

Implementation and Evaluation of the EasyVote Tallying Component and Ballot

Jurlind Budurushi
and Melanie Volkamer
Computer Science Department
Technische Universität Darmstadt
Darmstadt, Germany
Email: name.surname@cased.de

Karen Renaud
School of Computing Science
University of Glasgow, UK
Email: karen.renaud@glasgow.ac.uk

Marcel Woide
Psychology Department
Technische Universität Darmstadt
Darmstadt, Germany
Email: marcel.woide@cased.de

Abstract—The German federal constitutional court ruled, in 2009, that elections had to have a public nature. EasyVote, a promising hybrid electronic voting system for conducting elections with complex voting rules and huge ballots, meets this requirement. Two assumptions need to hold, however. The first is that voters will verify the human-readable part of the EasyVote ballot and detect discrepancies. Secondly, that electoral officials will act to verify that the human-readable part of the ballot is identical to the machine-readable part, and that they, too, will detect discrepancies. The first assumption was tested in prior work, so in this paper we examine the viability of the second assumption.

We developed an EasyVote tallying component and conducted a user study to determine whether electoral officials would detect discrepancies. The results of our user study show that our volunteer electoral officials did not detect all of the differences, which challenges the validity of the second assumption.

Based on these findings we proceeded to propose two alternative designs of the EasyVote ballot: (1) In contrast to the original EasyVote ballot, the human-readable part highlights only the voter’s direct selections in orange, i.e. votes that are automatically distributed by selecting a party are not highlighted; (2) The second alternative includes only the voter’s direct selections and highlights them in orange. Both alternatives reduce the number of required manual comparisons and should consequently increase the number of discrepancies detected by election officials. We evaluated both alternatives in an online survey with respect to ease of verification and understandability of the cast vote, i.e. verifying that the human-readable part contained the voter’s selections and understanding the impact (distribution of votes) of the corresponding selections.

The results of the online survey show that both alternatives are significantly better than the original EasyVote ballot with respect to ease of verification and understandability. Furthermore, the first alternative is significantly better than the second with respect to understandability of the cast vote, and no significant difference was found between the alternatives with respect to ease of verification of the cast vote.

I. INTRODUCTION

The German saying “different countries, different customs” holds true for elections, which can be very different between and even within countries. Some elections, like parliamentary elections in Estonia or Germany have very simple voting rules and small ballots. Voters can select 1 out of n -candidates,

where n is a relatively small number between two and 20. Other elections, like parliamentary and European elections in Luxembourg, parliamentary elections in Belgium and local elections in Germany (e.g. Bavaria, Bremen, Hamburg, Hesse), have very complex voting rules and huge ballots. In this paper we focus on the local elections in Hesse, because we were able to access original materials, e.g. ballots, tallying software and training presentations, used in the 2011 elections. In these elections voters can cast up to 93 votes¹ depending on the size of the district; usually more than ten parties and more than 450 candidates participate, which results in huge ballots, nearly the size of an A0² sheet of paper (Size: 27” x 35”). Furthermore, voters can select a party (votes are automatically assigned to the candidates of the selected party according to the list order), and cross out candidates they do not like. They can perform vote splitting (cast votes for candidates of different parties) and cumulative voting (cast up to three votes for each candidate). Such complexity introduces challenges regarding both vote casting and tallying processes. In the vote casting process, voters might unintentionally spoil their vote, due to the complex voting rules. Furthermore, the tallying process is very time intensive and likely to be error prone, because of the combination of complex voting rules and huge ballots.

In order to address these challenges and improve the situation for both voters and poll workers, in particular for local elections in Hesse, Volkamer *et al.* [2] proposed an electronic voting system, called EasyVote. The EasyVote system can be briefly described as follows: 1) Voters prepare their ballots on a voting device, which prints their selections. The printed ballot contains voters’ selections in a human- and machine-readable (a plaintext QR-Code) format. 2) Voters deposit their ballots into the ballot box. 3) Ballots are tallied automatically, by scanning the QR-Codes on the printouts.

Budurushi *et al.* [3] evaluated a number of electronic voting systems with respect to their feasibility for use in elections with complex voting rules and huge ballots. They

¹This number depends on the the number of available seats, which also limits the number of candidates nominated by a party.

²A0 according to [1].

report that, with respect to the *public nature of elections*³ and secrecy legal requirements, the EasyVote system supported the complex local elections in Hesse better than the other systems. Henning *et al.* [4] analysed the EasyVote system from a legal perspective and showed that it complied with German requirements for local elections in Hesse.⁴ Both analyses [3], [4] rely on the following assumptions being true: (1) Voters will act to verify the correctness of the human-readable part of their ballots; (2) Voters will detect discrepancies; (3) Electoral officials will verify that the human-readable matches the machine-readable part (QR-Code); (4) Electoral officials will detect discrepancies. However, before EasyVote can be used in practice, the validity of these assumptions has to be verified. With respect to the first and second assumptions, Budurushi *et al.* [5] showed that the number of voters that verified their printouts and detected discrepancies could be increased significantly if voters were provided with pre-printed, “just-in-time” verification instructions.

Thus, in the first part of this paper we focus our attention on the actions of electoral officials during the tallying process. We implemented a tallying component prototype based on the EasyVote system. The tallying process itself could, in general, be achieved using different techniques: (1) by scanning the printouts with different scanners manufactured by different manufacturers (trust distribution), or (2) by scanning printouts and performing either risk-limiting audits described in [6] and [7], or the Bayesian method described in [8], or (3) by scanning each ballot and comparing the human-readable printout with the details on the screen (generated from the QR-Code). We implemented the latter process, as this complies with the legal requirements [4]. We do not know whether the other techniques are aligned with the public nature of elections, because, to the best of our knowledge, no legal analysis has been conducted yet. Since electoral officials have to scan a large number of individual ballots, one after the other, the accuracy of the process becomes important and therefore should be evaluated. Accuracy is particularly important, because it relies on human attention, which is notoriously unreliable [9], [10]. This is especially the case when the prevalence of the target to be noticed is low [11], [12], when the searcher has to look for multiple different targets at the same time [13] and when the size of the area to be searched is large [14]. All of these are true for the EasyVote ballots so it seems important to test the impact of this well-known human limitation on the checking required during the EasyVote tallying process. Therefore in a user study, we evaluated the *accuracy* of the EasyVote tallying component by intentionally introducing manipulated printouts, i.e. printouts where the human-readable part did not match the machine-readable part (the data stored in the QR-Code).

³This principle was introduced by the Federal Constitutional Court of Germany in 2009, and states that it must be possible for the citizen to verify the essential steps in the election act and in the ascertainment of the results reliably and without special expert knowledge, i.e. each election step must be transparent for the voter.

⁴As the legal evaluation is in German, we outline here the most important conclusions: (1) Voters can verify their vote without any specialist knowledge. (2) Voters are not required to rely on the system’s integrity. (3) The system enables an automatic tally of single votes, and also a full manual tallying of votes, similar to the traditional one. (4) The human-readable part is the deciding factor regarding the tallying process. (5) The system strengthens the principle of the “public nature of elections”, since on the one hand voters can better understand the impact of their selections, and on the other hand the tallying process might be faster and more accurate than the traditional one.

Note that the goal was to evaluate the *accuracy* of the actions of electoral officials during the implemented tallying process, thus we assumed a compromised vote casting component and an honest and correctly implemented EasyVote tallying component. The results of this study show that this way of effecting the tallying in EasyVote is not fully accurate as we rely on human ability to detect differences and our participant “electoral officials” did not detect all the manipulations we introduced during their scanning and verification process. The study also revealed that it will be necessary either to improve the EasyVote system or to relax the legal requirements.

Based on these findings, in the second part of this paper we focused on improving the process and proposed two alternative EasyVote ballot designs: (1) In contrast to the original EasyVote ballot, the human-readable part highlights the voter’s direct selections in orange, i.e. votes that are automatically distributed by selecting a party are not highlighted; (2) The second alternative includes only the voter’s direct selections and highlights them in orange. Both alternatives reduce the number of required manual comparisons and should consequently increase the number of discrepancies detected by the poll workers. We evaluated the alternatives in an online survey with respect to *ease of verification* and *understandability* of the cast vote, i.e. verifying that the human-readable part contains the voter’s selections and understanding the impact (distribution of votes) of the corresponding selections. The results of the online survey show that the alternatives are significantly better than the original EasyVote ballot with respect to ease of verification and understandability of the cast vote. Furthermore, the first alternative is significantly better than the second with respect to understandability of the cast vote, and no significant difference was found between the alternatives with respect to ease of verification of the cast vote.

II. BACKGROUND

We first explain the traditional tallying process in the local Hesse elections. The paper ballots used in the traditional local elections in Hesse are shown and elaborated on in Figure 1. The traditional tallying process in the local elections in Hesse comprises two phases. Both phases are led by an electoral official who gives instructions to other electoral officials and observes the process. In the first phase, at the end of the election day, electoral officials perform the following steps:

- Open the ballot boxes, count the total number of cast ballots and compare it with the total number of marked voters in the electoral register.
- Divide the ballots into four categories: 1) Only party header is marked 2) Candidates and/or a party header are marked 3) Invalid 4) Not assignable to 1), 2) or 3).
- Check that ballots are assigned to the correct category.
- Divide and count the 1st category by parties (first intermediate result).
- Discuss and assign each single ballot of the 4th to the 1st, 2nd or 3rd category.
- Manually recompute the intermediate election result, based on the 1st and 3rd category.

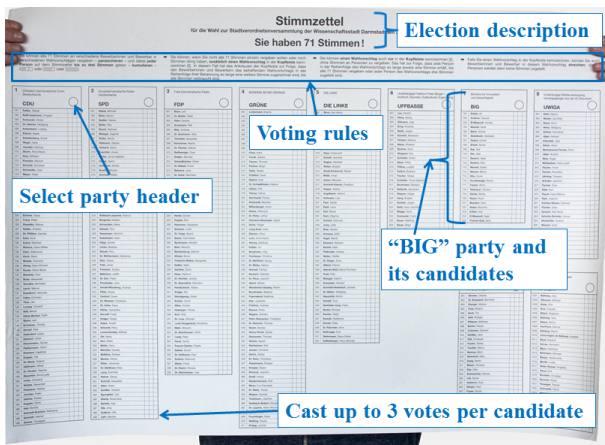


Fig. 1: Paper ballot of the local elections in Hesse in 2011. (Size: 27" x 35")

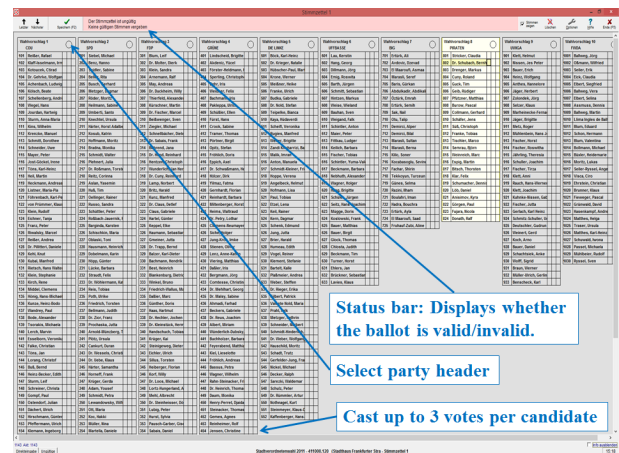


Fig. 2: Ballot entering and recording interface of PC-Wahl.

The second phase of the tallying process takes place the day after the election. This phase is supported electronically by special purpose software. The software used by traditional local elections in Hesse is called PC-Wahl.⁵ During this phase only ballots from the 2nd category, i.e. ballots that contain marked candidates and/or a party header, are tallied. Electoral officials perform the following steps:

III. IMPLEMENTATION

In this section we introduce and describe the different steps of the implemented EasyVote tallying process. The EasyVote ballots that need to be tallied are shown in Figure 3. Afterwards, we present the interfaces of the implemented prototype.

- Electoral officials enter the intermediate result from the first phase.
- First five ballots are entered and recorded into the PC-Wahl interface (Figure 2).
- Manually tally the first five ballots.
- Compare the electronic result with the manual result.⁶
- Enter and record the rest of the ballots into the corresponding PC-Wahl interface.
- Electronically compute the final election result, and sign the printed disposition.

The process of entering and recording ballots via the corresponding PC-Wahl interface is performed by three electoral officials. One electoral official narrates the marks from the ballot and a second enters them into the PC-Wahl interface. A third electoral official verifies that the first and second electoral officials have performed this correctly.

Note that electoral officials who participate in the second phase of the tallying process are employees of the corresponding electoral office and/or municipality. Hence, they have relatively high technical expertise. Furthermore, they participate in a theoretical training workshop regarding the PC-Wahl software. The workshop lasts approximately 30 minutes, and electoral official can practice if they wish to, in order to ensure their competence.

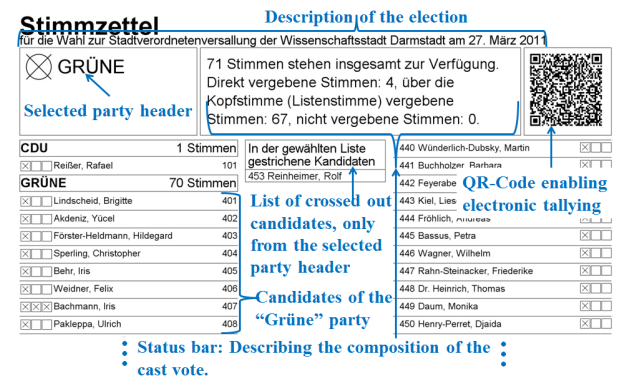


Fig. 3: The EasyVote paper ballot.

A. Tallying Process

The implemented EasyVote tallying process comprises the following steps: (1) Open the ballot boxes, count the total number of cast ballots and compare it with the total number of marked voters in the electoral register. (2) Scan each individual ballot. (3) Electronically compute the final election result, and sign the printed disposition.

Since the EasyVote ballots are electronically prepared and printed in a pre-defined layout, format and font, the ballots could feasibly be scanned by using Optical Character Recognition (OCR) scanners. However, for scanning each individual ballot we decided to use QR-Codes scanners, as originally proposed by Volkamer *et al.* [2], based on the following general advantages of QR-Code scanners:

- QR-Code scanners provide a much higher error correction level and therefore are more accurate.

⁵<http://www.pcwahl.de/>.

⁶This check only serves as a self-control for electoral officials, rather than checking the correctness of PC-Wahl.

- QR-Code scanners can be used for all type of ballots (universal encoding), while OCR scanners need to be configured and maintained for each type of ballot.

Hence, the process of scanning and counting an individual ballot, shown in Figure 4, consists of the following steps: (1) Pick up a ballot. (2) Scan the QR-Code. (3) Verify and confirm that the scanned information matches the human-readable part of the ballot. (4) Repeat process with the next ballot.

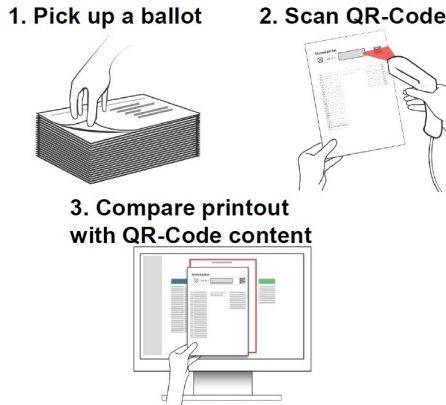


Fig. 4: Scanning and counting ballots with EasyVote.

Note that if we used OCR scanners the human-readable part is also the machine-readable part. This prevents the vote casting component from manipulating the machine-readable part, because voters would be able to detect the manipulation. However, in order to ensure the correctness of the scanning/counting process, electoral officials are still required to fully verify/examine the scanned ballot against the printout (EasyVote ballot). If we assume that electoral officials are required to detect *all* possible discrepancies, it makes no difference whether these are introduced by the vote casting or tallying components.

B. Interfaces of the Prototype

The EasyVote tallying component proposed by Volkamer *et al.* [2] uses two monitors (two different interfaces) for the tallying process. The first monitor, presented in step three on Figure 4, displays and enables the verification of each individual scanned ballot. The second monitor displays intermediate election results after scanning, verifying and confirming each individual ballot. This enables electoral officials and the general public to verify that each individual ballot is correctly added to the election result.

Figure 5 presents the implemented interface for the first monitor, while Figure 6 presents the implemented interface for the second monitor.

IV. USER STUDY - ACCURACY EVALUATION

In this section we describe the user study, in which we evaluated the prototype with respect to accuracy. The goal of the study was to find out if the implemented EasyVote tallying component is 100% accurate, i.e. that discrepancies where the QR-Code does not match the human-readable part

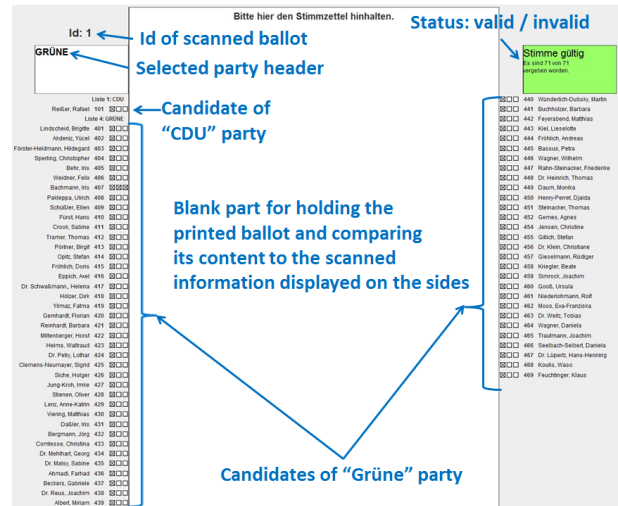


Fig. 5: Scanning and verifying the content of the current ballot.

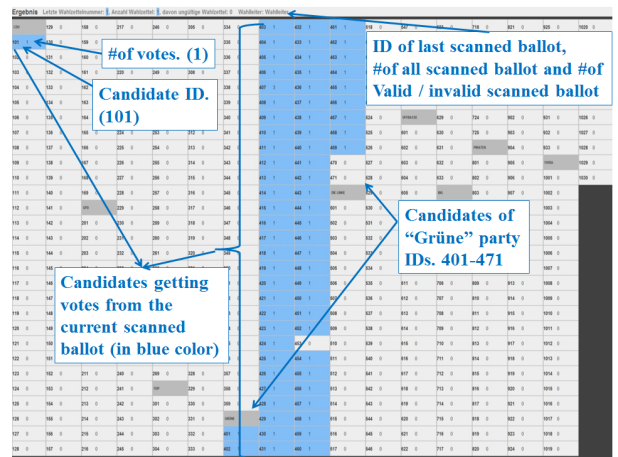


Fig. 6: Overview on the intermediate election result.

can always (in any case and by any participant) be detected. We intentionally introduced manipulated printouts, in order to check if participants detected the discrepancies.

A. Preliminary Considerations and Materials

In the user study we only focused on the process of scanning an individual ballot and verifying that the human-readable part matches the machine-readable part. Although by verifying intermediate results we might also be able to detect discrepancies, we assume that if participants cannot detect all discrepancies during the scanning and verifying process, they will also not detect further discrepancies while verifying intermediate results. Thus, for this study we assumed a compromised vote casting component and, an honest and correctly implemented EasyVote tallying component. Note that in practice the tallying component is not assumed trustworthy, as different mechanisms can be used to detect a malicious tallying component, for instance the tallying component provides a cryptographic commitment after each scanned ballot or a hash chain, or by videotaping both monitors at the same

time. Afterwards, random checks can be performed to ensure the correctness of the election result.

Furthermore, one of the most well-known challenges in the area of usable security is that you cannot communicate the primary goal of the study to participants without biasing them [15]. If participants know the primary goal of the study, they may act in a manner perceived as appropriate, and change their behaviour [16]. Therefore we told all participants in the user study that the goal was to evaluate the usability of the EasyVote tallying component. This was necessary so that the participants would not be biased in their behaviour.

The materials required to conduct the user study are listed here. For the materials from the local elections in Hesse we collaborated with the local authorities.

- Training workshop presentations for the PC-Wahl software.
- 189 original electronically filled in ballots from the local elections in Hesse 2011. They were split as follows: 94 from the 1st, 89 from the 2nd and 6 from the 3rd category.⁷
- The implemented EasyVote tallying component.
- Training workshop presentations for the EasyVote system. We created these presentations based on those for the PC-Wahl software.
- 189 EasyVote ballots. These ballots were electronically created, and duplicated the 189 traditional ballots.
- Five EasyVote test ballots to be used during the training phase: Three ballots with candidates and party header marked, and two ballots that also contained crossed out candidates. Two of the five ballots required corresponding corrections by the participants.

B. Study Design

In order to evaluate the accuracy of the implemented EasyVote tallying component we manipulated the QR-Codes of the EasyVote ballots. Hence, when scanning the QR-Code of a manipulated ballot participants should detect a discrepancy between the EasyVote ballot and the data displayed on the screen. As we do not aim to change, but rather to improve the tallying process for local elections in Hesse, participants were required to tally only ballots of the 2nd category, i.e. a total of 89 ballots that contain votes assigned to candidates and/or a selected party header.

C. Manipulations: Introducing Discrepancies

While manipulating the QR-Codes of the EasyVote ballots is technically trivial, we first had to solve the following challenges: 1) Identify all possible manipulations that lead to a difference between the printed human-readable part on the ballot and the data displayed on the monitor; 2) Select an adequate set of manipulations; 3) Introduce an adequate number of manipulations, in order to not directly reveal the study goal; 4) Decide how to randomly add manipulations to

ballots; 5) Decide how to introduce the manipulations into the ballot set randomly.

By performing a systematic analysis we identified 36 possible manipulations that we classified in the following five manipulation categories: 1) Changing only vote distribution (7 manipulations); 2) Change candidate names (14 manipulations); 3) Changing party, including its candidates (11 manipulations); 4) Invalidating a valid ballot (2 manipulations); 5) Validating an invalid ballot (1 manipulation).

In order to select a reasonable set of manipulations, we defined the following criteria: 1) Detecting the manipulation requires a full and careful comparison of the EasyVote ballot and monitor; 2) Manipulation should be hard to detect. This led us to the following adequate manipulation set:

- Remove votes from a candidate and assign them to another candidate (1st manipulation category).
- Remove votes from a candidate and do not re-assign them (1st manipulation category).
- Remove a candidate and insert another candidate instead (2nd manipulation category).
- Remove a candidate (2nd manipulation category).
- Remove a party, including its candidates (3rd manipulation category)

This set also covers the manipulations used in previous studies, refer to [17] and [18].

Furthermore, since we were restricted by the number of ballots used in this study we manipulated only 5 out of the 89 ballots. In this way we covered all manipulation categories and introduced a reasonable number of manipulations relative to the number of ballots, such that participants would not guess the primary study goal. We randomly selected 5 ballots and introduced the manipulations according to a random permutation. Finally, we randomly introduced the manipulated ballots into the set of all ballots. Note that each group was confronted with the same manipulations, but in a different random order.

D. Experimental Design and Procedure

11 participants were randomly allocated to four different groups. Three groups consisted of three participants, and one group of two. Each group had to perform the following steps:

- Read and sign the agreement form for participating to the study.
- Participate in the training workshop.
- Tally the 2nd category ballots with the implemented prototype.
- Debrief.

Furthermore, we randomly assigned participants of a group the following tasks: 1) Scanning (one participant had to scan the ballot); 2) Verifying (two participants had to verify that the human-readable part matches the machine-readable part). As the last group consisted only of two participants, one of the participants was randomly assigned to perform both tasks.

⁷Refer to section II for the description of the different categories.

Note that the EasyVote tallying process proposed by Volkamer *et al.* [2] requires only two electoral officials. However, we used the same setting as in the traditional local elections in Hesse, thus assigning three instead of two participants (electoral officials) to each group. The last group consisted only of two participants, because one of them did not show up.

E. Experimental Setup and Ethical Considerations

All experiments took place in our department. The venue was equipped with tables, chairs and a projector. The projector was used during the presentations in the training workshops. All groups were provided with the necessary hardware equipment, monitor(s), a computer on which the tallying software was installed, and a printer.

An ethics commission at our university provides ethical requirements for research involving humans. These requirements were met. All participants were told that all data would be stored anonymously and used only for the purposes of the experiment.

F. Recruiting and Sampling

The participants were recruited via e-mail, advertising in social networks and flyers. The experiment had 11 randomly selected participants (6 female, 5 male), age 19-57 years: 7 students from different subject areas and 4 employees of our university. All participants were naïve, with respect to the content, since none had worked as an electoral official before. Three different incentives encouraged participation: First, the employees of our university were interested in science and wanted to support our research. Second, 3 were psychology students, who are required by their department to participate in 30 hours of research studies. We compensated them with the appropriate amount of hours. For the rest of the participants we provided €10 per participant.

It is important to note that most of the participants were university students who are very familiar with technology. While they may not be representative of the larger “electoral officials” population, they probably serve a best-case scenario for what tallying performance could be.

G. Results

In this section we report the results regarding the dependent variable “detected” that reflects the accuracy of the implemented EasyVote tallying component. Table I summarises the results of the study. “True” means that the discrepancy was detected and corrected by the participants, while “False” means that the discrepancy was not detected.

TABLE I: Summary of the accuracy evaluation.

Manipulation categories [*]	Group 1 / Position	Group 2 / Position	Group 3 / Position	Group 4 ^{**} / Position
1	False / 1	True / 34	True / 6	True / 59
2	True / 83	False / 75	True / 68	True / 8
3	True / 51	False / 36	True / 88	False / 89
4	False / 9	True / 67	True / 25	False / 3
5	True / 87	True / 46	False / 54	True / 36

* Refer to section II for the description of the different categories.

** This group consisted only of two participants.

The results of the accuracy evaluation show that none of the groups detected all introduced discrepancies. Furthermore, the results indicate that detecting a discrepancy does not depend on the position, or on whether others have previously been detected, or on the specific manipulation category.

Note that due to these results, which already show that the implemented EasyVote tallying component does not achieve 100% accuracy, we decided not to continue the user study, i.e. not to include further groups (participants) enabling us to achieve an adequate sample size that would allow to perform various statistical tests.

V. ONLINE SURVEY - EASYVOTE BALLOT DESIGN

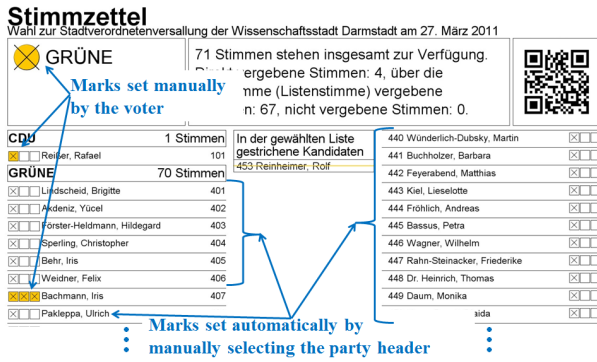
In this section we describe our online survey and present the results. This survey is motivated by the results of the user study reported in the first part of the paper. Hence, the goal was to identify an alternative EasyVote ballot design. On the one hand it ought to reduce the number of required manual comparisons and consequently increase the number of discrepancies detected by poll workers. On the other hand it enables voters easily to verify their cast vote. We also report on recruitment and sampling of participants.

A. Alternative EasyVote Ballots

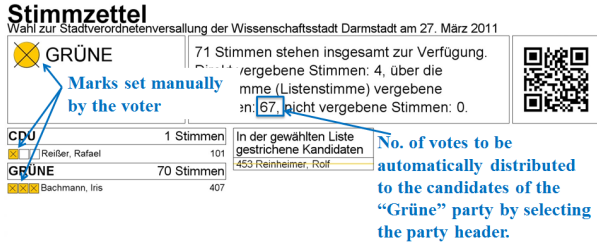
In the survey we presented participants with two possible EasyVote ballot designs (see Figure 7). In contrast to the original EasyVote ballot, both alternatives introduce colour as a new dimension. According to Braun and Silver [19], the colour red conveys the highest level of perceived hazard followed by orange, black, green and blue. Furthermore, Young and Wogalter [20] found that with respect to memory times print highlighted with orange was better remembered than non-highlighted text. Moreover, since red is problematic for a significant percentage of the male population due to colour blindness, orange seemed the best choice.

The first alternative, in contrast to the original EasyVote ballot, highlights the voter’s manual selections in orange. The second alternative simplifies things even further, since it eliminates everything except the voter’s manual selections and these are still highlighted in orange. Hence, automatically distributed votes, i.e. remaining votes that are assigned to the candidates of a party by selecting the party header, are not printed. The size of the printout remains the same, independent of the voter’s selections.

Furthermore, in contrast to the original EasyVote ballot, the machine-readable part (QR-Code) encodes only the voter’s manual selections. Thus, the “adapted” EasyVote tallying component implements the algorithm to automatically distribute votes independently of the voter’s manual selections, rather than only relying on the data stored in the QR-Code. Both alternatives reduce the number of required manual comparisons for both voters and electoral officials. However, in order to ensure the correctness of the election result, we suggest that electoral officials check the automatic distribution of votes for a random set of ballots, i.e. verify the complete ballot displayed/interpreted by the tallying component, rather than only voter’s manual selections.



(a) The first alternative.



(b) The second alternative.

Fig. 7: The alternative EasyVote ballot designs

B. Design and Procedure

The survey consisted of four parts and was structured as follows: (1) Participants were introduced to the local elections in Hesse. They were asked whether they had previously cast a vote in local Hesse or similar elections, and how often they participated in local elections. (2) Participants were told how many invalid votes were cast in the local elections in Hesse in 2011. This percentage, (5.5%)⁸ was much higher than the German federal elections in 2013 (on average 2.7%)⁹. Then they were introduced to the EasyVote vote casting process. (3) They were asked some general questions to assess the comprehensibility of the EasyVote vote casting process. (4) Participants were given a textual description of a cast vote, and confronted with the original and the two alternative ballots. All reflected the cast vote described in the text. Participants were asked to rank the ballot types (original and alternatives) with respect to ease of verification and understandability of the cast vote, i.e. verifying that the human-readable part contains the voters selections and understanding the impact (distribution of votes) of the corresponding selections. We also collected some demographic data (nationality, age, gender and education).

C. Recruiting and Sampling

The participants were recruited via e-mail, advertising in social networks, flyers and by personal contact. 87 subjects participated (35 female, 48 male, 4 others) between the ages of 19-75 years. We removed 14 participants (3 female, 9

male, 2 others) aged 22-75, because they did not answer all questions with respect to the vote casting process with the EasyVote voting system. The remaining 73 subjects (32 female, 39 male, 2 others) aged 19-65 comprised one participant with apprenticeship, four with a Ph.D. degree, five with middle school qualification, seven with a B.Sc. degree, seven with a technical college qualification, eight with a vocational education, 15 with a Diploma/M.Sc. degree and 26 with a high school qualification. Most (63) were Germans, four were Austrians, 2 were Turkish, one Swiss and one did not provide information about nationality. No incentives were provided, thus participation was purely voluntary.

D. Results

Table II summarises the results with respect to understandability of cast vote and Table III with respect to ease of verification.

TABLE II: Understandability of cast vote.

EasyVote Ballot	Times of ranking		
	First place	Second place	Third place
Original	5	27	41
First alternative	41	30	2
Second alternative	27	16	30

TABLE III: Ease of verification of cast vote.

EasyVote Ballot	Times of ranking		
	First place	Second place	Third place
Original	6	18	49
First alternative	32	40	1
Second alternative	35	15	24

In order to measure the difference between the original and the alternative designs of the EasyVote ballot we used the Wilcoxon non-parametric test. The test shows a significant difference between the first alternative and the original EasyVote ballot with respect to understandability, $Z=-6.722$; $p < 0.01$ and ease of verification, $Z=-6.722$; $p < 0.01$. A significant difference is also found between the second alternative and the original EasyVote ballot with respect to understandability, $Z=-2.891$; $p < 0.01$ and ease of verification, $Z=-4.205$; $p < 0.01$. Additionally, the first and second alternatives differ significantly regarding understandability, $Z=-3.673$; $p < 0.01$ with a higher rank sum for the first alternative (1993.50). No significant difference was found between both alternatives regarding ease of verification.

Furthermore, we evaluated participants' statements, on a five-point Likert scale, concerning the advantages of the EasyVote system compared to the traditional elections in Hesse. Approximately 92% of the participants agreed or fully agreed that the EasyVote system would support voters in such complex elections, such as the local elections in Hesse. 64% of the participants would be happy to use the EasyVote system at the next local elections in Hesse. Around 80% of the participants recognised or fully recognised the advantages of the EasyVote system compared to traditional local elections in Hesse, and think that the EasyVote system is a first step in the right direction to introduce technology in the context of legally-binding elections. Only one participant did not perceive any advantages with respect to using the EasyVote system.

⁸<http://www.statistik-hessen.de/K2011/EK1.htm>, last accessed 10.08.2014 (in German).

⁹http://www.bundeswahlleiter.de/en/bundestagswahlen/BTW_BUND_13/ergebnisse/landesergebnisse/106/, last accessed 10.08.2014.

VI. CONCLUSION AND FUTURE WORK

The focus of our research is on electronic voting systems for elections with complex voting rules and huge ballots that meet the German constitutional requirements, including the principle of the public nature of elections. This principle requires that voters should be able to verify all essential steps of the election without technical knowledge. Therefore, in this paper we considered the EasyVote [2] hybrid voting system, which is supposed to meet those requirements. Because of the public nature of elections, we focused on the tallying process in which ballots are scanned individually and each ballot is verified as correct before being tallied.

In the first part of this paper, we reported the results of a user study carried out to evaluate the accuracy of the implemented EasyVote tallying process. The main finding is that the implemented tallying process cannot guarantee a 100% accurate election result since participants did not notice all manipulations. Such human errors could be avoided by automatically scanning all EasyVote ballots, i.e. implementing a different tallying process. Furthermore, trust could be increased either by risk-limiting audit techniques or by using several independent scanners/tallying components. However, this would decrease the extent to which the public nature principle is implemented. This result shows that just because a voting system meets the public nature requirement it does not mean that discrepancies are detected or that underlying fraud is necessarily revealed.

In the second part we reported the results of an online survey, which evaluated two alternative EasyVote ballots designs. Both alternatives were shown to reduce the number of manual comparisons required and can be expected to increase the number of discrepancies detected by the election officials. The results of the online survey show that the first alternative design, where voters' manual selections are additionally highlighted in orange, differs significantly with the original EasyVote ballot with respect to understandability and ease of verification of the cast vote. Furthermore, the first and second alternatives differ significantly regarding understandability. No significant difference was found between the alternatives with respect to ease of verification.

Thus, for future interdisciplinary research we will study the reliability of mechanisms which comply with the principle of the public nature of elections. We plan to repeat the user study with the new EasyVote ballot design (first alternative), and also to propose different techniques to improve detection accuracy. Another open research question is to discover what an acceptable rate of errors is, if indeed we have to accept that some errors will remain undetected.

ACKNOWLEDGMENT

This paper has been developed within the project 'VerkonWa' - Verfassungskonforme Umsetzung von elektronischen Wahlen - which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Science Foundation), and the project 'BoRoVo' Board Room Voting - which is funded by the German Federal Ministry of Education and Research (BMBF) under grant no. 01IS12054. The authors assume responsibility for the content.

REFERENCES

- [1] International Organization For Standardization, *ISO 216:2007: Writing paper and certain classes of printed matter – Trimmed sizes – A and B series, and indication of machine direction*. ISO, 2007.
- [2] M. Volkamer, J. Budurushi, and D. Demirel, "Vote casting device with VV-SV-PAT for elections with complicated ballot papers," in *International Workshop on Requirements Engineering for Electronic Voting Systems*. Proceedings of the IEEE, 2011, pp. 1–8.
- [3] J. Budurushi and M. Volkamer, "Feasibility analysis of various electronic voting systems for complex elections," in *International Conference for E-Democracy and Open Government 2014*, May 2014.
- [4] M. Henning, M. Volkamer, and J. Budurushi, "Elektronische Kandidatenauswahl und automatisierte Stimmermittlung am Beispiel hessischer Kommunalwahlen," *Die Öffentliche Verwaltung (DÖV)*, no. 20, October 2012.
- [5] J. Budurushi, M. Woide, and M. Volkamer, "Introducing precautionary behavior by temporal diversion of voter attention from casting to verifying their vote," in *Workshop on Usable Security (USEC)*, Feb. 2014.
- [6] M. Lindeman and P. B. Stark, "A gentle introduction to risk-limiting audits," *IEEE Security and Privacy*, vol. 10, no. 5, p. 42, 2012.
- [7] M. Lindeman, P. S. B., and V. Yates, "Bravo: Ballot-polling risk-limiting audits to verify outcomes," in *Electronic Voting Workshop/Workshop on Trustworthy Elections (EVT/WOTE'12)*. Bellevue, WA: USENIX, 6-7 August 2012.
- [8] R. Rivest and E. Shen, "A Bayesian method for auditing elections," in *Proceedings of the 2012 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE'12)*. Bellevue, WA: USENIX, 6-7 August 2012.
- [9] D. J. Madden and S. R. Mitroff, "Aging and top-down attentional control in visual search," 2010, institute for Homeland Security Solutions Research Brief. <https://www.ihssnc.org>.
- [10] J. M. Wolfe, T. S. Horowitz, and N. M. Kenner, "Cognitive psychology: Rare items often missed in visual searches," *Nature*, vol. 435, no. 7041, p. 439440, 2005.
- [11] A. N. Rich, M. A. Kunar, M. J. Van Wert, B. Hidalgo-Sotelo, T. S. Horowitz, and J. M. Wolfe, "Why do we miss rare targets? Exploring the boundaries of the low prevalence effect," *Journal of Vision*, vol. 8, no. 15, p. 15, 2008.
- [12] J. M. Wolfe, T. S. Horowitz, M. J. Van Wert, N. M. Kenner, S. S. Place, and N. Kibbi, "Low target prevalence is a stubborn source of errors in visual search tasks." *Journal of Experimental Psychology: General*, vol. 136, no. 4, p. 623, 2007.
- [13] T. Menneer, K. R. Cave, and N. Donnelly, "The cost of search for multiple targets: Effects of practice and target similarity," *Journal of Experimental Psychology: Applied*, vol. 15, no. 2, pp. 125–139, 2009.
- [14] B. Zenger and M. Fahle, "Missed targets are more frequent than false alarms: A model for error rates in visual search." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 23, no. 6, p. 1783, 1997.
- [15] C. Kuo, A. Perrig, and J. Walker, "Designing an evaluation method for security user interfaces: lessons from studying secure wireless network configuration," *Interactions*, no. 3, pp. 28–31, 2006.
- [16] A. Sotirakopoulos, K. Hawkey, and K. Beznosov, "'I did it because I trusted you': Challenges with the study environment biasing participant behaviours," in *SOUPS Usable Security Experiment Reports (USER) Workshop*, Microsoft in Redmond, WA, July 14-16 2010.
- [17] S. P. Everett, *The usability of electronic voting machines and how votes can be changed without detection*, Std., 2007, doctoral dissertation, Rice University, Houston, TX.
- [18] B. A. Campbell and M. D. Byrne, "Now Do Voters Notice Review Screen Anomalies? A Look at Voting System Usability," in *Proceedings of the 2009 Conference on Electronic Voting Technology/Workshop on Trustworthy Elections*, ser. EVT/WOTE'09. Berkeley, CA, USA: USENIX Association, 2009, pp. 1–1. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1855491.1855492>
- [19] C. C. Braun and N. C. Silver, "Interaction of signal word and colour on warning labels: differences in perceived hazard and behavioural compliance," *Ergonomics*, vol. 38, no. 11, pp. 2207–2220, 1995.

- [20] S. L. Young and M. S. Wogalter, "Comprehension and memory of instruction manual warnings: Conspicuous print and pictorial icons," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 32, no. 6, pp. 637-649, 1990.