Exploring Reactions to Scantegrity: Analysis of Survey Data from Takoma Park Voters and Election Judges

October 13, 2011

Abstract

In 2009 the voters of Takoma Park, Maryland used the Scantegrity II voting system to cast votes in—and later verify—that year's municipal election. In this paper we analyze survey responses from voters and election judges. This study is novel because it studies a diverse group of voters and election judges in a binding election—the first governmental election to use an end-to-end verifiable voting system with ballot privacy. It provides insight into the experience of a binding election, into the use of an end-to-end-verifiable system in general, and Scantegrity in particular.

Among the findings of this study are that the majority of voters and poll workers surveyed expressed favorable opinions about the system, and that most voters claimed to understand the system. Additionally, there was no notable correlation between survey responses and any of: age, annual household income, computer usage and education level.

1 Introduction

On election day in November 2009, we surveyed voters and election judges about their experiences using and administering the new voting system. In this chapter we report our findings, providing an assessment of how voters and election officials react to an E2E system in a binding governmental election. Because we study voters and poll workers in a real election—as opposed to in a simulated election—these findings more accurately reflect the true experiences of voters and poll workers.

Our study provides insight into the experience of a binding election, into the use of an end-to-end-verifiable system in general, and Scantegrity in particular. We hypothesized that most voters would react as if the system were a traditional optical scan system and be satisfied with the system as seen in other studies, and that the verification mechanism would provide meaningful feedback to increase confidence in the system. We examine the following questions:

1. Is the ability to verify votes valued by voters?
2. Is the additional layer of verifiability suitably transparent to voters and poll workers?
3. Does receipt creation impact the voter experience?
4. Do voters accept the benefits of a system even if they do not necessarily completely understand the underlying technology that drives it?
5. Does the system impact voters of any particular demographic (such as with regard to computer expertise, income or education)?

This election study is the first to study an E2E system with ballot privacy as used in a binding governmental election. Additionally, among studies of E2E systems it is the first to survey such a diverse group of voters in a binding election, and the first to survey election judges in a binding election. With regard to other election studies, it is among only a handful that study voters in a binding election. Binding elections necessarily constrain research methodologies because only one voting system may be used and errors must not be intentionally introduced. Therefore some statements cannot be addressed as well as we would like to and some other statements are out of scope (*e.g.,* a comparison study). Nevertheless, a study of a binding election offers significant advantages over that of the mock election and other studies, by surveying real voters under real-world conditions which can yield observations more representative of the true voter or poll worker experience.

Although the reaction to Scantegrity was positive, especially by voters, the survey also captures the consequence of a number of procedural missteps and shortcomings of the system implementation. We did not find evidence that aspects of the cryptographic protocol negatively affected voter perception. We found little evidence that demographic factors affected voter experience.

Section 4 discusses how we collected our data. We present results in Section 5, and a discussion in Section 6.

2 Scantegrity II

Scantegrity II is an end-to-end publicly verifiable voting system. Each voter can verify that her ballot is counted correctly using a receipt that preserves privacy. In addition, anyone can see that the verified votes are counted correctly. Scantegrity II is an add on verification system, which is used with optical scan equipment.

The voting procedure for Scantegrity II is similar to the voting procedure of a optical scan voting system. The notable difference is that each oval next to a candidate's name used to mark the voter's vote, contains a random confirmation code. This code is printed in invisible ink, and can only be revealed using a decoder pen. When the voter makes her selection on the ballot, she uses the provided decoder pen to mark her selection, and the confirmation code for her choice is revealed. Note that the voter only sees the confirmation code for the candidate she selected, and does not see the confirmation codes for any other candidate on the ballot. If she wishes to verify her vote, she will write her confirmation code on a detachable receipt, which contains the ballot's serial number. If the voter chooses not to verify her vote, then she can ignore the confirmation code, and continue to cast her ballot as usual. If the voter makes a mistake on her ballot, she may ask the election official for another ballot, an her first ballot will be marked as "Spoiled". After the voter finishes voting, she removes the receipt with the written confirmation codes from the ballot, scans her ballot through the optical scanner, and takes her receipt home.

The confirmation codes are independently and randomly assigned to the ballots. This means that the confirmation code for candidate A on Ballot 1 is not the same as the confirmation code for candidate A on Ballot 2. Thus, the voter is able to share her codes, while keeping her vote secret. The confirmation codes cannot be changed or misprinted without detection because they are committed before the election. Before voting, a voter may elect to audit a ballot, and will receive two ballots. The voter chooses which ballot to use for auditing and which ballot to use for voting. The ballot that is chosen for auditing is marked as "Audit Ballot" by the poll official to ensure it is not included in the final vote tally. To audit a ballot, the voter reveals all the confirmation codes on the ballot, and takes the ballot home.

After the polls close, the election authority posts a list of confirmation codes for the positions marked on each ballot it received onto a public bulletin board. Voters who wrote down their confirmation codes can verify that the codes are correct for their ballot number and that there were no additions or deletions to their codes. If the voter finds that the codes she wrote down in the voting booth do not match the codes posted on

the bulletin board for her ballot, she may file a dispute. To file a dispute, the voter enters the dispute with her valid confirmation code online. If the confirmation code is valid, then the dispute will be investigated. If the voter chose to do a print audit on a ballot, the voter checks that all the confirmation codes listed on the audited ballot match the confirmation codes listed on the public bulletin board for that ballot.

After the election, anyone can audit the election. This is accomplished by using software to check the data for correctness and verify the fall tally is correct. The software used to check the integrity of the election can be written by anyone.

3 Related Work

Sherman, *et al.* [**?**, **?**] report on focus groups and a survey similar to ours conducted at a preparatory mock election. In the mock election, Scantegrity team members worked side-by-side with election officials to demonstrate capabilities of the system, and the surveys from voters during this election were positive. Carback, *et al.* [**?**] focus primarily on the technical and administrative aspects of the deployment of Scantegrity at Takoma Park, including lessons learned and briefly touching on survey findings.

There have been several mock and binding organizational elections using E2E systems in recent years [**?**, **?**, **?**, **?**], but the focus of those studies was on practical implementation and engineering decisions. The binding governmental election using the coercible RIES [**?**] system had a similar focus. In March 2009, the president of the Universite Catholique de Louvain was elected with the Internet voting system Helios [**?**, **?**]. Unfortunately, there has been little effort in gauging reaction from voters and election officials. Measuring impact on these users is critical to determining feasibility of end-to-end voting as a paradigm for real elections, and we obtain these measures through surveys and observations.

There are few user studies on E2E systems. Most studies are preliminary usability studies such as the student projects at UMBC (on Punchscan), MIT (on 3Ballot) [**?**], and Univ. of Surrey, England (on PAV). These studies focus on user interface. We are not aware of studies which focus on public acceptance, public reaction, and administrative challenges.

Using expert review, laboratory studies, and a field experiment with 1540 participants, Herrnson, *et al.* [**?**] found that voting system interface and ballot styles had an impact on voter satisfaction, the need for help, and voters abilities to cast their ballots as intended. He also found that verification technologies typically had a negative impact on voter experience. Results of this experiment varied by voter demographics and voting experience.

Examining social issues, Newkirk [**?**] found that public opinion remained remarkably stable between 2004 and 2008. During that time, Direct Recording Electronic (DRE) systems were the top-rated systems for voter trust, followed closely by precinct count optical scan (pcos) systems. Voters rated vote-by-mail, central count optical scan, and internet voting less trustworthy.

Norris [**?**] describes a telephone survey of registered voters in Maryland in which voters provide strongly positive opinions about the usability and accuracy of touch-screen voting. Voters were also positive about the reliability, trustworthiness and count-accuracy of touch-screen machines, while admitting that the systems could be corrupted by malware.

Public confidence in elections was rated highly in these studies, second only to banks. More confidence was voiced for elections than medical providers (including hospitals and clinics), universities and schools, large corporations, and the government. Given the impact of public opinion on the decisions of policymakers who purchase voting systems and oversee other matters related to the administration of elections, it is important to study public reactions to voting systems. This is particularly true for E2E systems, which change the user experience by providing receipts, and change voter expectations by enabling additional verifiability.

4 Methodology

We now describe the research procedures used to collect data during the election. Our research protocols and questionnaires were approved by UMBC's Institutional Review Board, as required for experiments with human subjects. Voters were polled from 8am to 5pm, and voting hours were from 8am to 8pm. The study participants comprised election judges who administered the election and voters who voted in the election.

4.1 Research Protocol

Our research team was not permitted in or around the voting area. Instead, we were allowed to use the area typically designated for exit polling. A surveyor was posted at the main exit, and additional surveyors also covered other exits throughout part of the day.

As each voter left the polling location, a surveyor asked the voter if she would be willing to fill out a questionnaire. If yes, the researcher handed the voter a conventional clipboard with two one-sided questionnaires: a voter field test questionnaire and a demographics questionnaire. Form numbers linked the field test and demographics questionnaires filled out by the same voter.

Voters could visit the on-line verification web site after polls closed. We wrote an on-line questionnaire, but Takoma Park requested we put it on the bottom of the ballot check results page. No voters filled out the on-line questionnaire.

Election judges were also given an election judge field test questionnaire and demographics form. At the end of the day, each judge was provided with an addressed, stamped envelope and the survey form. We requested that each judge fill out and mail back the form.

Election judges operated the system entirely by themselves, and select members of our research team were designated to fill vendor roles. These team members worked separately from the surveyors and did not perform any surveyor functions.

The demographics forms and questionnaires were sequentially numbered when given to participants. The sequential numbering allowed us to correlate demographic information with survey responses through the serial number while maintaining the anonymity of the participants. After collecting the forms we scanned the demographic and field study questionnaires separately. Using the sequential ID numbers we entered each response into one row of a spreadsheet using the mark position number to denote the selections of each respondent. Afterward, we imported the spreadsheet into an SQLite database,[1] and used the R statistical language to process the form data.[2]

We did not collect any personal identifying information for voters or election judges. We did however, serially number the forms handed out and the demographics forms were keyed to demographics forms. In some cases, where judges or voters fall into specific demographic patterns, or when the time the voters filled out the survey is known, it may be possible to identify some respondents.

4.2 Research Instruments

We gave out three types of forms to participants. Voters received a demographics form and a voter field study questionnaire. Election judges received the same demographics form, and an election judge field study questionnaire.

[1] `http://sqllite.org/`
[2] `http://r-project.org`

4.2.1 Demographics Form

We gave voters and election officials the same demographics form, consisting of 12 questions on standard demographic information and information about prior experience with different types of voting systems. The questions asked respondents about their sex, age, race, languages spoken, education level, computer usage, participation in previous elections, whether any mistakes had been made while voting in previous elections, previous voting systems used, physical challenges, and annual household income. See Figure 1 to view the demographics form.

4.2.2 Voter Field Study Questionnaire

The voter questionnaire comprised 17 questions. The first 12 used a 7-point Likert (Strongly Disagree to Strongly Agree) scale, and included a "not applicable" option. [3] The remaining questions (14 through 18) asked respondents how many times they made mistakes, whether they attempted to audit ballot printing, whether they asked for assistance, whether they had any difficulties voting, and whether they had comments about the process. Figure 2 shows the questionnaire used during the election.

4.2.3 Election Judge Field Study Questionnaire

The election judge questionnaire was two-sided. The first side featured fourteen 7-point Likert scale questions, which asked whether the system was easy to administer, whether it was easy for voters to mark ballots, whether it was easy for voters to correct mistakes, whether it was easy for voters to record code numbers, whether the voting system was easy for voters to use, whether voters appeared comfortable using the system, whether the respondent had confidence ballots were correctly printed, whether the official data recorded votes as intended, whether the final tally correctly included votes as case, whether the receipt revealed how the voter voted, whether votes were private and remained so, whether the option to verify the vote increases their confidence, and whether they had confidence in the voting system. The second side of the sheet asked free-form questions about difficulties administering the voting system, difficulties observed that voters had using the voting system, suggestions to improve the voting process, and additional comments. Figures 3 and 4 show the judge questionnaires.

4.3 Interpretation of Raw Data

Due to the nature of our study, participants were often unclear in their responses and we were unable to seek clarification from them. We had to interpret some of the responses, and we endeavored to do so in a clear and consistent fashion:

- If a respondent marked multiple answers on questions, either the darkest response or the response most like the other responses was accepted.
- If the respondent crossed out the unintended response to make his or her intent clear, we took that response instead of the darkest.

We recorded in our spreadsheet a column for each response describing any intent issues and how they were resolved.

[3] A typo on the survey skipped the numbering on the voter questionnaire from 2 to 4. This error is preserved in this paper.

Demographics Questionnaire

Feel free to skip any question that you prefer not to answer.

What is your sex?
O male **O** female

How old are you?
O 12-17 **O** 18-24 **O** 25-34 **O** 35-49 **O** 50-64 **O** 65-74 **O** 75+

What racial/ethnic group best describes you? (select all that apply)
O White **O** Black **O** Asian **O** Hispanic/Latino **O** Multiracial
O Other: _____ **O** I prefer not to provide this information.

What language do you speak at home? (select one)
O English **O** Spanish **O** Other:_____

What is the highest level of education you have completed?
O some high school **O** high school diploma or GED **O** some college, no degree
O 2-year degree **O** 4-year degree **O** some post-graduate work, no degree
O graduate or professional degree (*e.g.*, MS, PhD, MD, JD)

On average, how often do you use a computer?
O never **O** once every two weeks **O** 1-3 times per week
O 4-6 times per week **O** 7-9 times per week **O** 10+ times per week

In how many previous government elections (city, state, and/or federal) have you voted?
O 0 **O** 1 **O** 2 **O** 3+

In previous governmental elections, have you ever made a mistake a received a fresh ballot?
O yes **O** no

Are you, or have you ever been, a poll worker?
O yes **O** no

Before today, which voting technologies have you used? (select all that apply)
O none **O** paper **O** touch screen **O** punch card **O** lever machine
O Other: _____

What physical challenges do you face? (select all that apply)
O none **O** limited eyesight **O** blindness **O** limited hearing **O** deafness
O tremors **O** limited motor control **O** limited mobility
O other: _____ **O** I prefer not to provide this information.

Which category best describes your total annual household income?

Figure 1: The demographics form used during the municipal election.

Field Study Questionnaire 1

Darken the oval completely for the choice that best fits you answer.

For Questions 1-19, please indicate how strongly you agree or disagree with the following statements about the voting system you just used.

	strongly disagree 1	2	3	4	5	6	*strongly agree* 7
1. It was easy to mark my ballot.	O	O	O	O	O	O	O
2. It was easy to correct mistakes.	O	O	O	O	O	O	O
3. It was easy to record my codenumbers and keep a receipt.	O	O	O	O	O	O	O
4. It was easy to scan my ballot.	O	O	O	O	O	O	O
5. It was easy to use the locked clipboard.	O	O	O	O	O	O	O
6. I feel the locked clipboard adds security to the system.	O	O	O	O	O	O	O
7. Overall, the voting system was easy to use.	O	O	O	O	O	O	O
8 I feel comfortable using the system.	O	O	O	O	O	O	O
9. I am confident that the ballots were correctly printed.	O	O	O	O	O	O	O
10. I am confident the *official data* will include my intended vote.	O	O	O	O	O	O	O
11. I am confident the *final tally* will correctly include my vote as cast.	O	O	O	O	O	O	O

Figure 2: The voter questionnaire used during the municipal election.

Field Study Questionnaire for Poll Workers

For Questions 1-20, please indicate how strongly you agree or disagree with the following statements about the voting system you just used.

	strongly agree 1	2	3	4	5	*strongly disagree* 6	*not applicable* 7 N/A
1. It was easy to administer Scantegrity.	O	O	O	O	O	O	O O
2. It was easy for voters to mark ballots.	O	O	O	O	O	O	O O
3. It was easy for voters to correct mistakes.	O	O	O	O	O	O	O O
4. It was easy for voters to record codenumbers and keep receipts.	O	O	O	O	O	O	O O
5. It was easy for voters to scan ballots.	O	O	O	O	O	O	O O
6. Overall, the voting system was easy for voters to use.	O	O	O	O	O	O	O O
7. Voters appeared comfortable using the system.	O	O	O	O	O	O	O O
8. I am confident that the ballots were correctly printed.	O	O	O	O	O	O	O O
9. I am confident the *official data* will record votes as intended.	O	O	O	O	O	O	O O
10. I am confident the *final tally* will correctly include votes as cast.	O	O	O	O	O	O	O O
11. I am confident that a receipt by itself does not reveal how the voter voted.	O	O	O	O	O	O	O O

Figure 3: Page 1 of the election judge questionnaire used during the municipal election.

12. I am confident votes are and will remain private. **O O O O O O O O**

13. The option to verify a vote on line increases my confidence in the election results. **O O O O O O O O**

14. I have confidence in this system. **O O O O O O O O**

PLEASE CONTINUE ON THE OTHER SIDE

15. Please describe any difficulties administering this voting system.

16. Please describe any difficulties you observed voters having with this voting system.

17. Do you have any suggestions for improving the voting process?

18. Any additional comments?

Figure 4: Page 2 of the election judge questionnaire used during the municipal election.

4.4 Actions In the Polling Site

Our team was allowed to observe election day events, but we were not permitted to serve as election judges nor to interfere with the elections process. Additionally, only two representatives were permitted in the voting area at any given time.

Two members of our team acted as technical support when needed, fulfilling the role that a vendor would during election day.[4] Before the election the technical support team was directed to set up the scanning stations under supervision by an election judge. After the election they were asked to disconnect the scanning stations and to collate the memory sticks for tabulation by the election night tabulation software. The technical support team members did not interact with participants in the survey, and did not instruct the election judges regarding the survey.

5 Results

Descriptive statistics provide evidence which supports our hypothesis that the voting system was positively received by voters. Election judges, however, noted the effort made by voters on several different aspects of the system, and made suggestions for improvements. Voter comments indicated that voters would like more instruction on how to use the voting system.

5.1 Voter Surveys

1723 people voted on election day, of whom 276 (16%) filled out surveys. 36 (13%) left the demographics form blank and filled out the questionnaire only. Conversely, 5 (1.8%) respondents left the questionnaire blank but filled out the demographics form. Many did not answer a subset of questions on one or both forms. There were 235 (85.1%) respondents who answered some questions on both sheets, 240 (86.9%) who answered the demographics, and 271 (98.2%) who answered the questionnaire.

We provide bar charts for voter responses to the questionnaire in Figures 5 and 6. All of the Likert scales appear bimodal, indicating that respondents were highly opinionated. This is not something we saw in the mock election test study about this system with volunteer users [**?**]. Figure 7 shows an alternate view of the Likert questions designed to show correlation. Here we see that general agreement/disagreement of respondents on each question varies. Q2 (EasyToCorrectMistakes) and Q12 (IntendToVerify) appear to be equally distributed between the high and low values of the Likert scale.

Voter responses support our general hypothesis of satisfaction for the new voting system. The response to Q13 (HaveConfidence) showed 230 out of 268 participants (85.8%) marked at or above 5 (Somewhat Agree) on the scale, with a mean of 5.78 and a median of 7. Table 1 summarizes this information for the rest of the likert questions.

When compared with Q10 (VerificationGivesMoreConfidence), Q12 (IntendToVerify) indicates that respondents said they were more confident in the system because of the receipts, but fewer were willing to check their ballot online at home. This suggest that while voters may find value in the receipts, they might not take advantage of it, which is further supported by our observation that only 81 receipt checks were made by voters after the election (see Section **??**). It also provides evidence that most voters will accept systems that they do not necessarily completely understand.

To understand if voter demographics affect voter experience we use ordinary least squares (OLS) regression of a combined variable we call satisfaction over the demographics factors we collected. To create the satisfaction variable, we combined the average responses to Q1, Q4-11, and Q13 for each respondent. We dropped Q2 because it appears that most voters misread the question to put NA because they did not make a

[4] In Maryland, technical support representatives from the election vendor are available to election judges at each polling site on election day.

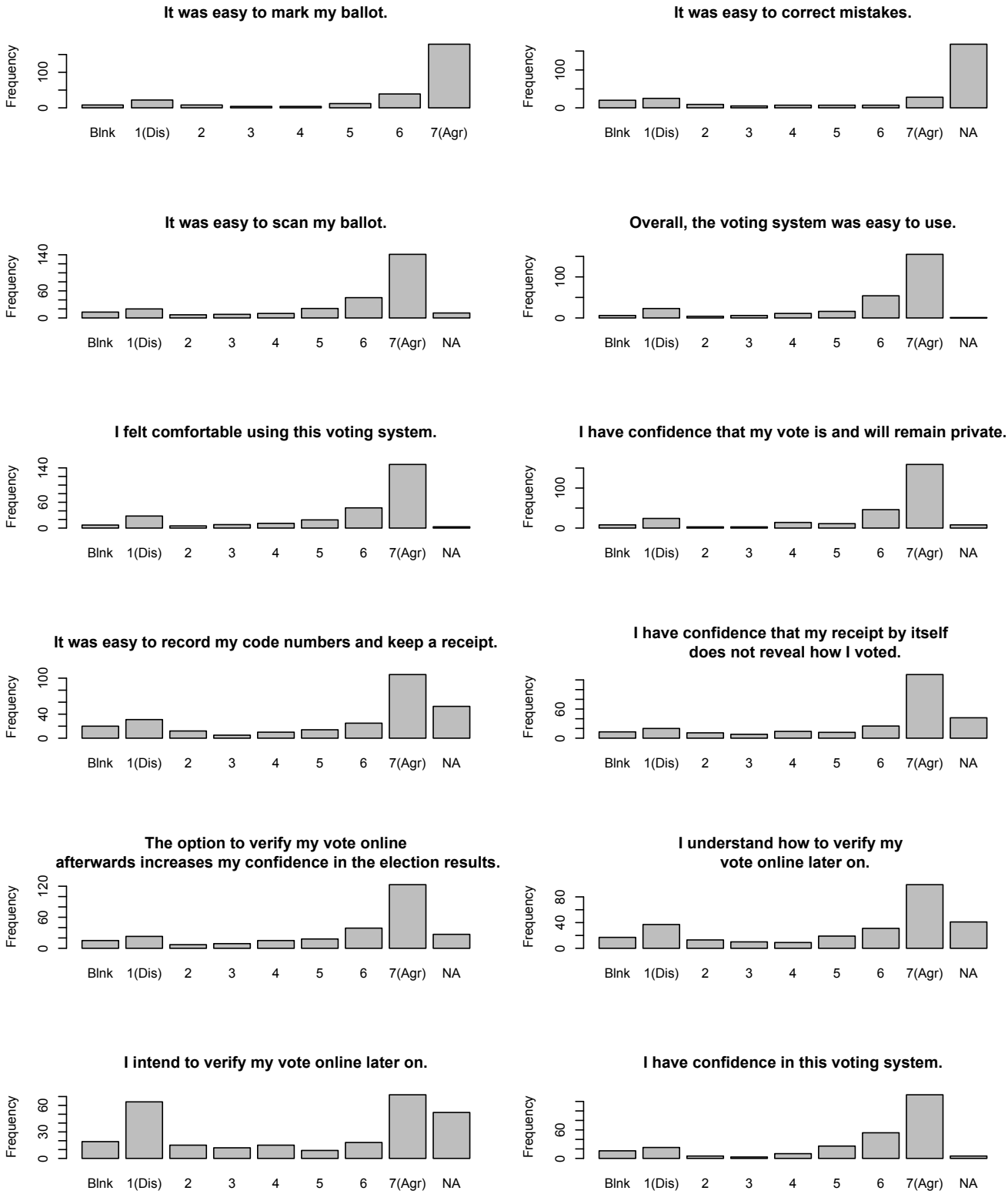


Figure 5: Voter reactions to Scantegrity at Takoma Park.

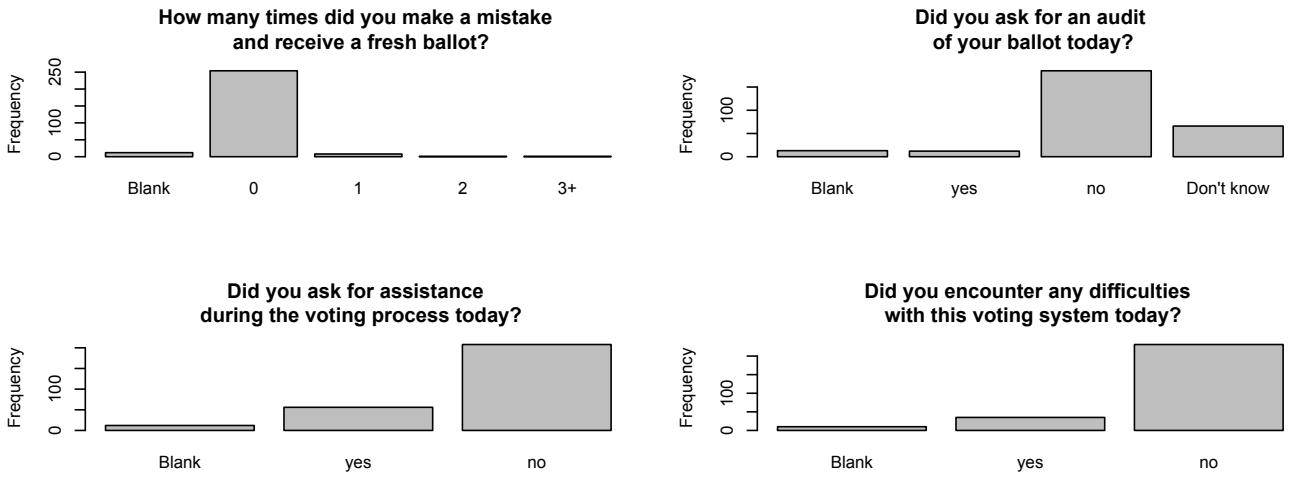


Figure 6: Voter reactions to Scantegrity at Takoma Park (cont).

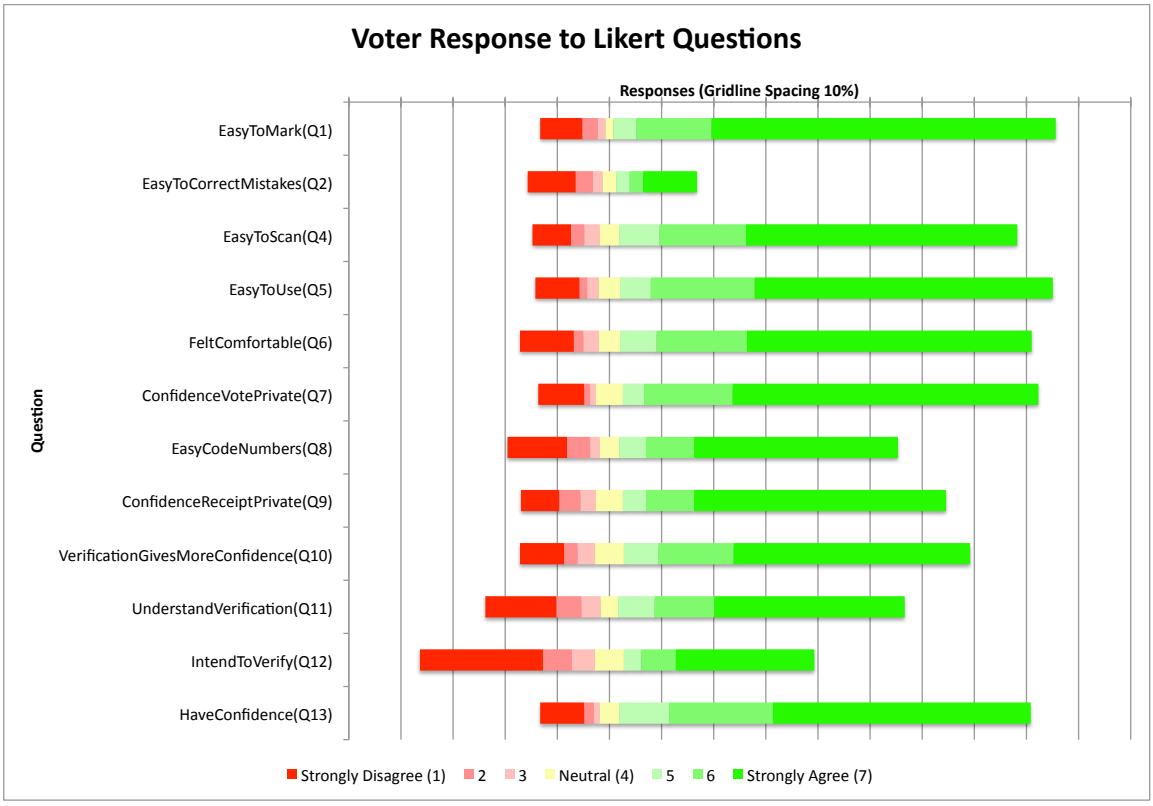


Figure 7: Horizontal bar charts showing distributions around the neutral position of the Likert questions. The width of each bar represents the total number of respondents for that question, and each bar is divided into subsections whose width represent the respondents in that category.

	≥ 5 (Agree)	$= 4$ (Neutral)	≤ 3 (Disagree)	Mean	Median	Std. Dev.	N
Q1(EasyToMark)	230 (85.82%)	4 (1.49%)	34 (12.69%)	6.02	7	1.86	268
Q2(EasyCorrect)	42 (47.73%)	7 (7.95%)	39 (44.32%)	4.08	4	2.52	88
Q4(EasyScan)	207 (82.14%)	10 (3.97%)	35 (13.89%)	5.79	7	1.87	252
Q5(EasyUse)	225 (83.64%)	11 (4.09%)	33 (12.27%)	5.88	7	1.84	269
Q6(FeltComfy)	214 (80.45%)	11 (4.14%)	41 (15.41%)	5.71	7	1.98	266
Q7(ConfPriv)	216 (83.08%)	14 (5.38%)	30 (11.54%)	5.92	7	1.86	260
Q8(EasyCodes)	145 (71.43%)	10 (4.92%)	48 (23.65%)	5.28	7	2.30	203
Q9(PrivReceipt)	168 (76.02%)	14 (6.33%)	39 (17.65%)	5.65	7	2.05	221
Q10(ConfVerify)	180 (76.92%)	15 (6.41%)	39 (16.67%)	5.59	7	2.00	234
Q11(UnderstandVerify)	149 (68.35%)	9 (4.13%)	60 (27.52%)	5.06	6	2.34	218
Q12(IntendToVerify)	99 (48.29%)	15 (7.32%)	91 (44.39%)	4.13	4	2.59	205
Q13(HaveConfidence)	219 (84.23%)	10 (3.85%)	31 (11.92%)	5.82	7	1.86	260

Table 1: Voter responses to Likert scale questions about Scantegrity at the municipal election.

mistake (168 out of 256 respondents put NA), and we dropped Q11 because it deals with expected behavior and not satisfaction with the process. The cronbach's α of the remaining selected questions was .97 (N = 142).

The resulting dependent satisfaction variable had a mean of 5.69[5] (StdDev = 1.7, N = 271). Because the data was negatively skewed (-1.94) and had high kurtosis (5.92) we analyzed the cube (x^3) of the values (skew = -.86, kurtosis = 2.69).

Age, education, computer use, and income were coded as ordinal data starting at 0 for the smallest category (see Figure 1 for the specific categories and range of each variable). Gender was coded to 1 for female, 0 for male. We coded 1 if the respondent was black non-hispanic, a former election judge, reported any disabilities (except hearing), or used any of a touch screen, punch card, or lever machine system.

We expected positive effects for education, computer use, being a former election judge, black non-hispanics, and use of any type of voting machine. People who have more education and use the computer more often might be more likely to understand the system, and thus see its value. Black non-hispanics have been observed to react positively to optical scan systems in other studies [**?**]. People who have been election judges or who have used different voting equipment might be more familiar with deficiencies in existing equipment and might also be more likely to see value in the ability to take home a receipt.

We expected negative effects for age, non-english speakers, and people with disabilities. We believed older voters would dislike the internet-enabled part of the system, and that people with disabilities would dislike the paper. Non-english speakers are at a natural disadvantage. We did not expect to see significant impact by gender or income.

Table 5.1 provides results for the effects of demographic variables on satisfaction. The model produces a mix of agreement and disagreement with our hypotheses.

The model shows a statistically significant positive effect for females and negative effect for income, and we did not expect either of these variables to have statistically significant effects. The income effect is smaller and weaker compared to the other significant variables, and could be the result of a sample size that is too small. It is unclear why women would react more favorably to the system.

While the effect of punch cards was statistically significant in the expected direction, the effect of touch screens was in the opposite direction. We suspect this is a result of the population being highly computer

[5] A table depicting mean scores of the satisfaction index for all demographic factors is available in figure 13 in appendix **??**.

Coefficients:	Estimate	Std. Error	t value	Pr($>\|t\|$)	
Intercept(Const)	274.162	62.588	4.380	1.98e-05	***
Gender	33.535	15.760	2.128	0.01733	*
Age	-3.515	9.516	-0.369	0.35612	
Black Non-Hispanic	13.448	25.109	0.536	0.29644	
Non-English	-39.987	33.042	-1.210	0.11387	
Education	1.419	5.808	0.244	0.40366	
Computer Use	-1.254	8.083	-0.155	0.43844	
Former Judge	-19.533	21.545	-0.907	0.18290	
Income	-6.885	5.096	-1.351	0.08915	.
Disabilities	-26.540	28.462	-0.932	0.17615	
Touch screen	-47.516	25.960	-1.830	0.03440	*
Punch card	48.322	18.425	2.623	0.00472	**
Lever	-7.497	18.135	-0.413	0.33989	
N	199				
Res. SE	106.771				
Mult. R^2	0.0945				
Adj. R^2	0.0361				
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	

Table 2: Voter satisfaction with Scantegrity across selected characteristics. Entries are OLS coefficients, standard errors, t-values, and probabilities. All entries are one-tailed, except Intercept(Const).

literate (almost 73% reported using the computer 10+ times a week) and already accustomed to the existing touch screen DRE system currently in use in Maryland.

The rest of the variables in the model did not show significant effects on satisfaction. The age, black non-hispanic, non-english, education, and disabilities variables show effects in the expected direction. The computer use, former judge, and lever variables did not show effects in the expected direction. The former judge variable, in particular, showed a strong negative effect. We believe these voters were more likely to notice less polished aspects of the system.

We further explored the demographics effects using correlation analysis, and we found women tended to agree more strongly than men on Q1(EasyToMark) ($\chi^2 = 10.14$, $p = 0.001$, $df = 1$), Q4(EasyScan) ($\chi^2 = 4.18$, $p = 0.041$, $df = 1$), Q5(EasyUse) ($\chi^2 = 7.64$, $p = 0.006$, $df = 1$), Q6(FeltComfy) ($\chi^2 = 5.0$, $p = 0.025$, $df = 1$), and Q7(ConfPriv) ($\chi^2 = 9.25$, $p = 0.002$, $df = 1$). We did not find correlation of the survey questions with age, race, education, income, computer usage, or experience in previous elections. There were some correlations among demographic data. We include these in Appendix A. Correlation tables of the demographic data with the questionnaire are in Appendix B.

5.1.1 Voter Comments

51 voters wrote comments on the questionnaires, often pointing out confusion about various aspects of the process:

1. Many were unaware of the verification option.

2. Some did not realize they were supposed to write down confirmation numbers.

ID	Response to Question (1=strongly disagree, 7=strongly agree, 8=N/A, 9=no response)													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
4	4	6	3	3	3	4	4	7	6	6	7	6	6	5
5	3	5	9	9	4	3	3	9	9	6	7	7	1	4
6	5	5	8	2	3	4	4	7	6	6	7	6	3	4
10	4	4	3	4	3	4	4	4	4	4	4	4	4	4
11	4	3	2	5	4	3	3	7	7	7	7	7	4	6
mean	**4.0**	**4.6**	**2.7**	**3.5**	**3.4**	**3.6**	**3.6**	**6.3**	**5.8**	**5.8**	**6.4**	**6.0**	**3.6**	**4.6**
median	**4.0**	**5.0**	**3.0**	**4.0**	**3.0**	**4.0**	**4.0**	**7.0**	**6.0**	**6.0**	**7.0**	**6.0**	**4.0**	**4.0**
number of responses			5		42%									
number of judges			12											

Figure 8: Election judge reactions to Scantegrity at Takoma Park.

3. Some found the pens confusing to use: they did not realize that the pens would expose confirmation numbers, and they did not know which end to use.
4. Some found confirmation numbers were hard to read.
5. Some did not understand how to mark an IRV ballot.
6. Some did not know how to place the ballot into the scanner.
7. One had no difficulty but wondered if seniors or people who speak neither English nor Spanish might have difficulties.
8. One wondered if the government might be able to discern his vote by linking his IP address used during verification with his ballot serial number and noting the time that he was issued a ballot.
9. Many suggested that it would have been helpful to have better instructions, including instruction while they waited in line.

5.2 Summary of Election Judge Surveys

There were 12 election judges on election day, of which 5 (42%) responded to our survey. The Judges were much more mixed in their reaction to the new system than were the voters in our sample. Figure 8summarizes responses to the Likert questions on the election judge survey.[6]

The judges noted the following difficulties:

1. There was too much information.
2. Some voters did not understand what to do, including how to create a receipt.
3. Some voters did not understand how to mark an IRV ballot.
4. The privacy sleeve was hard to use with one hand.

[6]We chose not to show election judge demographics to protect privacy.

5. The double-ended pens created confusion.

6. Voters, poll workers, and the Scantegrity team have different needs.

7. One wondered whether Scantegrity was worth the extra trouble.

The judges offered the following suggestions:

1. Simplify the ballot.

2. Provide receipts so that voters do not have to copy confirmation numbers.

3. Develop better pre-election voter education.

6 Discussion

In Chapter 1 we explain our interest in exploring how voters and election officials react to Scantegrity and its use of codes to represent votes. In particular, we are interested in determining if the ability to verify votes is valued by voters, and also whether this additional layer of verifiability is suitably transparent to voters and poll workers. The data from our sample indicate that voters have high levels of appreciation for Scantegrity, that there is no evidence the extra functionality detracts from voter experience, and that voters say the verification function it provides increases confidence in election results. Election officials were less enthusiastic about the system but they did not report that the system overhead too intensive.

We also obtained answers in more specific areas. The system was not too complex to use and administer in the context of elections at Takoma Park. Election judges responding to the survey found it workable. Comments from election judges and voters either spoke well of or ignored the verification option and focused on other issues they perceived to be problematic in the election. Large majorities of voters found various aspects of the system easy to use. 145 out of 203 respondents (71.4%) agreed to the statement "It was easy to record my code numbers and keep a receipt."

The small number of questions on each ballot worked in the system's favor. Increasing the number of questions on the ballot would likely decrease impressions of usability, although this is, to varying degrees, the case for any system.

The extra work involved to participate in verification does not appear to negatively impact voter experience. Levels of confidence for the system were high, indicating that the impact could be comparable to other systems.

Voters appeared to accept the system even if they did not understand it. This study shows high levels of support for the verification receipt, even when voters indicated they did not understand the cryptographic mechanisms behind it. The understand question correlated with confidence less well than other metrics on the survey, which indicates that voters had confidence and found the system easy to use even if they did not necessarily understand parts of the system (see Figure 7).

Respondents did say that they appreciated the extra security as the presence of the verification option was reported by voters to increase their confidence in the system (77% agreed to Q10, $N = 234$). This contrasts with the number of people who reported intent to verify (48.3%, $N = 205$), and more so with 81 checks that actually occurred on the ballot data. It appears that, even though voters appreciated the technology, they did not necessarily care to use it.

It is nontrivial to address if certain voting populations will be disadvantaged by this system because participants might not notice or report problems or difficulties they have with the system. We did not find statistically significant evidence that non-computer users, less education, black non-hispanic, or other factors had an undesirable impact on voter satisfaction. We did find that women tended to agree more strongly than

men on several ease of use questions, although men still overwhelming agreed strongly. We also found higher income and prior use of a touch screen to negatively impact satisfaction.

Mathematically and technically speaking, we did get enough voters to verify their ballot for this particular election (see Chapter **??**). Unfortunately, we cannot make the argument that we will always have enough verifiers to ensure election security. However, we also cannot find any evidence in the data to support the argument that we will not. This statement should be investigated in future work.

6.1 Known Limitations & Recommendations

There are a few limitations with the data we collected. The most significant is that due to the way procedures were implemented many voters were not aware of the receipt function. The receipts were put in each poll booth for voters to find. While instructions were available, the receipts were not explained to the voters as they were given a ballot, and many respondents complained to the surveyors and in comments about this issue. Our most important recommendation is to find ways to increase awareness and the number of individuals who will take home a receipt. The best way to accomplish this goal is to find a way to produce and provide the receipts automatically as part of the election process, which we discuss in Chapter **??**.

Another issue is that our survey sampled few voters with disabilities. It is important to understand how this new model will affect these users.

Q2 (EasyToCorrectMistakes) did not yield any useful information. Most voters marked NA or left this question blank, which was the correct response if they did not make any mistakes. We would not ask this in future surveys unless the voter had to experience the process of correcting a mistake. Q12 asks about intention and not experience. In the future we will instead use two binary questions, *e.g.,* "Did you record a receipt?" and "Do you intend to use your receipt to verify your vote online tomorrow?"

Because the study is observational, we were unable to address the question of the effects of the confirmation number receipts as well as we would have liked. Respondents to the questionnaire reported that the presence of a receipt increased their confidence in the results, but how many would have high confidence in the results if they had also used a system without a receipt? A comparative study which looks closely at this issue is a next step for this research.

There were several miscellaneous technical problems throughout the voting day. In the morning one scanner's power was tripped. Later in the day a few voters reported that the scanners had trouble pulling the ballots out of the privacy sleeves. One voter dropped his receipt into the scanner. One scanner jammed but was quickly fixed. A respondent who was the victim of, or witnessed, any of these issues was likely to have a negative response, and it is impossible to control for these types of issues in a real world environment. We recommend that technical measures be taken to minimize these issues, and we agree with the election judges that steps should be taken to simplify the ballot and to develop better pre-election voter education resources.

We believe that, while voters appreciate the ability to verify, it is important that every aspect of the system work well. Much attention was payed to small details, such as the ballot layout, in this implementation of Scantegrity. Instructions for IRV and how to record a receipt were checked carefully for clarity. Takoma Park used an official staff member to create translations of these instructions. The scanner recognized ballots being placed into the system with a beep. These seemingly unimportant issues led to a reasonably smooth election, but more work could have been done to improve the voter and election judge experiences.

7 Conclusion

The findings in this study indicate that voters were positive about the system, that they valued the security provided, that the extra work of optionally noting down confirmation codes did not negatively impact the

voter experience, and that they accepted it in spite of not understanding its inner workings completely. Voter comments indicated that there had been some confusion about specific aspects of the system—such as the double-sided pen used to mark the ballot and the privacy sleeve–and that voters would value additional instruction on use of the voting system. Election judge responses indicated that the system was not too hard to administer. These responses also indicated that election judges viewed more problems with the system, providing important recommendations for future use. Except for (a) the fact that women tended to agree more strongly than men on the Likert-scale questions, and (b) a correlation between ethnicity and the intention to verify online, we did not observe significant correlation between demographic data and responses. We did observe a correlation among most of the questions, however, leading to the conclusion that most voters were highly satisfied with the system.

For future work, a clear measure of the confidence increase (or decrease) the receipt provides is necessary. A comparison study between Scantegrity and a commercial optical scanning system is an obvious next step. Another area to explore is whether enough voters will use the receipts.

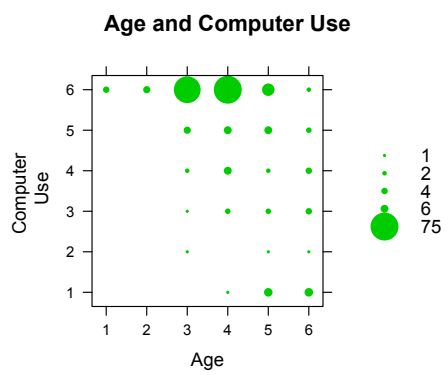
We believe that, while voters appreciate the ability to verify, it is important that every aspect of the system work well. Much attention was payed to small details, such as the ballot layout, in this implementation of Scantegrity. Instructions for IRV and how to record a receipt were checked carefully for clarity. Takoma Park used an official staff member to create translations of these instructions. The scanner recognized ballots being placed into the system with a beep. These seemingly unimportant issues led to a reasonably smooth election.

The fact that election officials were more mixed in reaction to the system is important. These individuals saw parts of the system that were not polished. A polished, professional implementation would yield a more positive response from election judges.

A Relationships between Demographic Data

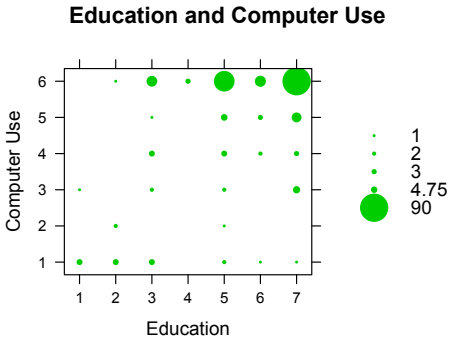
As part of our analysis, we found several correlations between the demographic data. These may assist in explaining our results. Bubble charts of the major

Age and Computer Use. We found that age was negatively correlated with computer usage (ρ = -.5, p = 4.4E-16, n = 233). Figure 9(a) shows age plotted against computer use in a bubble chart. The chart indicates that computer usage fans out in older respondants, but is high and universal among younger respondants.

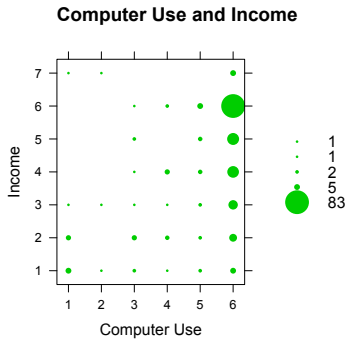


(a) Comparing Age and Computer Usage.

(b) Comparing Age and Income. Note that 7 in this graph is "Don't know," and 6 is $100k+.



(c) Comparing Education and Computer Usage.



(d) Comparing Computer Usage and Income.

Figure 9: Different correlations among the demographic data.

Age and Income. Age negatively correlated with income (ρ = -.4, p = 3.6E-9, n = 202). Figure 9(b) clearly shows a possible anomaly in the sample as the 2 youngest groups report unusually high income. We expected income to go up with age and decline after retirement age.

Education and Computer Use. Education positively correlated with Computer Use (ρ = .3, p = 3.16E-6, n = 232). Figure 9(c) shows a small number of respondants who do not use the computer regardless of education.

Computer Use and Income. Computer Use positively correlated with Income (ρ = .47, p = 3.1E-12, n = 202). Figure 9(d) shows few high income earners with low computer usage rates.

Race/Ethnicity. Due to low frequency of some categories, race and ethnicity appeared to correlate with every other demographic measure except DQ8 (PrevMistakes) and DQ9 (PollWorker). We combined Hispanic, Multi, and Black (traditionally underrepresented groups) categories and tested our group against Whites. The results still correlated with age, education, computer use, and income.

Other. We found a few other weak correlations. Education correlated with Income (ρ = .28, p = 5.3E-5, n = 202). Voting in previous elections correlated with education (ρ = .15, p = .022, n = 234), Computer use (ρ = .17, p = .01, n = 233), and Income (ρ = .14, p = .04, n = 203).

B Additional Data Tables

	Q1	Q2	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Q1 (EasyToMark)	276	85	250	266	263	257	201	219	232	216	203	253
Q2 (CorrectMistakes)	85	276	84	87	86	85	78	77	80	77	75	82
Q4 (EasyScan)	250	84	276	250	248	242	188	205	220	204	192	237
Q5 (EasyUse)	266	87	250	276	264	259	202	220	233	217	203	253
Q6 (FeltComfy)	263	86	248	264	276	256	199	217	232	215	203	252
Q7 (ConfPriv)	257	85	242	259	256	276	195	214	227	212	199	246
Q8 (EasyCodes)	201	78	188	202	199	195	276	190	189	180	166	194
Q9 (PrivReceipt)	219	77	205	220	217	214	190	276	198	192	175	211
Q10 (ConfVerify)	232	80	220	233	232	227	189	198	276	204	190	225
Q11 (UnderstandVerify)	216	77	204	217	215	212	180	192	204	276	191	208
Q12 (Intend to Verify)	203	75	192	203	203	199	166	175	190	191	276	195
Q13 (HaveConf)	253	82	237	253	252	246	194	211	225	208	195	276

(a) N values for questions 1-13 on the voter survey.

	Q1	Q2	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Q1 (EasyToMark)	NA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q2 (CorrectMistakes)	0.00	NA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00
Q4 (EasyScan)	0.00	0.00	NA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q5 (EasyUse)	0.00	0.00	0.00	NA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q6 (FeltComfy)	0.00	0.00	0.00	0.00	NA	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q7 (ConfPriv)	0.00	0.00	0.00	0.00	0.00	NA	0.00	0.00	0.00	0.00	0.05	0.00
Q8 (EasyCodes)	0.00	0.00	0.00	0.00	0.00	0.00	NA	0.00	0.00	0.00	0.00	0.00
Q9 (PrivReceipt)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	0.00	0.00	0.00	0.00
Q10 (ConfVerify)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	0.00	0.00	0.00
Q11 (UnderstandVerify)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	0.00	0.00
Q12 (Intend to Verify)	0.00	0.04	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	NA	0.00
Q13 (HaveConf)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA

(b) P values for questions 1-13 on the voter survey.

Figure 10: Remaining tables for Spearman's ρ correlation.

Nominal	Q1	Q2	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
DQ1 (Gender)	10.14	1.20	4.18	7.64	5.00	9.25	0.01	1.92	2.09	0.05	0.49	2.82
DQ8 (Mistakes?)	0.88	1.84	0.33	0.92	1.66	0.36	0.87	0.60	2.06	1.83	2.27	1.54
DQ9 (PollWorker?)	1.72	0.52	3.74	2.90	2.14	1.97	0.02	0.34	0.00	1.91	1.77	0.49
DQ3 (Race)	13.37	6.01	13.24	20.55	15.63	8.90	6.68	6.73	11.52	15.21	6.01	16.92
DQ3 Reduced	0.45	2.36	1.85	0.02	0.03	0.04	0.35	0.25	2.90	3.82	7.40	0.06
P Values												
DQ1 (Gender)	0.00	0.27	0.04	0.01	0.03	0.00	0.94	0.17	0.15	0.82	0.49	0.09
DQ8 (Mistakes?)	0.35	0.17	0.56	0.34	0.20	0.55	0.35	0.44	0.15	0.18	0.13	0.21
DQ9 (PollWorker?)	0.19	0.47	0.05	0.09	0.14	0.16	0.90	0.56	0.97	0.17	0.18	0.48
DQ3 (Race)	0.04	0.42	0.04	0.00	0.02	0.18	0.35	0.35	0.07	0.02	0.42	0.01
DQ3 Reduced	0.50	0.12	0.17	0.88	0.87	0.84	0.55	0.62	0.09	0.05	0.01	0.80
DF Values												
DQ1 (Gender)	1	1	1	1	1	1	1	1	1	1	1	1
DQ8 (Mistakes?)	1	1	1	1	1	1	1	1	1	1	1	1
DQ9 (PollWorker?)	1	1	1	1	1	1	1	1	1	1	1	1
DQ3 (Race)	6	6	6	6	6	6	6	6	6	6	6	6
DQ3 Reduced	1	1	1	1	1	1	1	1	1	1	1	1
Ordinal												
DQ2 (age)	0.00	0.14	0.13	0.13	0.12	0.09	0.03	0.00	0.07	0.14	0.09	0.01
DQ5 (education)	-0.07	-0.19	-0.09	-0.03	0.00	0.07	-0.01	0.05	-0.06	-0.09	-0.12	-0.04
DQ6 (ComputerUse)	-0.06	-0.21	-0.08	-0.14	-0.10	-0.06	-0.11	0.05	-0.08	-0.16	-0.12	-0.02
DQ7 (PrevElections)	0.03	0.15	0.00	0.07	0.07	0.02	0.10	0.03	0.05	0.02	0.00	0.00
DQ12 (Income)	-0.06	-0.21	-0.09	-0.10	-0.08	-0.13	-0.09	-0.01	-0.17	-0.13	-0.27	-0.02
P Values												
DQ2 (age)	0.98	0.23	0.06	0.06	0.07	0.20	0.68	0.96	0.29	0.06	0.25	0.93
DQ5 (education)	0.30	0.10	0.17	0.68	0.95	0.33	0.95	0.48	0.36	0.20	0.12	0.60
DQ6 (ComputerUse)	0.40	0.08	0.27	0.03	0.12	0.39	0.17	0.47	0.29	0.03	0.10	0.76
DQ7 (PrevElections)	0.68	0.21	0.95	0.28	0.32	0.78	0.18	0.66	0.48	0.77	0.96	0.99
DQ12 (Income)	0.39	0.11	0.23	0.14	0.29	0.07	0.29	0.94	0.02	0.11	0.00	0.76
N Values												
DQ2 (age)	230	72	219	229	229	222	171	186	200	186	180	219
DQ5 (education)	229	72	217	228	228	221	171	187	199	185	178	218
DQ6 (ComputerUse)	228	69	216	227	227	220	169	184	197	183	177	217
DQ7 (PrevElections)	230	72	218	229	229	222	171	186	199	185	179	220
DQ12 (Income)	199	59	189	197	198	193	146	161	173	161	158	189

Figure 11: Demographics correlation tables.

	Sample		City
	Number	Percent	Percent
Gender			
Male	92	38.33%	47.1%
Female	142	59.17%	52.9%
No Answer	6	2.50%	
Age			
18-24	4	1.67%	11.5%
25-34	5	2.08%	21.3%
35-49	78	32.50%	36%
50-64	92	38.33%	19.7%
65-74	35	14.58%	6%
75+	22	9.17%	5.6%
No Answer	4	1.67%	
Race/Ethnicity			
White	157	65.42%	75.1%
Black	37	15.42%	12.3%
Asian	7	2.92%	3.6%
Hispanic/Latino	10	4.17%	12.5%
Multiracial	2	0.83%	2.4%
Other	6	2.50%	6.5%
Refused	14	5.83%	
No Answer	7	2.92%	
Languages			
English	224	93.33%	87%
Spanish	13	5.42%	14.2%
Other	13	5.42%	10.4%
No Answer	4	1.67%	
Education			
Some High School	5	2.08%	8.4%
High School/GED	7	2.92%	15.5%
Some College	27	11.25%	16.1%
2-year degree	3	1.25%	3.9%
4-year degree	62	25.83%	22.4%
Some Graduate	20	8.33%	N/A
Graduate/professional	111	46.25%	27.2%
No Answer	5	2.08%	
Household Income			
$0-$19,999	16	6.67%	13.98%
$20,000-$39,999	22	9.17%	15.67%
$40,000-$59,999	19	7.92%	15.93%
$60,000-$79,999	29	12.08%	12.82%
$80,000-$99,999	26	10.83%	11.32%
$100,000+	91	37.92%	30.29%
Do not know	7	2.92%	
No Answer	30	12.50%	

Figure 12: Demographics of the sample compared to similar demographics from the US Census.

Gender	
NA	5.457606
Male	5.432483
Female	5.934415
Age	
NA	5.230486
18-24	5.56875
25-34	6.52
35-49	5.678515
50-64	5.838784
65-74	5.643374
75+	5.886848
Race	
NA	5.378509
White	5.873701
Black	5.749088
Asian	3.996296
Hispanic/Latino	6.39127
Multiracial	7
Other	3.777778
Language	
NA	5.300486
English	5.779209
Spanish	6.41358
Other	4.650641
Education	
NA	5.411856
some high school	6.766667
high school	5.285714
some college	5.843101
2-year degree	6.933333
4-year degree	5.645023
Some Grad	5.557059
MS,PhD,MD,JD	5.753706
Computer Usage	
NA	5.455225
never	5.694271
once every two weeks	6.966667
1-3 times per week	5.175325
4-6 times per week	6.215986
7-9 times per week	5.643567
10+ times per week	5.724197

Prev. Elections	
NA	5.380486
0	5.718519
1	5.063889
2	6.042361
3+	5.76415
Mistakes?	
NA	5.355556
Yes	6.404762
No	5.738316
Poll Worker?	
NA	5.372153
Yes	6.089146
No	5.682165
Voting Tech	
NA	5.481424
none	3.08125
paper	5.746735
touch screen	5.791268
punch card	5.933564
lever machine	5.827418
other	5.365
Physical Challenges	
NA	5.348608
none	5.778869
limited eyesight	5.153333
blindness	6.944444
limited hearing	6.454762
deafness	NaN
tremors	6.285714
limited motor control	5.804762
limited mobility	6.57381
other	6.766667
Will not give	7
Income	
NA	5.450617
$0-$19,999	6.16069
$20,000-$39,999	5.685552
$40,000-$59,999	5.79239
$60,000-$79,999	5.834906
$80,000-$99,999	5.948444
$100,000+	5.668878

Figure 13: Mean satisfaction values for each demographic factor. The satisfaction factor is a combination of responses on questionnaire questions 1, 4, 5, 6, 7, 8, 9, 10, 11, and 13.