E-Vote Your Conscience: Perceptions of Coercion and Vote Buying, and the Usability of Fake Credentials in Online Voting

Louis-Henri Merino*, Alaleh Azhir†, Haoqian Zhang*, Simone Colombo*,
Bernhard Tellenbach‡, Vero Estrada-Galiñanes*, Bryan Ford*
*EPFL †MIT ‡Armasuisse

Abstract—**Online voting is attractive for convenience and accessibility, but is more susceptible to voter coercion and vote buying than in-person voting. One mitigation is to give voters *fake voting credentials* that they can yield to a coercer. Fake credentials appear identical to real ones, but cast votes that are silently omitted from the final tally. An important unanswered question is how ordinary voters perceive such a mitigation: whether they could understand and use fake credentials, and whether the coercion risks justify the costs of mitigation. We present the first systematic study of these questions, involving 150 diverse individuals in Boston, Massachusetts. All participants "registered" and "voted" in a mock election: 120 were exposed to coercion resistance via fake credentials, the rest forming a control group. Of the 120 participants exposed to fake credentials, 96% understood their use. 53% reported that they would create fake credentials in a real-world voting scenario, given the opportunity. 10% mistakenly voted with a fake credential, however. 22% reported either personal experience with or direct knowledge of coercion or vote-buying incidents. These latter participants rated the coercion-resistant system essentially as trustworthy as in-person voting via hand-marked paper ballots. Of the 150 total participants to use the system, 87% successfully created their credentials without assistance; 83% both successfully created and properly used their credentials. Participants give a System Usability Scale score of 70.4, which is slightly above the industry's average score of 68. Our findings appear to support the importance of the coercion problem in general, and the promise of fake credentials as a possible mitigation, but user error rates remain an important usability challenge for future work.**

1. Introduction

Remote electronic (online) voting systems promise convenience and increased voter turnout [1], [2]. Online voting is particularly useful to overseas voters [3] or in crises such as a pandemic [4]. One important development in electronic voting is universal verifiability, which allows anyone (not just election officials and observers) to verify that votes have been tallied correctly, while protecting voter privacy [5]–[9].

Individual verifiability measures [10], [11] attractively enable voters to verify that their votes are cast as intended. Individually-verifiable *receipts* unfortunately make voting more susceptible to voter coercion [12], [13].[1] An abusive partner or other coercer might demand the voter's receipts, for example [14]. These receipts could also enable unscrupulous well-funded actors to buy votes at scale through anonymously-funded smart contracts [15]. To resist such attacks, a secure election system must prevent coercive adversaries from knowing whether a voter complied with their demands, even if the voter is willing to comply, e.g., in return for financial compensation [12], [16].

Most online voting systems lack coercion resistance [6], [17], [18]. *Deniable re-voting* permits a voter to override a coerced vote with a new vote cast later [19]–[21], but is vulnerable to coercers who can supervise voters or hold onto their credentials or voting devices until the election closes [14]. Estonia's online voting system employs deniable re-voting [19], [22], [23], but lacks universal verifiability.

A strategy proposed by Juels, Catalano and Jakobsson (JCJ) [12] enables voters to create, alongside their real voting credential, fake credentials which cast votes that do not count. Fake credentials present usability concerns, however, such as whether voters can distinguish their real credential from fake ones, or can create a fake credential while under coercion [22], [24]. While prior work has discussed the usability of fake credentials [22], [24], [25], only Neto et al. [26] performed a user study on this topic. Their study involved only university-affiliated participants, however. Further, the voting process they studied lacked individual verifiability: voters could not check whether the purportedly "real" credential they were issued was in fact real (as opposed to fake). Prior work also leaves other unanswered questions, such as whether ordinary voters even comprehend the coercion threat or believe it is important.

To fill this gap, we conducted a study with 150 individuals, recruited at a suburban park in Boston, Massachusetts, to examine whether voters might plausibly find coercion-resistant online voting with fake credentials to be usable and trustworthy. 120 of these participants underwent

1. We use the term "coercion" broadly to indicate any form of undue influence, including vote buying and voter intimidation.

a credentialing process to obtain real and fake credentials.[2] Participants then cast a mock vote using, at least, their real credential. The remaining 30 participants engage with the same system, but without any exposure to fake credentials. Participants conclude the study by completing a survey asking them to share their experiences and views of the system, as well as their perspectives on and any experiences with coercion in general. Our institutional review board approved the study; Section 3 discusses ethics considerations.

The credentialing process is an interactive user-interface prototype of Trust-limiting In-Person Registration (TRIP) [27], a voter-verifiable registration system for coercion-resistant online voting via fake credentials. Unlike prior coercion-resistant systems either deployed [1], [2] or subject to user studies [26], TRIP ensures that a compromised registrar cannot undetectably manipulate elections by secretly keeping real credentials for themselves, leaving voters with only fake credentials. To achieve verifiability without producing receipts usable for coercion, TRIP's registration kiosk produces interactive zero-knowledge proof transcripts, all of which are valid and checkable, but are *true* proofs only in real credentials, and are false proofs in fake credentials. Study participants used TRIP to create real and fake credentials, then used an Android device we supplied to cast a mock non-political vote. A random subset of participants were silently exposed to a "compromised" kiosk that issued only fake credentials. Because individual verifiability depends on voters being able to detect a compromised kiosk – not just in theory but *in practice* – our study sheds light for the first time on whether ordinary voters can effectively obtain verifiability *and* coercion resistance at once.

Using the data we collected from the study, we address the following four central questions:

1) What are voters' perceptions of and experiences with the coercion threat in general?
2) How likely will voters trust a coercion-resistant online voting system, versus other voting methods?
3) Can voters use a voter-verifiable credentialing process to create their real and fake credentials, and identify deviations to ensure voter verifiability?
4) Can voters understand and use fake credentials, casting their intended votes with their real one?

This is not a longitudinal study, so it cannot assess issues such as whether voters can recall the distinction between their real and fake credentials over an extended time period. This is one of several limitations we detail in §7.

This paper makes the following key contributions:

- The first study to systematically investigate whether voters find coercion-resistant online voting with fake credentials to be understandable, usable and trustworthy.
- The first study that assesses whether voters can use a credentialing process that requires voters to identify and report a misbehaving kiosk to ensure voter verifiability.

2. The study used the term "test credentials" instead of "fake credentials" to avoid negative connotations that fake credentials are invalid or inherently bad. The registration process suggests to users that "test credentials" may also be used for purposes other than coercion resistance, such as to test the voting system, or to share with friends or family for educational purposes.

Our prototype, which includes a user interface mockup of TRIP and an Android application that simulates activation and voting, is available at github.com/dedis/trip-usability.

2. Background

This section introduces coercion-resistant voting systems, TRIP, and the metrics we used in our study.

2.1. Online Voting Systems

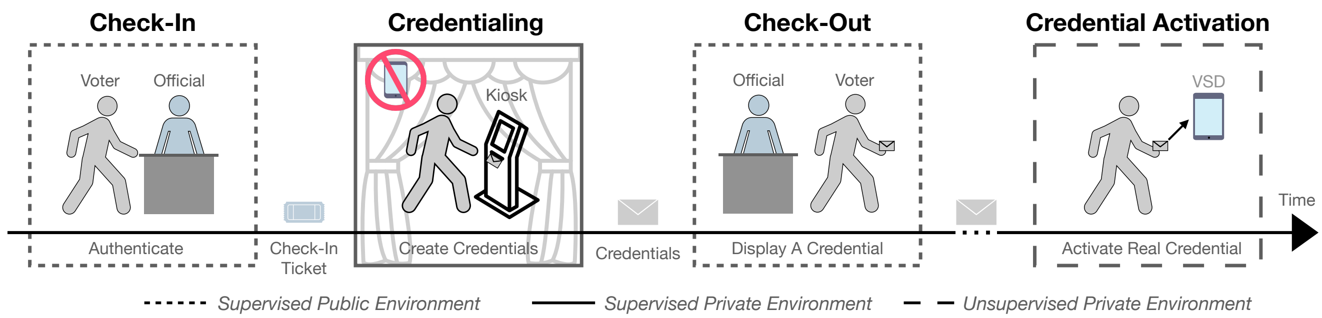
Nearly all online voting systems [5], [6], [17] strive for a minimum of verifiability and voter privacy: establishing vote confidentiality, while offering (publicly) verifiable election results. Substantial research seeks to achieve other desirable properties as well, such as cast-as-intended [10], [11], where voters are convinced that their encrypted ballot contains their intended vote. JCJ [12] proposed and formally defined coercion resistance as another desirable system property: in brief, the inability of an adversary to confirm whether a coerced voter has complied with their demands, *even if the voter wishes to do so*. While JCJ suggested the use of fake voting credentials, other works [19], [28], [29] have since proposed other strategies due to usability concerns with fake credentials [25]. This paper focuses on these concerns.

A typical online voting system interacts with voters, an election authority, and observers, and involves setup, registration, voting, and tallying phases. Performance-oriented systems research tends to focus on tallying: reducing the costs of shuffling and counting ballots, including the removal of fake ballots in coercion-resistant voting systems. For usability, however, registration and voting are the more crucial stages because they directly involve voters. This study focuses primarily on registration, where voters engage with the election authority to generate their voting materials.

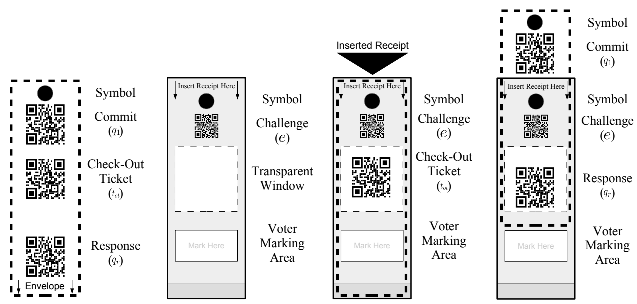
2.2. Trust-limiting In-Person Registration

Since the introduction of coercion resistance, most works have focused on the tallying process [30]–[34], leaving a key challenge unresolved: the development of a usable voter registration system that issues *voter-verifiable* real and fake credentials. Voter verifiability plays a key role in preventing a compromised election authority from monopolizing real credentials and issuing only fake ones to voters.

The Trust-limiting In-Person Voter Registration (TRIP) system [27] addresses this problem by leveraging physical presence of voters across four registration phases (Fig. 1a): check-in, credentialing, check-out, and activation. At check-in, voters identify themselves to a registration official, obtaining a check-in ticket that gives them access to a privacy booth for credentialing. Comparable to the "ballot selfie" problem with in-person voting, recording devices could compromise coercion resistance in this critical stage, so TRIP assumes that voters cannot use electronic devices in



(a) **TRIP Voter Workflow.** The voter (1) checks in with an official by authenticating themselves to receive a check-in ticket, (2) enters a supervised private environment to create their credentials using a kiosk that is unlocked by their check-in ticket, (3) checks out with an official by displaying the public part of any one of their paper credentials (Check-Out Ticket – Fig. 2c), and (4) activates their credentials on their voter supporting device (VSD).



(a) **Receipt** (b) **Envelope** (c) **Transport** (d) **Activate**

Figure 2: **TRIP Paper Credential**. Figures (a) and (b) present the paper credential's elements and Figures (c) and (d) present the paper credential's transport and activate states.

the booth.[3] In the booth voters find a kiosk, envelopes, and a pen. Voters create their real credential in four steps:

1) The voter scans their check-in ticket.
2) The kiosk prints a QR code and symbol on receipt paper.
3) The voter picks any envelope matching the printed symbol, and presents the QR code on it to the kiosk's scanner.
4) The kiosk prints two more QR codes on the receipt.

These steps establish an interactive zero-knowledge proof between the kiosk and the voter, ensuring the correctness of the real credential. The voter verifies that the kiosk adheres to these steps while the voter's device—later, during activation—verifies the proof's cryptographic validity.

To finalize their credential, the kiosk instructs voters to tear off the receipt (Fig. 2a), insert it into the envelope (Fig. 2b), and memorably mark the resulting combination – which we term the *real paper credential* – to distinguish it from future fake credentials. When the receipt is fully inserted and only its middle QR code is visible, the credential is in its *transport state* (Fig. 2c).

Voters may then create fake credentials in two steps:

1) The voter chooses and scans any envelope.
2) The kiosk then prints the receipt all at once.

To finalize the fake credential, voters again tear off and insert the receipt inside the envelope, marking the credential distinctively. Voters may create any number of fake credentials, limited only by their time in the privacy booth.[4]

These two steps, while also establishing an interactive zero-knowledge proof, compromises the proof's soundness compared to the four-step process. Despite the proof's soundness being violated, the proof's cryptographic validity remains unaffected. Therefore, the voter's real credential is cryptographically concealed among their fake credentials, differentiated only by the voter's own distinct markings.

Upon exiting the booth, voters proceed to check-out where they present any one of their credentials, real or fake, to the registration official. The official scans the receipt's middle QR code, visible through the envelope's transparent window, to complete the in-person portion of the process.

Sometime later, voters activate their real credential on any device they trust by placing the credental in the Activate state (Fig. 2d) and scanning it. After activation, the voter discards the now-unusable paper credential. Voters who have no device they trust may activate their real credential on a device of a trusted friend or family member.[5] Voters may give away or sell their fake paper credentials, or activate them on a device that is under a coercer's control.

TRIP's design does not limit the lifetime of voting credentials. Voters may therefore reuse their activated credentials to cast (both real and fake) votes in multiple successive elections, thereby amortizing the convenience cost of in-person registration. Election authorities may impose an expiration date on voting credentials by policy, perhaps aligning with the renewal cycle of identification documents.

3. The degree to which this rule is enforced is a policy decision, important but orthogonal to TRIP's design. Standard practice comparable to in-person voting would be merely to forbid the use of devices in the booth. Stronger enforcement might require voters to deposit electronic devices in a locker before entering the booth, at an obvious cost in convenience.

4. Whether to impose any particular time limit is a policy decision. We anticipate that voters spending an inordinate time registering should be a rare situation, manageable informally by an election official asking "is anything wrong?" at some point and gently escalating only as needed.

5. In hopefully-rare cases where a voter cannot hide a paper credential from the coercer, or have no access to any device they trust, TRIP offers an extension where voters can leave the booth with only fake credentials, while delegating their real vote to a designated proxy such as a political party. We do not explore this extension in the present study, however.

TRIP is designed to replicate the trust assumptions inherent in in-person voting to achieve verifiability and coercion resistance. In traditional in-person voting, the election authority mitigates the risk of coercion by providing voters with a privacy booth free from any coercer's influence. This supervised environment creates an untappable communication channel between voters and the election authority, enabling voters to vote their conscience. Voters, in return, collectively protect the integrity of the election by reporting inconsistencies with their ballot prior to submitting it [35]. TRIP, being designed for online voting, leverages a similar untappable channel but at registration time rather than voting time. While a privacy booth traditionally hides a voter's choices, in TRIP it hides knowledge of how many credentials the voter created and which one is real.[6] Voters can therefore convey their intentions to the election authority later while casting a vote online—by choosing to cast their vote using a real or fake credential. TRIP relies on interactive zero-knowledge proofs to ensure the integrity of the election process. Voters need not understand zero-knowledge proofs, but to verify the process's integrity, they *do* need to distinguish the four-step process of creating a real credential from the two-step process of creating a fake credential, and to report any inconsistencies they encounter.

2.3. Standardized Usability Metrics

In this study, we use the System Usability Scale (SUS) and the User Experience Questionnaire (UEQ) to measure participants' perceptions of system usability and user experience.

System Usability Scale (SUS). This scale [36] is most often used to measure usability, requiring only ten prompts. We altered the first prompt from "I would like to use this system frequently" to "I would like to use this voter registration system whenever I renew my identifications documents (i.e., every 5-10 years)" to align with TRIP's intended usage.

User Experience Questionnaire (UEQ). This questionnaire [37], unlike SUS, assesses not only traditional usability factors but also aspects of user experience, thereby ensuring the system meets user needs. User experience covers a participant's emotional, cognitive, and physical responses before, during, and after system usage [38]. The UEQ consists of 26 contrasting attributes, with participants expressing their level of agreement by selecting a value between 1 and 7, then re-calibrated to a score between -3 to 3. The responses allow for the measurement of six scales [37], such as Perspicuity, Efficiency, Dependability and Novelty, in addition to the individual scores obtained for each attribute, such as security and practicality. Furthermore, Hinderks et al. [38] introduce a Key Performance Indicator (KPI) extension to the UEQ, termed UEQ PKI. This extension allows participants to express their views on the significance of each of the six UEQ scales via six additional questions.

6. The design TRIP paper [27] includes a formal proof that coercers cannot distinguish between real and fake credentials, or identify the number of fake credentials created by a voter while inside the booth.

3. Methodology

This section presents our study's design, detailing its objectives, workflow, materials and evaluation metrics.

Ethical Considerations. Before participating, each individual received an information sheet outlining the study's scope, their rights, the expected tasks, and the data to be collected. We collected data from participants' interaction with our devices and credentials, supplemented with information from a survey. We respected participants' privacy by supplying all required materials, and by not asking for any form of ID, although showing ID would be standard in real voter registration. Furthermore, we informed participants that this system is not linked to any real voting system.

3.1. Objectives

We start by detailing our study's central questions, along with our strategies for addressing them.

What are voters' perceptions and experiences with coercion and how do they rate their trust in this coercion-resistant voting system versus other voting methods? To answer the former, we ask participants whether they have experienced or know of someone who has experienced coercion. The wording "or someone you know" is designed to encourage disclosure from those who might otherwise be hesitant to share their experience. We also ask to rate the likelihood of various coercion and vote-buying scenarios, along with potential perpetrators. For the latter, we ask participants to rate the trustworthiness of this coercion-resistant voting system in comparison to other voting methods.

Can voters use a credentialing process to create their real and fake credentials, and identify deviations during credentialing to ensure voter verifiability? To measure success rate, we observe the number of participants who complete the process successfully, without requiring a facilitator. While during the survey, we measure participants' experience and perceptions of usability using both the SUS and UEQ questionnaires. To measure the malicious kiosk detection rate, we intentionally make the kiosk misbehave and observe if participants report the anomaly.

Can voters understand and use fake credentials, while still casting their intended votes using their real credential? We gather data about participants' understanding of fake credentials in two phases: first via an interactive quiz on the kiosk, then later via survey questions. We also present an instructional video at the beginning of the process to prepare and educate the voter. Since real and fake credentials function identically, we direct participants to vote using their real credential, allowing us to assess whether they can distinguish their real credential from their fake credentials.

3.2. Study Setup & Workflow

We now describe the study's setup and workflow, detailing its location, the setup of the workstation and the participant workflow. The facilitator's scripts are available in §C.1.

Figure 3: **Study Setup**: Starting on the left chair, participants (1) review the study's information sheet and sign the consent form, (2) watch an instructional video on the laptop, (3) are handed a check-in ticket from the facilitator, (4) move to the right chair and interact with the kiosk to create their credential(s), (5) return to the left chair for check-out, (6) activate their credential(s) and cast mock votes on the supplied mobile device, and, (7) complete the study's exit survey on either our laptop or touchscreen tablet.

Location and Recruitment. To enhance demographic diversity while ensuring a neutral environment, we conduct our study in a suburban park. We verbally invited passerbys to participate in the study over three months, and showcased flyers/posters at our study location to encourage engagement. We limited participants to only those who said they had prior experience with voter registration.

Study Setup. Our setup (Fig. 3) consists of a table with two chairs on one side for the participant and a park bench on the other side for the facilitator. On the right side of the table, we have the credentialing process with the kiosk, envelopes and receipt printer, while on the left side, we have the check-in, check-out, activation and voting processes.

Participant Enrollment. Whenever an eligible participant agrees to participate, we guide them to the left side of the table to review the information sheet and sign the consent form. The facilitator addresses any logistical questions, and refrains from discussing the objectives of the study beyond "assessing the usability of an online voting system." The participant is then randomly assigned to one of five groups that we detail in §3.3. To enhance the ecological validity of our study [39], the facilitator asks the participant to envision themselves at a government office, as scripted in §C.1 in the extended version of this paper [40].

Instructional Video & Check-In. The facilitator then starts playing a video (§3.4) for the participant to watch, the content of which depends on the participant's assigned group. After the video concludes, the facilitator hands the participant a check-in ticket and directs them to the booth.

Credentialing and Check-Out. To simulate an authentic booth experience, the facilitator does not interact with the participant in this stage, intervening only upon request. Upon completing the credentialing process, the participant returns to the left side of the table and hands one credential (real or fake) to the facilitator. The facilitator scans the QR code through the visible window and returns the credential.

Activation and Voting. For activation, the facilitator asks the participant to envision that they are now at home, and to use the study's supplied mobile device to activate their real credential and cast a mock vote. The participant may optionally activate fake credential(s) as well and cast fake votes by clicking on 'cast vote with another credential'. Since the app was designed solely for the purpose of this study, it is stateless, storing only one credential at a time. If asked, we explain that in a real setting, one could unlock a specific credential using its associated user-generated PIN.

Survey and Compensation. Once finished, the participant completes the exit survey and is then compensated with $20. The entire session typically takes around 30 minutes.

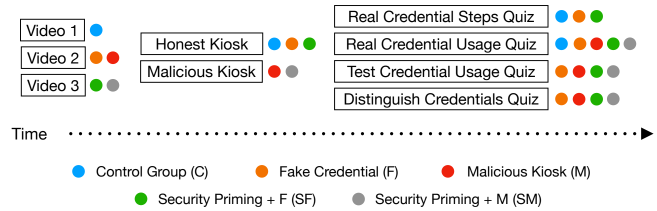


Figure 4: **Study Flow.** Participants first watch one of three instructional videos, then interact with either an honest or malicious kiosk that displays the corresponding quizzes. The distinguish credentials quiz is present for those who created at least one fake credential.

3.3. Study Groups

To answer our key questions (§3.1), we conduct a between-subjects study, randomly assigning study participants to five equal-size groups. Each group experiences a distinct variant of the process. To assess the usability of fake credentials in particular, our study exposes one control group only to real credentials, while exposing the other four groups to real and fake credentials. To assess the system's effective voter verifiability, we expose two of the latter groups to a "malicious" kiosk that attempts to "steal" the voter's real credential by silently guiding the user through the creation of a fake credential instead of a real one.[7] To familiarize participants with the correct voting process we rely on instructional videos, as detailed in §3.4. To study the tradeoffs between being more or less explicit about security threats and risks in educational materials, we use two contrasting types of instructional videos. We therefore obtain the following five groups, each consisting of 30 participants:

- **Control Group (C)**: Exposure only to the real credential known as "voting credential" and instructional video 1.

7. The deviation involves reordering steps 2 and 3, with step 3 now being before 2. Other deviations are possible and perhaps worth studying in the future, but most variations are either more overt and obvious (e.g., skipping the creation of a real credential) or detectable by the voter's personal device on activation (e.g., the cryptographic validity of the zero-knowledge proof).

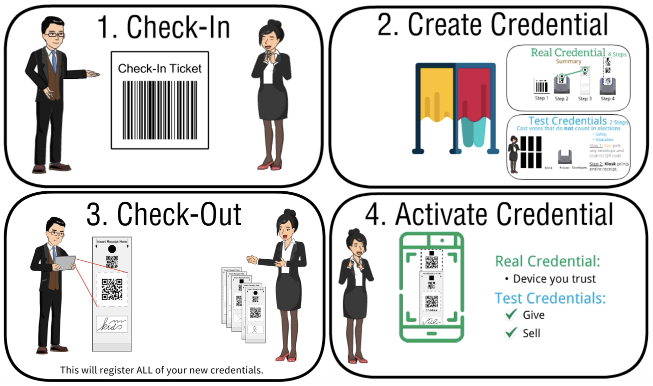


Figure 5: **Steps Overview.** Various screenshots from the instructional video, which demonstrate each phase.

- **Fake Credential Group (F)**: Exposure to real and fake voting credentials, and instructional video 2.
- **Malicious Kiosk Group (M)**: Exposure to real and fake credentials, instructional video 2, and a malicious kiosk.
- **Security Priming + F (SF)**: Exposure to real and fake voting credentials, and instructional video 3.
- **Security Priming + M (SM)**: Exposure to real and fake credentials, instructional video 3, and a malicious kiosk.

3.4. Materials

This study uses instructional videos to educate voters, while quizzes and a survey assist us in evaluating their understanding according to the study flow presented in Figure 4.

Instructional Videos. We introduce three videos, available at github.com/dedis/trip-usability, each illustrating their assigned conditions: control group (video 1), no security priming (video 2) and security priming (video 3). Video 1, lasting 3 minutes and 6 seconds, demonstrates only the creation of a "voting" credential. In contrast, videos 2 and 3, with durations of 5m 50s and 6m 25s respectively, illustrates the creation of real and fake credentials (Fig. 5). Videos 2 (Fig. 10a) and 3 (Fig. 10b) both highlight the differences in creating real and fake credentials. However, video 3 shows, under a conspicuous "BEWARE" sign, how to detect a "hacked" kiosk. Video 3 thus makes a key threat explicit, at the risk of a more unsettling or "scary" presentation.

We use videos as our primary source of instructional material, as individuals learn better from dynamic visuals compared to static images [41]. We select whiteboard animated videos due to their positive impacts on retention, engagement and enjoyment, even when conveying complex material [42]. To enhance the effectiveness of our videos, we incorporate the dynamic drawing principle, adopt a first-person perspective, and include narration with subtitles [43].

Quizzes. To evaluate participants' understanding of concepts and foster active learning, the kiosk exposes participants to up to four unannounced, multiple-choice quizzes (§C.3 in the extended version [40]). The selection of quizzes depends on the participant's group assignment and actions. For all groups, we evaluate participants' understanding that they must keep and activate their real credential on a trusted device. For the treatment groups (F, M, SF, SM), we examine participants' recall of the stated purposes of fake credentials. For groups C, F, and SF, we also assess comprehension of when to select and scan an envelope prior to creating their real credential. Finally, for those electing to create fake credential(s), we evaluate their recall on how to differentiate their real credential from fake credentials.

Exit Survey. In the exit survey (§D in the extended version [40]), we first ask participants their demographic information. Participants then rate their experience, including what they liked most and liked least, along with completing the SUS, UEQ, and UEQ PKI questionnaires. Participants continue by answering whether they noticed anything odd with the credentialing process. Participants exposed to fake credentials answer questions related to those credentials, such as whether they can recall their usage. Participants then rate their trust on various voting methods: three variants of in-person voting, and three variants of remote voting, including one for this voting system. The survey finishes by asking participants to describe their perceptions and experiences with coercion and vote-buying in their own lives. The average time spent on the survey is 17 minutes.

3.5. Statistical Methods

Throughout this paper, we use an alpha level of 0.05 to establish statistical significance. For our between-subjects comparison on non-parametric data (e.g., ordinals), we use the Kruskal-Wallis one-way ANOVA, while for pairwise comparisons, we use the Dunn test. For our within-subjects comparisons on non-parametric data, we use the Friedman rank sum test, while for pairwise comparisons, we use the Durbin-Conover test. To control for the family-wise error rate when conducting pairwise statistical comparisons, we use the Holm-Bonferroni method. We apply Shapiro-Wilk to assess whether the current data follows a normal distribution. We employ Cohen's d to measure effect sizes, which estimate the degree of difference between two groups. Typically, an effect size of about 0.2 is considered small, while an effect size near 0.8 is considered large. The term "participants" represent the average views or experiences of those involved in the study.

4. Participants' Perceptions and Experiences

This section details the demographic profile of our participants and discusses their experiences with and perceptions of coercion. Moreover, we explore their views on this coercion-resistant system versus other voting methods.

4.1. Demographics

We recruited 150 participants, aged 19 to 83, with an average and median age of 44 and 36.5, respectively. Figure 6 depicts this age distribution. In the extended version [40], Figure 11a provides a breakdown of the participants' age distribution

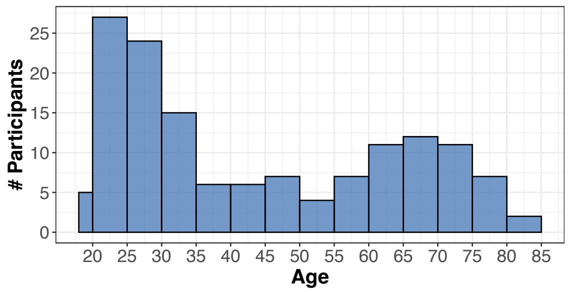


Figure 6: **Distribution of Participant Ages**. The minimum, median, mean and maximum ages among all groups were 19, 36.5, 44 and 83, respectively. Six participants did not disclose their age.

across the five study groups, while Figure 11b provides a breakdown of their gender, ethnicity and education. We observed distinct participation patterns: seniors primarily during the day, middle-aged individuals after work, and younger individuals (18-35) throughout the day. On initial encounter, many individuals had the first impression that they could register with us for real online voting; we clarified in our recruitment script (§C.1) that this is only a usability study on online voting with mock, non-political, elections. Participants' time availability was the primary recruitment challenge; we estimated the study to last around 30 minutes, and found, after the fact, that a typical participant took 35 minutes. A few individuals declined to participate due to their opposition to online voting, reconfirming well-known difficulties in overcoming selection bias in user studies.

4.2. Coercion and Vote Buying

We now discuss participants' perspectives and experiences with coercion and vote buying based on their survey responses. Participants rated the following four coercion scenarios on a 7-point Likert scale [44] (summarized from §D.8 in extended version [40]): (C)oercing by threatening harm (C-Forceful), (P)urchasing absentee ballots (P-Ballot)[8], purchasing proof such as a voter taking a selfie with their ballot containing the coercer-dictated choices (P-Selfie), and an app offering compensation to vote as directed (P-App). They also rated the following four potential sources of coercion on the same scale: a party operative (S-Party), an authority figure (S-Authority), a family member (S-Family), and an employer (S-Employer). Ratings are then converted to a score between -3 to 3, where -3 is 'completely unlikely' and 3 is 'completely likely'.

Perceived Scenarios (Fig. 7a). Among the coercion scenarios, participants find P-Selfie to be the most likely scenario with a mean score of 0.51 and a median of 1 (somewhat likely). A quarter (24%) of participants rated P-Selfie as extremely likely, with this scenario being perceived as having a statistically significantly higher likelihood compared to the other three coercion scenarios. In contrast, participants perceived C-Forceful as the least likely scenario with a mean score of -0.34 and a median of -1 (somewhat unlikely). We also see that the distribution of responses for C-Forceful almost inversely mirrors that of P-Selfie. The calculated Cohen's d effect sizes further quantify these differences: 0.41 when comparing P-Selfie with C-Forceful, 0.39 with P-Ballot and 0.26 with P-App. The remaining two coercion scenarios, P-Ballot and P-App also bear a visual similarity in their distributions, demonstrating two distinct peaks at somewhat likely and extremely unlikely.

Perceived Sources (Fig. 7b). Participants perceive S-Family to be the most likely source of coercion (mean score of 0.53 and a median rating of somewhat likely), with 21% of participants rating it as extremely likely. Moreover, S-Family is statistically significantly higher than the three other coercion sources. In contrast, S-Authority is the least likely source with a mean score of -0.25. The calculated Cohen's d effect sizes are: 0.38 when comparing S-Family with S-Authority, 0.34 with S-Employer, and 0.25 with S-Party. For both S-Employer and S-Party sources, participants have two dominant contrasting opinions, with one cohort seeing the source as somewhat likely and the other as very unlikely.

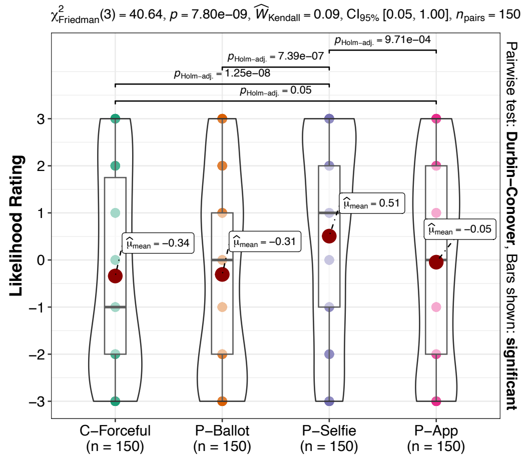
Coercion Instances (Fig. 7c). A quarter of participants (26%) report experiencing or knowing of someone who has experienced at least one form of voter coercion, while two-thirds (67%) report no such experiences; the remaining 7% preferred not to say. In line with participants' views, the most commonly reported source of coercion (15%) is S-Family. Such instances include multiple accounts of spousal oversight during voting, leading to dictated voting choices. However, despite being viewed as the least likely coercion scenario, S-Forceful is the most reported coercion scenario (12%). Around 10% identify S-Employer and S-Party as sources of coercion. Reported incidents include unions dictating votes, co-workers pressuring attendance at undesired political rallies, and prominent party members pressuring members to vote along party lines.

Discussion. The substantial reporting rate of recognized coercion instances at 26%, with family members cited as the predominant coercion source (15%), highlights the need for coercion-resistant strategies to counter *persistent* adversarial oversight. In such circumstances, the practice of deniable revoting, such as the one in use in Estonia [22], falls short. Family members, typically having substantial unrestricted access, can potentially cast a vote on behalf of the targeted relative just before the election concludes. Such a coercer can also realistically stay with the relative or retain their device or voting credential until the voting process ends.

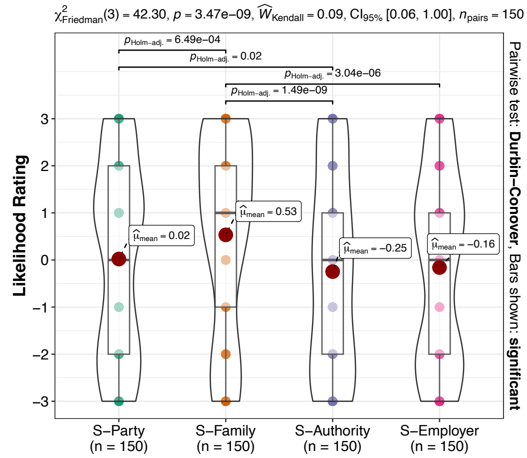
4.3. Trust in Voting Methods

We now investigate the level of trust the 120 participants in treatment groups perceive in our coercion-resistant system versus other voting methods. Participants rated their level of trust in the following voting methods on a 7 point Likert Scale (summarized from §D.7): (I)n-person voting with

8. Illustrated by the North Carolina ballot fraud incident [45].



(a) **Coercion Scenarios.** Participants perception of likely coercion scenarios; ballot selfies received the highest average likelihood, while forceful coercion received the lowest.



(b) **Coercion Sources.** Participants perception of likely sources of coercion; family members received the highest average likelihood rating, while authority figures received the lowest.

Scenario	Happened	Not Happened
C-Forceful	12%	81%
P-Ballot	8%	85%
P-Selfie	6%	86%
P-App	3%	88%
Total Count	44	511
Unique Count	26	112
Source		
S-Party	10%	81%
S-Family	15%	77%
S-Authority	7%	85%
S-Employer	11%	83%
Total Count	63	491
Unique Count	31	108
Unique Participants	39	98

(c) **Coercion Instances.** Participants state whether these items have happened to them or someone they know. The remaining participants chose not to disclose.

Figure 7: **Participants' Views and Experiences with Coercion Scenarios, and Coercion Sources**

Voting Method	I-Ballot	I-BMD	I-DRE	R-Mail	R-Online	IR-Online
Mean	1.38	1.68	1.18	0.72	0.82	1.25
Median	2	2	1.5	1	1	2
Trust Rating	*Participants (%)*					
Trusted (3)	35	36	26	22	20	20
Highly	25	33	24	24	22	35
Somewhat	12	11	18	15	17	16
Neutral (0)	12	10	17	12	18	17
Somewhat	7	5	7	7	9	5
Highly	6	3	4	12	6	4
Untrusted (-3)	3	3	3	8	7	3
Voting Method	*Pairwise Statistical Outcomes (Fig. 9b)*					
I-Ballot		×	×	↑	↑	×
I-BMD	×		↑	↑	↑	↑
I-DRE	×	↓		×	×	×
R-Mail	↓	↓	×		×	×
R-Online	↓	↓	×	×		×
IR-Online	×	↓	×	×	×	

TABLE 1: **Voting Methods.** This table presents the trust ratings for various voting methods, as given by the 120 participants exposed to fake credentials, and a summary of the statistical outcomes (Fig. 9b). × means no statistical difference, ↓ means statistically significantly lower, and ↑ means statistically significantly higher.

hand-marked paper ballot (`I-Ballot`), in-person voting with a ballot marking device (`I-BMD`), in-person voting with a direct electronic device (`I-DRE`), (R)emote voting via mail-in ballot (`R-Mail`), a fully-remote voting system where both voter registration and voting are online (`R-Online`), and the online voting system they just experienced with in-person voter registration (`IR-Online`). Table 1 presents our summarized results while Figure 9 in Appendix A contains more complete statistical results.

Summarized Results. Analyzing the trust scores and the pairwise statistical outcomes across voting methods, we identify three cohorts. Participants place the highest trust in `I-BMD` and `I-Ballot`, and least in `R-Mail` and `R-Online`, with `I-DRE` and `IR-Online` in-between.

IR-Online. Participants generally regard `IR-Online` as 'somewhat trustworthy', as indicated by a mean rating of 1.25. However, the median participant sees it as 'highly trustworthy', as indicated by a greater median rating of 2. This ranks `IR-Online` marginally higher than `I-DRE`, having a mean trust rating of 1.18, yet remains below `I-Ballot`, exhibiting a higher average of 1.38. Moreover, although we find no statistically significant difference in trust ratings between the treatment groups, the median trustworthiness rating for the security priming groups SF and SM is 'somewhat trustworthy', contrasting with groups F and M's 'highly trustworthy' rating (Fig. 9d). We also find no statistically significant difference in trust ratings for `IR-Online` and `R-Online` between the control group ($n = 30$) and the treatment groups ($n = 120$) (Mann-Whitney U test: $p = 0.969$).

Participants Experiencing Coercion. We examine the trust ratings from the 33 participants (22%) not in group C who report personal or known experiences of coercion. These results, depicted in Figure 9c, show statistically significant lower trust scores for `R-Mail` and `R-Online` when compared to `I-BMD`, indicating a general distrust towards remote voting methods. Based on Cohen's d, the effect sizes between `I-BMD` and `R-Mail`, as well as `I-BMD` and `R-Online`, are 0.76 and 0.69 respectively. Simultaneously, the score of `I-Ballot` dropped to match `IR-Online`, while that of `I-DRE` rose to match `I-BMD`. This could possibly be influenced by the perceived risk of ballot selfies.

Discussion. These findings demonstrate the system's promising potential to attain trustworthiness scores on par with in-person voting. Presently, the system's overall score exceed those of in-person voting with a direct recording device, employed in 11.5% of U.S. jurisdictions in 2020 [46],

easy	positive	17	new	neutral	3
interesting	positive	13	streamlined	positive	3
complicated	negative	8	long	negative	2
good	positive	7	fine	neutral	2
simple	positive	7	fast	positive	2
straightforward	positive	4	complex	negative	2
great	positive	4	convenient	positive	2
smooth	positive	4	cumbersome	negative	2
confusing	negative	4	efficient	positive	2

TABLE 2: **Distribution of Single-Word Summaries with Sentiment.** This table lists the words used more than once by participants to describe their impressions of the system along with each word's associated sentiment (positive, neutral or negative).

and approaches the trust levels associated with in-person hand-marked paper ballots, used in 68.1% of jurisdictions.

Additional Findings. In §A.1, we discuss our findings when we contrast the other voting methods with each other, particularly the unanticipated lower trust for `I-Ballot` compared to `I-BMD`. We also discuss the surprisingly low trust rating for `R-Mail`, despite its use by 46% of voters in the 2020 U.S. presidential election [47], and the high trust rating in `I-BMD`, despite only 6.6% of participants capable of detecting a *manipulated* BMD-printed ballot [35]. Moreover, we present our findings when considering all 150 participants, particularly the statistically significantly higher trust in `IR-Online` compared to `R-Online`.

5. Usability of Registration and Voting

In this section, we present the usability results for TRIP, a voter-verifiable registration process that outputs real and fake credentials and relies on voters to identify deviations during credentialing to meet voter verifiability. We begin with qualitative results, discussing the system aspects participants most liked and disliked. We then present the observed use errors and the number of participants who successfully completed the registration and voting process without facilitator intervention. We continue with the System Usability Scale and User Experience Questionnaire scores, contrasting these against benchmarks and other voting-related studies. We conclude with our findings concerning the participants who encountered our purposely designed malicious kiosk.

Groups. The control group (C) establishes the baseline against which we evaluate the four experimental groups (F, M, SF, and SM). Group F serves as the expected norm, with group M being relevant in the event of a malicious kiosk.

5.1. Qualitative Results

We present the results from participants' single-word summaries and the system aspects they liked or disliked the most. For the single-word summaries, we classified each word based on sentiment, marking them as positive, neutral or negative. Table 2 presents the words that occur more than once. We also devised categories that best encapsulate the items expressed and split them into two groups: most liked and most disliked (Table 3).

Groups	(C)ontrol	(F)ake Creds.	(M)alicious Kiosk	SF	SM
Rating			***System Approval Rating***		
Positive	73%	63%	60%	56%	53%
Neutral	13%	20%	3%	17%	23%
Negative	13%	17%	37%	27%	23%
			System Usability Scale		
Score	69.6	70.4	69.9	67.3	62.7
SD	18.6	18.6	17.4	19.8	21.9
N	29	28	29	30	29
Percentile	55.1	57.8	56.1	47.8	35.0
Usability	69.5	69.8	68.4	67.7	62.5
Learnability	69.8	73.2	75.9	65.8	63.4
Scale	***User Experience Questionnaire (Score vs. Benchmark)***				
Attractiveness	↑	↑	↓	↓	↓↓
Perspicuity	↑	↓	↓	↓	↓↓
Efficiency	↑↑	↑↑	↑	↓	↓
Dependability	↓	↑	↑	↓	↓↓
Stimulation	↑↑	↑↑	↑	↑	↓
Novelty	↑	↑↑	↑↑	↑↑	↑
Category			***Most Liked (%)***		
Ease of Use	20	27	27	43	17
Instructions	27	20	13	13	27
Remote Voting	23	10	20	10	13
Security	3	20	10	17	23
QR Codes	7	13	17	3	10
Other (Positive)	3	10	10	7	10
Other	17	0	3	7	0
Category			***Most Disliked (%)***		
Process Complexity	27	23	30	23	23
Credential Handling	20	23	7	17	30
Confusion	7	17	17	13	17
Security / Coercion	7	10	23	13	7
In-Person Reg.	7	3	7	0	3
(None)	17	13	10	23	10
Other	17	10	7	10	10
Types			***Use Errors (# Events)***		
Tore Receipt	3	1	0	0	1
Almost Tear	3	4	0	1	0
Envelope Pick	1	2	0	1	1
Discarded Real Cred.	0	0	0	0	1
Activate Difficulty	3	5	6	5	10
Total (48)	10	12	6	7	13
Levels			***Use Errors (# Events)***		
Mistakes	10	7	5	7	11
Violations	0	5	1	0	2
Types		***Use Errors (# Required Facilitator Events)***			
Kiosk	4	1	0	0	2
Activation	2	3	3	2	2
Study-Wide	0	0	0	0	1
Reporting			***Kiosk Reporting Rate***		
Facilitator	0%	0%	10%	0%	47%
Survey	10%	7%	20%	7%	57%

TABLE 3: **Usability Results Overview.** The top section up to "Like Least" represents participants' perception of system usability and user experience. The middle section represents their use errors and facilitator interventions during the study. The final distinct table represents the kiosk reporting rate. ↑↑ is top 25% of benchmark studies, ↑ is top 50%, ↓ is bottom 50% and ↓↓ is bottom 25%.

We observe that the control group perceived the system most positively. As the level of engagement increased, subsequent groups showed a consistent decline in positive ratings. Meanwhile, neutral and negative ratings vary between 13 to 27%, with the exception of group M (a 37% negative rating). This surge partly derives from participants' suspicion of the kiosk's unexpected (and incorrect) behavior, expressing their surprise with words like "hacked" and "suspect"—terms absent from other groups' feedback, including SM. Moreover, while group M shows the fewest participants valuing security among the treatment groups, group SM shows the most. Despite these observations, we cannot confirm statistical differences between the groups and the ratings (Fisher's Exact Test, $p = 0.24$).

We find a roughly equal divide between those who find the system easy to use and those who consider it complex. We also observe a roughly equal split between those who appreciated the system's instructional guidance and those who disliked the handling (e.g., scanning, storage) of paper credentials. We also find that given the lack of online voting in the study location, 15% of participants appreciate the ability to vote remotely. A minority (4%) expressed dissatisfaction with the need for in-person registration. Finally, we observe that participants in the treatment groups valued the system's security at least threefold compared to the control group, with half of these comments praising fake credentials.

5.2. Use Errors

During the study, we record observational notes for each participant, including use errors and facilitator interventions. We focus on use errors fundamental to the credentialing process, while we discuss interface and device-induced use errors in §A.2. We classify use errors into two categories: mistakes and violations. We characterize mistakes when participants' plan or intended action is flawed, typically resulting from misinterpretation of instructions. In contrast, violations arise when participants intentionally disregard or skip instructions. We now present a discussion of process-induced use errors we observed, as reported in Table 3.

Credentialing. During the real credential creation process with an honest kiosk (C, F, SF), 5 participants (6%) prematurely tore off the receipt after the kiosk printed the first QR code—a mistake that required us to intervene during activation. An additional 8 participants (9%) attempted to tear off the receipt but when they encountered resistance from the receipt printer, they rectified their mistake by consulting the kiosk display guiding them to pick and scan an envelope. In terms of envelope selection, 5 participants (6%) initially opted for an envelope that did not correspond with the symbol on the receipt; the kiosk alerted them to this mistake and they all successfully scanned a correct envelope on their second attempt. A single participant misinterpreted the "discard check-in ticket" screen after creating their real credential and mistakenly threw away their credential.

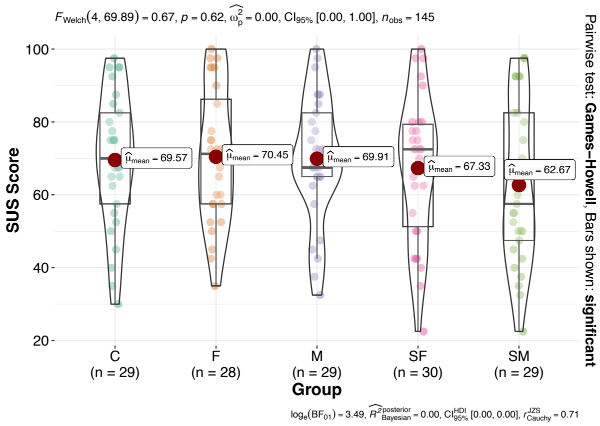
Activation. The majority of use errors occurred during the activation phase, after "leaving the government office", with approximately 19% of participants encountering difficulties and 8% requiring facilitator assistance. Participants commonly skipped the on-screen instructions, a violation error. Participants therefore removed the entire receipt from the envelope when the device prompted to scan three QR codes. These participants likely associated the "three QR codes" to be the three QR codes on the receipt rather than the two QR codes from the receipt and one QR code from the envelope. For those participants who read the instructions yet still faced challenges, mistakes often involved scanning only a subset of QR codes at once instead of all three simultaneously, and placing the credential on the device's screen instead of on the table to be scanned by the device's camera. Considering these activation errors, it naturally follows that credential handling emerged as the second most disliked aspect of the system among participants.

Success Rate. 95% of participants succeeded in registering and creating their credentials without assistance from the facilitator. This result appears to support the practicality and usability of a coercion-resistant voter-verifiable registration process like TRIP's, despite its complexity. Success rate drops when we include errors later in the voting pipeline, however. 92% of participants activated their credential without help. In combination, 87% accomplished both registration and activation by themselves, with 19 participants (13%) needing assistance. Counting participants who mistakenly used their fake credential to cast their vote further reduces the success rate to 83%, as we discuss in §6.2.

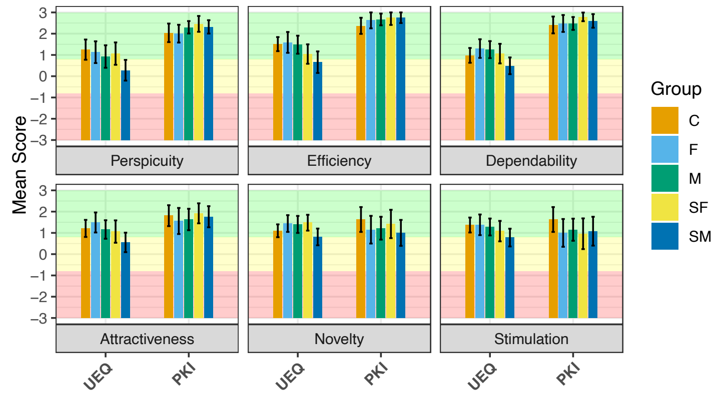
Statistical Analysis. We categorized participants into four age groups: 18-30, 31-45, 46-65, and 65+. Analysis of use errors across these age groups using the chi-squared test revealed no statistically significant differences ($p = 0.8791$). Further examination employing logistic regression to assess the interaction effects between age groups, ethnicity, education, and gender on use errors also indicated no statistically significant associations as determined by the Wald test.

System Improvements. The study results suggest several potential ways to improve the success rate. A receipt detection mechanism could enable the kiosk to restart credential creation if the user prematurely tears off the receipt. Activation issues might be reduced by reading the QR codes incrementally, instead of expecting three readable QR codes in one image. Redisplaying the activation instructions upon unsuccessful activation may help users who skip the instructions. An animation showing how to use the device's rear-facing camera to scan credentials might also help.

Study Comparisons. We compare our success rate with three usability studies containing seven variants of voting systems: the only other JCJ-style coercion-resistant voting study we are aware of by Neto et al. [26], a study by Acemyan et al. [48] on three voter-verifiable systems (Helios [6], Prêt à Voter [49], Scantegrity II [50]), and a recent study on STAR-Vote [51], a secure, auditable and transparent ballot marking device for in-person voting. The Neto et al. [26] between-subjects study involved 80 university-affiliated individuals aged 18-39 acquiring their real credential using one of the following three variants and then casting both real and fake votes: (1) in-person acquiring a pen drive, (2) remotely via email, and (3) in-person with a password set by the voter. The success rate for the first two variants is 100% and the last one is 85%. Acemyan et al. [48] conducted a within-subjects study with 37 diverse participants on Helios, Prêt à Voter and Scantegrity II where Helios is an voter-verifiable online voting system while Prêt à Voter and Scantegrity II are coercion-resistant, voter-verifiable, in-



(a) **System Usability Scale Across Groups**



(b) **User Experience Questionnaire Mean Scores Across Scales**

Figure 8: **Usability Scales**

person voting systems. The success rate for Helios is 60%, Prêt à Voter is 60%, and Scantegrity II is 50%. The study on STAR-Vote [51] reported a success rate of 93%.

Discussion. TRIP outperforms three out of the seven variants, including prominent coercion-resistant, voter-verifiable, in-person voting systems Prêt à Voter and Scantegrity II, which TRIP most closely resembles as a coercion-resistant, voter-verifiable, in-person registration system. Despite the remaining four variants surpassing TRIP, they either fall short in achieving one or more of these properties—illustrating the challenge in designing a usable coercion-resistant system—or have a homogeneous study population. TRIP substantially improves this success rate to 83%, up from 60% (Prêt à Voter) and 50% (Scantegrity II), and closer to the 93% success rate that STAR-Vote achieves.

To our knowledge, NIST has established standards for electronic voting [52], but no official success rate exists. Only a NIST internal document [53] recommends a success rate of 98% (known as Total Completion Score) but where 98% only needs to fall within the 95% confidence interval.

5.3. Perceived Usability and User Experience

We now present the system usability scale and the user experience questionnaire scores and compare them to benchmarks, along with the systems mentioned in §5.2.

System Usability Scale. In Figure 8a, we present the System Usability Scale (SUS) scores across our study groups, while removing 5 participants due to inconsistent answers (agreeing to most positive and negative items), as suggested by Sauro [54]. We observe that groups C, F, and M demonstrate similar average and median scores while groups SF and SM exhibit lower average scores, with group SM having a score decrease of 10% over the control group. Despite these observations, Welch's one-way ANOVA reveals no statistically significant differences (p = 0.62). Nonetheless, we can compare our SUS mean scores with a benchmark consisting of 446 studies involving a range of products and services, from business/consumer software to hardware devices [54]. These 446 studies reveal an average SUS score of 68, with standard deviation of 12.5. Groups C, F, and M achieve marginally superior SUS scores, ranking in the 55.1st, 57.8th and 56.1st percentiles, respectively. Conversely, groups SF and SM underperform with scores ranking in the 47.8th and 35.0th percentiles. Works such as Bangor et al. [55] have proposed adjective ratings based on the SUS score. These ratings would classify group F—representing the expected common case in realistic settings—as "Acceptable," and achieving a "Good" adjective rating.

SUS: Study Comparisons. We now compare group F's SUS score with Neto et al. and Acemyan et al.'s study [26], [48]. While Neto et al.'s variant 2 and 3 achieve a higher mean SUS score of 77.5 (81st percentile) and 77.4 (81st percentile), respectively, these were not statistically significantly higher than group F's score (Welch one-sided t-Test, $p = 0.094$ and $p = 0.051$). Neto et al., unfortunately, did not administer the SUS questionnaire to participants for variant 1, which is the most comparable to TRIP, in that it involves the in-person delivery of voter materials. From Acemyan et al.'s [48] study[9], Group F's mean SUS score of 70.4 was statistically significantly higher than Prêt à Voter's score of 61 (Welch one-sided t-test, $p < 0.05$), and Scantegrity II's score of 59 (Welch one-sided t-test, $p < 0.05$) but lacked conclusive statistical difference from Helios' score of 76 (Welch two-sided t-test, $p = 0.17$).

SUS: Discussion. The comparable percentile scores and the absence of statistical difference between groups C, F, and M suggest that introducing fake credentials do not change participants' perceptions of usability. Further, even with voters engaging in a nontrivial 4-step protocol, F achieved

9. These numbers are not present in the text [48]; we extract these numbers from their figure 4 and assume a 95% confidence interval.

an Acceptable rating and achieved statistically significantly higher SUS scores than Scantegrity II and Prét à Voter.

User Experience Questionnaire. Figure 8b and Table 7 illustrate the UEQ scores for each scale (Attractiveness, Dependability, Efficiency, Novelty, Perspicuity, and Stimulation), along with the UEQ Key Performance Indicator (KPI) scores, which represent participants' perception of an ideal registration experience. Traditional usability aspects encompass Efficiency, Perspicuity, and Dependability, while Novelty and Stimulation relate to user experience. Upon comparison, participants view this system as slightly surpassing their expectations in Novelty, Stimulation, and Attractiveness. However, it falls short in the traditional usability aspects by approximately a full point. Nonetheless, scores above -2 and 2 are extremely rare due to differing opinions and answer tendencies [56] (e.g., avoidance of extreme responses). Typically, as depicted in Figure 8b, a positive evaluation is a score above 0.8, neutral evaluation is a score between -0.8 and 0.8, and negative evaluation is a score less than -0.8 [56]. Based on this metric, Group F has a positive evaluation for each of the scales.

UEQ: Benchmark. Similar to SUS, to better assess usability, we need to compare the UEQ scores with other studies. We first compare our scores to the UEQ benchmark [56], consisting of 21,175 participants from 468 diverse studies. While group F's scores for efficiency, stimulation and novelty rank in the top quartile, the perspicuity scores are between the 50th and 75th percentiles. This suggests that despite participants finding it more challenging to familiarize themselves with this system compared to the benchmark, they still complete their tasks without excess effort, while perceiving the system as both engaging and innovative.

UEQ: Study Comparisons. We now compare our scores with other voting-related studies, although the number of studies that administer UEQ is more limited. One comparable study conducted by Marky et al. [57] focuses on voter verifiability—though not coercion-resistance—through the use of code voting [58] as prominently used in the Swiss Online Voting System [17]. This approach provides voters with physical materials, typically by mail, to cast a vote so as to prevent potential vote alteration by the voter's device. Marky et al. tests three voting code variants—manual codes, QR codes, and tangibles—with 18 participants. Our interest lies in the QR-codes variant, as it is most similar to TRIP, and is also the option favored in Marky et al.'s study. Unlike in TRIP, their scores for user experience (stimulation and novelty) are neutral, although their scores for the traditional usability aspects (perspicuity, efficiency and dependability) are all positive. Marky et al.'s study also achieves higher perspicuity and efficiency scores, 2.2 vs. 1.13, and 2 vs. 1.59, respectively, while TRIP achieves a marginally higher dependability score 1.3 vs. 1.2.

5.4. Detecting a Malicious Kiosk

We present our findings from participants exposed to our malicious kiosk (groups M and SM). We assess their ability to identify the kiosk's misbehavior during their interaction with the kiosk and via a survey question probing if they detected any irregularities while creating their real credential (Appendix D.6 in [40]). We administered the same survey question to groups F and SF as a control to identify false positives rates. Table 3 presents the reporting rates.

Kiosk Reporting Rate. Among group M participants, 10% reported the kiosk's misbehavior to the facilitator, whereas this rate significantly increased to 47% for group SM. Further, this difference in reporting rate is statistically significant (Chi-squared, $\chi^2 = 8.2079, p < 0.01$; Cramér's V $= 0.4068$).[10] Instructional video 3, shown to group SM, thus substantially increased the reporting rate.

Survey Reporting Rate. A greater percentage of participants reported the kiosk's misbehavior in the exit survey: 20% for group M and 57% for group SM. We still observe a statistically significant difference in the reporting rates between groups M and SM (Chi-squared test, $\chi^2 = 7.0505, p < 0.01$; Cramér's V $= 0.3771$). Unlike the kiosk reporting rate, however, the survey reporting rate was non-zero for groups C, F, and SF, averaging at 8%; these participants primarily cited confusion about the process.

Real-World Scenario. According to participants' responses to our survey questions, 85% are willing to watch an instructional video before participating in voter registration in a real world context. However, only 59% indicated willingness to "lock up" their personal devices in a locker before entering the booth and retrieve them afterwards.

Study Comparison. Bernhard et al. [35] conducted a study among diverse participants to examining the rate at which voters could detect malicious ballot manipulation from ballot marking devices (BMDs). Both our and Bernhard et al.'s studies involved participants operating under the election authority's supervision, interacting directly with an official device: the BMD in their case and the kiosk in ours. In both studies, we assess whether participants can visually discern anomalous behavior from these devices. In their study, without any guidance, only 6.6% of 31 participants reported the error to the facilitator. When participants were asked before submission if they had carefully reviewed their ballot, 12.9% of 31 reported the error to the facilitator. In contrast, the reporting rate for TRIP is 10% without security priming, and 47% with priming.

6. Usability of Fake Credentials

This section evaluates the comprehension and usability of fake credentials among the 120 participants exposed to them.

6.1. Understanding Fake Credentials

We examine voter comprehension of fake credentials in two distinct stages: first with a pop quiz presented by the kiosk, then later in the survey. This dual-stage assessment measures understanding of passive instruction and efficacy

10. For Cramér's V, values around 0.1 are typically considered small, around 0.3 indicate a moderate effect size, and values of 0.5 or greater signify a strong effect.

Quiz Attempts	1	2	3	Continue (Missing Correct Options)	Continue (Incorrect Option)	Count
Real Credential Steps	65	21	1	0	3	90
Real Credential Usage	148	2	0	0	0	150
Test Credential Usage	84	18	4	7	7	120
Distinguish Credentials	91	1	0	0	0	92

TABLE 4: **Quiz Results.** This table represents the quiz attempts; "Continue" signifies that these participants could not pass the quiz after the third attempt. "Missing Correct Options" indicates that these participants only selected correct answers but did not simultaneously choose all correct answers. "Incorrect Option" indicates that these participants selected the incorrect option in one of the quiz attempts and could not rectify it by the third attempt. §3.4 discusses participants' assignment to quizzes.

of active instruction. The quiz poses the question "Which of the following purposes can you use a test credential for?" (§C.3). Not selecting "To cast a vote that counts in election" underscores the crucial understanding that fake credentials cannot be used to cast a real vote; the remaining three correct options must be concurrently selected to pass the quiz.

Quiz Results: Test Credential Usage. We find that 70% of participants answer the quiz correctly on their first attempt (Table 4). An additional 15% succeed on the second attempt, including 3 who initially chose the incorrect answer. The third and final attempt sees the remaining 15% (18 participants), split as follows: 4 participants answer correctly, 7 fail to select all correct options simultaneously, and 7 choose the incorrect option. In this quiz stage, 92% of participants avoid selecting the incorrect answer in all attempts, thereby demonstrating an understanding of fake credentials.

Exit Survey. Among the 8% of participants (3 from second attempt + 7 from third attempt) who incorrectly selected "To cast a vote that counts in an election" in the quiz, five of them correctly identified the use of fake credentials in the survey. In the remaining five, four could not recall the purpose of fake credentials, and one only wrote "voting". We infer that 4% (5 participants) likely finished the study without a clear understanding of fake credentials, with one participant thinking that fake credentials can be used to cast real votes. Among the participants who did not select the incorrect option, 89% remembered, in their free-text responses, the use of fake credentials to resist coercion. The remaining 11% wrote about their use to educate others, test the system, and even profit by selling their fake credentials.

6.2. Usability of Real and Fake Credentials

Considering that both real and fake credentials are used to cast votes in the same way—assessed previously in §5.2—the success of using real or fake credentials ultimately hinges on the voter's ability to distinguish between the two. In the study, out of the 120 participants exposed to fake credentials, 92 chose to create one or more fake credentials.

Credential Distinguish Quiz. We initially verify participants' understanding of how to distinguish their credentials by administering the Credential Distinguish quiz (§C.3). Remarkably, 99% of participants select the correct option "Only myself with my pen markings" on their first attempt, thereby confirming their apparent comprehension (Table 4).

Distinguishing Credentials. We assess participants' *actual* ability to distinguish their credentials by asking them to cast a vote using their real credential after completing voter registration. Among the 92 participants, 90% (83) accurately identified their real credential. For the 10% (9 participants) who failed to do so, we explore the reasons behind this.

One participant discarded their real credential during registration, as mentioned in §5.2. We examined each set of credentials the remaining eight participants created to ascertain whether their real credentials had distinct markings, and found that they did. For details about how study participants generally marked their credentials, see Appendix A.3. Although credential marking was apparently not the cause of these errors, many other potential causes remain that we could not identify, such as memory lapses, misreading instructions, environmental distractions, or even deliberately disobeying our instructions (e.g., viewing the facilitator as a potential coercer, which "by our own game" might suggest voting with a fake credential in the facilitator's presence).[11]

To gain deeper insights into individuals' ability to recall sensitive data, we consider password-related studies [59], [60], which indicate a wide range of retention rates from 23% to 98%. With a 90% success rate, this rate lies in the upper end of this spectrum and is similar to Déjà Vu [60], a study conducted on using images for authentication. Voters marking their own credentials incorporate several known retention-enhancing strategies, such as user-generated content and favoring visual imagery over text. The Déjà Vu study provides encouraging evidence that imagery-based memory degradation is significantly lower than that of PIN/passwords: after one week, only 1 out of 20 participants failed to login with images while 7 to 6 individuals failed to login with PIN and passwords, respectively.

We expect that the duration over which TRIP users must remember their credential markings should normally be much shorter than the requirements for long-term passwords: from leaving the privacy booth until credential activation on their device. This duration may be only minutes if the voter brought their voting device with them, hours or at most days for a credential the voter activates at home or elsewhere. Voters who do forget which credential is real may re-register at any time to obtain fresh credentials.

Confidence in Distinguishing Credentials. Participants also rated how confident they were in recalling their real credential on a 7 point scale, where 1 is "No confidence" and 7 is "Extremely confident." 87% of participants gave a confidence rating of 5 or more, with 55% of participants giving the extremely confident rating. 7% rated a 4 (neutral), and ratings 1 and 2 each got one participant.

Fake Credentials In Reality? We ask participants about their willingness to create fake credentials if such a system

11. We informally observed a few participants hiding their credentials under the table out of the facilitator's sight, which might suggest such a "facilitator as adversary" perspective.

existed; 53% of participants affirm that they would. Testimonials vary from "because I have been in situations where others forced me to vote", "to argue less with people voting for another candidate" to "[...] a fun souvenir", and "I'm in a demographic where I cannot imagine having someone trying to solicit my vote, if they did, I would simply tell them no without fear." Additional testimonials, including a taxonomy, are available in Appendix A.4.

Discussion. Despite introducing an unforeseen and unprecedented concept to 120 participants, only 4% (5 participants) appear to have finished the study without a reasonable grasp of the use of fake credentials. Furthermore, although voters were not required to generate fake credentials, 76% of participant chose to create at least one. This proactive engagement is further expressed by 53% of participants who are willing to create fake credentials in a real-world context. Lastly, in spite of the identical nature of fake and real credentials, 90% of participants successfully distinguished between the two when aiming to activate and cast a real vote, on par with a study of using images for authentication.

7. Discussion

This section primarily offers key takeaways of our findings, building on the many results we have presented earlier. We also expand on our study's limitations.

The Need For Coercion-Resistance. Our study reveals that a quarter of participants have either personally experienced coercion, or is aware of someone who has. These findings highlight the necessity of coercion-resistance to uphold free and fair elections. In essence, votes should not just be private but incorruptible. This importance is further reflected with these same participants rating (Fig. 9c) statistically significantly lower trust levels for both mail-in and online voting methods compared to in-person voting (with BMD).

Usability of Coercion-Resistant Voting Systems. Designing a usable coercion-resistant voting system is a difficult task. As seen in §5.2, voting systems without coercion resistance, such as STAR-Vote [18], exhibit much higher success rates than systems that address coercion, such as Prêt à Voter [61] and Scantegrity II [50] (93% versus 60% and 50% respectively). This disparity has motivated other coercion-resistant strategies with less intricate tasks such as deniable re-voting [19], [21], where voters only receive a single voting credential and perform the same vote casting process to override their previous vote. However, this strategy is vulnerable to last-minute coercion, as well as to domestic coercion—the primary source of coercion in our findings—where the coercer simply keeps the voting credential or device. The 83% success rate achieved by TRIP shows promise in narrowing the success gap between coercion-resistant and non-coercion-resistant systems, though this gap remains an important usability challenge for future work.

The Impact of Security Vigilance on User Experience. Our findings show an interesting tradeoff between user experience and security vigilance. Participants from group SF typically rated system usability and user experience lower than group F, despite the only change being an educational brief in the instructional video about the rare possibility of a compromised kiosk (§C.2). This increase in discomfort, however, came with a statistically significant improvement in detecting a malicious kiosk in group SM compared with group M (47% vs. 10%), with no increase in false positives. Meanwhile, twice as many participants in SM compared to M liked the system's security the most. These observations confirm that finding the right balance between security education and user comfort is an important challenge.

Limitations. This study has several important limitations. First, it did not evaluate long-term questions, such as whether voters can effectively store, manage, and properly use credentials on their devices after activation to vote in successive elections over periods of years. While the problem of remembering which paper credential is real technically ends at credential activation time in TRIP, the problem of remembering which *device* holds the real voting credential – or of remembering which account, password, or PIN guards it – remains. Thus, the detailed design and evaluation of long-term storage systems for coercion resistance with fake credentials remains an important area for future work. Systematically evaluating the usability of long-term credential storage systems may be a particularly difficult challenge, as studies addressing this challenge would necessarily be longitudinal in nature, requiring participants to remain involved over an extended period. We know of no deployed voting system that could facilitate such a study.

The TRIP system's accessibility to voters with disabilities is also important but beyond the scope of this study. In the future, a thorough design for accessibility and corresponding study should include voters with disabilities, which would likely yield further insights. Although we aim for diversity, the study being conducted in a single geographical location may affect the generalizability of our findings. Finally, the study did not replicate a government office setting, which might influence differences in participants' views and behaviors. Similarly, the study does not replicate circumstances where a voter is truly under coercion.

8. Related Work

In previous sections (§5.2 & §5.3), we compared our results with other usability studies. This section therefore provides a broader overview of related work.

A few works [24], [25], [62] have considered usability issues with fake credentials and proposed possible implementations, such as using smart cards. These works did not conduct systematic studies with real users, however.

Civitas [63], strengthens the verifiability of the JCJ coercion-resistance scheme by having voters interact with multiple registration officials. There are no user studies of Civitas, however, and such a study may be difficult due to the logistical challenges presented by multiple registrars, and the absence of a well-defined voter-facing design.

Other works have looked into using other countermeasures to resist coercion [22], such as deniable re-voting and masking. The masking approach [58] provides each voter with a unique value b, known only to them, which is then used to offset their cast vote. However, a usability concern arises as it remains uncertain if voters can recall and apply their assigned value b during vote casting [22].

An important security-related study in voting is one by Distler et al. [64], which examines the impact of displayed security mechanisms on user experience. They found that participants exposed to these mechanisms, e.g., messages like "Encrypting your vote", have a better user experience than those who were not exposed to these mechanisms. Their research thus highlights the importance of the visibility of security measures. In contrast to their work, we study the involvement of voters to achieve security measures.

9. Conclusion

This paper presented a study with 150 individuals to evaluate their experiences and perceptions of coercion, along with their views of a coercion-resistant system with a voter-verifiable registration system resilient to coercion via the use of fake credentials. A quarter of participants had either personally faced coercion or know of someone who did. Remarkably, 96% of participants understand the concept of fake credentials, with 90% successfully casting a vote using their real credential. Moreover, over half of the participants exposed to fake credentials indicated their willingness to use them in reality. These findings show promise in narrowing the usability gap between voting systems with and without coercion-resistance, and confirm the need for continued research into making verifiable coercion-resistant voting systems more usable.

Acknowledgements

The authors would like to thank the study participants for their invaluable insights and their time, as well as the reviewers for their helpful feedback. This project was supported in part by the armasuisse Science and Technology, the AXA Research Fund, and the US ONR grant N000141912361.

References

[1] P. Ehin, M. Solvak, J. Willemson, and P. Vinkel, "Internet voting in Estonia 2005–2019: Evidence from eleven elections," *Government Information Quarterly*, vol. 39, no. 4, p. 101718, Oct. 2022.

[2] P. van den Besselaar, A.-M. Oostveen, F. De Cindio, and D. Ferrazzi, "Experiments with E-Voting Technology: Experiences and Lessons," Social Science Research Network, Scholarly Paper 1433569, 2003.

[3] D. Jefferson, A. D. Rubin, B. Simons, and D. Wagner, "A Security Analysis of the Secure Electronic Registration and Voting Experiment (SERVE)," 2004.

[4] R. Krimmer, D. Duenas-Cid, and I. Krivonosova, "Debate: Safeguarding democracy during pandemics. Social distancing, postal, or internet voting—the good, the bad or the ugly?" *Public Money & Management*, vol. 41, no. 1, pp. 8–10, Jan. 2021.

[5] J. Benaloh, "Verifiable Secret-Ballot Elections," Ph.D. dissertation, 1987.

[6] B. Adida, "Helios: Web-based Open-Audit Voting." in *USENIX Security Symposium*, vol. 17, 2008, pp. 335–348.

[7] V. Cortier, D. Galindo, R. Küsters, J. Müller, and T. Truderung, "SoK: Verifiability Notions for E-Voting Protocols," in *2016 IEEE Symposium on Security and Privacy*, 2016.

[8] R. Küsters, J. Liedtke, J. Müller, D. Rausch, and A. Vogt, "Ordinos: A Verifiable Tally-Hiding E-Voting System," in *2020 IEEE European Symposium on Security and Privacy*, 2020.

[9] Y.-X. Kho, S.-H. Heng, and J.-J. Chin, "A Review of Cryptographic Electronic Voting," *Symmetry*, vol. 14, no. 5, p. 858, May 2022.

[10] J. Benaloh, "Ballot Casting Assurance via Voter-Initiated Poll Station Auditing," in *2007 USENIX/ACCURATE Electronic Voting Technology Workshop*, Boston, MA, Aug. 2007.

[11] V. Cortier, A. Debant, P. Gaudry, and S. Glondu, "Belenios with cast as intended," in *Voting 2023 - 8th Workshop on Advances in Secure Electronic Voting*, May 2023.

[12] A. Juels, D. Catalano, and M. Jakobsson, "Coercion-Resistant Electronic Elections," in *Towards Trustworthy Elections: New Directions in Electronic Voting*, Berlin, Heidelberg, 2010, pp. 37–63.

[13] J. Benaloh and D. Tuinstra, "Receipt-free secret-ballot elections," ser. STOC '94, May 1994.

[14] S. Park, M. Specter, N. Narula, and R. L. Rivest, "Going from bad to worse: From Internet voting to blockchain voting," *Journal of Cybersecurity*, vol. 7, no. 1, Feb. 2021.

[15] P. Daian, T. Kell, I. Miers, and A. Juels, "On-Chain Vote Buying and the Rise of Dark DAOs," Jul. 2018.

[16] R. Küsters, T. Truderung, and A. Vogt, "A Game-Based Definition of Coercion-Resistance and Its Applications," in *23rd IEEE Computer Security Foundations Symposium*, 2010, pp. 122–136.

[17] "Protocol of the Swiss Post Voting System: Computational Proof of Complete Verifiability and Privacy," Tech. Rep. 1.0.0, 2021.

[18] S. Bell, J. Benaloh, M. D. Byrne, D. Debeauvoir, B. Eakin, P. Kortum, N. McBurnett, O. Pereira, P. B. Stark, D. S. Wallach, G. Fisher, J. Montoya, M. Parker, and M. Winn, "STAR-Vote: A Secure, Transparent, Auditable, and Reliable Voting System," in *2013 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections*.

[19] D. Achenbach, C. Kempka, B. Löwe, and J. Müller-Quade, "Improved Coercion-Resistant Electronic Elections through Deniable Re-Voting," in *USENIX Journal of Election Technology and Systems*, 2015.

[20] K. Krips and J. Willemson, "On Practical Aspects of Coercion-Resistant Remote Voting Systems," in *E-Vote-ID 2019: Electronic Voting*, 2019, pp. 216–232.

[21] W. Lueks, I. Querejeta-Azurmendi, and C. Troncoso, "VoteAgain: A scalable coercion-resistant voting system," in *29th USENIX Security Symposium*, 2020, pp. 1553–1570.

[22] O. Kulyk and S. Neumann, "Human Factors in Coercion Resistant Internet Voting–A Review of Existing Solutions and Open Challenges," in *Proceedings of the Fifth International Joint Conference on Electronic Voting*. TalTech press, 2020, p. 189.

[23] D. Springall, T. Finkenauer, Z. Durumeric, J. Kitcat, H. Hursti, M. MacAlpine, and J. A. Halderman, "Security Analysis of the Estonian Internet Voting System," in *ACM SIGSAC Conference on Computer and Communications Security*, 2014.

[24] S. Neumann and M. Volkamer, "Civitas and the Real World: Problems and Solutions from a Practical Point of View," in *Seventh Conference on Availability, Reliability and Security*, 2012.

[25] E. Estaji, T. Haines, K. Gjøsteen, P. B. Rønne, P. Y. A. Ryan, and N. Soroush, "Revisiting Practical and Usable Coercion-Resistant Remote E-Voting," in *E-Vote-ID 2020: Electronic Voting*, 2020.

[26] A. S. Neto, M. Leite, R. Araújo, M. P. Mota, N. C. S. Neto, and J. Traoré, "Usability Considerations For Coercion-Resistant Election Systems," in *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems*, ser. IHC 2018, Oct. 2018, pp. 1–10.

[27] L.-H. Merino, S. Colombo, R. Reyes, A. Azhir, H. Zhang, J. Allen, B. Tellenbach, V. Estrada-Galiñanes, and B. Ford, "TRIP: Trust-Limited Coercion-Resistant In-Person Voter Registration." [Online]. Available: https://arxiv.org/abs/2202.06692

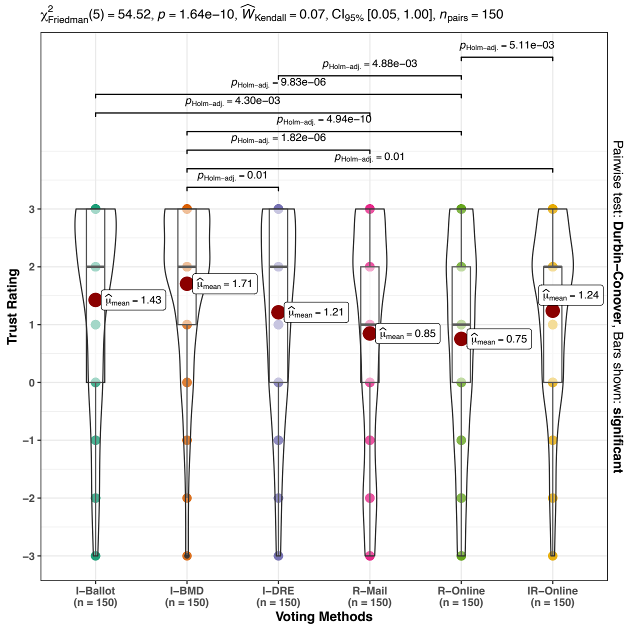
[28] D. Chaum, R. T. Carback, J. Clark, C. Liu, M. Nejadgholi, B. Preneel, A. T. Sherman, M. Yaksetig, and F. Zagorski, "VoteXX: A Remote Voting System that is Coercion Resistant," Oct. 2020.

[29] M. Backes, M. Gagné, and M. Skoruppa, "Using mobile device communication to strengthen e-Voting protocols," in *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, 2013.
[30] T. Krivoruchko, "Robust Coercion-Resistant Registration for Remote E-voting," in *Proceedings of the IAVoSS Workshop on Trustworthy Elections*, 2007.
[31] O. Spycher, R. Koenig, R. Haenni, and M. Schläpfer, "A New Approach towards Coercion-Resistant Remote E-Voting in Linear Time," in *Proceedings of the 15th International Conference on Financial Cryptography and Data Security*, 2011, pp. 182–189.
[32] R. Araújo, S. Foulle, and J. Traoré, "A Practical and Secure Coercion-Resistant Scheme for Internet Voting," in *Towards Trustworthy Elections: New Directions in Electronic Voting*, 2010.
[33] J. Benaloh, T. Moran, L. Naish, K. Ramchen, and V. Teague, "Shuffle-Sum: Coercion-Resistant Verifiable Tallying for STV Voting," *IEEE Transactions on Information Forensics and Security*, 2009.
[34] S. G. Weber, R. Araujo, and J. Buchmann, "On Coercion-Resistant Electronic Elections with Linear Work," in *The Second Conference on Availability, Reliability and Security*, 2007.
[35] M. Bernhard, A. McDonald, H. Meng, J. Hwa, N. Bajaj, K. Chang, and J. A. Halderman, "Can Voters Detect Malicious Manipulation of Ballot Marking Devices?" in *2020 IEEE Symposium on Security and Privacy*, May 2020, pp. 679–694.
[36] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation In Industry*. CRC Press, 1996.
[37] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 2008, pp. 63–76.
[38] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, and J. Thomaschewski, "Developing a UX KPI based on the user experience questionnaire," *Computer Standards & Interfaces*, 2019.
[39] J. Kjeldskov and M. B. Skov, "Studying Usability In Sitro: Simulating Real World Phenomena in Controlled Environments," *International Journal of Human–Computer Interaction*, 2007.
[40] L.-H. Merino, A. Azhir, H. Zhang, S. Colombo, B. Tellenbach, V. Estrada-Galiñanes, and B. Ford, "E-Vote Your Conscience: Voter Perceptions on Coercion and Vote Buying, and the Usability of Fake Voting Credentials for Online Voting." in *2024 IEEE Symposium on Security and Privacy (SP)*. [Online]. Available: https://arxiv.org/abs/2404.12075
[41] R. Ploetzner, S. Berney, and M. Bétrancourt, "When learning from animations is more successful than learning from static pictures: Learning the specifics of change," *Instructional Science*, 2021.
[42] S. Türkay, "The effects of whiteboard animations on retention and subjective experiences when learning advanced physics topics," *Computers & Education*, vol. 98, pp. 102–114, Jul. 2016.
[43] R. E. Mayer, L. Fiorella, and A. Stull, "Five ways to increase the effectiveness of instructional video," *Educational Technology Research and Development*, vol. 68, no. 3, pp. 837–852, Jun. 2020.
[44] R. Likert, "A technique for the measurement of attitudes." *Archives of Psychology*, vol. 22 140, pp. 55–55, 1932.
[45] "Four people plead guilty in North Carolina ballot probe of 2016 and 2018 elections," NBC News, 2022.
[46] "Verifier - Verified Voting 2024," https://verifiedvoting.org/verifier/.
[47] S. Atske, "Pew Research Center - The voting experience in 2020," Nov. 2020.
[48] C. Z. Acemyan, P. Kortum, M. D. Byrne, and D. S. Wallach, "Usability of Voter Verifiable, End-to-end Voting Systems: Baseline Data for Helios, Prêt à Voter, and Scantegrity II," vol. 2, no. 3, 2014.
[49] P. Y. A. Ryan, D. Bismark, J. Heather, S. Schneider, and Z. Xia, "PrÊt À Voter: A Voter-Verifiable Voting System," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 662–673, 2009.
[50] D. Chaum, R. Carback, J. Clark, A. Essex, S. Popoveniuc, R. L. Rivest, P. Y. A. Ryan, E. Shen, and A. T. Sherman, "Scantegrity II: End-to-End Verifiability by Voters of Optical Scan Elections Through Confirmation Codes," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 611–627, Dec. 2009.
[51] C. Z. Acemyan, P. Kortum, M. D. Byrne, and D. S. Wallach, "Summative Usability Assessments of STAR-Vote: A Cryptographically Secure e2e Voting System That Has Been Empirically Proven to Be Easy to Use," *Human Factors*, pp. 866–889, Aug. 2022.
[52] NIST, "Voluntary Voting System Guidelines Version 2.0," Tech. Rep., Feb. 2021.
[53] ——, "Usability Performance Benchmarks For the Voluntary Voting System Guidelines," Tech. Rep., Aug. 2007.
[54] J. Sauro, *A Practical Guide to the System Usability Scale*. Measuring Usability LLC, 2011.
[55] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.
[56] M. Schrepp, "User Experience Questionnaire Handbook Version 10."
[57] K. Marky, M. Schmitz, F. Lange, and M. Mühlhäuser, "Usability of Code Voting Modalities," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.
[58] P. Y. A. Ryan and V. Teague, "Pretty Good Democracy," in *Security Protocols XVII*, 2013.
[59] A. S. Brown, E. Bracken, S. Zoccoli, and K. Douglas, "Generating and remembering passwords," *Applied Cognitive Psychology*, vol. 18, no. 6, pp. 641–651, 2004.
[60] R. Dhamija and A. Perrig, "Deja Vu: A User Study Using Images for Authentication," in *9th USENIX Security Symposium (USENIX Security 00)*, 2000.
[61] D. Khader, Q. Tang, and P. Y. Ryan, "Proving prêt à voter receipt free using computational security models," in *2013 Electronic Voting Technology Workshop*, 2013.
[62] S. Neumann, C. Feier, M. Volkamer, and R. Koenig, "Towards a practical JCJ/Civitas implementation," in *Informatik Angepasst an Mensch, Organisation Und Umwelt*, 2013.
[63] M. R. Clarkson, S. Chong, and A. C. Myers, "Civitas: Toward a Secure Voting System," in *2008 IEEE Symposium on Security and Privacy*, May 2008, pp. 354–368.
[64] V. Distler, M.-L. Zollinger, C. Lallemand, P. B. Roenne, P. Y. A. Ryan, and V. Koenig, "Security - Visible, Yet Unseen?" 2019.
[65] P. B. Stark, "There is no Reliable Way to Detect Hacked Ballot-Marking Devices," *Election Law Journal Rules, Politics and Policy*, vol. 19, no. 3, 2019.
[66] A. W. Appel, R. A. DeMillo, and P. B. Stark, "Ballot-Marking Devices Cannot Ensure the Will of the Voters," *Election Law Journal: Rules, Politics, and Policy*, vol. 19, no. 3, pp. 432–450, Sep. 2020.

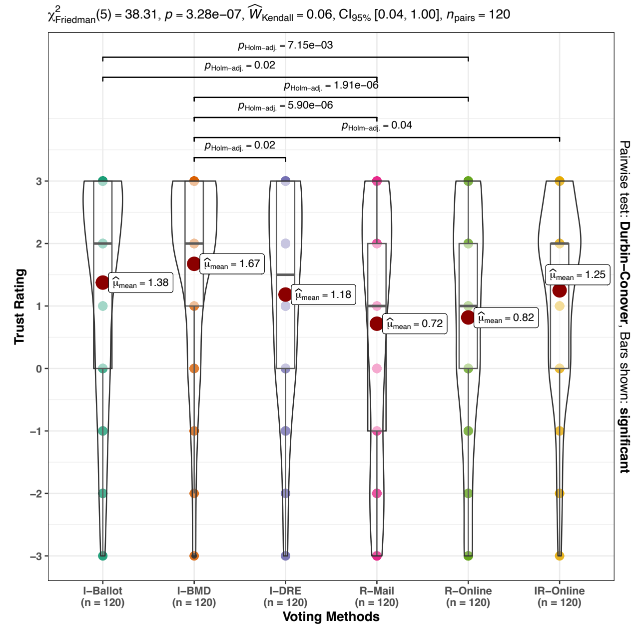
Appendix A. Miscellaneous Questions

A.1. Trust Rating in Other Voting Methods

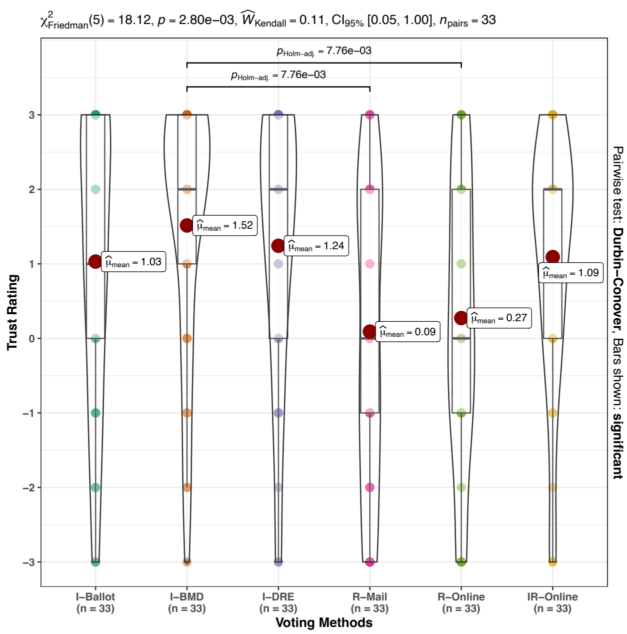
Most Trusted. Participants express the most trust towards `I-BMD`, with 69% deeming it as highly trustworthy or higher, and assigning it an average trust score of 1.68 out of 3 (Table 1). Moreover, `I-BMD` is statistically significantly higher than that of all other voting methods, except for `I-Ballot` (Fig. 9b), which earns the same level of trust from 60% of participants and an average trust score of 1.38.
Least Trusted. Participants express the least trust in `R-Mail`, only 46% considering it highly trustworthy or better (mean score of 0.72). Participants also view `R-Online` similarly, yielding similar mistrust (mean score of 0.85).
All Participants. Figure 9a depicts the trust ratings across voting methods from all 150 participants (including the control group), along with our statistical results. Participants trust in-person registration for online voting statistically significantly more than they trust a fully-online voting system. Participants further emphasized this during the study by



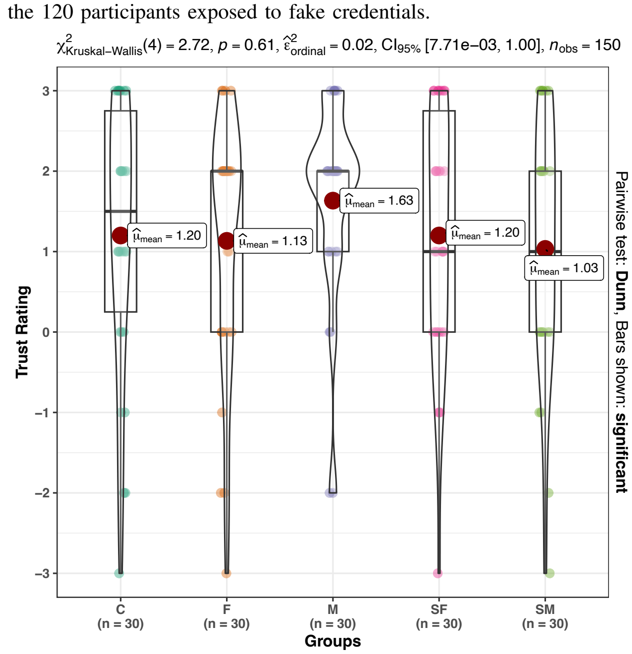
(a) **Voting Methods Trust Rating from All Participants.** All 150 participants' trust ratings across voting methods.



(b) **Voting Methods Trust Rating from Participants Exposed to Fake Credentials.** The trust ratings for various voting methods, as given by the 120 participants exposed to fake credentials.



(c) **Voting Methods from Coercion-Happened Participants.** The trust ratings for various voting methods, as given by the 39 participants who experienced coercion or know someone who has experienced coercion.



(d) **IR-Online Voting Method Across Groups.** The participants' trust ratings across groups for the online voting with in-person registration voting method.

Figure 9: **Voting Methods.**

asking us whether we would require individuals to verify their eligibility, as we had not required this in the study.

Discussion. 46% of voters cast an absentee ballot in the 2020 U.S. presidential election [47]. Despite this, mail-in voting has the second-lowest trust score (Fig. 9a) or the lowest trust score (Fig. 9b & Fig. 9c). Participants expressed a lack of confidence in the U.S. postal service's capacity to deliver mail reliably. Despite concerns with BMDs [65], [66], including a study finding that only 6.6% of participants can detect a *manipulated* printed ballot [35], `I-BMD` received the highest trust score across voting methods.

A.2. Use Errors

We discuss process-induced errors in §5.2. This section discusses device- or interface-induced errors, which include premature scanning of check-in tickets, unawareness to answer the quiz, and unfamiliarity with the device's camera.

Premature Scanning. Of the 150 participants, 56 scanned their check-in ticket before the kiosk asked them to do so. This occurred 51 times in groups C, F and SF. 47 (84%) corrected their mistake without our intervention upon recognizing the correct intended action, such as pressing "Continue" or "Begin". No participants in groups M or SM required our intervention. This pattern occurred mostly in groups C, F and SF due to an instructional slide, for which groups M and SM were not privy due to the kiosk's malicious setting. Upon viewing this instructional slide's four steps, starting with "Scan Check-in ticket", participants scanned their ticket instead of pressing "Begin" to initiate the process. When we were asked to intervene, we guided participants towards the intended action, typically by pressing "Begin", although 84% of participants did not require help. These errors may result from a mix of environmental factors, such as inadequate contrast between the button and the screen in an outside environment, alongside behavioral factors. Such behavioral factors may include participants scanning their check-in ticket before reading the instructions, perhaps due to over-confidence after seeing the instructional video. To address these issues, we propose permitting a combination of inputs: e.g., interpreting the scanning of the check-in ticket during an instructional slide as an indication that the participant is prepared for subsequent steps.

Mobile Device-induced. An additional 12 participants had initial difficulty with activating their credential as they either placed the QR codes directly on the device's screen or used the front-facing camera. In each instance, we intervened to instruct the participant to use the back-facing camera. Participants who continued to need assistance beyond this guidance are reported in the main body of the paper.

Kiosk-Induced. Three participants faced technical difficulties with the kiosk device, necessitating our intervention. Two of these instances arose from participants not realizing they had to complete a quiz before proceeding, while one participant had trouble locating the QR code scanner.

A.3. Credential Markings

We examine how participants mark their credentials to distinguish their real credentials from their fake ones. We categorized the set of credentials for 88 participants (four sets were unaccounted for), first by looking at the type of marking and then level of differentiation based on patterns across credentials. We find six general types of markings:

- Envelope Symbol: No pen markings.
- Number: Random numbers.
- Scribble: Indecipherable writings, including signatures.
- Symbols: Predominantly shapes such as stars or squares but sometimes including smiley faces or animal drawings.
- Text: Words like "Real" or "Test" or participants' initials.
- Type Change: Alternation between two or more of the above categories across the real and test credentials.

Differentiator	Envelope Symbol	Number	Scribble	Symbol	Text	Type Change	Count
Explicit	0	0	0	0	30	2	32
Implicit	0	0	0	5	4	10	19
Indistin-guishable	5	1	10	8	9	4	37
Count	5	1	10	13	43	16	88

TABLE 5: **Credential Markings.** Type of markings that participants used to differentiate their fake credentials from their real credential. "Type Change" signifies participants use distinct types, "Scribble" includes signatures, and "Envelope Symbol" indicates participants used no markings, instead relying on distinct envelope symbols. Four sets of envelopes are unaccounted for.

Our findings, detailed in Table 5, show that 49% of participants used *Text* to differentiate their credentials, followed by *Type Change* at 18%. Furthermore, 42% of participants marked their credentials indistinguishably, making it impossible to differentiate the real credential from the fake ones. Participants may lie about their real credential by marking their real one as "Fake" and their fake one as "Real" but participants in our study were not influenced to do so.

A.4. Fake Credentials in Reality?

We examine whether participants exposed to fake credentials would be willing to create them alongside their real credential if such a system existed in the real world. 53% would do so, citing the following reasons:

- Security (27 participants): "Because I have been in situations where others forced me to vote", "On the off chance someone tries to force me to vote in a particular way".
- Convenience (15 participants): "no harm in creating a test credential", "might come in handy", and "to argue less with people voting for another candidate".
- Education (7 participants): "It might be useful to show my students what the process was like", "To see that it worked and educate family and friends", and "It would be good for teaching people how to vote. Also might be a fun souvenir".
- Other (12 participants): "I would want to create as many test credentials as possible before the market became flooded with test credentials", and "I'd like to test my device setup".
- No Reason Given (2 participants)

The following were reasons for not creating a test credential:

- Unnecessary (39 participants): "I'm in a demographic where I cannot imagine having someone trying to solicit my vote, if they did, I would simply tell them no without fear," and "Not interested in the uses."
- Cumbersome or Confusing (11 participants): "I don't want to take the risk of being confused between my real credentials and the fake one. I can use videos or websites to teach someone else how to vote," and "More to lose, and I would mix them up."
- No Reason Given (7 participants).

Appendix B.
Meta-Review

The following meta-review was prepared by the program committee for the 2024 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers.

B.1. Summary

This paper presents a field-experiment and usability test of the TRIP (Trust-limiting In-Person Registration) protocol: a voter-verifiable registration system for coercion-resistant online voting via the idea of fake credentials. The core contribution of the paper is to provide a sense of how usable the TRIP system is with an experiment that attempts to capture as wide of a voter-pool as possible. The paper details the protocol, discusses the recruitment and study strategy, and provides a detailed analysis of how users interacted with the system, what confusion points and errors occurred, and how these results might be incorporated into a TRIP-enabled system in the future.

B.2. Scientific Contributions

- Independent Confirmation of Important Results with Limited Prior Research
- Provides a Valuable Step Forward in an Established Field
- Provides a New Data Set For Public Use
- Creates a New Tool to Enable Future Science
- Establishes a New Research Direction

B.3. Reasons for Acceptance

1) The paper presents a valuable step forward in an established field by presenting results from a field-experiment for a potential online voting scheme. The authors do a good job of detailing the user study and rigorously testing the TRIP mechanism.
2) The paper also offers researchers insight into leveraging fake credentials for coercion resistance and outlines potential pain points of this mechanism in real-world deployments

B.4. Noteworthy Concerns

1) The results illustrate that some voters may encounter difficulties using fake credentials, with 17% of users encountering an issue and 10% of users mistakenly voting with their fake credential. These findings could adversely affect the system's deployability and underline the need for further research to mitigate and reduce such use errors.
2) It is unclear how requiring in-person registration may impact voter turnout. This presents a practical consideration for the real-world deployment of this system that must be examined in future research.
3) The system's formal threat model excludes coercion threats that could be realistic in practice, such as side-channel attacks, electronic surveillance via wearable devices a voter is successfully coerced to sneak into the registration booth, or a coercer waiting to strip-search the voter immediately upon leaving the booth to confiscate all of their voting credentials. The authors suggest a potential countermeasure for this last case, but all of these risks beyond the formal threat model – and potential mitigations for them – remain important areas for further study.