Why Johnny Checks but Doesn’t Alert: Reporting as the Missing Step in Verifiable Internet Voting

Tobias Hilt
SECUSO
Karlsruhe Institute of Technology
Karlsruhe, Germany
tobias.hilt@kit.edu

Christian Mack
SECUSO
Karlsruhe Institute of Technology
Karlsruhe, Germany
christian.mack@kit.edu

Benjamin Maximilian Berens
SECUSO
Karlsruhe Institute of Technology
Karlsruhe, Germany
benjamin.berens@kit.edu

Melanie Volkamer
SECUSO, Karlsruhe Institute of Technology
Karlsruhe, Germany
melanie.volkamer@kit.edu

Abstract

End-to-end verifiable Internet voting promises that voters can remotely check whether their ballot was recorded correctly and that all ballots were tallied as cast. However, in order to achieve an adequate level of security, voters actually need to perform the first check. Our research focuses on the cast-then-audit approach for this check. We use related work to improve this approach in particular by providing a step-by-step guide. We conducted a deceptive online user study ($N = 437$) to compare our improved system with a baseline version from an actual election. We also measured the usability and participants confidence in using such systems. Our findings show that participants from the improved system perform significantly better than the baseline w.r.t. manipulation detecting and reporting capabilities. Furthermore, we show that it is important to distinguish between detection and reporting to understand how to further increase the overall security.

CCS Concepts

• **Security and privacy → Usability in security and privacy**; Social aspects of security and privacy; • **Applied computing →** Voting / election technologies.

Keywords

Security, Voting, Verifiability, Usability Study, cast-then-audit

ACM Reference Format:
Tobias Hilt, Christian Mack, Benjamin Maximilian Berens, and Melanie Volkamer. 2026. Why Johnny Checks but Doesn’t Alert: Reporting as the Missing Step in Verifiable Internet Voting. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI ’26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3772318.3791035

CHI ’26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791035

1 Introduction

Elections are a central element of democratic societies, and an increasing number of election organisers are considering the introduction of an internet voting channel. Within the internet voting community, there is broad consensus that any internet voting system must provide *end-to-end (E2E) verifiability*, since typical physical safeguards, such as election observers or paper-based auditing, are absent in remote voting contexts. E2E verifiability [3] ensures that voters can check whether their own ballot was recorded correctly (*individual verifiability*) and that all recorded ballots are tallied correctly (*universal verifiability*). Crucially, it also enables voters to detect whether the voting device behaved honestly, an especially relevant property because internet voting takes place on general-purpose machines outside governmental control and therefore susceptible to compromise [13, 17, 42].

The central challenge of individual verifiability lies not in the cryptography, but in the human factors: voters must be willing and able to carry out verification, detect anomalies, and, most importantly, report them.

Prior HCI and usable-security research has started to explore these issues. However, the research did not differentiate between voters who merely *noticed* anomalies and those who actually *reported* them: two distinct behaviours that should be considered separately, as only *reported* manipulations can be acted upon by election officials. In addition, prior work has paid little attention to how voters’ *perceptions of trustworthiness and risk* evolve when faced with manipulations. There are different approaches for archiving individual verifiability in internet voting, from which the so-called *cast-then-audit* approach was examined the least.

We identify four systemic gaps in prior work (see Section 2): (1) the absence of a clear separation between *detecting* and *reporting* manipulations, (2) little assessment of how perceived trust and risks evolve during manipulations, (3) lack of a realistic *baseline* reflecting how voters encounter verification with the cast-then-audit approach in practice, and (4) no structured guidance to support voters in navigating critical election processes in the cast-then-audit approach.
In this paper, we address these gaps with four contributions:

- **External reporting channel:** We separate detection from reporting by introducing an independent mechanism outside the potentially compromised voting system.
- **Trust and risk dynamics:** We assess perceived trustworthiness and risks throughout the study, examining how manipulations and subsequent debriefing affect voter confidence.
- **Realistic baseline:** We recreate the interfaces and voter materials of a real cast-then-audit election, establishing a representative benchmark for detection and reporting.
- **Step-by-step guide:** Building on insights from usable security on structured instructions and decision aids, we design and evaluate a voter guide that supports voters throughout the complete election process, from casting to verifying and potentially reporting.

We examine these contributions in a large-scale online user study ($N = 437$) featuring two simulated elections: an unaltered referendum and a manipulated party vote. Beyond detection and reporting, we also examine how perceived trustworthiness and risk evolve, shedding light on whether manipulations erode voter confidence or whether clear guidance and reporting channels sustain it. Our results show that the guide significantly improves detection of verification prevention attacks and increases reporting rates across manipulation types.

2 Background and Related Work

2.1 E2E verifiable Internet Voting and Individual Verifiability

E2E verifiability means that the entire election process, including all its steps, can be checked and verified independently of the system providers [32]. E2E verifiability is usually divided into two interrelated components: individual verifiability and universal verifiability. Individual verifiability ensures that every voter can check whether their own ballot left the voting device as they intended (cast-as-intended) and whether the ballot was received and stored by the voting server without alteration (recorded-as-cast). Universal verifiability, in contrast, describes the ability of anyone to check that all ballots stored in the ballot box were tallied correctly (tallied-as-recorded). Our focus is on individual verifiability. Several different approaches have been developed to achieve individual verifiability. These can be grouped into the following main concepts:

Audit-or-Cast. Often referred to as the Benaloh challenge, audit-or-cast schemes let a voter decide, after encryption, whether to *challenge* the produced ciphertext (have it opened and checked) or to *cast* it as their ballot. Because the device cannot predict which encryptions will later be challenged, systematic manipulation becomes detectable with high probability: any attempt to alter choices risks being revealed when a challenged instance is decrypted and shown to be incorrect [1, 2, 20, 29].

Return codes. Return-code systems provide each voter with personalised codes (delivered over an out-of-band channel such as postal mail) that correspond to their available selections. After the vote is formed on the device, the system displays the codes it believes match the voter's chosen options; the voter compares these on-screen codes with the pre-distributed reference sheet. If the codes match, the voter gains assurance that the encrypted ballot reflects their intent and was not altered by a malicious client [12, 14, 15]. This method has been used in Swiss cantonal e-voting trials (e.g., 2011 National Council) [39, 50].

Tracking Codes. Tracking-code approaches (e.g., Hyperion [7], Selene [41]) assign each voter a unique tracking code that is either generated by the system or created by the voter themselves. This tracking code is stored together with the encrypted ballot. After tallying, all ballots are published alongside their tracking codes on a public bulletin board. Voters can then verify whether their specific tracking code corresponds to the choice they originally intended, thereby ensuring inclusion and correctness [7, 9, 41].

Cast-then-Audit. Cast-then-audit uses two independent devices: a *primary* device to prepare and cast the encrypted ballot, and a *secondary* device to verify that the stored ballot encodes the voter's intended selections [16, 32, 37, 44]. After casting, the system presents cryptographic verification data—commonly as a QR code—which the secondary device scans to retrieve the ballot and reconstruct a human-readable representation (e.g., an image of the stored ballot) for the voter to confirm. Estonia introduced cast-then-audit verification nationwide in 2011 as part of its binding internet voting system [16, 44], and Germany employed it for the first time in a major binding election during the 2023 "Sozialwahl", the country's third-largest election with over 22 million eligible voters [19, 43].

While these approaches provide different ways of achieving individual verifiability, their effectiveness in practice depends on whether voters actually use the verification steps notice and report manipulations. Moreover, even though cast–then–audit explicitly requires two devices, the need for a secondary device is not unique to this approach. Even in approaches where verification can theoretically be performed on the same device (e.g., the return-code based approach, cast-or-audit approach or tracking-code based approach), a second, independent device becomes necessary once a voter detects an anomaly and wants to report it securely. From a security perspective, the device used for casting a vote should not also be trusted for verifying or reporting it, as this would concentrate trust in a single potentially compromised device. Thus, the requirement for a second device is a broader property of individual verifiability in remote electronic voting rather than a limitation of cast–then–audit specifically.

2.2 Manipulation Studies

Research on individual verifiability has often investigated voters' ability to detect manipulations during the voting or verification process. Two main categories of manipulations are typically distinguished [18, 24]: **vote tampering** manipulations and **Verification Prevention** manipulations. In **Vote Tampering**, the system alters the content of the ballot so that the stored version no longer matches the voter's original choice. In **Verification Prevention** attack, the process is disrupted in such a way that voters are unable to complete verification, for instance by withholding the information needed to confirm their ballot. Across manipulation studies, how a manipulation was detected was measured in various ways. In one study, participants had to call a hotline and describe their problem [47], in another they clicked a dedicated "Yes/No" button [31], and in others anomalies were only inferred from free-text survey responses [33]. The majority of empirical work on manipulation studies was focused on return-code approaches (e.g. [20, 24, 30, 31, 47, 49]). In contrast, the cast-then-audit approach remains comparatively underexplored, with only two major user studies to date explicitly examining voters' ability to notice manipulations. Marky et al. [31] conducted a lab study comparing five cast-as-intended mechanisms, including cast-then-audit. Their design introduced a vote tampering manipulation and asked participants within the verifier interface whether the displayed ballot matched their intent. Clicking "No" was treated as reporting, thereby conflating detection and reporting, since voters could not independently initiate a report. Moreover, the interfaces were author-created mockups rather than reproductions of real election systems, limiting ecological validity. Under these conditions, 64% of participants noticed the manipulation. Hilt et al. [18] performed an online study using the commercial POLYAS system, examining both vote tampering and verification prevention. Detection differed sharply: 96% for vote tampering but only 24% for verification prevention. As in Marky et al. [31], detection and reporting were not cleanly separated: detection was inferred from survey responses or use of a reporting channel, but the number of active reports was not disclosed. Participants were also nudged to verify, and only those who had completed verification in a prior (unmanipulated) trial were included in the manipulated condition, likely inflating detection rates. Although based on a real voting platform, the system setup did not mirror materials used in a binding election, and thus no realistic baseline was established. Taken together, prior studies highlight the difficulty of measuring manipulation detection reliably. Detection and reporting are often intertwined, verification is frequently presupposed despite low real-world verification rates (e.g., 10% in Estonia [48] and 25% in the German GI election [11]), and realistic baselines based on actual election materials are largely missing. These limitations restrict our understanding of how voters interact with cast-then-audit systems under adversarial conditions.

2.3 Detecting versus Reporting

Much of the verifiable e-voting literature focuses on enabling voters to *detect* manipulation (e.g., via verification codes or receipts), but far less attention is given to the equally critical step of *reporting* it. Detection alone is insufficient: unless voters escalate anomalies through a reliable channel, election officials cannot respond. Research in usable security consistently shows a substantial gap between noticing and reporting. Phishing websites, for example, are visited an average of 27 times before anyone reports them [34]. Fraud and scams are similarly underreported, with as few as 13% of UK victims and under 7% in U.S. FTC data ever submitting a report [26]. Users often hesitate because they are unsure whether an issue "counts," uncertain how or where to report, or lack confidence in their judgement [8, 28, 46]. Even when reporting mechanisms exist, unclear instructions or lack of feedback frequently suppress reporting. This distinction is especially important in electronic voting. E2E verifiability relies on voters as the final checkpoint: they must verify that their ballot reflects their intent. Yet none of the individual verifiability approaches discussed earlier provide a dedicated reporting mechanism. Instead they implicitly assume that voters will know how to contact election officials on their own, which is unrealistic given known underreporting patterns. We argue that a dedicated *out-of-band* reporting channel is essential. Such a channel must be independent from the potentially compromised voting system so that reports cannot be intercepted or suppressed.

3 Methodology

This section outlines the study methodology, including participant recruitment, experimental design, manipulations, study conditions, and data collection. We describe the structure of the user study, how manipulation detection and reporting were operationalised, and the procedures participants followed across two election scenarios. A pre-study was conducted to ensure clarity and technical robustness.

3.1 Research Questions

Our study focuses on two central outcomes: whether voters can *detect* manipulations and whether they subsequently *report* them. We distinguish between these outcomes because noticing an anomaly and taking the step to report it represent different behaviours.

First, we test whether introducing usability and security improvements, including simplified terminology, enhanced feedback, stronger security cues, and a structured step-by-step guide, helps participants detect and report manipulations compared to a realistic baseline system.

- **RQ1**$_{Detection}$: Do the system improvements increase participants' likelihood to *detect* manipulations?
  - **RQ1.1**$_{Detection}$: Higher detection of **Vote Tampering**?
  - **RQ1.2**$_{Detection}$: Higher detection of **Verification Prevention**?
- **RQ2**$_{Reporting}$: Do the system improvements increase participants' likelihood to *report* manipulations.
  - **RQ2.1**$_{Reporting}$: Higher reporting of **Vote Tampering**?
  - **RQ2.2**$_{Reporting}$: Higher reporting of **Verification Prevention**?

Second, we examine whether the structured step-by-step guide itself contributes beyond the other improvements:

- **RQ3**$_{Detection}$: Does the guide improve voters' ability to *detect* manipulations?
  - **RQ3.1**$_{Detection}$: Effects on detection of **Vote Tampering**?
  - **RQ3.2**$_{Detection}$: Effects on detection of **Verification Prevention**?
- **RQ4**$_{Reporting}$: Does the guide improve voters' likelihood to *report* manipulations?
  - **RQ4.1**$_{Reporting}$: Effects on reporting of **Vote Tampering**?
  - **RQ4.2**$_{Reporting}$: Effects on reporting of **Verification Prevention**?

Finally, to complement these security-focused outcomes, we examine participants' overall perceptions of the system. This includes both usability and confidence (i.e., perceived trustworthiness and perceived risks to vote integrity and secrecy) in the election process.

- **RQ5**$_{Usability\ and\ Confidence}$: How do different system versions compare in terms of *perceived usability* and voters' *confidence* in the system?

3.2 Study Groups

Figure 1 provides an overview of the study groups and their mapping to hypotheses and research questions.

To investigate these research questions, we implemented three versions of a verifiable internet voting system. Each system included its own voting interface, verification website, and voter-facing materials (e.g., election invitation, instructions):

1. Baseline System. Modelled after the system used in the 2024 election of the GI, the German association for computer science. It included the original voting and verification interfaces, as well as invitation materials adapted from that real-world election. Only minor adjustments were made to accommodate our study's two election formats (referendum and party election).

2. Improved System. Incorporated the full set of usability and security improvements (see Section 4), including simplified terminology, enhanced feedback, stronger security cues, and a structured step-by-step guide. This represents the "best-case" system that combines all improvements.

3. Improved System without Step-by-Step Guide. Retained all technical and design improvements but omitted the guide. Since the guide contains critical information (e.g., how to report problems), the election invitation was modified slightly to include those missing instructions. This system serves to isolate the effect of the guide itself.

Each of these systems was paired with one of the two manipulation types mentioned in related work (**Vote Tampering**, **Verification Prevention**), resulting in six study groups (3 systems × 2 manipulations). The following subsection will explain in detail, how these manipulation types were simulated. This design allowed us to directly test our hypotheses on detection and reporting (**RQ1**$_{Detection}$, **RQ2**$_{Reporting}$), examine the specific contribution of the guide (**RQ3**$_{Detection}$, **RQ4**$_{Reporting}$), and compare perceived usability and confidence across all systems (**RQ5**$_{Usability\ and\ Confidence}$).

3.3 Studied Manipulation Types

To investigate how different types of manipulation affect voters' ability to detect and report irregularities, we implemented two manipulation types that had been highlighted in prior cast–then–audit research, particularly by Hilt et al. [18]. The first represents a **blatant** anomaly (vote tampering), while the second constitutes a more **subtle** attack (verification prevention).

Vote Tampering. This manipulation simulates a compromised voting device altering the ballot content before it is cast and stored in the digital ballot box. Voters can detect this manipulation if they complete the verification step on a second device, since the verification website displays the stored ballot for comparison against their intended selection. In our study, this was implemented by modifying the verification website, so that it always displays an altered ballot: the originally selected option was consistently replaced with a predefined alternative (e.g., if Option 1 was chosen, the verifier displayed Option 2).

Verification Prevention. This manipulation targets the verification process itself by suppressing the information needed for checking by withholding the QR code required for checking the ballot with a second device. This type of manipulation is harder to detect, as it relies on voters recognising that an expected verification feature is missing. In our study, participants were presented with a version of the system that skipped the verification step entirely. At the end of the voting process, the system displayed a closure message falsely reassuring that verification had already been completed, accompanied by a green check-mark:

> *"Thank you for casting your vote. Your vote has been transferred to the digital ballot box. Your vote was already successfully checked for you. You can close this window now."*

Figure 2 shows the difference between the unmanipulated finalisation page, which provides voters with the information needed for verification, and the manipulated version, which falsely reassures voters that verification has already been completed.

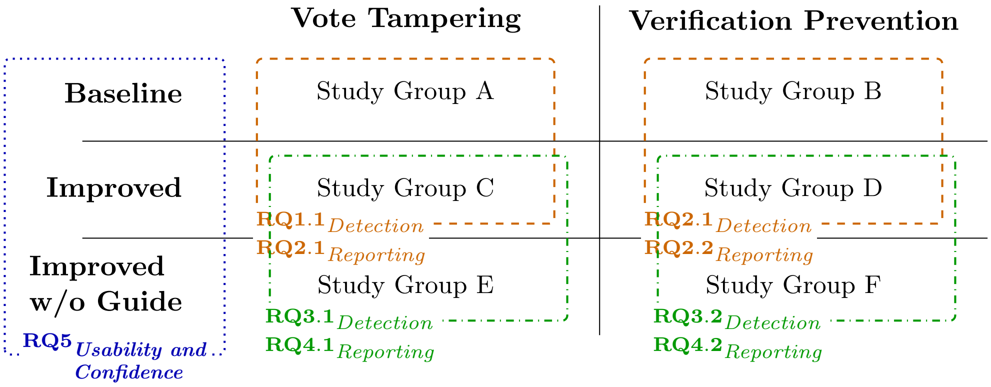
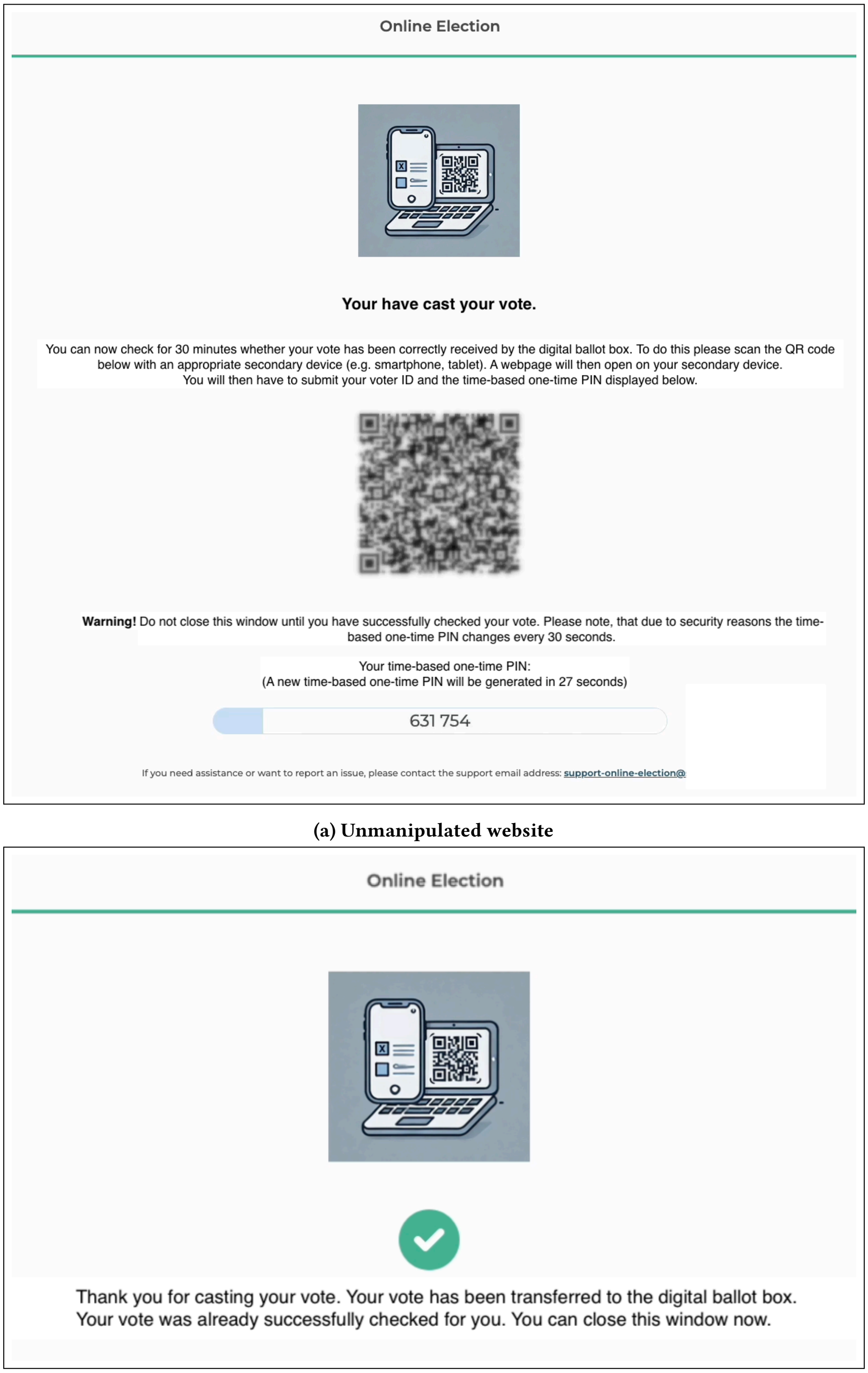


Figure 1: Study design with six groups across three systems (Baseline, Improved, Improved w/o Guide) and two manipulation types (Vote Tampering, Verification Prevention). Each group (A–F) is mapped to the relevant research questions (RQ1–RQ5).



Online Election

Your have cast your vote.

You can now check for 30 minutes whether your vote has been correctly received by the digital ballot box. To do this please scan the QR code below with an appropriate secondary device (e.g. smartphone, tablet). A webpage will then open on your secondary device. You will then have to submit your voter ID and the time-based one-time PIN displayed below.

Warning! Do not close this window until you have successfully checked your vote. Please note, that due to security reasons the time-based one-time PIN changes every 30 seconds.

Your time-based one-time PIN:
(A new time-based one-time PIN will be generated in 27 seconds)

631 754

If you need assistance or want to report an issue, please contact the support email address: support-online-election@

(a) Unmanipulated website

Online Election

Thank you for casting your vote. Your vote has been transferred to the digital ballot box. Your vote was already successfully checked for you. You can close this window now.

(b) Manipulated website

Figure 2: Differences between finalisation screens in the unmanipulated and manipulated systems. (a) The unmanipulated website informs voters that they have cast their vote, provides a QR code and a one-time PIN for checking the ballot with a second device. (b) In the manipulated website, this verification information is withheld. Instead, a green checkmark and short message falsely reassure voters that their vote was already transmitted and successfully checked.

3.4 Study Procedure

Figure 3 provides an overview of the main steps of the study. Participants completed two voting tasks (referendum and party election), each followed by a survey, with additional elements such as informed consent, debriefing, and manipulation detection/reporting integrated into the flow. Details of each step are described below.

Recruiting over Clickworker

Participants were recruited through the online platform Clickworker. Eligibility criteria required participants to be fluent in German, reside in Germany, and be at least 18 years old. To reduce potential bias, we excluded individuals who had previously taken part in similar studies from the same research team. Participants were informed that the study aimed to assess the usability of different internet voting formats.

Informed Consent and Study Overview. After recruitment, participants received a personalized link to the online questionnaire hosted on SoSci Survey. Upon entering the study, they were randomly assigned to one of six study groups (see Section 3.2). Participants were first presented with an informed consent form and a data protection statement. A textual and visual overview of the study procedure was provided, with the current step highlighted in green throughout the questionnaire to aid orientation. Participants could only proceed after confirming their consent to participate and agreeing to the terms of data collection.

First Voting Task: Referendum. Participants were first introduced to a simulated referendum election. They received a download link to retrieve their personalized voting materials, which included:

- An election invitation containing the link to the voting system.
- A role card outlining their assigned persona, voting preference, and voter ID.
- For participants in the *Improved* condition: a step-by-step voting guide.

Participants were instructed to read the materials carefully and follow the role card instructions to maintain vote secrecy. They then accessed the voting platform and cast their vote. In line with real-world implementations of individual verifiability (e.g., Estonia), vote verification was optional. However, participants in the *Improved* and *Improved w/o guide* systems received textual cues in their invitation materials encouraging them to verify their vote. These cues included an analogy and a norm-based nudge, as proposed by Olembo et al. [36]. Verification could only be performed using a second device and the system-specific verification tool.

Post-Vote Survey (Part 1). After completing the first voting task (and optional verification), participants returned to the online questionnaire. They completed the System Usability Scale (SUS) to assess the usability of the voting system [4]. Participants were also asked whether they had verified their vote using a second device, and if not, asked to explain their reason in open text. Finally, they rated the voting system on three five-point Likert scales capturing confidence-related perceptions: trustworthiness, perceived risk to vote integrity, and perceived risk to vote secrecy.

Second Voting Task: Party Election Format. The second election simulated a party election[1] (1-out-of-n voting format) and followed the same procedural structure as the first election. Again, participants downloaded personalized election materials, cast their vote, and (optionally) performed verification. The key difference was the simulated manipulation that was introduced and which depended on participants' assigned group: (a) **vote tampering**, where the verification result differed from the cast ballot, or (b) **verification prevention**, where the system falsely claimed the vote had already been checked.

Post-Vote Survey (Part 2). Immediately after the second vote, participants rated the system again on the same three confidence-related dimensions. This allowed us to capture participants' immediate reactions to the manipulations. Then, participants were asked whether they noticed anything unusual during the voting process, following the approach used by Hilt et al. [18]. Those who confirmed were asked they reported it - and if not, why.

Debriefing, Final Assessment, and Study Completion. The final part of the survey included a debriefing screen that revealed the true purpose of the study. Participants were informed that the election systems they interacted with had been intentionally manipulated and were provided with an explanation of the specific manipulation applied to their study group. The debriefing also emphasized the importance of detecting and reporting manipulations

[1] In a party election format voters do not cast their votes for specific candidates from a party but instead for the political party itself. This is for example the case for the so-called "Zweitstimme" for the election for the German parliament [5].

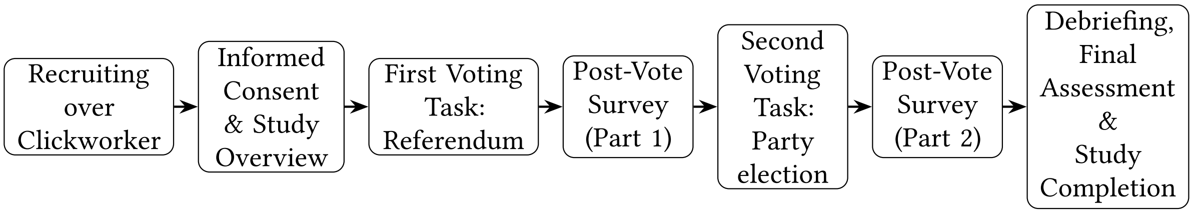

Recruiting over Clickworker
Informed Consent & Study Overview
First Voting Task: Referendum
Post-Vote Survey (Part 1)
Second Voting Task: Party election
Post-Vote Survey (Part 2)
Debriefing, Final Assessment & Study Completion

Figure 3: Overview of study procedure.

in the context of verifiable internet voting systems. Immediately following the debriefing, participants were asked once more to rate the system on the same three confidence-related dimensions. This allowed us to assess whether the debriefing influenced participants' retrospective evaluations. Finally, participants received a unique confirmation code to verify their participation and receive their compensation.

3.5 Pre-Study

Before launching the main study, we conducted a pre-study with 18 participants — three participants for each of the six experimental conditions. The goal was to identify potential usability issues, misunderstandings in the instructions or materials, and any technical problems across the different voting platforms and manipulation types. During the pre-study, no major issues were observed. Participants were able to navigate the survey, voting systems, verification websites, and reporting tools without confusion or unexpected behaviour. As a result, no structural changes were made to the study design. However, we used the recorded completion times from the pre-study to update our recruitment materials. Specifically, the estimated study duration was adjusted to 30–45 minutes to reflect the average range observed in the pre-study.

3.6 Ethics and Data Protection

The study was reviewed and approved by our institution's ethics board. The research protocol followed the ethical guidelines of our institution, including full transparency about the study's data collection practices and participant rights. We developed a data protection and privacy notice in consultation with the institutional data protection office. This document detailed the scope and purpose of data collection and was presented to participants at the beginning of the survey. Participation could only proceed after participants gave their informed consent by actively confirming agreement with the data protection terms.

3.7 Recruitment

We conducted an a priori power analysis (G*Power; $\alpha = .05$, power $1 - \beta = .95$) for a chi-square test of independence with three groups (df = 2), assuming a *medium* effect size in the sense of Cohen (i.e., $w = 0.30$). We additionally planned Bonferroni corrections for families of categorical comparisons; the power analysis therefore targeted the corrected α. The analysis yielded a minimum total sample size of $N = 428$. To accommodate exclusions and ensure sufficient power per cell, we targeted $N = 460$ participants overall. Based on pre-study timings, the expected duration was $30 - 45$ minutes. Participants received 10 € for completing the study, which exceeds the minimum hourly wage in Germany.

3.8 Data Collection

Survey responses. Participants responded to a range of items in the online questionnaire, including standardized scales (e.g., SUS), Likert-scale ratings on perceived trust and risk, and open-ended responses about anomalies or manipulation detection. These responses were stored securely in pseudonymized form and linked to participants' condition assignment via an anonymized voter ID.

Interaction logs. In addition to survey data, we logged participant behaviour on selected study websites to better understand how they interacted with the voting ecosystem. This included the verification website, FAQ page, and reporting website. Each participant received personalized links containing an anonymized reference ID. These links enabled us to track user behaviour (e.g., visits, button clicks, reporting actions) without collecting personally identifiable information. Interaction data were collected as state-action logs with timestamps, allowing us to reconstruct key user actions, such as whether participants attempted verification, accessed help pages, or submitted a report. For this study, logging was limited to the websites we controlled. Importantly, no behavioural or interaction data were collected from the actual voting platform itself, which was provided by POLYAS. Due to their privacy policy and system constraints, we were unable to implement tracking on their system.

3.9 Data Analysis Approach

All analyses were conducted in Python (pandas, NumPy, SciPy, statsmodels; plotting with Matplotlib). We report exact two-sided p-values, with $\alpha = .05$ as the significance threshold. Where Bonferroni adjustment was applied, we report adjusted p-values.

Binary outcomes. Dichotomous variables (*manipulation detection*; *reporting* among all participants; *reporting among detectors*; *completion* of vote+verification) were analysed using $2\times k$ or 2×2 contingency tables. Omnibus and pairwise tests used chi-square tests of independence; Fisher's exact test was substituted for 2×2 tables with any expected cell count < 5. Multiple pairwise tests within a family were Bonferroni-adjusted. Effect sizes were reported as Cramér's V (chi-square) or odds ratio (Fisher).

Continuous/ordinal outcomes. For completion time, SUS scores, and individual Likert items from the perception scales (trustworthiness, risk to secrecy, risk to integrity), we first examined the assumptions. To this end, we employed Shapiro-Wilk tests for normality and Levene's test for homogeneity of variances. Subsequently, Kruskal-Wallis tests were applied to all analyses on these variables. Non-parametric contrasts used Mann–Whitney U without multiplicity correction due to the large number of planned exploratory contrasts; in the latter case, interpretation emphasised effect sizes (Cohen's d for parametric, rank-biserial r_{rb} for non-parametric).

Open-text responses. Qualitative responses from the survey and the reporting website were coded inductively by two independent coders using an open coding approach, identifying themes for the responses (see 6.1). Coding discrepancies were resolved through discussion until full consensus was reached across all items.

Classification of detection and reporting. Detection and reporting outcomes were coded independently by two researchers. For *detection*, we analysed participants' responses in the post-election questionnaire. Participants who indicated that they noticed something unusual were asked to describe what they observed. These open-text descriptions served as the basis for determining whether the manipulation had been correctly identified. A response was classified as a detection if it referred to the manipulation in the participants study group. For *reporting*, we analysed the messages submitted through the reporting website. A message was classified as a report if it contained a reference to the manipulation that occurred in the participant's study condition. Note for both classifications, participants did not need to use specific words, such as "manipulation", instead a reference, such as "my vote changed" or "the qr code is missing" was sufficient. Disagreements in either classification were discussed until full consensus was reached.

3.10 Data Cleaning and Sample

We initially recruited 457 participants. Participants were excluded if they (i) did not complete both elections, (ii) did not complete the immediate post-vote survey as instructed, or (iii) were identified as extreme outliers (>1.5×IQR) on either completion time or SUS score. After applying the excluding criteria, the final sample comprised 437 participants (Group A: 73; Group B: 73; Group C: 73; Group D: 77; Group E: 71; Group F: 70; see Figure 1 for an overview).

3.11 Limitations

Our study has several limitations. Participants were recruited via an online panel and compensated for their time. This may have motivated them to complete tasks, such as performing verification, that some voters might skip in real elections if not compelled by payment. While this incentive structure may inflate verification rates, the relative differences across conditions remain valid, as all participants were subject to the same study constraints.

Our study design focused on reporting as an independent, out-of-band process. We did not examine scenarios where a compromised voting device could also suppress reporting attempts, as this would have required linking reports to individual devices and potentially collecting personal identifiers. While this choice avoids privacy concerns, it also omits an additional layer of manipulation that could occur in practice.

The step-by-step guide represents one specific instantiation of structured voter support. Alternative designs, such as shorter, more compact versions or longer, more explanatory formats, might lead to different usability, detection, or reporting outcomes. Future work should explore such variations.

The participant pool consisted of German-speaking individuals recruited from an online panel. This demographic and linguistic focus may limit the generalisability of results to other populations and electoral contexts. Furthermore, we simulated a parliamentary election in Germany, which has not yet been realized with a remote electronic voting channel. The results may differ for different systems and cultural contexts. For example, in countries such as Estonia or Switzerland, where internet voting has been present for some time, people might be more likely to detect verification prevention manipulations, as they are already more used to the intended flow of the internet voting processes.

Participants received the voting materials via a download link in the online questionnaire. The download consisted of a compressed .zip file containing two folders. The first one represents the material voters would receive in a real election and contained the election invitation and, for the *improved* group, the step-by-step guide. The second one contained only the role card to clearly separate it from the election materials. In a real election the envelope with the election material would be delivered in printed form via postal mail for security reasons. However, as this study was conducted entirely online, providing physical paper materials was not feasible.

The wording of the "Report fraudulent activities"[2] button may have decreased participants likelihood to press this button and submit a report as "fraudulent activities" is a strong claim, one would not make lightly. We discussed this extensively beforehand and deliberately chose this specific phrasing as we wanted to single out reports that believed that actually something bad went wrong compared to "only" a technical issue happening. Nonetheless, future work could explore how different phrasing of this action button might influence participants willingness to report an issue.[3]

We did not directly assess whether participants understood the instruction that printed materials should be prioritised over any information displayed on the device. This is a limitation, as the effectiveness of the step-by-step guide depends on voters recognising the printed instructions as the trusted source when conflicts arise. Moreover, our study examined only one specific implementation of each manipulation type within the cast–then–audit approach. The exact wording of a verification-prevention attack may strongly influence detection and reporting. An attacker could, for example, explicitly claim that the printed materials are outdated or invalid due to a recent update. Such a message would test whether voters actually apply the intended priority rule, and might alter detection rates. Future work should therefore explore how voters interpret conflicting instructions and how manipulation wording affects their behaviour.

The cast-then-audit approach for individual verifiability requires the use of two devices: one for casting a vote and one for verification. While this may appear to be a limitation of our specific system, it reflects a broader requirement of individual verifiability, as explained in Section 2. Nevertheless, the need for a secondary device introduces practical and accessibility challenges. To maximise the likelihood of compliance in our study, we restricted participation to users on desktop or laptop computers, increasing the probability that a second internet-capable device was available. In real elections, however, many voters may cast their vote on a smartphone and not have access to another trusted device for verification or reporting, which may limit the applicability of our findings. These voters could still use the device of a relative, friend or neighbour, but this would need to be communicated properly and would introduce further usability constraints. Future work should examine how device availability and socio-economic factors affect voters' ability to complete all stages of individual verifiability.

4 System Components and (Design) Improvements

To help readers unfamiliar with internet voting gain a quick overview, we provide a short video walkthrough of the system components and their interactions.[4] Readers interested in the technical and design rationale will find a detailed description of each component below, including its function, interface, and the enhancements we implemented.

[2] In the study we used the German term "Probleme melden".

[3] None of the participants referenced this point in the open text responses, but we also did not specifically question them about the wording.

[4] https://secuso.org/2026-01-23_CHI-Why_Johnny_Checks_but_Doesnt_Alert/video/video.mp4

The system components and materials used in this study are based on the material used by Hilt et al. [18]. We systematically improved these components based on at least one of the "lacks" identified by Kulyk et al. [22], which describe reasons why voters may fail to verify their vote. Each component is described below in three steps: role in the process, interface, and design improvements over Hilt et al. [18].

4.1 Election Invitation Letter

The full election invitation is attached as Appendix A. The invitation letter begins by drawing a parallel to traditional paper-based voting. This analogy is intended to encourage and explain why voters are able to check that their digital vote was not altered before casting. This "analogy cue" was inspired by Olembo et al.'s work on motivating voters to perform verification [36]. In addition, a "norm cue" reinforces the expectation that checking one's vote is a standard and responsible action, that voters who wish to protect democracy perform, further nudging voters to check their ballot. By integrating these textual cues we also address one of the recommendations from Marky et. al, that future studies should "provide information why verification is needed" [31], in an effort to bolster the rate in which voters check their ballot. Research by Olembo et al. showed that the term "verifiability" is often perceived as too abstract [35]. As consequence, we used a simplified and and more intuitive wording in the invitation: we replaced "verify" with "check", and referred to the verification website as the "website to check your ballot". To increase accessibility, key points were presented in a bulleted format to reduce cognitive load and support rapid scanning. In contrast to the approach by Hilt et al. [18], we deliberately avoided describing the voting and verification process in full detail within the invitation letter. Instead, participants were referred to a separate step-by-step guide that provides structured and detailed support, as described in the following subsection. Our goal was to keep the invitation concise and allow participants to concentrate more fully on the guide when performing the task.

Summary of Improvements to the Election Invitation:

- **Simplified terminology.** Replaced abstract terms like "verify" with intuitive language such as "check your ballot".
- **Motivational cues.** Included analogy and norm-based cues to encourage vote checking.
- **Reference to guide.** Directed voters to the step-by-step guide for process details, reducing cognitive load.

4.2 Step-by-Step Voting Guide

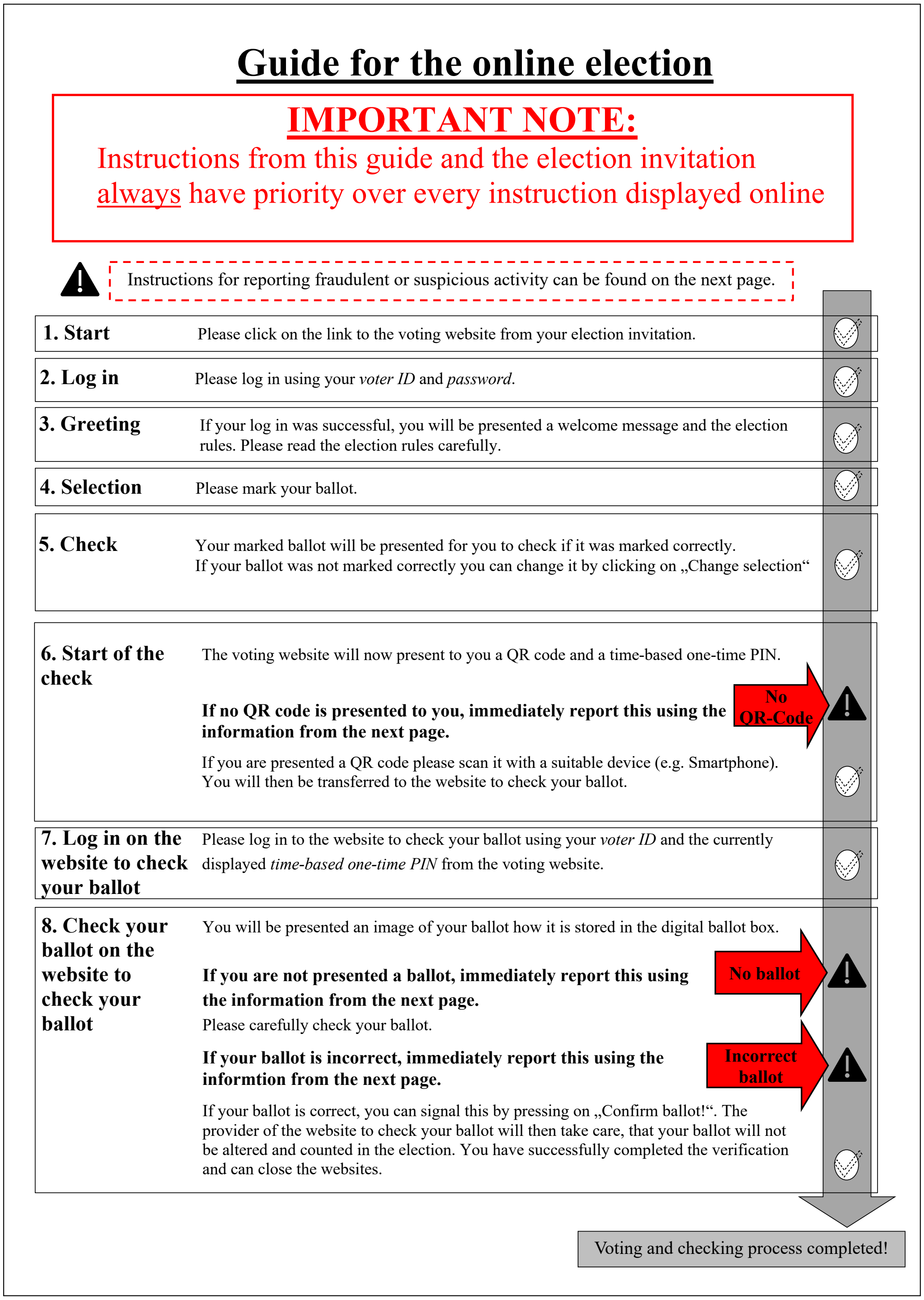
To support participants in completing verification, we provided a concise, step-by-step guide. Prior work in usable security has shown that clear, actionable instructions can improve user adherence to security tasks. For example, more structured and comprehensible SSL warnings have been found to significantly increase user compliance [10]. Similarly, research on incident-response playbooks demonstrates that structured, stepwise guidance helps users to follow complex procedures more reliably, even in high-stress security contexts [45]. Usability principles for secure system design also highlight the value of keeping instructions minimal and consistent, ensuring visibility of system status and reducing cognitive load on users [27]. Together, these findings motivated the design of our guide as a series of discrete checks, each expressed with a short title, explicit If-Then constructions ("If your ballot is incorrect: report"), and a visual cue (checkmark or cross). To our knowledge, no prior user study or real-world implementation of the cast-then-audit approach has provided voters with a comprehensive, structured guide for completing the voting and verification process. Inspired by the improved code-sheet concept developed by Kulyk et al. for return-code-based voting systems [24], and by similar documents used in Swiss elections, we developed a tailored step-by-step guide for our setting, as shown in Figure 4. Our guide was designed to support voters through all stages of casting and checking their vote, while reinforcing manipulation awareness and promoting reporting of irregularities. The guide begins with a red security notice stating that instructions from the election invitation and the guide take priority over any information presented online. This was meant to establish these materials as trusted sources and to mitigate the risk of users being misled by deceptive or manipulated websites.

The main section of the guide walks participants through eight discrete steps, each representing one screen in the voting or verification interface. Each step includes (1) a concise title (e.g., *Log in*, *Selection*, *Check*), (2) a short instruction describing the action to be performed and (3) a checkbox that voters could use to track their progress. At critical steps — specifically, those where manipulations might occur — the guide includes explicit If-Then constructions, providing a clear next step. For example, step six warns voters to report immediately if there is no QR code present, using a bold message and an attention-grabbing red arrow pointing to the symbol introduced in the guide's preamble. Similarly, step eight advises users to report if the ballot shown on the verification website is missing or incorrect. In each case, the warning refers users to the dedicated instructions on the backside of the guide.

The backside provides detailed instructions for reporting suspicious activity (see Figure 5). Voters are explicitly told not to report from the same device they used for voting, and are given several reporting channels: a QR code linking to the reporting website (to encourage the use of a second device), a dedicated support email address, a phone number, and the "Report fraudulent activities" button available on the verification website. In addition, a summary is included to clarify which events warrant reporting, such as missing QR codes or incorrect ballots.

Summary of Advantages of the Step-by-Step Guide:

- **Novel support format.** To our knowledge, this is the first structured step-by-step guide tailored for a cast-then-audit voting system.
- **Security priming.** A clear priority notice and warning icon establish the guide as the primary, trusted information source.
- **Usability-focused structure.** Steps are broken down into short, labelled actions with checkboxes to help stay oriented.
- **Manipulation awareness.** Manipulation-prone steps include visual cues and reporting prompts to increase vigilance.
- **Redundant reporting channels.** The back page consolidates all reporting mechanisms and emphasises the use of a second device for safety.



Guide for the online election

IMPORTANT NOTE:
Instructions from this guide and the election invitation always have priority over every instruction displayed online

⚠ Instructions for reporting fraudulent or suspicious activity can be found on the next page.

1. Start Please click on the link to the voting website from your election invitation. ✓

2. Log in Please log in using your *voter ID* and *password*. ✓

3. Greeting If your log in was successful, you will be presented a welcome message and the election rules. Please read the election rules carefully. ✓

4. Selection Please mark your ballot. ✓

5. Check Your marked ballot will be presented for you to check if it was marked correctly.
If your ballot was not marked correctly you can change it by clicking on „Change selection“ ✓

6. Start of the check The voting website will now present to you a QR code and a time-based one-time PIN.

If no QR code is presented to you, immediately report this using the information from the next page. No QR-Code ⚠

If you are presented a QR code please scan it with a suitable device (e.g. Smartphone). You will then be transferred to the website to check your ballot. ✓

7. Log in on the website to check your ballot Please log in to the website to check your ballot using your *voter ID* and the currently displayed *time-based one-time PIN* from the voting website. ✓

8. Check your ballot on the website to check your ballot You will be presented an image of your ballot how it is stored in the digital ballot box.

If you are not presented a ballot, immediately report this using the information from the next page. No ballot ⚠

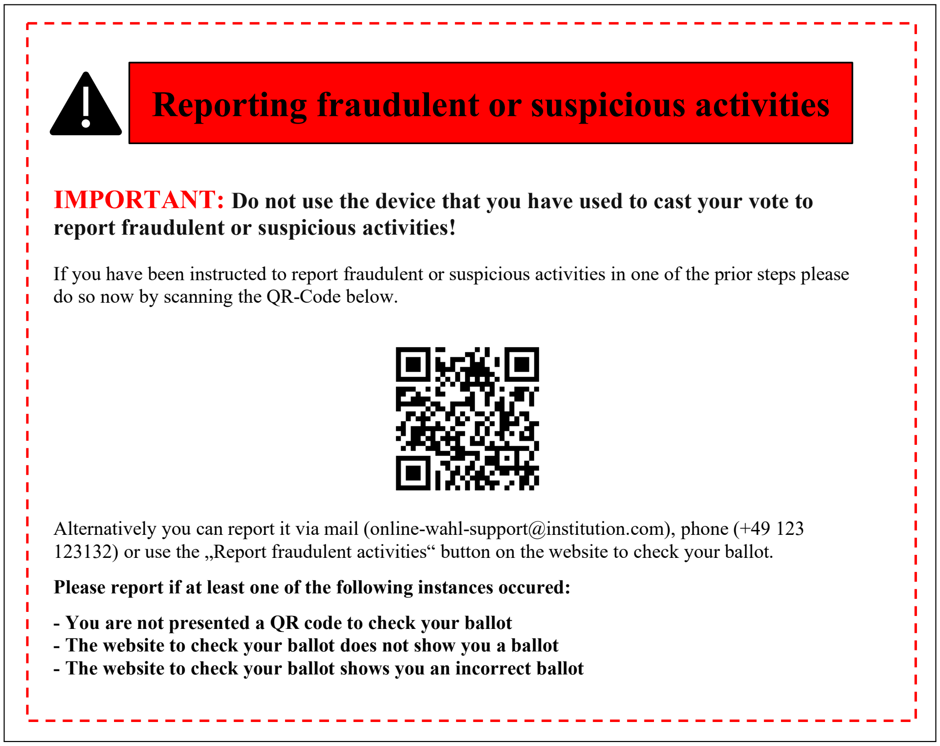
Please carefully check your ballot.

If your ballot is incorrect, immediately report this using the informtion from the next page. Incorrect ballot ⚠

If your ballot is correct, you can signal this by pressing on „Confirm ballot!“. The provider of the website to check your ballot will then take care, that your ballot will not be altered and counted in the election. You have successfully completed the verification and can close the websites. ✓

Voting and checking process completed!

Figure 4: The detailed step-by-step guide leading voters through all voting and verifying steps. The guide was provided as a single sheet to ensure readability and to emphasize its role as an independent, authoritative reference throughout the voting process. It outlines each stage from login and ballot marking to verification, including explicit instructions on how to respond to anomalies (e.g., missing QR code, missing ballot, incorrect ballot). Clear warnings, red markers, and "If-Then" instructions direct voters to immediately report suspicious behaviour using the reporting channel described on the following page. This design ensured that voters had unambiguous, accessible guidance at every stage of the process.



Reporting fraudulent or suspicious activities

IMPORTANT: Do not use the device that you have used to cast your vote to report fraudulent or suspicious activities!

If you have been instructed to report fraudulent or suspicious activities in one of the prior steps please do so now by scanning the QR-Code below.

Alternatively you can report it via mail (online-wahl-support@institution.com), phone (+49 123 123132) or use the „Report fraudulent activities“ button on the website to check your ballot.

Please report if at least one of the following instances occured:

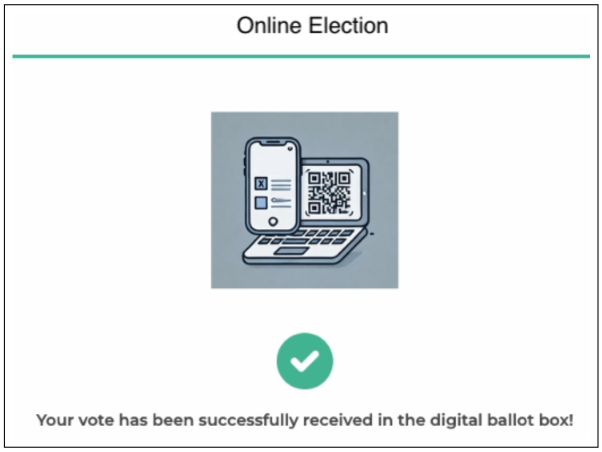
- You are not presented a QR code to check your ballot
- The website to check your ballot does not show you a ballot
- The website to check your ballot shows you an incorrect ballot

Figure 5: The backside of the step-by-step guide, providing voters with instructions for reporting fraudulent or suspicious activities. It emphasizes that reporting must not be done on the same device used to cast the vote and offers multiple reporting channels, including QR code, email, phone, and a website button. The guide also specifies conditions that should trigger reporting (e.g., no QR code displayed, no ballot shown, or ballot content incorrect). The QR code displayed here was modified for submission purposes to preserve anonymity.

4.3 Voting Website Interface

Although we made only minor changes to the voting interface itself, these adjustments were motivated by prior usability issues and aimed at improving clarity and consistency. Most notably, we reduced explanatory text throughout the interface. Our design assumed that voters would rely primarily on the election invitation and the accompanying step-by-step guide for trusted information, rather than the interface itself. This approach helped us maintain a clean user interface (UI) and emphasised that procedural trust should be placed in the materials provided before voting.

We removed a green check mark and accompanying confirmation text that previously appeared immediately after casting a vote (see Figure 6). Positioned prominently at the top of the page, the element was visually large and occupied the upper third of the interface. This design risked misleading voters into believing the process was complete, even though the verification step had not yet taken place. Removing it helped prevent a false sense of closure and ensured that the interface better aligned with the intended verification flow. To ensure consistency with other components, we avoided abstract terminology such as "verify" and "verifiability" within the interface. These were replaced with more accessible alternatives like "check your ballot", in line with wording used in both the guide and invitation letter. We also corrected minor inaccuracies, such as referring to a PIN as a "password".



Online Election

Your vote has been successfully received in the digital ballot box!

Figure 6: Original version of the finalisation page of the voting website showing a large green check mark and confirmation text at the top of the page, which could falsely suggest to voters that the process was complete before verification occurred.

Summary of Improvements to the Voting Website Interface:

- **Terminology consistency.** Replaced abstract terms like "verify" with more intuitive phrasing such as "check your ballot".
- **Accuracy.** Corrected misleading labels (e.g., changing "password" to "PIN").
- **Removed false reassurance.** Eliminated the green checkmark and post-vote confirmation text that could prematurely signal completion.

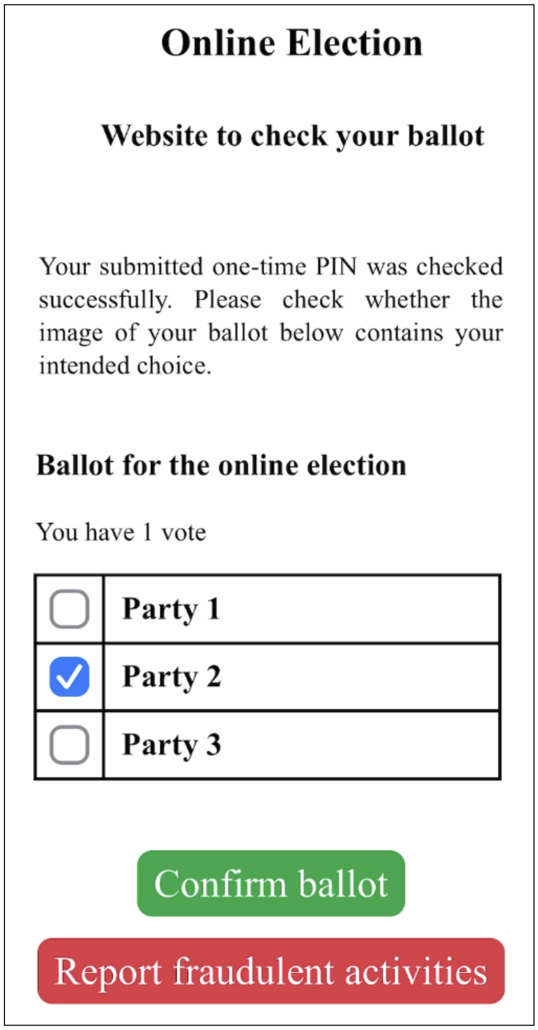
4.4 Ballot Checking Website

The ballot checking website was adapted from the publicly available codebase used in the 2024 GI election [6]. While the underlying functionality remained intact, we introduced several user interface improvements aimed at enhancing usability and reducing ambiguity for voters.

From a security perspective, we removed the pre-filled voter ID input field that had been present in the original implementation. Automatic field filling could facilitate a so-called "clashing attack", in which a voter might be shown a ballot cast by someone else who selected the same option [25].[5] By requiring voters to manually enter their own voter ID, we ensured that they would actually verify their identity, and thereby reduced the risk of passive confirmation errors. Once voters are logged in the ballot checking website shows an image of their ballot how it is stored in the digital ballot box. This image is modifiable in an effort to make vote selling more complicated, as a simple screenshot of the verified ballot does not provide proof of the voting decision. If the image of the ballot is altered, a pop-up notification informs the voter that this change does not affect the ballot itself, as only an image of the ballot is modified. Importantly, only the image can be altered, but not the ballot itself. The intention of this feature was to transport the message that a simple screenshot would not be sufficient to reliably prove how one voted, which might deter people from trying to sell their vote.

A key addition to the interface was the inclusion of two action buttons shown after the ballot was displayed (see Figure 7). These buttons were designed to clarify the available next steps and to provide an immediate channel for responding to anomalies, and thereby offer a clearer decision pathway, reduce ambiguity, and support both trust and accountability in the verification process:

- **"Confirm ballot"**: Clicking this button led voters to a final confirmation screen, stating that their ballot had been checked and would now be securely stored and counted. The original interface lacked such sense of closure, which could leave voters uncertain about whether further steps were required.
- **"Report fraudulent activities"**: This button redirected participants to the dedicated reporting website. Its inclusion was intended to lower the barrier for reporting and to provide voters with a fast and contextually appropriate way to respond if their ballot did not reflect their intended vote.



Online Election

Website to check your ballot

Your submitted one-time PIN was checked successfully. Please check whether the image of your ballot below contains your intended choice.

Ballot for the online election

You have 1 vote

☐	Party 1
☑	Party 2
☐	Party 3

Figure 7: Revised version of the ballot checking website. After scanning the QR code and entering the one-time PIN from the finalisation page of the voting system, the ballot checking website displays an image of the voter's recorded ballot and provides two clear action buttons: a green button to confirm the ballot and a red button to report fraudulent activities.

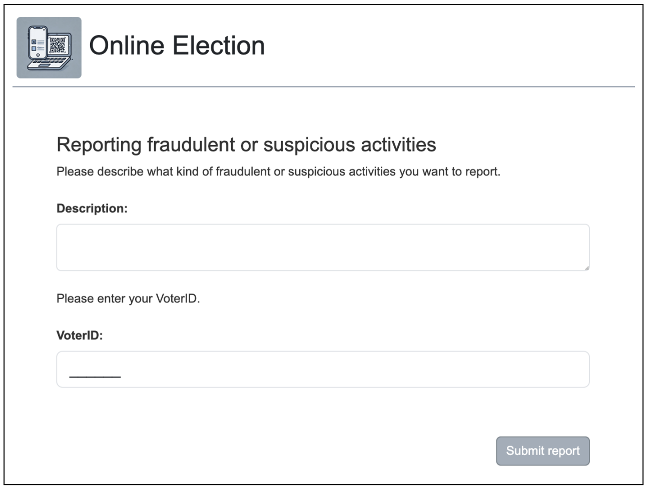
Summary of Improvements to the Ballot Checking Website:

- **Removal of pre-filled voter ID.** Required manual entry to impede potential clashing attacks and increase voter attention to identity confirmation.
- **Explicit completion feedback.** Added a "Confirm ballot" button and a final screen to clearly signal successful completion of the verification process.
- **Integrated reporting access.** Included a "Report fraudulent activities" button for a seamless transition to the reporting site in case of anomalies.
- **Improved interface clarity.** General UI enhancements for readability and consistency with other components.

[5] Note we did not specifically test for the detection of clashing attacks but we did not want to introduce misaligned user expectations that could potentially enable clashing attacks.

4.5 Problem Reporting Website

We introduced a dedicated problem reporting website to provide voters with a channel independent from a potential malicious voting system for reporting irregularities. To our knowledge, this is the first instantiation of such a mechanism in a user study of individual verifiability with the cast-then-audit approach. Figure 8 shows the reporting website.



Reporting fraudulent or suspicious activities

Please describe what kind of fraudulent or suspicious activities you want to report.

Description:

Please enter your VoterID.

VoterID:

Figure 8: Reporting website used in the study. The site provides a simple form where voters can describe the suspicious activity and enter their voter ID before submitting the report. This interface was designed to lower the barrier for reporting by offering a clear, dedicated channel separate from the voting process. Voters reached this page by clicking the red "Report fraudulent activities" button on the ballot checking website, or alternatively by scanning the QR code / typing the URL provided in the step-by-step guide.

The reporting website was hosted on a separate domain from the voting platform. This separation was intended to signal independence and operational decentralisation to technically savvy voters – reinforcing the notion that the reporting channel is controlled by a different entity than the system that may be compromised. While this distinction may go unnoticed by most participants, it serves as an additional layer of perceived and actual trustworthiness.

The interface itself was intentionally minimal: voters were asked to enter their voter ID and describe the issue they encountered.[6] After submitting the form, they received a confirmation message indicating that their report had been successfully transmitted and would be reviewed by election officials.

Although technically simple, the reporting website plays a key role in closing the loop between manipulation detection and potential follow-up. It provides a concrete path for voters to act on suspicion or evidence of tampering, and it reinforces the message that their concerns will be taken seriously.

Summary of Advantages of the Problem Reporting Website:

- **Introduction of a reporting mechanism.** To our knowledge, this is the first dedicated reporting website used in a cast-then-audit individual-verifiability user study.
- **Operational separation.** The reporting system was hosted independently to strengthen perceived trustworthiness and signal decoupling from the voting system.
- **User reassurance.** A clear confirmation message was shown after submission to provide feedback and reinforce that reports are acted upon.

4.6 FAQ Website

As part of the overall voter support infrastructure, we included an FAQ page that was accessible to all participants. The FAQ content was adapted from the version used in the GI election, with minor adjustments to reflect the study setup. A screenshot of the FAQ is retrievable in the appendix A.2 and also in the before mentioned video, illustrating the voting and study processes.

The FAQ was linked in the election invitation and aimed to serve as a fallback information source for voters who encountered confusion or uncertainty during the voting or verification process. It addressed common questions about accessing the voting system, eligibility to participate and technical troubleshooting. It is important to note that the FAQ was required in the *baseline* condition because the original GI election invitation did not explain how verification should be performed or how voters could report irregularities and instead referred all open questions to the FAQ. Moreover, the baseline verifier did not include a "report fraudulent activities" button, leaving participants without any reporting option. To avoid introducing systematic differences between conditions that might affect detection or reporting, and because the FAQ also contained general information (e.g., on data storage) that could be relevant to participants, we included an identical FAQ in the *improved* and *improved w/o guide* conditions as well.

Importantly, the FAQ was not presented as the primary channel for instructions or procedural guidance for the improved systems. Instead, it complemented the official invitation letter and the step-by-step guide, both of which were framed as the most authoritative sources.

Summary of Adjustments to the FAQ Page:

- **Content alignment.** Updated terminology and procedures to match those used in our study (e.g., "check your ballot" instead of "verify").
- **Role clarification.** Positioned the FAQ as a fallback, secondary support layer behind the invitation and step-by-step guide.

5 Results

We first provide a brief overview of our main findings before presenting detailed analyses for each dependent variable. Across measures, the *improved* system substantially increased detection and reporting rates compared to the *baseline*, particularly for the more subtle **verification prevention** manipulation. The addition of the step-by-step guide notably improved both detection and reporting for **verification prevention** manipulation, while having little to no effect for **vote tampering**. Usability analyses showed that

[6] Note, we requested a short description of the issue to differentiate genuine manipulation reports from reports related to technical problems or general uncertainty.

both improved systems were more effective (more participants completed verification) but required similar or slightly longer completion times. Trustworthiness generally declined over the course of the study, except for the *improved* system, which regained trust after debriefing. Perceived risks increased over time for most systems, with the *improved* system again standing out by showing a post-debriefing reduction in both integrity and secrecy risks.

The following subsections present each dependent variable in turn. To aid readability, results are structured as: (1) a descriptive summary of observed trends (with figures/tables where applicable), (2) inferential statistical results, and (3) a short *Takeaway* paragraph highlighting the main conclusions for that variable.

5.1 Manipulation Detection

Table 1 summarises the proportion of participants who correctly identified each manipulation type across the three systems (*baseline*, *improved*, and *improved w/o guide*).

Table 1: Percentage of participants detecting each manipulation type, by system.

	Vote Tampering	Verification Prevention
Baseline	61.6%	26.0%
Improved	74.7%	70.0%
Improved w/o Guide	76.7%	41.6%

RQ1.1$_{Detection}$ and **RQ1.2**$_{Detection}$: Comparing the *baseline* and *improved* systems, a significant improvement in detection was observed for **Verification Prevention** ($p < .001$, $V = 0.426$). For **Vote Tampering**, the difference was not significant ($p = .135$, $V = 0.125$), although the improved system achieved a higher raw detection rate (75% vs. 62%).
RQ3.1$_{Detection}$ and **RQ3.2**$_{Detection}$: Comparing the *improved* system with the *improved w/o guide* system revealed no significant effect of the guide for **Vote Tampering** ($p = .925$, $V = 0.008$); interestingly, the system without the guide had a slightly higher detection rate (77% vs. 75%). For **Verification Prevention**, however, the presence of the guide was associated with a significantly higher detection rate ($p = .001$, $V = 0.272$, small-to-medium effect).

Takeaway Detection

The *improved* system substantially enhanced detection for **Verification Prevention** manipulations, and the step-by-step guide provided a further benefit for this manipulation type.

5.2 Reporting

We measured reporting from two complementary perspectives:
Reporting$_{total}$ - proportion of all participants in a group who submitted a manipulation report, and
Reporting$_{detected}$ - proportion of participants who reported the manipulation among those who had detected it.

Table 2 presents reporting rates for both measures across the three systems and manipulation types.

Table 2: Proportion of participants who reported the manipulation, by system and manipulation type. Percentages in parentheses (marked with *) indicate the proportion among those who had detected the manipulation (*Reporting*$_{detected}$).

	Vote Tampering	Verification Prevention
Baseline	8.2% (13.3%*)	2.7% (5.3%*)
Improved	67.7% (83.0%*)	58.6% (81.6%*)
Improved w/o Guide	58.9% (73.2%*)	20.8% (43.8%*)

RQ2.1$_{Reporting}$ and **RQ2.2**$_{Reporting}$: Comparing the *baseline* and *improved* systems, *Reporting*$_{total}$ increased sharply for both **Vote Tampering** ($p < .001$, $V = 0.599$) and **Verification Prevention** ($p < .001$, $V = 0.593$).
RQ4.1$_{Reporting}$ and **RQ4.2**$_{Reporting}$: For **Vote Tampering**, differences in *Reporting*$_{total}$ between the *improved* and *improved w/o guide* system were not significant ($p = .363$, $V = 0.076$). However, for **Verification Prevention** the guide led to a significantly higher reporting rate ($p < .001$, $V = 0.373$).

When considering only those participants who detected the manipulation (*Reporting*$_{detected}$), effects became even more pronounced. Between the *baseline* and *improved* systems, reporting among detectors increased dramatically for both **Vote Tampering** ($p < .001$, $V = 0.695$) and **Verification Prevention** ($p < .001$, $V = 0.700$). Comparing the *improved* and *improved w/o guide* systems, there was no significant difference for **Vote Tampering** ($p = .651$, $V = 0.118$), but the guide again significantly improved reporting for **Verification Prevention** ($p = .0012$, $V = 0.393$), suggesting it is particularly helpful when the manipulation is less obvious.

We further analysed participants' explanations for not reporting despite having detected a manipulation (Table 3) and identified the following four themes (apart from a few miscellaneous responses):

- Correction: participants believed they had already resolved the issue by editing the image of the ballot.[7]
- Unawareness: participants were unaware that reporting was possible or did not know how to do it.[8]
- Irrelevance: participants considered reporting unnecessary, often because they saw the anomaly as unimportant in the study context.
- Misinterpretation: participants attributed the anomaly to benign causes, such as user error, bugs, or connection issues.

Overall, 45 participants in the *baseline*, 21 in the *improved*, and 44 in the *improved w/o guide* systems detected but did not report manipulations. Importantly, none of the participants in the *improved* system indicated that they were unaware of the possibility to report, whereas several in the other two systems explicitly stated that they would have wanted to report but did not know how. Across all three systems, nearly half of the non-reporting cases were due to misinterpretation, most often the assumption that the anomaly was caused by their own mistake.

[7] The ballot checking website only displayed an image of the ballot stored in the digital ballot box, which was modifiable (see Section 4.4). Importantly, altering the image did not affect the ballot in the digital ballot box, hence a vote could not be changed.

[8] As explained in Section 4.6 participants from the *baseline* were only able to learn about how detected manipulations can be reported if they visited the FAQ. In total 50 (34.2%) participants from the *baseline* visited the FAQ, compared to 6 (4.3%) from *improved* and 16 (10.7%) from *improved w/o guide*.

Table 3: Identified themes on reasons given by participants who detected but did not report a manipulation (absolute counts and percentages).

System	Correction	Unawareness	Irrelevance	Misinterpretation
Baseline (N = 45)	3 (6.7%)	12 (26.7%)	4 (8.9%)	22 (48.9%)
Improved (N = 21)	3 (14.3%)	0 (0.0%)	2 (9.5%)	10 (47.6%)
Improved w/o Guide (N = 44)	3 (6.8%)	9 (20.5%)	4 (9.1%)	19 (43.2%)

Takeaway Reporting

The *improved* system - and particularly the guide - substantially boosted reporting, especially for less obvious manipulations. Notably, in the systems without a guide, several participants detected a manipulation but felt unable to report it, whereas this problem did not occur in the *improved* system.

5.3 Usability and Confidence

To answer $\text{RQ5}_{\textit{Usability and Confidence}}$, we examined how the three systems differed not only in terms of *usability* but also in participants' *confidence in the election process*. We operationalised usability through the standard ISO dimensions-*effectiveness*, *efficiency*, and *satisfaction* — while confidence was captured via perceived *trustworthiness* of the system and perceived *risks* to vote integrity and secrecy. Together, these measures provide a complementary perspective on how voters experienced the systems beyond their ability to detect and report manipulations.

Usability. We first evaluated the three systems across effectiveness, efficiency, and satisfaction (Figure 9).

Effectiveness. Effectiveness was operationalised as the proportion of participants who completed both the voting and verification steps—i.e., reached the verification website page displaying their ballot.

An omnibus Chi² test indicated a statistically significant difference in effectiveness between the three systems, $\chi^2(2, N = 437) = 17.20$, $p < .001$, $V = 0.198$.

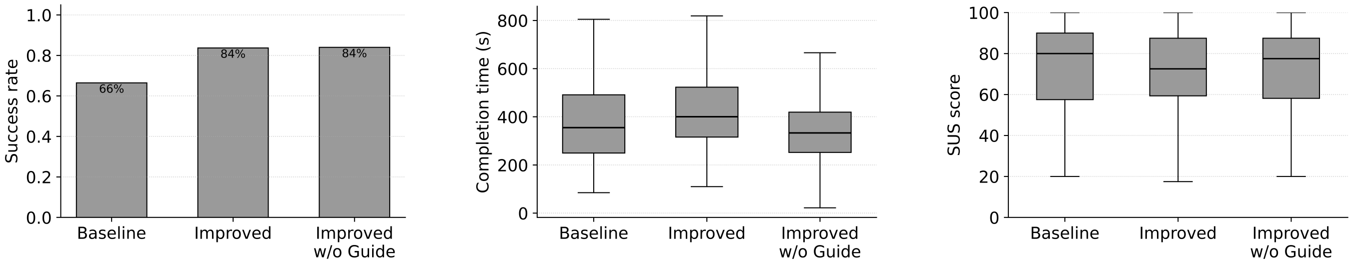
Post-hoc pairwise comparisons with Holm correction revealed that both the *improved* system ($M = 0.84$, $SD = 0.37$) and the *improved w/o guide* system ($M = 0.84$, $SD = 0.37$) achieved significantly higher effectiveness than the *baseline* system ($M = 0.66$, $SD = 0.47$; both $p_{\text{adj}} = .002$, $V = 0.191$ and $V = 0.196$, respectively). No difference was observed between the two improved systems ($p_{\text{adj}} = 1.000$, $V \approx 0$).

Efficiency. Efficiency was measured as the total time required to complete the full voting process, including reading the provided materials, casting the vote, and verifying the ballot. Only participants who completed verification were included in this analysis.

A Kruskal–Wallis test indicated a statistically significant difference in completion time between the three systems, $H(2) = 12.96$, $p = .0015$, $\varepsilon^2 = 0.032$.

Post-hoc pairwise comparisons with Holm correction showed no significant differences between the *baseline* system (M = 06:23, SD = 02:54) and the *improved* system (M = 06:50, SD = 02:26; $p_{\text{adj}} = .288$, $r = 0.133$) or between the *baseline* and *improved w/o guide* system (M = 05:41, SD = 02:07; $p_{\text{adj}} = .288$, $r = 0.112$). However, the *improved* system took significantly longer than the *improved w/o guide* system ($p_{\text{adj}} = .0034$, $r = 0.278$).

Satisfaction. Satisfaction was assessed using the System Usability Scale (SUS). Mean SUS scores were $M = 74.41$ ($SD = 19.57$) for the *baseline* system, $M = 70.89$ ($SD = 19.60$) for the *improved* system, and $M = 72.07$ ($SD = 19.03$) for the *improved w/o guide* system. A Kruskal–Wallis test found no statistically significant differences[9] between the three systems, $H(2) = 2.99$, $p = .224$, $\varepsilon^2 = 0.002$. Given the non-significant omnibus result, no pairwise post-hoc tests were interpreted.



(a) Effectiveness (success rate). **(b) Efficiency (completion time).** **(c) Satisfaction (SUS).**

Figure 9: Usability comparison across systems. (a) Effectiveness: Success rate increased from 66% in the *baseline* to 84% in both improved systems. (b) Efficiency: Completion times were similar across systems, with slightly longer times in the *improved* system due to reading the step-by-step guide. (c) Satisfaction: SUS scores remained comparable across all systems.

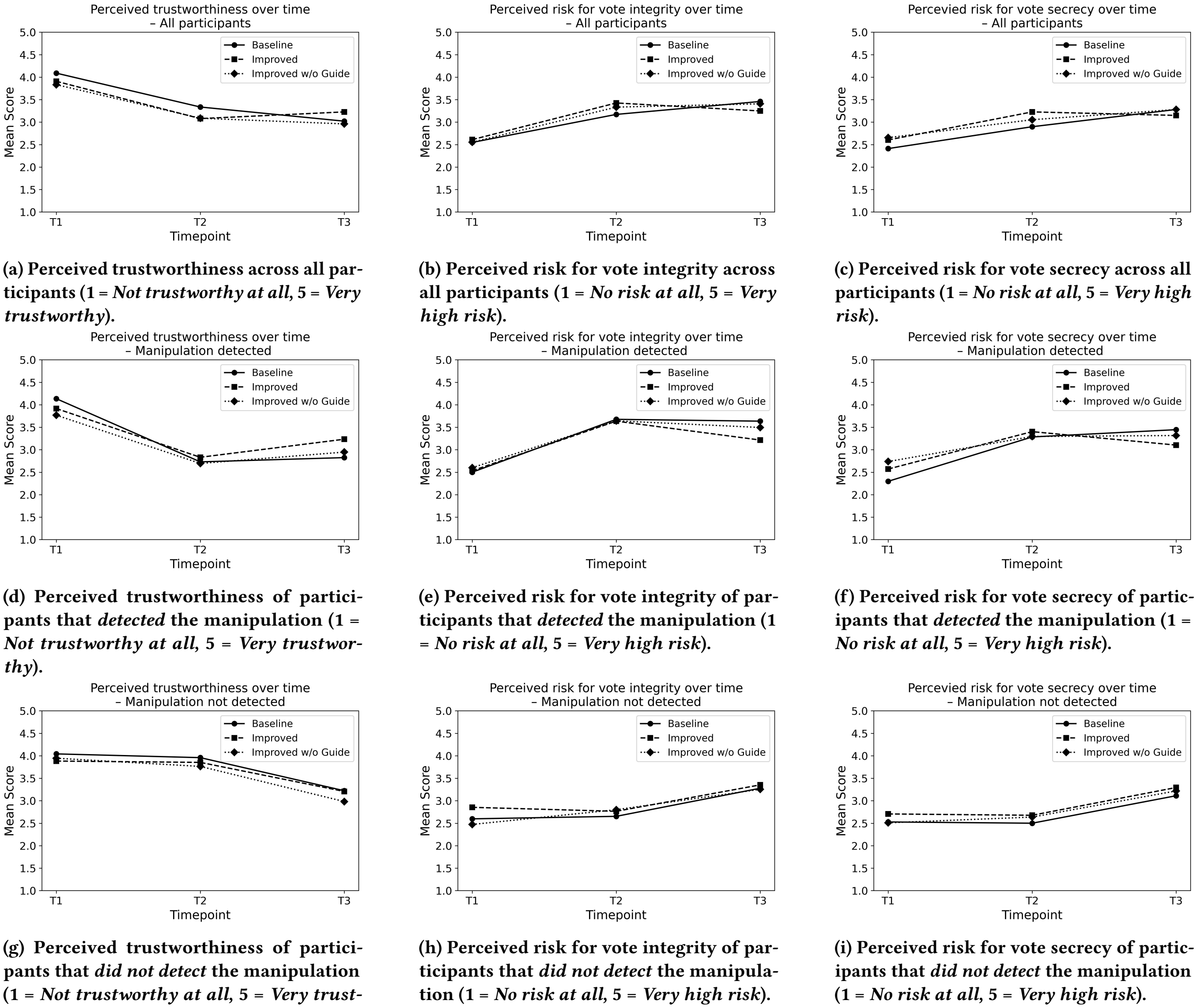
[9] Bootstrapped standard errors for group means were small, but observed mean differences were modest, resulting in overlapping confidence intervals and suggesting small differences in central tendency.

Takeaway Usability

Both improved systems significantly increased effectiveness (more participants completed vote + verification) relative to the *baseline*, while efficiency was broadly comparable—except that the *improved* system was slower than *improved w/o guide*. Satisfaction (SUS) was high and statistically indistinguishable across systems.

Confidence. We next examined perceived trustworthiness and perceived risks for vote integrity and secrecy at three time points: T1 after the first election, T2 after the manipulated election and T3 after debriefing. We measured confidence using five-point Likert scales (for perceived trustworthiness: 1 = *Not trustworthy at all*, 5 = *Very trustworthy*; for perceived risks: 1 = *No risk at all*, 5 = *Very high risk*). Figure 10 gives an overview of all participants (top row), those that detected the manipulation (middle row) and those that did not detect it (bottom row) w.r.t. to perceived confidence.

For each of the following metrics, we first describe the overall trends, then report statistical comparisons across systems for all participants, and finally examine differences between participants who detected the manipulation and those who did not at each of the three time points.



(a) Perceived trustworthiness across all participants (1 = *Not trustworthy at all*, 5 = *Very trustworthy*).

(b) Perceived risk for vote integrity across all participants (1 = *No risk at all*, 5 = *Very high risk*).

(c) Perceived risk for vote secrecy across all participants (1 = *No risk at all*, 5 = *Very high risk*).

(d) Perceived trustworthiness of participants that *detected* the manipulation (1 = *Not trustworthy at all*, 5 = *Very trustworthy*).

(e) Perceived risk for vote integrity of participants that *detected* the manipulation (1 = *No risk at all*, 5 = *Very high risk*).

(f) Perceived risk for vote secrecy of participants that *detected* the manipulation (1 = *No risk at all*, 5 = *Very high risk*).

(g) Perceived trustworthiness of participants that *did not detect* the manipulation (1 = *Not trustworthy at all*, 5 = *Very trustworthy*).

(h) Perceived risk for vote integrity of participants that *did not detect* the manipulation (1 = *No risk at all*, 5 = *Very high risk*).

(i) Perceived risk for vote secrecy of participants that *did not detect* the manipulation (1 = *No risk at all*, 5 = *Very high risk*).

Figure 10: Confidence measures across systems and three time points (T1 = after first election, T2 = after manipulated election, T3 = after debriefing), shown separately for participants who detected the manipulation (middle row) and those who did not (bottom row). Panels (a–c) show all participants, panels (d–f) only detectors, and panels (g–i) only non-detectors.

Perceived trustworthiness. Across all participants, *baseline* was perceived as most trustworthy at T1, followed by *improved* and *improved w/o guide*. At T2, trust declined for all systems, but *baseline* remained highest, with the two improved systems showing similar, lower ratings. At T3, trust in the *improved* system increased, whereas trust in *baseline* and *improved w/o guide* decreased further.

At T1 the omnibus test showed no overall differences between systems ($H = 5.851$, $p = .054$, $\varepsilon^2 = 0.009$), but *baseline* was rated significantly higher than *improved w/o guide* ($p = .015$, $r_{\text{rb}} = 0.149$). At T2, neither the omnibus test nor any pairwise comparison reached significance ($p > .05$). At T3, the omnibus test was again not significant ($H = 4.348$, $p = .114$, $\varepsilon^2 = 0.005$), but *improved* was rated significantly higher than *improved w/o guide* ($p = .048$, $r_{\text{rb}} = 0.129$). Considering detection status, at T1 the omnibus test among detectors was significant ($H = 6.958$, $p = .031$, $\varepsilon^2 = 0.018$). Within this group, *baseline* was rated higher than *improved w/o guide* ($p = .006$, $r_{\text{rb}} = 0.220$). Non-detectors showed no differences between the three systems ($p > .52$), and perceived trustworthiness did not differ between detectors and non-detectors ($p = .98$).

At T2, none of the omnibus tests for detectors or non-detectors were significant ($p \geq .61$), but participants who detected the manipulation reported lower trustworthiness than non-detectors ($p < .001$, $r_{\text{rb}} = 0.476$). This difference between detectors and non-detectors was also significant within each system ($p < .001$ for *baseline*, *improved*, and *improved w/o guide*).

At T3, the omnibus test among detectors was significant ($H = 6.113$, $p = .047$, $\varepsilon^2 = 0.015$). Pairwise comparisons showed that *improved* was rated significantly higher than *baseline* ($p = .0249$, $r_{\text{rb}} = 0.190$). Among non-detectors, no differences between systems emerged ($p = .48$). Detection status did not affect trustworthiness overall at T3 ($p = .38$), but the difference between detectors and non-detectors reached significance for *baseline*, with detectors rating it lower than non-detectors ($p = .035$, $r_{\text{rb}} = 0.197$). No such differences occurred for the two improved systems ($p \geq .38$).

Perceived risk for vote integrity. Across all participants, perceived integrity risk was similar across systems at T1. At T2, risk increased for all systems, with *improved* showing the highest perceived risk and *baseline* the lowest. At T3, risk decreased for *improved* but remained stable (*Improved w/o guide*) or increased slightly (*baseline*), resulting in *improved* having the lowest perceived risk for integrity at the final time point.

At T1, no omnibus or pairwise differences emerged ($H = 0.091$, $p = .956$). Both at T2 ($H = 3.059$, $p = .217$) and T3, no omnibus or pairwise test reached significance ($H = 2.999$, $p = .223$). Considering detection status, at T1 neither detectors ($H = 0.593$, $p = .743$) nor non-detectors ($H = 2.829$, $p = .243$) showed differences between systems. Integrity risk did not differ between detectors and non-detectors overall ($p = .311$), and no system-specific comparisons were significant ($p > .10$).

At T2, the omnibus tests for detectors ($H = 0.337$, $p = .845$) and non-detectors ($H = 0.588$, $p = .745$) were not significant . However, detectors reported a higher perceived risk for vote integrity risk than non-detectors overall ($p < .001$, $r_{\text{rb}} = 0.438$), a pattern that was significant within each system (*baseline*: $p < .001$, $r_{\text{rb}} = 0.462$; *improved*: $p < .001$, $r_{\text{rb}} = 0.431$; *improved w/o guide*: $p < .001$, $r_{\text{rb}} = 0.402$).

At T3, omnibus tests were again non-significant for detectors and non-detectors (detectors: $H = 7.443$, $p = .024$; non-detectors: $H = 0.321$, $p = .852$). However, *baseline* was rated higher than *improved* ($p = .0088$, $r_{\text{rb}} = 0.222$). Detection status did not result in difference to perceived risk for vote integrity overall ($p = .243$), but within *baseline* detectors reported significantly higher risk than non-detectors ($p = .041$, $r_{\text{rb}} = 0.188$). No such differences occurred for the improved systems ($p \geq .42$).

Perceived risk for vote secrecy. Across all participants, the perceived risk for vote secrecy at T1 was lowest for *baseline* and slightly higher for the two improved systems. At T2, it increased for all systems, with *improved* showing the highest ratings. At T3, the perceived risk decreased for *improved* but increased further for *baseline* and *improved w/o guide*.

At T1, the omnibus test was not significant ($H = 4.731$, $p = .094$), but *baseline* showed significantly lower perceived risk than *improved w/o guide* ($p = .034$, $r_{\text{rb}} = 0.134$). At T2, the omnibus test was significant ($H = 6.589$, $p = .037$). Pairwise comparisons showed that *baseline* was rated significantly lower than *improved* ($p = .010$, $r_{\text{rb}} = 0.170$). At T3, the omnibus test was not significant ($H = 1.302$, $p = .522$), and no pairwise differences emerged ($p > .29$). Considering detection status, at T1 neither detectors ($H = 9.353$, $p = .0093$) nor non-detectors ($H = 0.972$, $p = .615$) showed system differences. Only for detectors participants from the *baseline* perceived less risk compared to participants from *improved w/o guide* contrast was significant ($p = .0018$, $r_{\text{rb}} = 0.264$). Detectors and non-detectors did not differ overall at T1 ($p = .860$).

At T2, at T1 neither detectors ($H = 0.593$, $p = .743$) nor non-detectors ($H = 1.066$, $p = .587$) showed differences between systems. However, detectors reported significantly higher perceived risk for vote secrecy compared to non-detectors ($p < .001$, $r_{\text{rb}} = 0.374$). This difference was significant within each system (*baseline*: $p < .001$, $r_{\text{rb}} = 0.368$; *improved*: $p = .0004$, $r_{\text{rb}} = 0.388$; *improved w/o guide*: $p = .0007$, $r_{\text{rb}} = 0.323$).

At T3, the perceived risk did not differ between systems among detectors or non-detectors (detectors: $H = 4.695$, $p = .096$; non-detectors: $H = 0.916$, $p = .633$). Detection status did not influence the perceived risk overall ($p = .472$), but within the *baseline* system detectors reported higher secrecy risk than non-detectors ($p = .046$, $r_{\text{rb}} = 0.184$). No such differences occurred for the improved systems ($p \geq .33$).

Takeaway Confidence

Across all participants, confidence changed markedly over time. Trustworthiness declined after the manipulated election and only the *improved* system showed a recovery after debriefing. Perceived risks for vote integrity and secrecy increased after the manipulation, with only the *improved* system showing reduced risks again at T3. When considering detection status, strong differences emerged at T2: participants who detected the manipulation reported lower trustworthiness and higher perceived risks for integrity and secrecy than non-detectors, both overall and within each system. By T3, these detection-related differences largely disappeared, with only small residual effects remaining in the *baseline* system.

6 Discussion

Our study highlights three main findings. The most important is the clear distinction between detection and reporting of manipulations: only reported anomalies can be acted upon by election officials, but detection and reporting rates differ significantly. We also show the beneficial role of a step-by-step guide in improving both detection and reporting, particularly for subtle **Verification Prevention** manipulations. Finally, we find that our improvements had mixed effects on usability and confidence, adding minor efficiency costs but improving effectiveness and supporting trust recovery after manipulations. We discuss these findings in two parts: first, detection and reporting of manipulations and the role of the step-by-step guide, and second, usability and confidence. Each subsection ends with a short takeaway that synthesises the implications for research and real-world elections.

6.1 Detecting and Reporting Manipulations

Comparison to prior cast–then–audit studies (and why detection rates differ). Marky et al. examined only **vote tampering** and reported a 64% detection rate [31], very close to our baseline. As in their study, many participants either abstained from verification or did not recognise the anomaly, underscoring that just providing a verifier does not guarantee its effective use. Both our improved conditions yielded substantially higher detection rates, which likely stems from incorporating Marky's recommendation to explain *why verification is necessary*, as their participants often skipped verification entirely in the absence of such framing. We addressed this by incorporating textual cues of an analogy and norm based on research by Olembo et. al [36]. Hilt et al. investigated both **tampering** and **verification prevention** [18]. For **vote tampering**, they reported a strikingly high 96% detection rate, which can be explained by their methodological choice to restrict the manipulated voting phase to participants who had already verified in an earlier, unmanipulated voting phase and by explicitly instructing them to verify their vote in the study material. For **verification prevention**, by contrast, their detection rate was only 24%, compared to the 42% detection rate of the *improved w/o guide* system, which corresponds to the system they used. This difference is very likely attributable to the different study setup: Hilt et al. separated trials by two weeks, potentially diminishing memory of the verification process, while we chose immediate successive elections. We chose to perform the two elections successively to give participants a realistic chance to still have a complete understanding and recollection of the intended processes, especially for participants that did not receive the step-by-step guide.

*System improvements (**H1**$_{Detection}$, **H2**$_{Reporting}$).* Relative to a realistic baseline, our improved system substantially increased both *detection* and *reporting* of manipulations. The gains were especially strong for **verification prevention**, a difficult to detect manipulation that provides no conflicting content to inspect. Detection of **verification prevention** rose with a *medium* effect ($V = 0.426, p < .001$), while reporting increased with *large* effects for both manipulation types ($V = 0.599, p < .001$ for **vote tampering**; $V = 0.593, p < .001$ for **verification prevention**). These results align with usable-security research that examined design cues to change verification behaviour [36] and clearer feedback to discover risks that users would otherwise miss or dismiss [10]. In election processes, reporting is the decisive endpoint. A detected anomaly that is never escalated cannot trigger remediation.

*Added value of the guide (**RQ1**$_{Detection}$, **RQ2**$_{Reporting}$).* Comparing the *improved* with the *improved w/o guide* system isolates the effect of the step-by-step guide. For **vote tampering** no significant differences wrt. to detection and reporting were found. In contrast, for **verification prevention** the guide significantly improved detection ($V = 0.272, p = .001$) and reporting ($V = 0.373, p < .001$). This suggests the guide functions as a *decision aid at ambiguity points*. We assume that the If-Then-Else constructions (e.g., *Check your ballot! "If your ballot is incorrect: Report"*) combined with easy to follow instructions played a central role, confirming findings by Stevens et al. who demonstrated that providing structured playbooks enhanced participants incident response efforts [45]. We further assume that having clear wording and symbols (e.g. as a red arrow pointing to a warning symbol at points of potential manipulation), attributed to the beneficial effect of the step-by-step guide, following the recommendation by Felt et al. [10] of integrating opinionated design with visual cues. Analysing responses on why people chose to not report we found that nine participants from the *improved w/o guide* systems stated unawareness as their reason to not report. Notably, *no* participant in the *improved* system cited unawareness of reporting channels, further emphasizing the benefit of the step-by-step guide. Finally, it is important to note a structural disadvantage of the *baseline* condition: participants needed to consult the FAQ to learn how to report irregularities. If they accessed the FAQ on the same (potentially compromised) voting device a malicious client could also manipulate the FAQ content, potentially suppressing or altering reporting information. This could further decrease reporting rates in the *baseline* system and illustrates the value of providing trusted, device-independent guidance and reporting instructions.

The Detection–Reporting Gap (and why it matters). Consistent with broader usable-security literature on underreporting, e.g., phishing sites receiving many visits before the first report [34] or people being hesitant to report despite suspicion [8], we observe a noticeable gap between detecting and reporting a manipulation (e.g. 62% of participants of the *baseline* system *detected* the manipulation, but only 8% *reported* it). This issue becomes especially pressing when examining only voters who indicated that they have detected the manipulation: in the baseline system only 13% of people detecting the **tampering** also reported it and only 5% of people that detected the **prevention** manipulation reported it. The overall improvements, as shown by the *improved w/o guide*, substantially increased these rates to 73% and 44% respectively. Again, the addition of the step-by-step guide further improved these rates to 82% and 83% clearly demonstrating the beneficial effect of easy-to-follow instructions as shown by Stevens et al. [45].

Even with these improvements, reporting did not reach 100%. While Kulyk et al. originally framed their "lacks" as explanations for why voters fail to verify [22], our findings suggest the same categories may help explain why not all detectors proceed to report. For example, a *lack of perseverance* may mean some participants noticed the manipulation but got sidetracked before reporting; a *lack of compulsion* could reflect perceptions that reporting was unnecessary or too much effort; and a *lack of self-efficacy* may stem from the belief that reporting would be ineffective against a powerful adversary.

Qualitative explanations provided by participants support this interpretation. Some described *unawareness* of how or where to report, while others *misinterpreted* the anomaly as their own mistake or as a technical issue, replicating findings from broader security contexts, where users often recognise something might be wrong but fail to act without a clear, immediate next step [40].

Detecting (and consequently reporting) the **verification prevention** attack requires in-depth knowledge of the voting process, since one must notice subtle deviations from the intended protocol. As the time between elections can vary, it cannot be expected that the electorate retains such detailed knowledge. Without a step-by-step guide, lay users and non-expert voters have very little chance to detect this manipulation. The high detection rate (70%) in our *improved* system further emphasises the importance of structured guidance in surfacing this otherwise hard-to-detect manipulation.

Furthermore, as detection and reporting are distinct measurements with different implications (a problem can only be addressed if it is also reported), conflating them (as in prior e-voting user studies [18, 31]) risks overestimating real-world readiness.

Takeaway Detecting and Reporting Manipulations

There is a clear distinction between detecting and reporting manipulations: only reported manipulations can be acted upon by election officials, and our study shows that these numbers can differ substantially. Improving components of the electoral process (e.g. invitation, UI, verifier) substantially improves detecting and reporting abilities with the step-by-step guide playing a central role, especially for harder to detect verification prevention manipulations. Such manipulations require in-depth knowledge of the intended processes and flows which can not be expected by the average voter. Consequently, we recommend that future research on individual verifiability should include a detailed step-by-step guide. As past studies did not differentiate between detecting and reporting manipulations, we advise to revisit these studies to not falsely overestimate real-world readiness of these implementations based on (likely) inflated numbers. For election organisers, the importance of integrating a dedicated reporting channel is clear: Without actionable reporting, even detected anomalies may remain invisible to officials, jeopardising electoral integrity. We therefore recommend implementing an independent, out-of-band reporting channel accessible to all voters, and pairing it with a clear, step-by-step guide. Together, these measures can strengthen both the technical and human layers of individual verifiability.

6.2 Usability and Confidence

Usability (effectiveness, efficiency, satisfaction). Both, the *improved* and *improved w/o guide*, systems increased effectiveness, as more participants checked their ballot after casting, compared to the *baseline* system (both 84% vs. 66%), while efficiency (*baseline* 6:23; *improved* 6:50; *improved w/o guide* 5:41) and satisfaction scores (*baseline* 74.41; *improved* 70.89; *improved w/o guide* 72.07) remained broadly comparable. Participants from the *improved* system took slightly longer, than the participants from the *improved w/o guide* system, which was expected, as participants from the *improved* system needed to read an additional document (the step-by-step guide). These findings align with Petelka et al.'s work on phishing warning design, which demonstrated that relocating warnings closer to the areas under inspection (potential phishing links) and enforcing focused attention significantly improves compliance with security actions (click-through avoidance) [38]. In our context, structuring guidance so that relevant instructions are embedded at every decision point appears to similarly support completion of verification workflows without diminishing perceived usability.

Confidence (perceived trustworthiness and risks over time). Confidence proved dynamic across study phases. After the manipulated election, perceived trustworthiness dropped and perceived risks rose slightly across all systems; an expected reaction to adversarial events. Notably, only the *improved* system showed recovery at debriefing, with trustworthiness rebounding and the perceived risks, w.r.t. an undetected breach of vote integrity or vote secrecy, decreasing. This suggests that explicit, action-oriented instructions (e.g. what to look for, what to do, and how to respond) can support not only in-the-moment behaviour but also post-incident attitudes, echoing findings from Stevens et al. that actionable guidance can help restore confidence after failures [45]. Across all conditions, however, overall confidence decreased from the start to the end of the study, reflecting the lasting impact of exposure to manipulations.

The distinction between participants who detected the manipulation and those who did not became most pronounced at T2, with large effect sizes across all confidence measures. This is unsurprising: detectors directly experienced that the system could be tampered with and, consequently, expressed lower trust and higher perceived risks. Interestingly, both perceived risk for vote integrity and vote secrecy increased to a similar extent. This suggests that participants recognised that "something" was wrong but were not necessarily able to attribute the problem specifically to integrity or secrecy. If an attacker can change a vote, it is reasonable for voters to infer that the attacker might also learn it; hence increases in both risk dimensions are plausible.

Across systems, participants in the *baseline* system generally expressed higher confidence at time points T1 and T2. However, at T3 participants from the *improved* system expressed the highest confidence. Two factors may contribute to this pattern. First, participants in the two improved systems received more explicit warnings and instructions about how tampering might manifest and how they can be reported to the authorities. While this likely facilitated the substantially higher reporting rates, such warnings may have also made vulnerabilities more salient and thus reduced confidence. Comparing *improved* with *improved w/o guide* one can see that the existence of the step-by-step guide lead to more confidence, showing that the existence of expressive warnings may reduce confidence but coupled with a clear step-by-step guide providing aid for the situations these warning refer to leads to recovery. Second, the proportion of non-detectors was highest in the *baseline* system, and non-detectors consistently expressed more confidence than detectors. Both mechanisms can help explain why the *baseline* system appears more trusted overall. Taken together, these results highlight a potential trade-off: systems that guide users more explicitly may increase detection and reporting, but also risk reducing overall confidence by making threats more salient -unless coupled with clear, actionable guidance that helps voters navigate those threats confidently.

Security–usability trade-offs. The *improved* system trades a modest amount of *efficiency* for substantial gains where it matters most: detection and reporting of manipulations. While the *improved* system trades a modest amount of efficiency for higher detection performance, our findings also suggest a parallel trade-off in confidence: more explicit guidance can increase awareness of manipulation risks, which may slightly lower perceived trust in the system. Prior e-voting research shows that voters are willing to accept such trade-offs: Kulyk et al. found that participants preferred the most secure but least usable system over less secure alternatives [21]. Furthermore, in a different study Kulyk et al. [23] found that different security interventions ranging from light to "high alarming interventions" lead to reduced voter confidence but their willingness to further use the internet voting system in question did not decrease. Our results similarly indicate that small usability costs are outweighed by security benefits in high-stakes contexts such as elections. Given that national elections are relatively infrequent, the small additional effort required from voters appears acceptable if it meaningfully strengthens electoral integrity.

Takeaway Usability and Confidence

Improved micro-structure and clear instructions strengthened effectiveness without reducing satisfaction, and enabled participants in the *improved* system to regain confidence after a manipulated election. While explicit, action-oriented guidance substantially increased manipulation detection and reporting, it also made potential threats more salient, which can slightly reduce overall confidence, particularly among those who notice irregularities. Security-critical domains such as e-voting can tolerate minor efficiency costs and modest decreases in perceived confidence when these lead to markedly higher detection performance. Designing guidance that supports both reliable detection and sustained trust remains an important direction for future systems and deployments.

7 Conclusion

This paper examined individual verifiability in internet voting through a large-scale online user study with 437 participants, focusing on the cast–then–audit approach. By recreating a realistic baseline based on a real election, we were able to show how voters react to manipulations in a setting close to practice. We compared this baseline to an improved system with a detailed step-by-step guide supporting voters during casting and verification, as well as a system without the guide to isolate its effect.

Our results highlight three main points. First, *detection and reporting are distinct behaviours* and should be measured separately, since only reported anomalies can inform election officials. Second, a *step-by-step guide* enables voters to recognise and report manipulations that would otherwise go unnoticed. Third, while explicit warnings make threats more salient and can slightly lower perceived trust, pairing them with clear, actionable guidance helps mitigate this effect. The improvements introduced only modest efficiency costs, which appear acceptable in this high-stakes context given the substantial gains in detection, reporting, and confidence recovery.

For future research, the baseline we establish can serve as a point of comparison for future improvements. Past studies should also be revisited to correctly separate detection from reporting, and future work should consider step-by-step guidance for the different individual verifiability approaches, that have not yet been examined w.r.t. to the addition of step-by-step guidance, i.e. tracking code based approaches and cast-or-audit approaches. For practice, our findings underline the importance of independent reporting channels and clear guidance materials that support voters at ambiguity points.

Strengthening the security of internet voting is not only about adding more cryptography, but about ensuring that voters can, and will, use the mechanisms already in place.

Acknowledgments

This work was supported by funding from the project "Engineering Secure Systems" of the Helmholtz Association (HGF) [subtopic 46.23.01 Methods for Engineering Secure Systems] and by KASTEL Security Research Lab. Furthermore, this research was supported by Polyas by providing the voting systems free of charge. Finally, we would like to thank Philipp Matheis (KIT), Tobias Länge (KIT) and Tomasz Truderung (Polyas) for their valuable help in setting up the different voting systems and technical infrastructure.

References

[1] Claudia Z. Acemyan, Philip Kortum, Michael D. Byrne, and Dan S. Wallach. 2014. Usability of Voter Verifiable, End-to-end Voting Systems: Baseline Data for Helios, {Prêt} {à} Voter, and Scantegrity {II}. *USENIX Journal of Election Technology and Systems (JETS)* 2 (2014), 26–56. https://www.usenix.org/jets/issues/0203/acemyan ISBN: 9781931971140.

[2] Josh Benaloh. 2006. Simple Verifiable Elections. In *2006 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT 06)*. USENIX Association, Vancouver, B.C., 5–5. https://www.usenix.org/conference/evt-06/simple-verifiable-elections

[3] Josh Benaloh, Ronald Rivest, Peter Y. A. Ryan, Philip Stark, Vanessa Teague, and Poorvi Vora. 2015. End-to-end verifiability. doi:10.48550/arXiv.1504.03778 arXiv:1504.03778 [cs].

[4] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (1995), 4–7.

[5] German Bundestag. n.d.. *Election of Members of the German Bundestag.* https://www.bundestag.de/en/parliament/elections/arithmetic/arithmetic-199936 last accessed 28.11.2025.

[6] Michael Kirsten Christoph Niederbudde. 2025. *Github - polyas-core3-second-device-verification.* https://github.com/kastel-security/polyas-core3-second-device-verification last accessed 01.09.2025.

[7] Aditya Damodaran, Simon Rastikian, Peter B. Rønne, and Peter Y. A. Ryan. 2024. Hyperion: Transparent End-to-End Verifiable Voting with Coercion Mitigation. https://eprint.iacr.org/2024/1182 Publication info: Preprint..

[8] Verena Distler. 2023. The Influence of Context on Response to Spear-Phishing Attacks: an In-Situ Deception Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3544548.3581170

[9] Verena Distler, Marie-Laure Zollinger, Carine Lallemand, Peter B. Roenne, Peter Y. A. Ryan, and Vincent Koenig. 2019. Security - Visible, Yet Unseen?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300835

[10] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings: Comprehension and Adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association

for Computing Machinery, New York, NY, USA, 2893–2902. doi:10.1145/2702123.2702442

[11] Gesellschaft für Informatik e.V. 2025. *Gesellschaft für Informatik - past elections.* https://gi.de/ueber-uns/organisation/wahlen last accessed 08.09.2025.

[12] David Galindo, Sandra Guasch, and Jordi Puiggalí. 2015. 2015 Neuchâtel's Cast-as-Intended Verification Mechanism. In *Proceedings of the 5th International Conference on E-Voting and Identity - Volume 9269 (VoteID 2015)*. Springer-Verlag, Berlin, Heidelberg, 3–18. doi:10.1007/978-3-319-22270-7_1

[13] Ed Gerck, C. Andrew Neff, Ronald L. Rivest, Aviel D. Rubin, and Moti Yung. 2002. The Business of Electronic Voting. In *Financial Cryptography*, Paul Syverson (Ed.). Springer, Berlin, Heidelberg, 243–268. doi:10.1007/3-540-46088-8_21

[14] Rolf Haenni, Reto Koenig, Philipp Locher, and Eric Dubuis. 2017. CHVote System Specification.

[15] Rolf Haenni, Reto E. Koenig, Philipp Locher, and Eric Dubuis. 2017. CHVote Protocol Specification. Cryptology ePrint Archive, Paper 2017/325. https://eprint.iacr.org/2017/325

[16] Sven Heiberg and Jan Willemson. 2014. Verifiable internet voting in Estonia. In *2014 6th International Conference on Electronic Voting: Verifying the Vote (EVOTE)*. IEEE, Lochau / Bregenz, Austria, 1–8. doi:10.1109/EVOTE.2014.7001135

[17] Jörg Helbach and Jörg Schwenk. 2007. Secure Internet Voting with Code Sheets. In *E-Voting and Identity*, Ammar Alkassar and Melanie Volkamer (Eds.). Springer, Berlin, Heidelberg, 166–177. doi:10.1007/978-3-540-77493-8_15

[18] Tobias Hilt, Benjamin Berens, Tomasz Truderung, Margarita Udovychenko, Stephan Neumann, and Melanie Volkamer. 2024. Systematic User Evaluation of a Second Device Based Cast-as-Intended Verifiability Approach. In *Financial Cryptography and Data Security. FC 2024 International Workshops: Voting, DeFI, WTSC, CoDecFin, Willemstad, Curaçao, March 4–8, 2024, Revised Selected Papers.* Springer-Verlag, Berlin, Heidelberg, 33–49. doi:10.1007/978-3-031-69231-4_3

[19] Tobias Hilt, Oksana Kulyk, and Melanie Volkamer. 2023. German Social Elections in 2023. In *Informatik 2024 - Lock-in or log out? Wie digitale Souveränität gelingt.* Gesellschaft für Informatik, Bonn, Bonn, 10.18420/e. https://dl.gi.de/handle/20.500.12116/45434 ISSN: 1617-5468.

[20] Oksana Kulyk, Jan Henzel, Karen Renaud, and Melanie Volkamer. 2019. Comparing "Challenge-Based" and "Code-Based" Internet Voting Verification Implementations. In *Human-Computer Interaction – INTERACT 2019*, David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.). Springer International Publishing, Cham, 519–538. doi:10.1007/978-3-030-29381-9_32

[21] Oksana Kulyk, Stephan Neumann, Jurlind Budurushi, and Melanie Volkamer. 2017. Nothing Comes for Free: How Much Usability Can You Sacrifice for Security? *IEEE Security & Privacy* 15, 3 (2017), 24–29. doi:10.1109/MSP.2017.70

[22] Oksana Kulyk and Melanie Volkamer. 2018. Usability is not Enough: Lessons Learned from 'Human Factors in Security' Research for Verifiability. https://eprint.iacr.org/2018/683 Publication info: Published elsewhere. 3rd International Joint Conference on Electronic Voting (E-Vote-ID 2018).

[23] Oksana Kulyk, Melanie Volkamer, Niklas Fuhrberg, Benjamin Berens, and Robert Krimmer. 2022. German Voters' Attitudes Towards Voting Online with a Verifiable System. In *Financial Cryptography and Data Security. FC 2022 International Workshops: CoDecFin, DeFi, Voting, WTSC, Grenada, May 6, 2022, Revised Selected Papers* (Grenada, Grenada). Springer-Verlag, Berlin, Heidelberg, 335–350. doi:10.1007/978-3-031-32415-4_23

[24] Oksana Kulyk, Melanie Volkamer, Monika Müller, and Karen Renaud. 2020. Towards Improving the Efficacy of Code-Based Verification in Internet Voting. In *Financial Cryptography and Data Security : FC 2020 International Workshops, AsiaUSEC, CoDeFi, VOTING, and WTSC, Kota Kinabalu, Malaysia, February 14, 2020, [1st Asian Workshop on Usable Security, AsiaUSEC 2020, the 1st Workshop on Coordination of Decentralized Finance, CoDeFi 2020, the 5th Workshop on Advances in Secure Electronic Voting, VOTING 2020, and the 4th Workshop on Trusted Smart Contracts, WTSC 2020, held at the 24th International Conference on Financial Cryptography and Data Security, FC 2020; Kota Kinabalu; Malaysia; 14 February 2020 through 14 February 2020], Revised Selected Papers. Ed.: Matthew Bernhard.* Springer, Kota Kinabalu, Sabah, Malaysia, 291. doi:10.1007/978-3-030-54455-3_21 ISSN: 0302-9743.

[25] Ralf Kusters, Tomasz Truderung, and Andreas Vogt. 2012. Clash Attacks on the Verifiability of E-Voting Systems. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy (SP '12)*. IEEE Computer Society, USA, 395–409. doi:10.1109/SP.2012.32

[26] Michael LaFleur. 2025. *Why Underreporting Holds Back Fraud Prevention.* https://www.threatmark.com/why-underreporting-holds-back-fraud-prevention last accessed 18.08.2025.

[27] Luigi Lo Iacono, Matthew Smith, Emanuel Von Zezschwitz, Peter Leo Gorski, and Peter Nehren. 2018. Consolidating Principles and Patterns for Human-centred Usable Security Research and Development. In *Proceedings 3rd European Workshop on Usable Security*. Internet Society, London, England, –. doi:10.14722/eurousec.2018.23010

[28] Ioana Andreea Marin, Pavlo Burda, Nicola Zannone, and Luca Allodi. 2023. The Influence of Human Factors on the Intention to Report Phishing Emails. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3544548.3580985

[29] Karola Marky, Oksana Kulyk, Karen Renaud, and Melanie Volkamer. 2018. What Did I Really Vote For? On the Usability of Verifiable E-Voting Schemes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173750

[30] Karola Marky, Verena Zimmermann, Markus Funk, Jörg Daubert, Kira Bleck, and Max Mühlhäuser. 2020. Improving the Usability and UX of the Swiss Internet Voting Interface. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376769

[31] Karola Marky, Marie-Laure Zollinger, Peter Roenne, Peter Y. A. Ryan, Tim Grube, and Kai Kunze. 2021. Investigating Usability and User Experience of Individually Verifiable Internet Voting Schemes. *ACM Trans. Comput.-Hum. Interact.* 28, 5, Article 30 (Sept. 2021), 36 pages. doi:10.1145/3459604

[32] Florian Moser, Johannes Müller, Veronique Cortier, Alexandre Debant, Pierrick Gaudry, Anselme Goetschmann, Ralf Küsters, and Melanie Volkamer. 2024. A Study of Mechanisms for End-to-End Verifiable Online Voting. *BSI* – (2024), –.

[33] Christina Nissen, Oksana Kulyk, Melanie Volkamer, Lara Elisabeth Fredrich, and Helena Hermansen. 2024. Tracking Code-based Verification - Design and Evaluation. In *E-Vote-ID 2024*. Gesellschaft für Informatik, Bonn, 85–100. doi:10.18420/e-vote-id2024_06

[34] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. 2020. Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale. In *29th USENIX Security Symposium (USENIX Security 20)*, Vol. 29. USENIX Association, –, 361–377. https://www.usenix.org/conference/usenixsecurity20/presentation/oest-sunrise

[35] Maina M. Olembo, Steffen Bartsch, and Melanie Volkamer. 2013. Mental Models of Verifiability in Voting. In *E-Voting and Identify*, James Heather, Steve Schneider, and Vanessa Teague (Eds.). Springer, Berlin, Heidelberg, 142–155. doi:10.1007/978-3-642-39185-9_9

[36] M. Maina Olembo, Karen Renaud, Steffen Bartsch, and Melanie Volkamer. 2014. Voter, What Message Will Motivate You To Verify Your Vote?. In *Proceedings 2014 Workshop on Usable Security*. Internet Society, San Diego, CA, –. doi:10.14722/usec.2014.23038

[37] Olivier Pereira. 2021. Individual Verifiability and Revoting in the Estonian Internet Voting System. https://eprint.iacr.org/2021/1098 Publication info: Preprint. MINOR revision..

[38] Justin Petelka, Yixin Zou, and Florian Schaub. 2019. Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3290605.3300748

[39] Swiss Post. n.d.. *E-voting - Electronic vote casting for Switzerland.* https://www.post.ch/en/about-us/profile/swiss-post-and-politics/swiss-post-in-the-digital-world/e-voting-electronic-vote-casting-for-switzerland 18.08.2025.

[40] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3174086

[41] Peter Y. A. Ryan, Peter B. Roenne, and Vincenzo Iovino. 2015. Selene: Voting with Transparent Verifiability and Coercion-Mitigation. https://eprint.iacr.org/2015/1105 Publication info: Preprint. MINOR revision..

[42] Michael Schläpfer and Melanie Volkamer. 2012. The secure platform problem taxonomy and analysis of existing proposals to address this problem. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance (ICEGOV '12)*. Association for Computing Machinery, New York, NY, USA, 410–418. doi:10.1145/2463728.2463807

[43] Carsten Schürmann. 2024. Social Elections. In *E-Vote-ID 2024*. Gesellschaft für Informatik, Bonn, 127–139. https://dl.gi.de/handle/20.500.12116/45868

[44] Drew Springall, Travis Finkenauer, Zakir Durumeric, Jason Kitcat, Harri Hursti, Margaret MacAlpine, and J. Alex Halderman. 2014. Security Analysis of the Estonian Internet Voting System. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Scottsdale Arizona USA, 703–715. doi:10.1145/2660267.2660315

[45] Rock Stevens, Daniel Votipka, Josiah Dykstra, Fernando Tomlinson, Erin Quartararo, Colin Ahern, and Michelle L. Mazurek. 2022. How Ready is Your Ready? Assessing the Usability of Incident Response Playbook Frameworks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3491102.3517559

[46] Sarah Tabassum, Cori Faklaris, and Heather Richter Lipford. 2024. What Drives SMiShing Susceptibility? A U.S. Interview Study of How and Why Mobile Phone Users Judge Text Messages to be Real or Fake. In *Twentieth Symposium on Usable*

Privacy and Security (SOUPS 2024). USENIX Association, Philadelphia, PA, USA, 393–411. https://www.usenix.org/conference/soups2024/presentation/tabassum-sarah

[47] Paul Tim Thürwächter, Melanie Volkamer, and Oksana Kulyk. 2022. Individual Verifiability with Return Codes: Manipulation Detection Efficacy. In *Electronic Voting*, Robert Krimmer, Melanie Volkamer, David Duenas-Cid, Peter Rønne, and Micha Germann (Eds.). Springer International Publishing, Cham, 139–156. doi:10.1007/978-3-031-15911-4_9

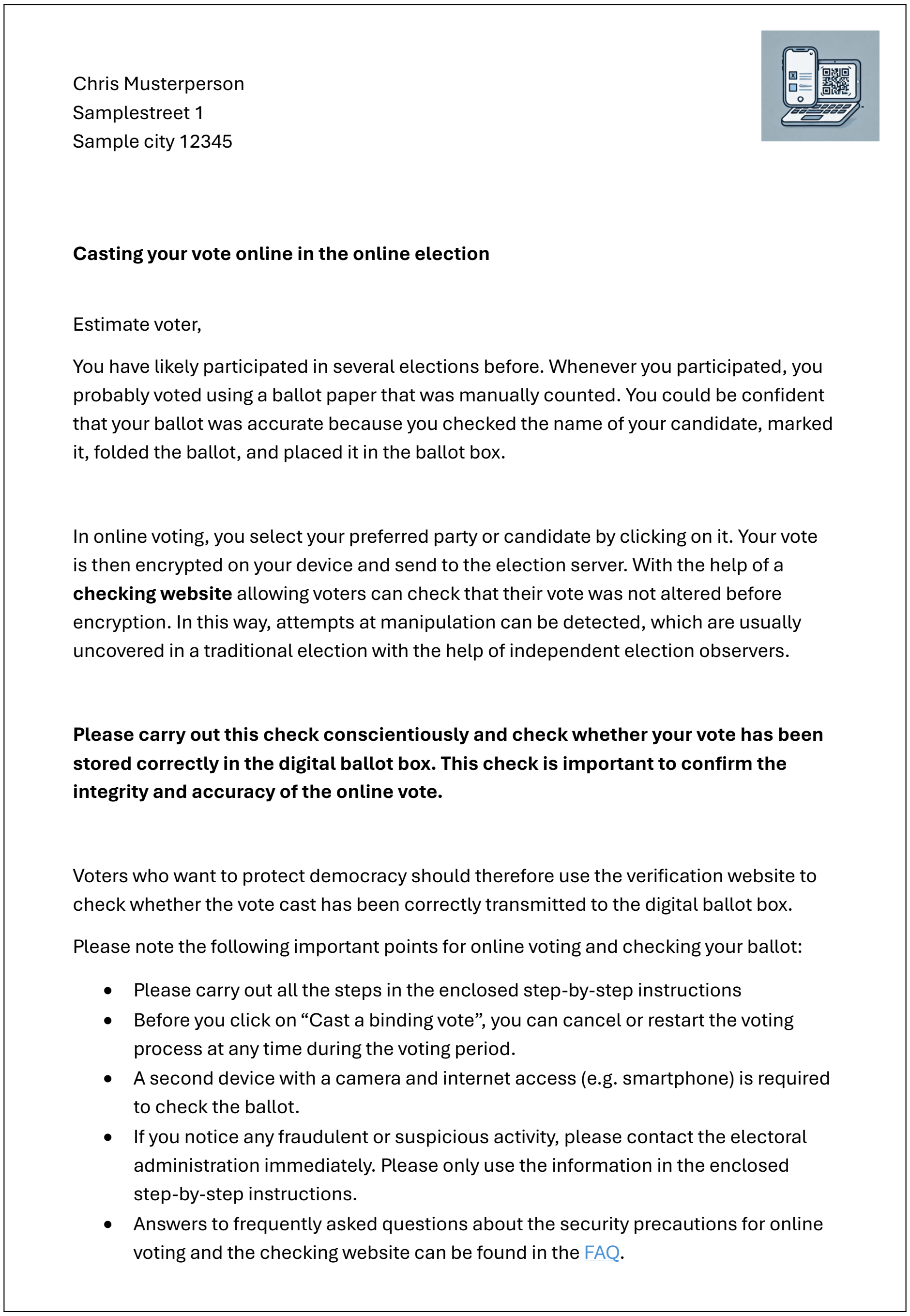
[48] valimised. n.d.. *Statistics about Internet voting in Estonia*. https://www.valimised.ee/en/archive/statistics-about-internet-voting-estonia last accessed 18.08.2025.

[49] Melanie Volkamer, Oksana Kulyk, Jonas Ludwig, and Niklas Fuhrberg. 2022. Increasing security without decreasing usability: A comparison of various verifiable voting systems. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, USA, 233–252. https://www.usenix.org/conference/soups2022/presentation/volkamer

[50] Anina Weber and Geo Taglioni. 2011. Swiss Elections to the National Council: First trials with e-voting in elections at federal level. doi:10.48550/arXiv.1109.2489 arXiv:1109.2489 [cs].

A Appendices

A.1 Election Invitation



Chris Musterperson
Samplestreet 1
Sample city 12345

Casting your vote online in the online election

Estimate voter,

You have likely participated in several elections before. Whenever you participated, you probably voted using a ballot paper that was manually counted. You could be confident that your ballot was accurate because you checked the name of your candidate, marked it, folded the ballot, and placed it in the ballot box.

In online voting, you select your preferred party or candidate by clicking on it. Your vote is then encrypted on your device and send to the election server. With the help of a **checking website** allowing voters can check that their vote was not altered before encryption. In this way, attempts at manipulation can be detected, which are usually uncovered in a traditional election with the help of independent election observers.

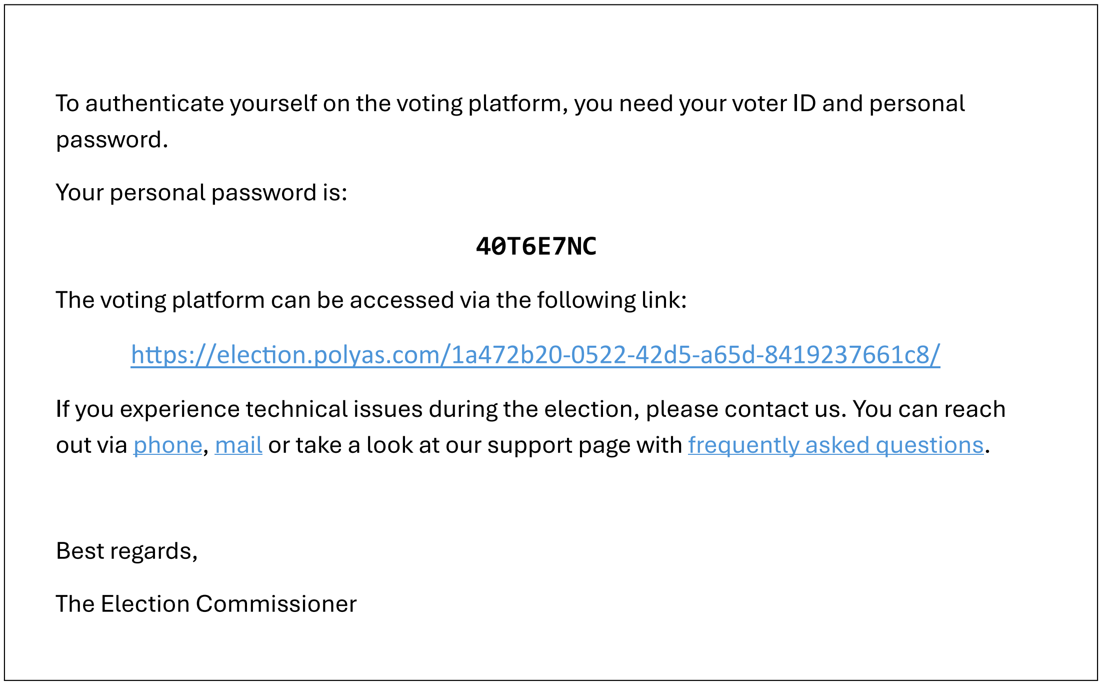
Please carry out this check conscientiously and check whether your vote has been stored correctly in the digital ballot box. This check is important to confirm the integrity and accuracy of the online vote.

Voters who want to protect democracy should therefore use the verification website to check whether the vote cast has been correctly transmitted to the digital ballot box.

Please note the following important points for online voting and checking your ballot:

- Please carry out all the steps in the enclosed step-by-step instructions
- Before you click on "Cast a binding vote", you can cancel or restart the voting process at any time during the voting period.
- A second device with a camera and internet access (e.g. smartphone) is required to check the ballot.
- If you notice any fraudulent or suspicious activity, please contact the electoral administration immediately. Please only use the information in the enclosed step-by-step instructions.
- Answers to frequently asked questions about the security precautions for online voting and the checking website can be found in the FAQ.

Figure 11: Page 1 of the improved election invitation letter. This full-page DIN A4 document introduced the voting process, highlighted the role of the verification website, and stressed the importance of checking the ballot for integrity. It also provided practical instructions, including the need for a second device, fraud reporting channels, and a link to the FAQ



To authenticate yourself on the voting platform, you need your voter ID and personal password.

Your personal password is:

40T6E7NC

The voting platform can be accessed via the following link:

https://election.polyas.com/1a472b20-0522-42d5-a65d-8419237661c8/

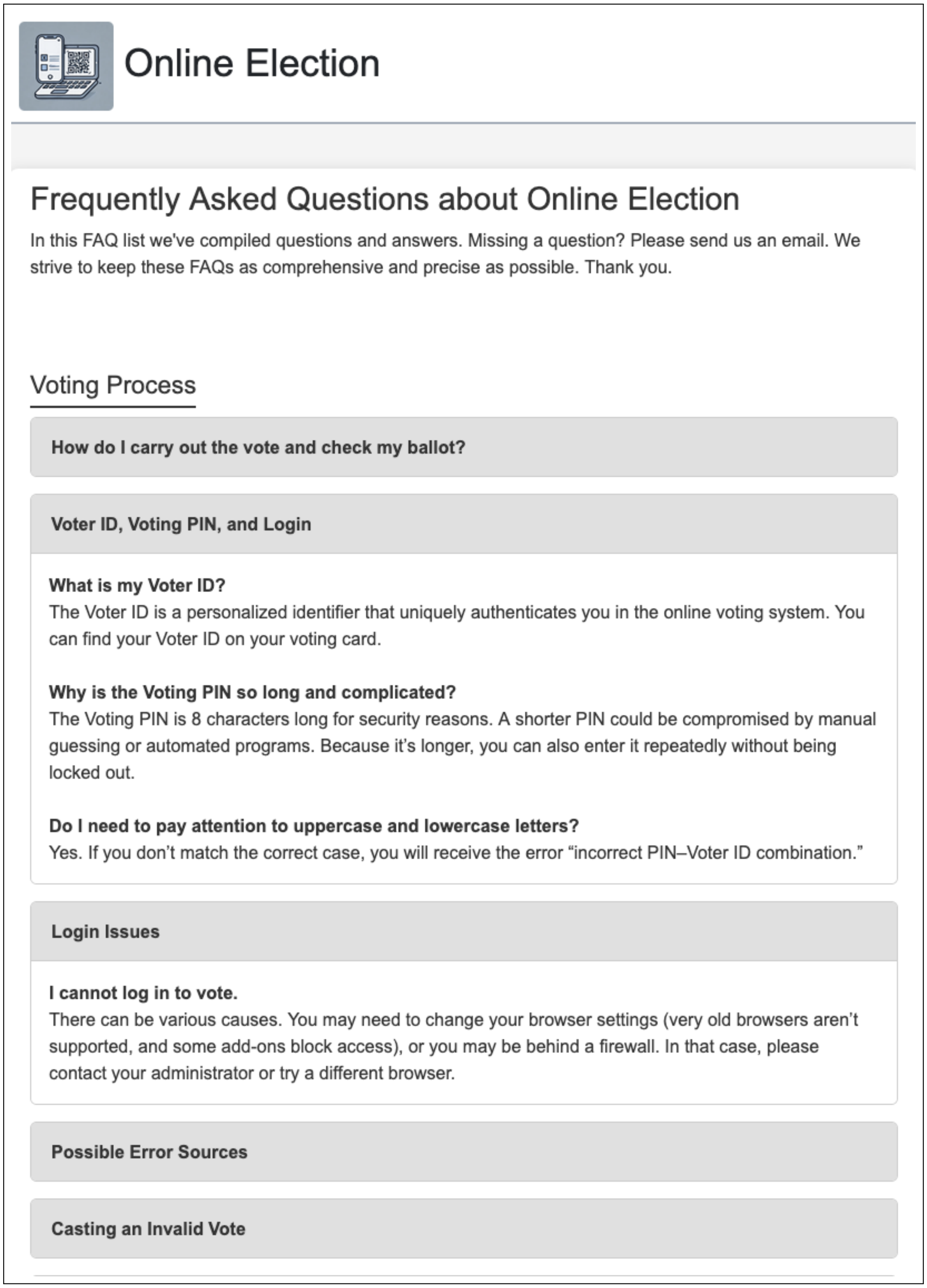
If you experience technical issues during the election, please contact us. You can reach out via phone, mail or take a look at our support page with frequently asked questions.

Best regards,

The Election Commissioner

Figure 12: Page 2 of the improved election invitation letter. This page provided voters with their personal login credentials, the official election link, and multiple support channels for technical assistance.

A.2 FAQ



Online Election

Frequently Asked Questions about Online Election

In this FAQ list we've compiled questions and answers. Missing a question? Please send us an email. We strive to keep these FAQs as comprehensive and precise as possible. Thank you.

Voting Process

How do I carry out the vote and check my ballot?

Voter ID, Voting PIN, and Login

What is my Voter ID?
The Voter ID is a personalized identifier that uniquely authenticates you in the online voting system. You can find your Voter ID on your voting card.

Why is the Voting PIN so long and complicated?
The Voting PIN is 8 characters long for security reasons. A shorter PIN could be compromised by manual guessing or automated programs. Because it's longer, you can also enter it repeatedly without being locked out.

Do I need to pay attention to uppercase and lowercase letters?
Yes. If you don't match the correct case, you will receive the error "incorrect PIN–Voter ID combination."

Login Issues

I cannot log in to vote.
There can be various causes. You may need to change your browser settings (very old browsers aren't supported, and some add-ons block access), or you may be behind a firewall. In that case, please contact your administrator or try a different browser.

Possible Error Sources

Casting an Invalid Vote

Figure 13: The FAQ page provided participants with practical information on the voting process, use of voter ID and PIN, troubleshooting login issues, and common error sources, thereby mirroring typical support materials in real-world elections.