using party<br>• Interface updates: Stronger instructions to review, highlighted undervote notifications, party on the review screen | 108 general population from Houston | **50%** noticed the manipulation of the review screen.<br>• **64%** of directed<br>• 44% of not directed<br>• 42% of partially directed |
| Acemyan et al. 2013 | • System: VoteBox<br>• Ballot: Custom ballot aimed at making the contests more salient | 125 Rice undergrads | **25%** of the participants noticed changes. |
| Gilbert et al. 2013 | • System: Prime III<br>• Ballot: Cartoon character names<br>• Interface differences: Review screen shows 4 contests at a time<br>• Added paper ballots for an additional verification opportunity. | 110 students and staff at a university | **90%** detected changes<br>• 74% detected changes on screen<br>• 16% detected changes on paper |
| DeMillo & Marks 2018 | • System: Printed paper ballot<br>• Ballot: Facsimile of blank ballot used in real election<br>• Change to ballot contents:<br>- Changed an entire contest to a different district<br>- Changed the candidates listed to those from another district<br>• 60% of the ballots were correct, 40% altered | 103 voters at a primary election in Tennessee | **25%** said the ballot was incorrect, claiming to detect changes. However:<br>• 13% (8 of 59) misidentified a correct ballot as incorrect (false negative)<br>• 56% (22 of 39) misidentified an incorrect ballot as correct (false positive) |

| **Study**<br>Author and year | **Methodology**<br>Variations in the basic plan | **Participants**<br># and type shown errors | **Results**<br>% detecting changes |
|---|---|---|---|
| DeMillo & Marks 2018 | • System: Printed paper ballot<br>• Observation of voters at a primary election with 18 contests | 87 voters | **52%** were observed reviewing their ballot selection cards for an average of 4 seconds |
| Byrne 2019 | • System: VSAP<br>• Ballot: mock election ballot<br>• Observations in 6 polling sites<br>• No interventions | 87 voters | **51%** (41 of 81) voters were observed verifying their ballot before casting.<br>**2:10** minutes average verification time |
| Bernhard et al. 2020 | • System: Modified electronic BMD interface<br>• Ballot: 13 contests based on the real 2018 ballot.<br>• Ballot style: bubble and summary, mimicking Hart InterCivic<br>• Changed a single contest on each ballot, switching, adding or removing the selection<br>• Interventions included signage, verbal scripts, and directed voting with a slate of candidates | 241 voters, recruited from the community | **Without intervention**<br>• **40%** were observed examining ballot<br>• **7%** reported an error<br>**With 6 interventions**<br>• Overall: **64%** examined ballot; **28%** reported the error<br>• Final variant (script + slate): **95%** examined ballot; **86%** reported the error |

## End-to-end cryptographic design

One of the promising innovations for election security is the concept of using cryptography to protect a ballot from end-to-end during the voting process with the ability to prove that the ballot was counted as cast by the voter.

Although most of the papers cover the technical cryptographic methods, there is a body of work that highlights the challenges of making a system both secure and usable. We focused on the work of two research teams because they each conducted several studies on related systems and topics.

One paper reported on usability studies of three different methods of performing the implementation in the Benaloh Challenge which allows a voter to either challenge or cast a ballot. They found that successful verification varied by method, from 61% to 81% with the rest mistakenly thinking that they had performed the verification correctly. They also found that voters misunderstood the nature of the verification, even when this was the focus of the test and participants were given explicit instructions to carry out the verification procedure (Marky et al., 2018).

The most visible end-to-end project in U.S. Elections was STAR-Vote, a project to design a "secure, transparent, auditable, and reliable" voting system for Travis County Texas (Bell et al., 2013). A team of researchers at Rice University conducted several studies to explore different aspects of end-to-end systems including formative design testing of elements of the STAR-Vote concepts (Acemyan & Kortum, 2014) and gathering baseline data from other systems to inform their work

Their studies comparing three different end-to-end systems reported that participants could not cast a ballot 42% of the time and could not verify their ballots 53% of the time. They found that a small number of design problems contributed to most of the failures, suggesting that improving the usability of the system is an important first step in making the end-to-end system suitable for use. They also pointed out that two of the systems would likely pose accessibility problems that would be difficult to overcome (Acemyan et al., 2014, 2015a).

These two studies, and others, show the difficulty of designing a system that changes the basic mental model for how a paper ballot is cast and call for better usability evaluations as the system is designed. This theme is echoed in Kulyk & Volkhamer, 2018) where they show that poor usability can affect adoption of good security practices and stress the importance of users being both aware of the actions required as well as why they are important.

One study at Rice (Acemyan et al., 2015b) focused on gaining a deeper understanding of voter mental models and whether voters were able to accurately connect the methods used in end-to-end systems with increased accuracy, security, or transparency of the systems. The researchers asked the voter participants to draw a diagram of each of the three voting systems they used. Typically, voters simply drew what they understood to be the steps in the process with some explanations of the goal of each step.

- They found that the mental models across systems were based on the specific steps and equipment, rather than any robust (or even accurate) understanding of the cryptographic models. For example, voters might know that "encryption" is a security procedure, but not how it helps secure ballots. Despite being able to describe the systems, participants did not think that they were more secure or accurate than a conventional paper-based voting method. On the other hand, a follow-up study (Acemyan & Kortum, 2016) examined perceptions of security and found that participants could tell the difference between meaningless security features (which they called "security theatre") and an end-to-end system using real security measures.

Even this small sample of the research on end-to-end systems shows the difficulty in introducing a new concept in voting and building mental models that help voters understand its value in order to be able to use the new features accurately.

This suggests that one of the keys to successful verification requires overcoming several hurdles. Voters must connect the act of verification with greater overall election integrity, even if they believe that a system error or malicious change is unlikely. They must know how to verify so they can perform the verification correctly. Finally, the ballot (and any associated verification tools) must be well-designed to make an accurate verification easy.

## Designing a ballot for verification

In a 2019 white paper, the OSET Institute (OSET, 2019) reviewed the printed ballots produced by current BMDs comparing the ballot layout along with other features of the systems. Using the OSET review along with information from the Pennsylvania Department of State certification information (PA DOS 2019), we summarize current ballot layout features below.

**Summary of ballot layout features on current voting systems (Sources: OSET, 2019 and PA DOS, 2019)**

| Voting system | Layout notes | Ballot paper | Encoding |
|---|---|---|---|
| Dominion ICX | Single-space lines<br>Mixed case text<br>Contest titles above indented selections<br>Up to 3 columns | Letter-sized paper | QR code at top of ballot |
| ES&S ExpressVote | Single-spaced lines<br>All capital letters<br>Contest titles to the left; selections to the right<br>Single column | 4.25" wide cardstock | Barcodes for each contest at top of ballot |
| Hart InterCivic Touch Duo | Single-spaced lines<br>Selections all capital letters<br>Ballot information in tabular columns for<br>Contest title<br>Selection<br>Order number<br>Party | Letter-sized paper | QR code at top of ballot for ballot type; selections read by OCR |
| Unisyn FVT | Single-spaced lines<br>All capital letters<br>Contest titles above indented selections<br>Single column | 3.5" wide roll-fed paper | Hollerith grid at bottom of ballot |
| VSAP (Los Angeles County | Space and separator line between each contest<br>Mixed case text<br>Selections bold<br>Contest titles above selections; candidate ID number<br>Up to 3 columns | Multiple custom paper sizes | QR code on the left side of ballot |

| Voting system | Layout notes | Ballot paper | Encoding |
|---|---|---|---|
| Voting Works | Space and separator line between each contest<br>Mixed case text<br>Selections bold<br>Contest titles above selections<br>Up to 2 columns | Letter-sized paper | |
| Prime III<br>Open Source<br>primevotingsystem.com | 1.5 spaced lines<br>Mixed case text<br>Contest titles with arrows pointing to selections | Letter-sized paper or larger | Optional QR code encoded with selections at top |
| ElectionGuard<br>(E2E demo system) | Space and separator line between each contest<br>Mixed case text<br>Selections bold<br>Contest titles above selections<br>Up to 2 columns | Letter-sized paper | QR code at top of ballot for ballot type.<br>Selections encoded using E2E |

In addition to these systems, the Hart InterCivic TouchWriter and the Clear Ballot ClearAccess, along with Prime III, create printed bubble-style ballots that look similar to the brand's hand-marked ballots.

There are few guidelines for the visual design of paper ballots. The VVSG 1.1 and 2.0 requirements (EAC, 2020) include requirements for text size and contrast, but little else that addresses the legibility of the ballot for human reading. There are also descriptive guidelines that can be inferred from the usability testing reports on the LA County VSAP, including the need for text that is at least 10 points large, white space, and separators between the contests.

The accessibility reports from the Pennsylvania Department of State certification testing (PA DOS, 2019) include anecdotal reports of attempts to read a summary ballot using personal assistive technology text reader applications including the Kurzweil KNFB Reader and Microsoft Seeing AI. The experience of participants in using these tools varied, but suggested that simpler layouts with space or punctuation separating contests

might work well enough to be useful, especially with voter education information about how to set up the tools for optimal reading of each tool.

Summers and Langford (2015) offered principles for designing for voters with low literacy that apply to the printed ballot as well as the marking interface:

- **Make it look easy to read.** Good visual cues help readers process text more easily. Indicate the main points or sections of text, and signal transitions clearly.
- **Create a linear flow.** Don't force readers to split their attention between different parts of the page or screen. Help them focus on one thing at a time, with good navigation from one part of the task to the next.
- **Support a narrow field of view.** Make sure everything they need to know is right there in the center of the page.

The legibility of the BMD printed ballots varies widely, and often is designed to solve, in part, the technical problem of fitting enough contests on a single piece of paper or making it easy for a scanner to read, rather than making it easy for voters to read.

These reports, however, are all basic design guidelines which do not go much further than the 2007 EAC ballot design guidelines, or the requirements included in the VVSG.

More importantly, there has been little research on what makes a printed ballot verifiable—that is what properties of the printed ballot support a motivated voter to make a more accurate check that the ballot reflects their intent.

Another gap in the research is understanding the nature of the verification task. Is verification, for example a memory task, relying on voters *remembering* their selections and comparing that memory to the printed ballot? Or a recognition task, focusing on voters *recognizing* that the selections on their printed ballot reflect their intent. In their seminal paper on heuristic evaluation, Nielsen and Molich (1990) called this "recognition vs. recall" saying that relying on recognition is easier because showing users a sample helps them retrieve the information from memory.

Finally, it is not clear what motivates voters to verify their ballot and how that relates to the strength of their intent in voting and whether their mental model include the possibility of not just identifying their marking errors, but also a need for a final check against system errors and malicious changes?

# Methods for researching voting

In addition to reviewing the literature for what is known about how voters interact with voting systems, we also looked for insights about the best way to conduct research aimed at understanding voter behaviors and mental models.

Differences in methodology we noted in conducting this review are:

- There was a wide variation in how many participants there were and how they were recruited.
- The voting systems used the studies included both systems in use in real elections and prototypes created for the purpose.
- The context of the research varied in how voters were instructed to mark their ballot, and how much the test setup mimicked a polling place.
- There were observational studies, surveys, and quantitative task-based studies.

## Recruiting approaches

There were several approaches to finding a sufficiently diverse group of participants, even in studies with generally small numbers. The research includes many studies with homogeneous participants who aren't representative of the voting population.

For example, Redish et al. (2008) used only two criteria to recruit their participants: eligibility to vote and ability to speak English. They found that this gave them good diversity in demographic characteristics like age, gender, race/ethnicity, education, and voting experience. Other studies used a naturally large population, such as people who had just voted in a real election or were present at the research location for their own reasons (De Jong et al., 2007; MacArthur et al, 2018; Ramchandrani & MacArthur, 2017).

In a large-scale study that aimed for a realistic polling place, Pandolfo et al. (2006) administered a reading level test to volunteer participants to be able to analyze the effect of a reading disability on voting, but they did not include or exclude people based on the results.

Another way of reaching diverse participants was to purposefully go to locations where different types of voters might be found. McGrew (2013) tested the use of a tablet by elderly voters by going to nursing homes and working with residents there. Similarly, Center for Civic Design has made a practice of using location-based intercepts for unscheduled sessions at libraries, community centers, government offices, parks, malls, churches, and other public places (Chisnell, 2015)

Other recruiting methods included recruiting on campus, either for a small honorarium or as part of a class assignment. In general, this approach seems to have been made for convenience rather than for any methodological advantage (Acemyan & Kortum, 2006).

We summarize below:

### Approaches to recruiting participants

| Approach | Advantages | Disadvantages |
|---|---|---|
| Scheduled sessions, recruited through ads, social media, or community partners | Ability to screen for behavior, ability or demographics for diversity. | Placement of notices is a critical decision in range of participants. |
| Intercepts at a real election | All participants will be active voters. | Limits on when research can be conducted.<br>Limited to the demographic of the polling location. |
| Intercepts at constructed events | Ability to reach a large, specific audience if the location is selected well. | Takes time and resources to set up and promote the event. |
| Intercepts at public spaces | Diverse participants, who might not reply to an ad or attend an event | Location must be chosen carefully to be conducive to research session needs.<br>Practical limits on the length of session. |
| University students | Easy to recruit | Not very representative or educationally diverse. |

## Locations for the research

The questions of how to research also extend to the location of the testing. Some studies have aimed for testing near a polling place or during an election period so that participants will be more likely to be thinking about elections. Despite some concerns that the political issues in an election will overwhelm the research, these studies show that user research can effectively be conducted near polling places, even on election day. (De Jong et al., 2007; Swierenga & Soucy, 2008; DeMillo et al., 2018; MacArthur & Quesenbery, 2018).

When conducting research on an actual election day was not possible, researchers used spaces set up to look like polling places or used locations similar to polling places. A test experience that is ecologically similar to a real voting experiences is preferred as using real world conditions do influence the quality of data gathered, but does create large challenges. For example, Pandolfo and colleagues (2006) recruited real poll workers for the study, but others used research assistants played the role of election judges, rather than recruiting actual poll workers to recreate a realistic polling place setting (Herrnson et al., 2020; Ramchandrani & MacArthur, 2017).

The Voting System for All People (VSAP, nd) project set up large mock polling places in specific communities to ensure that participants in their usability testing included people with disabilities, or who spoke other languages, using community partners to promote the events.

Other studies were simply conducted in a lab setting. These were often tests of specific aspects of the voting system, more focused on the design of the system than the overall experience. These included a test of a concept for tamper-resistant ballot boxes (Kortum, Acemyan & Belton, 2015), paper-prototype testing of a tablet interface (Summers et al., 2014), a series of tests of end-to-end voting system concepts at Rice University (Acemyan & Kortum, 2015; Acemyan et al., 2015) and many small-scale agile usability tests that supported product development in projects including the Los Angeles Voting System Assessment Project (VSAP).

We compare the choices below.

### Choices of locations

| Approach | Advantages | Disadvantages |
|---|---|---|
| Lab or lab-like space | Controlled environment for conducting research. | Neutral location could affect realism of behavior. |
| :Constructed "polling place" | Location can influence realistic behavior and reactions. | Environment can be chaotic, limiting the depth of some aspects of the research. |
| Located near a real election | Participants primed for thinking about voting. | Best for research that directly relates to the real election experience. |
| "Random" public space | Wide range of options and participants | Need to match limitations of the location to research needs. |

## Approaches for running a within-subjects study

Most of the studies included only one voting activity, so all participants had the same experience. This is a type of "within-subjects" study design where the same person tests all the conditions (i.e., all the user interfaces).

We were interested in the best practices for studies in which each participant would complete two voting activities, so they can be compared.

Studies that we found included Redish et al. (2008), which was a study of plain language and looked at different styles of ballot instructions. Other studies had participants vote on an electronic ballot marking interface and on a hand-marked paper ballot. (de Jong et al, 2007; Herrnson et al, 2012). These also had slightly different ballot contents for the two activities.xxx

Most studies alternated the order of the activities, to counterbalance any order effects on the results, and had a short debrief and break between them.

Everett, Byrne and Greene's research (2006) mixed these two approaches. In addition to *ballot type* (treated as a within-subject variable), they tested *candidate type* (realistic/fictional) and *information given* (slate/guide/no guide), treating both of them as between-subjects variables.

## Real, realistic, or fake ballot contents

Another aspect of voting research to consider is what the content of the ballot will be. Many demonstration ballots use famous people, ice-cream flavors or other light-weight decisions, but there was early recognition that these do not really provide the level of realism that is needed to simulate an election.

In response to this issue, and early research project at NIST examined ballots from around the country and designed a "medium complexity" ballot populated with real office titles, realistic (but not real) candidate names, and samples of some real ballot questions. Another approach is to base the test ballots on real ballots used in a recent election with randomly generated names. Or using real names of people who could potentially run for office in the future.

Many research projects have used the "realistic-but-not-real" approach, including tests of error detection on electronic ballot marking interfaces (Everett et al., 2006). ballot instructions (Redish et al., 2008), straight party voting (Herrnson et al., 2012), and ballot designs (Summers et a.l, 2014).

Others have taken different approaches. The use of real names and campus issues in Acemyan, Kortum and Payne (2013) shows that merely using real names is not enough to create realistic conditions. However, other studies have used comic characters or other amusing topics with successful results, arguing that it allows more focus on the interface being studied Gilbert et al., 2013).

In research at polling places, studies have used the same ballot just voted on in the real election (De Jong et al., 2007; DeMillo et al., 2018) or a mix of real and realistic contests (MacArthur & Quesenbery, 2018). The rationale for this approach is that the voter-participants are already primed with the current election, so have a more strongly held intent about how they vote.

### Choices of ballot contents

| Approach | Advantages | Disadvantages |
|---|---|---|
| Real (or only slightly altered) ballot | Strong realism.<br>Good for research immediately tied to an election. | Personal politics can impinge on the research. |
| Realistic-but-not-real | Ability to decide on length and type of contests and diversity of candidate names.<br>Avoids personal politics.<br>Good for simulating elections. | Unfamiliar names and parties make it harder for participants to make realistic selections. |
| Amusing contests (famous people, flavors) | Can be more engaging.<br>Good for research that does not rely on simulating a real election. | Lowest realism, can make it harder for participants to remember their selections. |

## Voter choice or instructed voting

A final variable in the research methodologies is the instructions given to voters about how to mark their ballot. The decision made for each study was always a mix of the relationship between the study goals and practical considerations in running the study. There was a wide range of variations in both the approach and how the instructions to participants were phrased.

### Variations in how participants make voting choices

| Approach | Advantages | Disadvantages |
|---|---|---|
| Detailed instructions (written or verbal) | All participants vote the same way, so errors and other deviations are obvious in the results.<br>Adds opportunity to include changes or corrections in the protocol. | Participants' ability to follow instructions can affect the result. |
| General guidance for making selections | Simpler instructions to communicate.<br>Allows voters some independence. | Harder to determine whether any errors or variance are intentional. |
| Consultative decision in advance | Participants are allowed to make their own choices in advance, communicating or consulting with the moderator | Adds time to the session.<br>Process might difficult for some participants. |
| Independent use of a voter guide | Participants make their own choices, marking a voter guide before voting. | Adds to complexity of materials handled while marking a ballot. |
| No instruction | Least restrictive | Least precise data. |

There were also two variations on how participants were given detailed instructions. Some studies suggested provided written instructions, in effect giving participants a voter guide and permission to use it in a naturalistic way.

Others had the moderator read the instructions aloud for each contest as the participant reached it. This approach is most helpful for participants with visual or cognitive disabilities or low literacy. It also ensures that the instructions are given just before they are used, so the marking task is focused on recognition of the desired candidate rather than recall of instructions.

As far as we can tell, studies used one approach or another, rather than using written instructions unless the voter could not read them effectively.

## Studies for the Los Angeles County voting system project

We cannot leave the topic of research methodologies without mentioning the work in Los Angeles County as they developed their new voting system, the Voting System for All People (VSAP, nd). It is an excellent best-practice example of a robust user-centered design process that produced strong evidence for their overall direction and final design decisions.

The testing was conducted by IDEO in collaboration with the Los Angeles County elections team from 2010 to 2016. Their studies include both small (5 participants) and large (150+ participants) studies, focus groups, surveys, and studies in mock polling places as well as lab studies. The participants were diverse in terms of race/ethnicity, gender, age, educational attainment, ability, voting experience, technology experience, and financial status, with awareness of and attention to including minority groups.

Reviewing the Phase 3 design research from 2015 - 2018, there are many insights and innovative approaches to testing voting systems.

- Studies used purposeful sampling, specifically targeting participants who represented groups of interest including 2nd language speakers, the elderly, voters with physical/learning disabilities.
- One study tested a "prototype" of the audio format by using a bilingual voice-over artist to act as the system audio, reading out the text for each screen in response to participant actions.
- They engaged community organizations to help recruit participants for studies at a mock election.

# Bibliography


Acemyan, C. Z., & Kortum, P. (2012). The relationship between trust and usability in systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2012, 56*(1), 1842-1846. https://doi.org/10.1177/1071181312561371

Acemyan, C. Z., Kortum, P., & Payne, D. (2013). Do voters really fail to detect changes to their ballots? An investigation of ballot type on voter error detection. *Proceedings of the Human Factors and Ergonomic Society Annual Meeting, 2013, 57*(1), 1405–1409. https://doi.org/10.1177/1541931213571314

Acemyan, C. Z., Kortum, P., Byrne, M., & Wallach, D. (2014). Usability of voter verifiable, end-to-end voting Systems: Baseline data for Helios, Prêt à Voter, and Scantegrity II. *Proceedings of EVT/WOTE14*, San Diego, California *and USENIX Journal of Election Technology and Systems, 2*(3). https://www.usenix.org/system/files/conference/evtwote14/jets_0203-acemyan.pdf

Acemyan, C. Z., & Kortum, P. (2014). To verify or not to verify, that is the question: Voters' use and expectations of voting verification methods. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 1295–1299. https://doi.org/10.1177/1541931214581270

Acemyan, C. Z., & Kortum, P. (2015). Does the polling station environment matter? The relation between voting machine layouts within polling stations and anticipated system usability. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 1066–1070. https://doi.org/10.1177/1541931215591299

Acemyan, C. Z., Kortum, P., Byrne, M. D., & Wallach, D. S. (2015a). From error to error: why voters could not cast a ballot and verify their vote With Helios, Prêt à Voter, and Scantegrity. *USENIX Journal of Election Technology and Systems (JETS), 3*(2), 1–25. https://www.usenix.org/system/files/conference/jets15/jets_0302-acemyan.pdf

Acemyan C. Z., Kortum P., Byrne M.D., & Wallach D.S. (2015b). Users' mental models for three end-to-end voting systems: Helios, Prêt à Voter, and Scantegrity II. *Proceedings of the 3rd International Conference on Human Aspects of Information Security, Privacy, and Trust, HAS 2015, Lecture Notes in Computer Science 9190*, 463-474. https://doi.org/10.1007/978-3-319-20376-8_41

Acemyan, C. Z., & Kortum, P. (2016). Can voters tell when their voting method Is secure? Effects of end-to-end security and security theater on perceptions of voting systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60*(1), 1220–1224. https://doi.org/10.1177/1541931213601285

Appel, A. (2018). Continuous-roll VVPAT under glass: an idea whose time has passed. October 19, 2018. https://freedom-to-tinker.com/2018/10/19/continuous-roll-vvpat-under-glass-an-idea-whose-time-has-passed/

Balebako, R., Schaub, F., Adjerid, I., Acquisti, A., & Cranor, L. (2015). The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices - SPSM '15*. https://doi.org/10.1145/2808117.2808119

Bell, S., Benaloh, J., Byrne, M., DeBeauvoir, D., Eakin, B., Fisher, G., Kortum, P., McBurnett, N. et al. (2013). STAR-Vote: A secure, transparent, auditable, and reliable voting system. *Proceedings of EVT/WOTE13*, Washington, DC and *USENIX Journal of Election Technology and Systems, 1*(1). https://www.usenix.org/system/files/conference/evtwote13/jets-0101-bell.pdf

Bernhard, M., McDonald, A., Meng, H., Hwa, J., Bajaj, N., Chang, K., & Halderman, J. A. (2020). Can voters detect malicious manipulation of ballot marking devices? *Proceedings of the 41st IEEE Symposium on Security and Privacy,* May 2020. (Preprint) https://mbernhard.com/papers/bmd-verifiability-sp20.pdf

Byrne, M. D. (2020). VSAP Mock Election Observation Report to the Los Angeles County Registrar-Recorder/County Clerk, January 7, 2020. http://chil.rice.edu/research/pdf/VSAP_MockElectionReport.pdf

Caltech/MIT Voting Technology Project. (2001). Voting: What Is, What Could Be. Report 1. https://vote.caltech.edu/reports/1

Campbell, B. A, & Byrne, M. D. (2009). Now do voters notice review screen anomalies? A look at voting system usability. *Proceedings of EVT/WOTE09*, Montreal, Canada. https://www.usenix.org/legacy/event/evtwote09/tech/full_papers/campbell.pdf

Center for Civic Design (2017). Best practices for ranked choice voting ballots and other materials. Ranked Choice Voting Resource Center, February 28, 2017 https://drive.google.com/file/d/0B3K2g6lIQMWsYkEzSkgtWHBQVDg/view

Chisnell, D. (2015). Talking to strangers on the street: Recruiting through intercepting people. *User Experience Magazine,15*(5). http://uxpamagazine.org/talking-to-strangers-on-the-street/

Chisnell, D. (2017). The epic journey of American voters. Civic Designing – Medium, March 22, 2017. https://medium.com/civic-designing/the-epic-journey-of-american-voters-ed07bd0e6c57

de Jong, M., van Hoof, J., & Gosselt, J. (2007). User research of a voting machine: preliminary findings and experiences. *Journal of Usability Studies, 2*(4), 180–189. http://uxpajournal.org/user-research-of-a-voting-machine-preliminary-findings-and-experiences/

DeMillo, R., Kadel, R., & Marks, M. (2018). What voters are asked to verify affects ballot verification: A quantitative analysis of voters' memories of their ballots. https://doi.org/10.2139/ssrn.3292208

Dit-Cosaque, E., Straub, C., Detmar W., & Vance, A. (2014). Examining trust in information technology artifacts: the effects of system Quality and Culture. *Journal of Management Information Systems, 24*(4), 73-100. https://doi.org/10.2753/MIS0742-1222240403

Egger, F.N. (2003). From interactions to transactions: designing the trust experience for business-to-consumer electronic commerce. 2003 Doctoral Dissertation, Technische Universiteit Eindhoven. https://doi.org/10.6100/IR572640

Election Assistance Commission (2020). Voluntary Voting System Guidelines. https://www.eac.gov/voting-equipment/voluntary-voting-system-guidelines/

Election Assistance Commission (2007). Effective Designs for the Administration of Federal Elections. https://www.eac.gov/election-officials/designing-polling-place-materials/

Evans, D., & Paul, N. (2004). Election security: perception and reality. *IEEE Security & Privacy, 2*(1), 24-31. https://doi.org/10.1109/MSECP.2004.1264850

Everett, S. P., Byrne, M. D., & Greene, K. K. (2006). Measuring the usability of paper ballots: efficiency, effectiveness, and satisfaction. *Proceedings of the Human Factors and Ergonomic Society Annual* Meeting, 2006, 50(24), 2547-2551. https://doi.org/10.1177/154193120605002407

Everett, S. P. (2007). The usability of electronic voting machines and how can be changed without detection. 2007 Doctoral Dissertation, Rice University. https://hdl.handle.net/1911/20601

Feng, J, Lazar, J & Preece, J. (2004). Empathy and online interpersonal trust: a fragile relationship, *Behaviour & Information Technology, 23*(2), 97-106. https://doi.org/10.1080/01449290310001659240

Garber, K. (2008). Why the New ES&S Digital Scanner Should Not Be Certified. Florida Fair Elections Center report. https://www.ffec.org/reports/Summary of DS200 Overvote Findings.pdf

Garg, V., Benton, K., & Camp, L. J. (2014). The privacy paradox: a Facebook case study. *2014 TPRC Conference* Paper. https://doi.org/10.2139/ssrn.2411672

Gilbert, J. E., Martin, A. M., Rogers, G., McClendon, J., & Ekandem, J. (2012). Hey, that's not who I voted for!: A study on touchscreen ballot design. *Interactions 19*(6), 34-39. https://doi.org/10.1145/2377783.2377792

Gilbert, J., Dunbar, J., Ottley, A., & Smotherman, J. M. (2013). Anomaly detection in electronic voting systems. *Information Design Journal, 20*(3), 194–206. https://doi.org/10.1075/idj.20.3.01gil

Glaeser, E., Laibson, L., Scheinkman, J., & Soutter, C. (2000). Measuring trust. *Quarterly Journal of Economics, 115*(3), 811-846. https://doi.org/10.1162/003355300554926

Greene, K. K., Byrne, M. D., & Goggin, S. N. (2013). How to build an undervoting machine: lessons from an alternative ballot design. *Proceedings of EVT/WOTE13*, Washington, DC and *USENIX Journal of Election Technology and Systems, 1*(1). https://www.usenix.org/system/files/conference/evtwote13/jets-0101-greene.pdf

Herrnson, P. S., Hanmer, M. J., & Niemi, R. G. (2012). The impact of ballot type on voter errors. *American Journal of Political Science, 56*(3), 716–730. https://doi.org/10.1111/j.1540-5907.2011.00579.x

Holmes, D. V. (2015). Checking your answers: an investigation of alternative review screen design for electronic voting systems. 2015 Doctoral Dissertation, Rice University. http://hdl.handle.net/1911/88131

Hooghe, M., Marien, S., & Pauwels, T. (2011). Where do distrusting voters turn if there is no viable exit or voice option? The impact of political trust on electoral behaviour in the Belgian Regional Elections of June 2009. *Government and Opposition, 46*(2), 245-273. https://doi.org/10.1111/j.1477-7053.2010.01338.x

Howell, S. E., Fagan, D. (1988). Race and trust in government: testing the political reality model. *Public Opinion Quarterly, 52*(3), 343–350. https://doi.org/10.1086/269111

IDEO (2015a.) VOX Research Report, User Inspiration Research, The holistic touch voting experience V1.2 Prototype 5.1.2/3.1.2. Los Angeles County, February 27, 2015. https://vsap.lavote.net/wp-content/uploads/2016/06/Research-Report-March-2015-V1.5.pdf

IDEO (2015b). VOX Research Report: User Inspiration Research V1.2 Prototype 5.1.1/3.1.1. Los Angeles County, February 5, 2015. http://vsap.lavote.net/wp-content/uploads/2016/06/Research-Report-Feb-2015-V1.4.pdf

Jarrett, C. (2015). "The psychology of voting, digested." *British Psychological Society Research Digest* blog. https://digest.bps.org.uk/2015/04/14/the-psychology-of-voting-digested/

Kimball, D. C., & Kropf, M. (2017). Analysis of overvotes in the 2016 California Senate primary election. https://civicdesign.org/wp-content/uploads/2016/04/Analysis-of-Overvotes-2016-Senatt-Kimball-Kropf.docx

Kortum, P., Acemyan, C. Z., & Belton, M. G. (2015). How hard can it be to place a ballot into a ballot box? Usability of ballot boxes in tamper resistant voting systems. *JUS - The Journal of Usability Studies, 10*(4), 129–139. https://uxpajournal.org/usability-ballot-boxes-tamper-resistant-voting-systems/

Kulyk, O., & Volkamer, M. (2018). Usability is not enough: Lessons Learned from "human factors in security" research for verifiability. https://eprint.iacr.org/2018/683

Laskowski, S. J., Autry, M., Cugini, J., Killam, W., & Yen, J. (2004). Improving the usability and accessibility of voting systems and products. NIST Special Publication (NIST SP) 500-256. https://www.nist.gov/publications/improving-usability-and-accessibility-voting-systems-and-products

Laskowski, S. J., Dawkins, S., Quesenbery, W., Chisnell, D., Summers, K., & Rinn, C. (2015). A roadmap for future usability and accessibility guidance. NIST and Center for Civic Design. https://civicdesign.org/wp-content/uploads/2015/05/Roadmap-V2-15-0715.pdf

Little, L., Storer, T., Briggs, P., & Duncan, I. (2008). E-Voting in an ubicomp world: trust, privacy, and social implications. *Social Science Computer Review, 26*(1), 44–59. https://doi.org/10.1177/0894439307307683

Lola, P., Eugene, W., Hall, P., & Gilbert, J. E. (2013). Balloting: speeding up the voting process. *Proceedings of the International Conference on Human-Computer Interaction, HCI 2013*, 373–377. https://doi.org/10.1007/978-3-642-39476-8_76

Los Angeles County (2010-2106). Voting Systems for All People: Research. https://vsap.lavote.net/research/

MacArthur, C. & Quesenbery, W. (2018). 2018 Maine Primary Grid-style Ballot Study Report. Center for Civic Design. https://civicdesign.org/wp-content/uploads/2017/07/RCV-Maine-Research-Report-18-0703-FINAL.pdf

McCarthy, J. D., Sasse, M. A., & Riegelsberger, J. (2003). The researcher's dilemma: evaluating trust in computer-mediated communication. *International Journal of Human-Computer Studies, 58*(6), 759-781. https://doi.org/10.1016/S1071-5819(03)00042-9

McGrew, G. (2013). Assessing usability of the iPad for ballot marking with voters living in group residential facilities (Working Paper #008). https://elections.itif.org/reports/AVTI-008-McGrew-2013.pdf

Marky, K., Kulyk, O, Renaud, K., & Volkamer, M. (2018). What did I really vote for? On the usability of verifiable e-voting schemes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM CHI '18*, 1-13. https://doi.org/10.1145/3173574.3173750

Morrell, J. (2019). Knowing It's Right: A two-part guide to risk limiting audits. Democracy Fund, May 22, 2019. https://www.democracyfund.org/publications/knowing-its-right

National Center for Education Statistics (NCES) (2003). National Assessment of Adult Literacy: Key Findings. https://nces.ed.gov/naal/kf_demographics.asp

National Conference of State Legislatures (2019). Checking the Election: Risk-Limiting Audits. *NCSL Legisbrief, 27*(26), July 2019. http://www.ncsl.org/research/elections-and-campaigns/checking-the-election-risk-limiting-audits.aspx

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *ACM CHI'90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 249-256. https://doi.org/10.1145/97243.97281

Norden, L., (2006). The Machinery of Democracy: Voting System Security, Accessibility, Usability and Cost. Brennan Center for Justice at NYU School of Law. https://www.brennancenter.org/publication/machinery-democracy

Norden, L., Kimball, D., Quesenbery, W., & Chen, M. (2008). Better Ballots. Brennan Center for Justice at NYU School of Law. https://www.brennancenter.org/sites/default/files/2019-08/Report_Better-Ballots.pdf

Norden, L., Kimball, D., & Quesenbery, W. (2012). Better Design: Better Elections. Brennan Center for Justice at NYU School of Law. https://www.brennancenter.org/sites/default/files/2019-08/Report_Better_Design_Better_Elections.pdf

Olembo, M.M., Bartsch, S., & Volkamer, M. (2013). Mental models of verifiability in voting. *Proceedings of the International Conference on E-Voting and Identity Vote-ID 2013*, 142–155. https://doi.org/10.1007/978-3-642-39185-9_9

OSET Institute (2019) Machine-marked printed vote records: recommended principles and guidelines. https://www.osetfoundation.org/research/2019/02/01/pvrprinciples

Pandolfo, A., Rosenzweig, E., & Selker, T. (2006). A methodology for testing voting systems. *The Journal of Usability Studies, 2*(1), 7–21. http://uxpajournal.org/a-methodology-for-testing-voting-systems/

Parent, M., Vandebeek, C., & Gemino, A. (2005). Building citizen trust through e-government. *Government Information Quarterly*, 22(4), 720-736. https://www.sciencedirect.com/journal/government-information-quarterly/vol/22/issue/4

Pennsylvania Department of State (PA DOS) (2019). Electronic Voting Systems Certified after January 1, 2018. https://www.votespa.com/About-Elections/Pages/New-Voting-Systems.aspx

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, *6*(1), 65-78.

Ramchandrani, T. & MacArthur, C. (2017). Usability Testing Report: Ranked Choice Paper Ballots. Center for Civic Design, December 11, 2017. https://civicdesign.org/wp-content/uploads/2017/07/RCV-Denver-Research-Report_18-0104-FINAL.pdf

Redish, J. C, Chisnell, D. E, Newby, E, Laskowski, S. J. & Lowry, S. Z. (2008). Report of Findings: Use of Language in Ballot Instructions, NISTIR 7556, December 2008. https://www.nist.gov/system/files/documents/itl/vote/NISTIR-7556.pdf

Riegelsberger, J., Sasse, M. A., & Mccarthy, J. D. (2005). The mechanics of trust: A framework for research and design. *Proceedings of the International Journal of Human-Computer Studies, 62*(3), 381-422. https://doi.org/10.1016/j.ijhcs.2005.01.001

Sherman, E. , Schiffman, L. Gl, & Thelen, S. T. (2012). Young voters' trust of information and) media sources: The 2008 U.S. presidential election. *Journal of Political Marketing, 11*(4), 246-264. https://doi.org/10.1080/15377857.2012.724294

Stark, P. B., & Wagner, D. (2012). Evidence-based elections. *IEEE Security & Privacy, 10*(5), 33–41. https://doi.org/10.1109/MSP.2012.62

Stewart, C. (2018). Voters think about voting machines. MIT Election Lab – Medium. July 5, 2018. https://medium.com/mit-election-lab/voters-think-about-voting-machines-b468a5a31b44

Summers, K, Chisnell, D., Davies, D, Alton, N., & Mckeever, M. (2014). Making voting accessible: designing digital ballot marking for people with low literacy and mild cognitive disabilities. *Proceedings of EVT/WOTE14*, San Diego, California and *USENIX Journal of Election Technology and Systems, 2*(2). https://www.usenix.org/system/files/conference/evtwote14/jets_0202-summers.pdf

Summers. K., & Langford, J. (2015). The impact of literacy on usable and accessible electronic voting. Proceedings of the *International Conference on Universal Access in Human Computer Interaction, UAHCI 2015: Access the Human Environment and Culture, 248-257.* https://doi.org/10.1007/978-3-319-20687-5_24

Summers, K., Langford, J., Rinn, C., Stevenson, J., Rhodes, E., Lee, J., & Sherard, R. (2017). Understanding voting barriers to access for Americans with low literacy skills. *Proceedings of the International Conference of Design, User Experience and Usability, DUXU 2017: Understanding Users and Contexts*, 294–312. https://doi.org/10.1007/978-3-319-58640-3_21

Swierenga, S., Soucy, K., (2008). Usability test report: New Hampshire. Usability in Civic Life Project. https://civicdesign.org/text-on-the-left-or-right/

Tepfer, D. (2018). Judge refuses to block certification of Stratford's election results. CT Post. November 26, 2018. https://www.ctpost.com/local/article/Judge-refuses-to-block-certification-of-13422622.php

Udani, A. & Kimball, D. C., (2017). Immigrant resentment and voter fraud beliefs in the U.S. electorate." *American Politics Research, 46*(3), 402-433. https://doi.org/10.1177/1532673X17722988

Verified Voting (nd). Risk limiting audits. https://www.verifiedvoting.org/resources/post-election-audits/risk-limiting-audits/

Wallach, D. (2008). Vote "flipping" on Hart InterCivic eSlate systems, ACCURATE blog, October 22, 2008. http://accurate-voting.rice.edu/2008/10/22/vote-flipping-on-hart-intercivic-eslate-systems/index.html

Willemson, J. (2018). Bits or paper: Which should get to carry your vote? Journal of Information Security and Applications, 38, 124–131. https://doi.org/10.1016/j.jisa.2017.11.007