Click and Cast: Assessing the Usability of Vote App

Laura Cristiano[1][0000−0003−0895−3466], Riccardo Longo[1][0000−0002−8739−3091], and Chiara Spadafora[2][0000−0003−3352−9210]

[1] Fondazione Bruno Kessler, Center for Cybersecurity, 38123 Povo, Trento, IT
{l.cristiano,rlongo}@fbk.eu
[2] University of Trento, Department of Mathematics, 38123 Povo, Trento, IT
chiara.spadafora@unitn.it

Abstract. This article provides a comprehensive examination, from the usability perspective, of Vote App, a coercion-resistant electronic voting system. We report on the results of the usability tests conducted, where participants had to cast a vote in an artificial election while using the anti-coercion mechanisms of Vote App. To evaluate the results, we follow the three usability metrics of effectiveness, efficiency and satisfaction.

Keywords: User-Centered Design · Usability Metrics · SUS · E-Voting

1 Introduction

In recent years, electronic voting (e-voting) systems emerged as a significant innovation for democracy. However, the success of an e-voting system depends on its usability - how easily, effectively and satisfactorily voters can use it.

In this article we report the results of a usability assessment of a mobile voting application called *Vote App* (see Section 3 and [19,4]), which was developed with the specific aim of guaranteeing coercion-resistance. Over 5 million Italians abroad currently have the right to a postal vote, with well-known logistic and security issues: coercion due to organized crime is a historically well-documented threat [8]. Informally, a voting protocol is coercion-resistant if voters cannot prove whether or how they voted, even if they can interact with the adversary while voting. Vote App achieves this property using the mechanism of *fake credentials* [16,6]. These credentials allow voters under the influence of a coercer to express their true votes while pretending to comply with the coercer's demands. When the voter is, or fears to be, subject to a coercion attack, they can autonomously create a ruse credential indistinguishable from a real one. This credential will not validate the corresponding ballot when votes are tallied.

To evaluate Vote App, we first conducted a pilot test before extending it to a larger event (see Section 4), allowing us to reach a broader sample. This approach offers several advantages, including the opportunity to refine the application based on the initial feedback. In this paper, we focus mainly on evaluating the usability of Vote App, rather than on user experience (UX), which was partially analyzed in [7]. As this work is ongoing, we plan to conduct iterative testing of both UX and usability in the future.

Ethics and Data Protection. The tests were designed to maintain anonymity, focusing exclusively on gathering feedback regarding Vote App, without collecting personal data. The pilot was conducted within our Research Center, and the Data Protection Officer advised that a privacy consent form was unnecessary. The main test took place at a public event, and also in this context a privacy consent form was deemed unnecessary.

1.1 Research Objectives

This paper addresses the following research question (RQ): *Is an anti-coercion electronic voting system usable by the general public in terms of effectiveness, efficiency, and user satisfaction?*

To address this, following the ISO 9241-11 [15] (see Section 1.2) definition of usability and its three metrics, we derive these three hypotheses:

- *Effectiveness Hypothesis (H1)*: Vote App will demonstrate a high level of effectiveness, calculated by the completeness and accuracy of the voting process without user errors.
- *Efficiency Hypothesis (H2)*: Vote App system will demonstrate high efficiency, reflected by minimal effort and time required by users to complete the voting process.
- *Satisfaction Hypothesis (H3)*: Users will report high satisfaction levels in using Vote App, as assessed by their subjective feedback on the overall voting experience.

1.2 Usability metrics

According to ISO 9241-11 [15], usability is defined as the *effectiveness*, *efficiency* and *satisfaction* with which users accomplish specific objectives within particular environments. The National Institute of Standards and Technology (NIST) recommends these three metrics as suitable for assessing e-voting systems [18].

Effectiveness is defined as the completeness and accuracy with which a user can achieve a specific task [21]. In electronic voting, effectiveness is the measure of voters' ability to successfully cast their vote for their intended candidate without encountering errors [5,12]. To determine effectiveness, the examiner can observe participants while they perform the task, use visual recording, or ask participants to self-report their progress. Additionally, Thinking Aloud [23] can be employed. This method allows participants to vocalize their thoughts as they engage with the system [23]. Error rates are also an effective way for verifying effectiveness, especially in the case of e-voting, as they are linked to the voter's intention and the actual outcome [21].

Efficiency assesses whether participants achieve their goals without using excessive resources, striking a balance between effort and time [21,12]. For e-voting, it is represented by the time voters spend casting their vote and the time required by the user to complete verification [21,5].

Satisfaction refers to how content and at ease users feel during their interaction with a system or process [5,12]. Satisfaction is the only subjective usability metric recommended by NIST. In e-voting, satisfaction reflects how pleased or fulfilled voters feel while participating in the election [5]. Satisfaction can be measured with standard methods like the System Usability Scale (SUS) questionnaire [12] or it can also be measured using a non-standardized questionnaire developed by the examiner, tailored to the specific purpose of the study [21].

The SUS [12] is a 10-item survey designed to collect subjective usability assessments from participants in a usability test, evaluating aspects like efficiency, intuitiveness, ease of use, and satisfaction. Participants express their agreement or disagreement with each statement using a five-point Likert scale: *strongly disagree* (1), *disagree* (2), *neutral* (3), *agree* (4) and *strongly agree* (5) [13].

Sections Overview In Section 2, we discuss related works concerning usability evaluation, providing an overview of existing studies. Section 3 presents briefly Vote App. Section 4 outlines the methodology we employed to conduct our usability tests. In Section 5 we analyze the results of the tests. Finally, in Section 6, we conclude by summarizing the lessons learned from our study and discussing potential directions for future research and development.

2 Related Work

Given the sensitivity of e-voting systems, it is particularly crucial to emphasize the importance of conducting usability testing with potential users [25,14] and to assess the usability of the system [2,10,11,17,22]. As highlighted by [24], voters' perception of voting technology significantly influences their intention to use it. Additionally, usability issues, as noted by [20], have the potential to lead to incorrect election outcomes. Acemyan et al. [1] compare the usability of three voting systems: Helios, Prêt à Voter, and Scantegrity II. Their evaluation reveals that only 58% of the participants were able to complete the voting process, with even lower completion rates observed during the verification phase. These usability findings underline the importance of not only guaranteeing the security of a system but also prioritizing usability. In [3] the authors assert that ensuring system usability is further complicated by the infrequency of elections. Voters are expected to cast their votes with nearly 100% success rates, despite often having no prior experience or training with the voting system. In the study conducted by [21], it is emphasized that human factors should be taken into account early in the design process of e-voting systems. In the article, the authors provide guidelines for assessing usability metrics of e-voting solutions.

3 The Tested E-Voting System

3.1 Description of Vote App

In this Section, we give a brief overview of the functionalities of Vote App (see Figure 1). For more information about the protocol, the cryptographic primitives used and the implementation details, we refer the reader to [4,19].

Voters[3] authenticate themselves via their national digital identity to obtain a voting credential [16] from the responsible authorities in order to cast ballots. These credentials allow voters under the influence of a coercer to express their true votes while pretending to comply with the coercer's demands. When the voter is, or fears to be, subject to a coercion attack, they can autonomously create a ruse credential indistinguishable from a real one. This credential will not validate the corresponding ballot when votes are tallied. To enhance usability of the scheme, in [4,19] the credential is given to the voter in the form of a five-digit PIN. When inserted during the voting phase, this PIN unlocks the valid credential needed to cast a valid vote. To create a ruse credential, it is sufficient to set up a ruse PIN. The correctness of the PIN can be verified via a Designated Verifier Non-Interactive Zero-Knowledge Proof (DVNIZKP), which proves the correctness of the associated credential. To enhance the security of the system, $\mathcal{V}$ can also use an external verification service, which ideally should be a pre-installed application with no internet connection, to check the correctness of the data managed by Vote App. For coercion resistance, we assume that the voter has control over their voting device. This assumption is valid since Vote App can be used from any device. Furthermore, everyone can access a Web Bulletin Board (WBB), from any device, in which all the public data regarding the election is published.

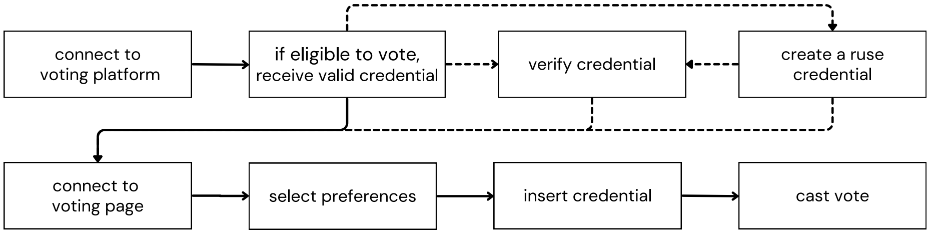

Fig. 1. User Flow of Vote App with and without coercion.

The voting protocol can be divided into three steps.

[3] The target population of Vote App is composed by Italian citizens eligible to vote in Italian elections.

Setup and Registration The voter $\mathcal{V}$ downloads Vote App from an official application store and installs it on their device. Once Vote App is installed, $\mathcal{V}$ can start the registration process:

- $\mathcal{V}$ opens Vote App and reads the high-level information displayed.
- $\mathcal{V}$ logs in with their digital identity.
- If the login is successful, Vote App displays to $\mathcal{V}$ the procedure to retrieve the PIN and to activate Vote App from another device.
- After a random waiting period, $\mathcal{V}$ receives a push notification alerting them that the PIN is ready to be retrieved.
- $\mathcal{V}$ logs in to Vote App and receives their PIN along with its visual representation, encoded as a string of emojis called "Private PIN Emojis". $\mathcal{V}$ is warned to save both the PIN and its visual representation.
- After a random waiting period, $\mathcal{V}$ receives a push notification alerting them that from now on they can verify, via the DVNIZKP, as many times as needed, the correctness of the PIN.
- At any time, $\mathcal{V}$ can use the external verification service to check the correctness of the emoji string composing "Private PIN Emojis".

PIN Management In order to resist coercion, $\mathcal{V}$ can set up one (or more) ruse PIN with which they can simulate a vote[4]. Moreover, a forged verification proof is made available to Vote App, so that the voter can deceive the coercer by pretending that this ruse PIN is a valid one. The procedure is the following:

- $\mathcal{V}$ opens Vote App and goes to the "app management" screen, in the tab "ruse PIN".
- $\mathcal{V}$ types in the ruse PIN they want to set up.
- After a random waiting period, $\mathcal{V}$ receives a push notification alerting them that the ruse PIN is ready to be retrieved. Note that this notification is exactly the same as the one sent during Registration, but in this case the ruse PIN is retrieved instead of the real one.
- $\mathcal{V}$ logs in to Vote App and receives their (ruse) PIN along with its visual representation, encoded as a string of emojis called "Private PIN Emojis". $\mathcal{V}$ is warned to save both the PIN and its visual representation. This screen is identical to the one $\mathcal{V}$ has seen during the registration step.
- After a random waiting period, $\mathcal{V}$ receives a push notification alerting them that from now on they can verify the PIN.

Differently from the previous step, in which the PIN that verifies the proof is the valid one, after the ruse PIN request, this proof is verified only by the ruse PIN. If $\mathcal{V}$ inserts the valid PIN, the verification will not be not successful. Nevertheless, the valid PIN remains the only one that allows to cast a valid vote.

In this step $\mathcal{V}$ can request a reminder of their valid PIN. In this case the procedure is as follows:

[4] During the voting phase, $\mathcal{V}$ can simulate a vote that will not be counted just by casting a vote with a random PIN. However this does not guarantee coercion-resistance, since, by doing this, they cannot produce any proof of PIN reception or verification.

- $\mathcal{V}$ opens the App and goes to the "app management" screen, in the tab "PIN reminder".
- $\mathcal{V}$ confirms their choice.
- The issuance of the PIN follows the same procedure as in the previous cases. From now on, until a new ruse PIN request is made, only the valid PIN will verify the DVNIZKP.

Ballot Casting and Individual Verification When the voting period starts, $\mathcal{V}$ can cast a vote with the following procedure:

- $\mathcal{V}$ opens Vote App and from its Homepage clicks on the "Vote" button.
- $\mathcal{V}$ expresses and confirms their preference.
- Vote App advises $\mathcal{V}$ that their preference is being encrypted.
- Vote App shows to $\mathcal{V}$ a string of emojis called "Ballot Emojis" that will be needed in the verification phase and asks $\mathcal{V}$ to enter a PIN.
- After PIN insertion, Vote App advises $\mathcal{V}$ that their vote is being sent, while also showing the "Private PIN Emojis". If the sequence differs from the expected one, $\mathcal{V}$ can report a problem via the "Report a Problem" button.
- $\mathcal{V}$ chooses one out of two values, called *control numbers*, shown by Vote App in a confirmation page to confirm their vote.
- Vote App displays a confirmation message to $\mathcal{V}$ that the voting has been cast and received.
- Vote App displays the hash of the vote and all the information needed by $\mathcal{V}$ to check that their vote has been correctly registered and that it has not been tampered with.
- Using these information, $\mathcal{V}$ connects to the WBB and checks that:
  - their vote has been correctly registered and confirmed, by checking that the hash given by Vote App is included in the list of registered hashes, along with a tick acknowledging correct confirmation.
  - their vote has not been tampered with by checking that the control number displayed is the same as the one previously chosen.

Vote App allows for re-voting: only the last vote cast, per PIN, will be kept and, if cast with the valid PIN, counted.

When the voting period ends, the WBB shows to every user the final tally. and everyone can download the proofs to check its correctness.

Almost every interface of Vote App includes a magnifying lens icon with more detailed information about the current step. To enhance usability, $\mathcal{V}$ can always access the User Manual of Vote App and, whenever $\mathcal{V}$ is advised to store some data, a "Share button" is included near the data to be saved.

4 Study design and procedure

Relying on the usability metrics outlined in Section 1.2, we conducted a pilot test and a main test, where the participants were asked to cast a vote in a fictional election using Vote App.

To evaluate the *effectiveness*, we assessed the task performance by counting how many participants successfully completed the task and the number of attempts required. Additionally, we encouraged participants to share their thoughts verbally during the test following the *Thinking Aloud* method [21].

For measuring *efficiency*, we recorded the time participants spent on each task, with a special focus on ballot completion.

To evaluate *satisfaction*, we administered the System Usability Scale (SUS) questionnaire at the end of each test session, tailored to our e-voting context (e.g., *"I think that I would like to use this system frequently"* became *"I think that I would use Vote App to vote in a real election"*). The final SUS score ranges from 0 to 100, with a score above 68 considered good [9]. We decided to use the SUS questionnaire because it is quick and easy to administer (taking only a few minutes to complete), and it has been widely used in e-voting user studies [21].

As suggested by [21], during the pilot test, we also asked participants four additional questions to gain deeper insights into their experience with Vote App.

The questions were the following: (i) *In your opinion, how was the voting process?*; (ii) *What do you think about the PIN feature?*; (iii) *Did you notice the presence of the emojis during the voting process? What do you think about them?*; (iv) *What is your final opinion about Vote App?*.

4.1 Context

We conducted the pilot test inside a Research Center for Cybersecurity. The Center's objective is to enhance cyber risk management, with a particular focus on digital identity and the quality of online services. The test was conducted by two cryptographers with mathematical background and a UX/UI expert with sociological background.

The main test was conducted during 2023 European Researchers' Night, an EU initiative launched in 2005, that aims to bridge the gap between science and the general public. The event spanned from 5 pm to midnight, providing enough time to conduct multiple test sessions. Two information security and mobile security experts, and a jurist joined the group conducting the test. The different backgrounds allowed us to discuss the complexities of e-voting from various angles and dispel any possible doubts.

4.2 Test organization

During each test session, participants had to complete a task, specifically to cast a vote in a fictional election. After completing the voting process, participants were invited to check that their vote was recorded and to verify the accuracy of the information received through Vote App. These features were available in the main test through the Verification Service and the WBB (see Section 3), but omitted during the pilot due to ongoing development work. Also, in order to mitigate the lack of information experienced during the pilot test, for the main test we prepared two explanatory posters. The first one provided a simplified explanation of the cryptography behind Vote App, specifically focusing on the creation and verification of voting credentials, the second one contained a step-by-step guide of Vote App.

Fictional Election In the pilot test, users were asked to vote for their preferred city in Italy to host EXPO 2030. Initially, participants were prompted to select a list and subsequently choose a candidate from their chosen list. The lists were: *Northern Italy*, *Central Italy* and *Southern Italy and Islands*. The candidate cities were: (i) for Northern Italy: *Trento, Milan, Turin, Venice*; (ii) for Central Italy: *Rome, Florence, Ancona, Perugia*; and (iii) for Southern Italy and Islands: *Naples, Bari, Tropea, Palermo*.

In the main test, to speed up the voting process, the election was a referendum. Six different questions were created for participants to choose from (e.g., Q: "Do you prefer cream-flavored or fruit-flavored ice cream?").

Scenario For the pilot test, we designed two distinct scenarios: one coercion-free and one with a coercer[5]. Six participants were pre-assigned to the coercion-free scenario, while the remaining four were assigned to the scenario with the coercer. In the scenario with the coercer, participants were approached by a fictional criminal who demanded that they vote for a specific city. The criminal's motivation was financial gain, as he and his gang stood to profit substantially if the chosen city emerged victorious in the election. Participants were required to initially cast a vote for the chosen city using the anti-coercion functionalities of Vote App, thereby sending an invalid vote. Subsequently, they could send a valid vote for their preferred city. This scenario was intentionally crafted to evaluate the effectiveness of the anti-coercion features integrated into Vote App.

In the main test, scenarios were not included to allow participation by multiple individuals simultaneously in a noisy environment. Each voter was given complete freedom to build their own scenario, choosing independently how many times to vote, whether to verify the PIN, and whether to set and vote with a ruse PIN. Almost 75% of the participants tested the anti-coercion functionalities.

User Authentication The focus of the test was not on the authentication, so this step was simulated in the tests. In the pilot, participants had to press on the button *Login with digital identity* but, instead of inserting their credentials, they were automatically logged in. In the main test, authentication was necessary in order to correctly assign voting credentials, thus voters were required to choose a pseudonymous identity from a pre-generated set of options. Each identity was represented by a drawn portrait of an animal, blatantly fictional first and last names, and a code for logging into Vote App.

[5] The coercer is an attacker who tries to influence or force the voter to follow their demands. Possible demands range from voting for a specific candidate or not voting at all.

4.3 Test session

Participants and Recruitment For the pilot, participants were recruited via email. We sent invitations to all Italian individuals working within the Center eligible for voting in Italian elections. We decided to exclusively involve Italian citizens. The reason behind this decision was twofold: firstly, Vote App was developed in Italian, and secondly, they possess a familiarity with the electoral process in Italy, which was what we wanted to test. The group taking part in the test was composed of 10 people with expertise that ranges from information technology and cybersecurity to cryptography. Approximately 30% possessed a Bachelor's or Master's degree, while 70% also held a PhD. In terms of age distribution, roughly 30% of the sample fell between the birth years of 1975 and 1990, and the remaining 70% were born between 1991 and 2001. It is important to acknowledge the potential for biases as their perspectives and experiences might not fully represent the broader spectrum of voters' capabilities.

For the main test, due to the nature of the event, participants were not formally recruited. Instead, we approached individuals who came to our stand and invited them to try the application and provide feedback. We were able to gain insights into the participant demographics through the official data obtained from the event organizers[6]. Approximately 25% did not hold a university degree, while 50% possessed a Bachelor's or Master's degree. Approximately 7% of participants held a PhD, with the remaining 18% having an educational qualification below the level of an middle-school diploma. In terms of age distribution, roughly 50% of the sample fell between the birth years of 1997 and 2003, 25% were born between 1991 and 1997, 13% were born between 2003 and 2009, and the remaining 12% were born between 1961 and 1991. Regarding employment status, the 67% of participants were students, while the remaining 33% were engaged in permanent employment (without further classification of occupation types). Approximately, 80 people participated in the main test.

Equipment. The tests have been conducted using Android smartphones provided by the Center, with Vote App already installed and opened on the login page. To facilitate note-taking, we gave each participant a writing support.

For the main test, two desktops were set up to ensure visibility of the WBB. Additionally, we hung the previously prepared posters near our stand, so that they were easily accessible for reference in case of any doubts. The list of potential referendum questions was printed on a card accessible by all. For the pseudonymous identities, we adopted a similar approach by printing them on paperboard and folding them in half, creating a resemblance to real ID cards. We even included a QR code on the back of the identity cards, which directed participants to a web page where the posters were uploaded.

[6] For confidentiality reasons, this data was not publicly available but was provided to us by the organizers exclusively for the purpose of this article.

Location. The pilot test was conducted in one of the meeting rooms of the Center. The environment was quiet and the participants had no external distractions. For the main test, our stand was clearly marked with signage. The environment was busy and noisy, with people exploring the exhibition stands and wondering around.

Test. We conducted multiple test sessions. In the pilot, each session involved only one participant, with approximate length between 20 to 30 minutes. In the main test, each session involved from 3 to 5 participants, with approximate length between 10 to 15 minutes.

1. *Introduction.* We offered the participant a brief overview of the project. Subsequently, we presented the test instructions and we recommended to maintain a good flow of comments while interacting with the system, adhering to the Thinking Aloud (see Section 1.2).
2. *Election Context.* In the pilot we read and explained the assigned scenario to the participant, giving them a printed copy for reference. In the main test there was no scenario, instead we gave to the participating group a list of referendum questions, from which they could choose which one to vote on.
3. *Authentication.* Each participant simulated the authentication step as explained before.
4. *PIN reception.* Vote App provided to each participant a PIN which they could write down.
5. *Voting.* Each participant accessed the ballot and cast their vote. In the main test, they were also able to view the registration of the vote on the WBB.
6. *Re-Voting, Ruse PIN and Verification.* Participants could set up a ruse PIN, re-vote with any PIN, verify a PIN. Only in the coercion scenario during the pilot re-voting with a ruse PIN was mandatory.
7. *Tabulation.* In the main test, upon completion of the voting phase by all the participants, we started the tabulation and displayed the results on the WBB.
8. *Post-Test Questionnaire.* At the end of each test session, we asked the participant to fill out the System Usability Scale (SUS) questionnaire. In the pilot this was mandatory, and we also asked participants to answer four additional questions (see Section 4).

5 Results evaluation

After both tests, we performed an in-depth analysis of the data gathered. This analysis included the evaluation of the three usability metrics established by ISO 9241-11 [15]. Additionally, we reviewed the feedback provided by participants during their interactions with Vote App, aiming to gain valuable insights into their user experience and identify areas for improvement.

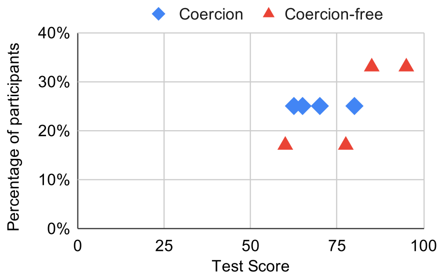

Fig. 2. SUS Results of the Pilot Test.

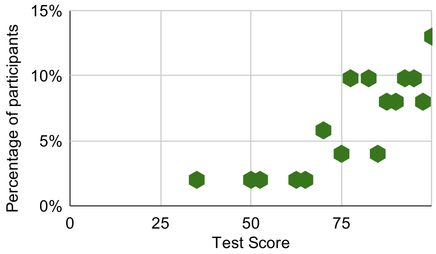

Fig. 3. SUS Results for the Main Test.

5.1 Pilot test

Quantitative results *Effectiveness (H1).* In both scenarios, every test participant successfully completed the required task. In the coercion-free scenario, each test participant accomplished the required task on the first attempt and without any errors. Conversely, in the scenario with coercion, only two participants managed to complete the task on their first attempt; the others required at least two more attempts to succeed. As suggested in [21], the Thinking Aloud method is not really reliable for determining effectiveness, but it can provide valuable insights into participants' feelings. Therefore, we collected users' comments while interacting with the prototype in the qualitative results. In terms of error rates, participants in the coercion-free scenario completed the task without any errors. Conversely, only half of the participants in the scenario with coercion managed to complete the task. The first error arose from a misunderstanding of the coercer's request (i.e., the participant did not understand that they had to vote with the ruse PIN for the forced choice to evade the coercer's request, see Section 4.2), while in the other case, the participant entered a random PIN instead of the ruse one, demonstrating a lack of understanding of the distinction.

Efficiency (H2). Due to some issues encountered with the (then still under development) prototype of Vote App, it was impossible to collect data on the time users spent on the task. We noted, however, that users in the scenario with coercion required more time to both comprehend and complete the task.

Satisfaction (H3). The total (average) score of the SUS questionnaire (see Section 1.2), in the coercion-free scenario, was 82.9/100, ranging from 60/100 as the lowest to 95/100 as the highest (see Figure 2). Conversely, in the scenario with coercion, the total score was of 69.3/100, with scores ranging from 62.5 to 80 out of 100[7]. From these results, it is clear that participants in the scenario with coercion provided lower scores (see Figure 2), which was also highlighted by the feedback gathered during the test. This could be attributed to either an ineffective presentation of the task, or to an ineffective app design concerning the anti-coercion features.

[7] Note that the lowest score is considered good, see [9].

Qualitative Results The qualitative data presented in this subsection has been acquired during each test session by employing the Thinking Aloud method, as well as through participant's answers to the questions we posed at the conclusion of the test (see Section 4). Additionally, we carefully observed how participants behaved while interacting with Vote App, in order to obtain more insights on the app. Here we summarize our analysis on the collected comments, observations and behaviours.

Positive feedback. The voting process was perceived as straightforward and efficient by all participants in both scenarios, who expressed satisfaction with the initial page providing a concise overview of the app. They also appreciated the clarity of the voting instructions and the option to consult electoral programs. Regarding the PIN, participants liked its brevity (only 5 digits) and found the PIN recovery and verification features to be highly beneficial and practical. Moreover, the setup mode for the ruse PIN proved to be intuitive. Participants highly valued the possibility to capture images and screenshots of Vote App. Although there were some negative comments regarding the usage of emojis, many participants verified that the "Private PIN Emojis" displayed alongside the PIN at the beginning of the protocol matched those shown during the voting process.

Negative feedback. After an in-depth analysis of the qualitative data collected, it became evident that the majority of negative feedback came from a lack of comprehensive explanations regarding the key features of Vote App. Initially, during the design phase of the test, we chose not to include extensive descriptions of the app in order to prevent the test from becoming too difficult. Given that this application requires complex functionalities that diverge significantly from standard apps, this decision ultimately proved ineffective. For example, it was not evident to participants that Vote App assigns a unique, valid PIN for each user, which remains unchanged. Additionally, they were unaware that they could vote multiple times during the voting process. The purpose of the ruse PIN functionality was also not fully understood, as some questioned the necessity of setting a ruse PIN to submit an invalid vote when they could simply enter a random one. Furthermore, when participants did use a ruse PIN, they often chose easily guessable combinations, such as "12345".

The emojis created confusion among some participants. Users that were familiar with emojis being used as a security measure, such as on platforms like Telegram, the messaging app, found this feature helpful and did not experience interface confusion. On the other hand, participants accustomed to encountering emojis primarily in messaging contexts found it challenging to understand their purpose within Vote App. Some perceived them as a distraction or mistook them for application errors, and a few remarked that they were difficult to remember. Recognizing this issue, a participant suggested to add the option of entering the PIN twice. This proposed solution aims to provide an additional layer of assurance for users who may not pay close attention to emojis or struggle to comprehend their function. By allowing users to input their PIN twice, they would have greater control and confidence in ensuring the accuracy of their

PIN entry. Another feature that caused confusion were the control numbers. In fact, upon encountering them, many participants wondered, *"What am I supposed to do now?"* or *"Why am I being asked to select these numbers?"*, which strengthens our initial thesis about a lack of app explanation.

Other suggestions. We also collected some suggestions regarding how to improve the user interface design of Vote App. During the voting process, it was not clear when and where the emojis will appear and a participant suggested to introduce an element signaling their arrival. Another feedback we received pertains to text descriptions: it would be best to keep them as neutral as possible, avoiding exclamation marks and words like "attention" or "remember".

Post Test Improvements. Based on our analysis on the pilot conducted, we have made several updates to improve user experience and security. These include modifying interfaces to clarify PIN and emoji information and providing more explicit instructions for setting a non-trivial ruse PIN.

5.2 Main test

Quantitative Results *Effectiveness (H1).* All the participants managed to successfully complete the voting process and send the vote to the WBB. Regarding the other features of the app (e.g., set up a ruse PIN, verify the PIN) we were not able to collect all the data as we did in the pilot due to the test context. Nevertheless, from the partial data gathered, we can state that participants were able to understand and use the verification service and to handle the ruse PIN request with lower error rates with respect to the pilot.

Efficiency (H2). All the participants managed to complete the test in 10 minutes, which was the time we expected for each test session to last.

Satisfaction (H3). In total, 50 out of the 80 participants in our demo compiled the SUS questionnaire. The total score was 84.9/100, with scores ranging from 35 to 100 out of 100 (see Figure 3).

Qualitative Results As for the pilot test, we present a brief summary of the qualitative data acquired via the Thinking Aloud method and our observations on the participants behaviours.

Positive Feedback. Based on participants' behavior, we observed that everyone wrote down their PIN, indicating the importance they attached to it, and the difficulty of remembering it. When we asked the participants who tested the anti-coercion functionalities to provide feedback on the ruse PIN, they did not have any negative or positive feedback. This suggests that we made a step towards improving explanations on Vote App, by communicating more efficiently the anti-coercion functionalities. Moreover multiple participants expressed positive feedback on emojis, stating that they were an interesting introduction and did not cause confusion. Moreover, their use in the voting process was well understood. The WBB introduction was well-received and participants found it to be useful and easy to use.

Negative Feedback. Some participants did not understand that the PIN could be retrieved if forgotten. Another participant mentioned that it was not clear that the PIN had to be entered each time voting. In terms of the use of emojis in general, one participant found them confusing since they were designed differently between the two different devices (i.e., the smartphone and the screen where the WBB was displayed). Regarding control numbers, participants thought that they were not well-explained and that it was unclear what they were for. Participants also questioned why they had to select one value, instead of letting the app doing it automatically.

Other suggestions. Some participants were confused by the usage of emojis to visualize the PIN and believed that the PIN was represented by them. The reason behind that is because the emojis of "Private PIN Emojis" were centrally placed on the PIN reception interface.

6 Conclusions

The goal of this study was to evaluate the usability of our proposed design for a coercion-resistant electronic voting system (RQ). The results obtained in this study highlight the importance of having an intuitive and user-centric design in an e-voting application. By providing clear instructions and ensuring accessibility, voters were able to make informed decisions and actively participate in the voting process. Moreover, the ability to view the registration of their vote on the WBB is an important step towards having a transparent voting procedure. The optional questions we posed to the willing voters, allowed us to gather valuable insights, further improving the overall usability.

The most evident result from our evaluation is the need to balance the amount of information being provided to the users in each interface. While some participants felt the explanations of the functionalities to be overwhelming, others found them useful and essential for a voting application, even if long. This shows that we still have a lot of work to do to find right balance, ensuring users can comprehend all aspects of the app without giving them excessive information. Regarding our initial research hypotheses, the results showed that participants successfully completed the voting task on their first attempt, although difficulties arose in the scenario with coercion during the pilot and insufficient feedback was collected in the main test regarding the anti-coercion features (H1). Both the pilot and main test demonstrated high efficiency, with participants completing the voting process within a few minutes (H2), and high satisfaction levels, with participants appreciating the system's ease of use and security features (H3).

This approach will help alleviate confusion and improve user experience, ultimately enhancing the effectiveness and trustworthiness of the voting system. As future works, we intend to address critical aspects of both Vote App and this Usability tests. Our study primarily involved students, which may limit the representativeness of our findings. Future research should expand the sample to include a more diverse population, such as different age groups, professions, and backgrounds. While our exploratory study lays the groundwork, it is difficult to properly assess a user reaction in a situation which involves fear. We plan to develop further tests to assess the anti-coercion functionality and to verify its effectiveness and reliability. Moreover we plan to provide more concise and user-friendly instructions to voters.

Acknowledgment This work has been partially supported by a joint laboratory between Fondazione Bruno Kessler and the Italian Government Printing Office and Mint and by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

References

1. Acemyan, C.Z., Kortum, P., Byrne, M.D., Wallach, D.S.: Usability of voter verifiable, end-to-end voting systems: Baseline data for helios, Prêt à voter, and scantegrity II. USENIX Journal of Election Technology and Systems (JETS) pp. 26–56 (Aug 2014), https://www.usenix.org/jets/issues/0203/acemyan
2. Acemyan, C.Z., Kortum, P., Byrne, M.D., Wallach, D.S.: 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 (2022)
3. Ali, S.T., Murray, J.: An overview of end-to-end verifiable voting systems. Real-World Electronic Voting: Design, Analysis and Deployment pp. 171–218 (2016)
4. Bitussi, M., Longo, R., Marino, F.A., Morelli, U., Sharif, A., Spadafora, C., Tomasi, A.: Coercion-resistant i-voting with short PIN and OAuth 2.0. In: E-Vote-ID 2023. Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn (2023), to appear.
5. Budurushi, J., Renaud, K., Volkamer, M., Woide, M.: An investigation into the usability of electronic voting systems for complex elections. Annals of Telecommunications **71** (04 2016). https://doi.org/10.1007/s12243-016-0510-2
6. Cortier, V., Gaudry, P., Yang, Q.: Is the JCJ voting system really coercion-resistant? Cryptology ePrint Archive, Paper 2022/430 (2022)
7. Cristiano, L., Spadafora, C.: Enhancing usability in e-voting systems: Balancing security and human factors with the hc3 framework. In: International Conference on Human-Computer Interaction. pp. 33–42. Springer (2024)
8. Desantis, V.: Il voto degli italiani all'estero: nuove criticità e vecchi problemi nella prospettiva del superamento del voto per corrispondenza. Federalismi.it **22**, 31–51 (2022), https://www.federalismi.it/nv14/articolo-documento.cfm?Artid=47672
9. Designer Italia 2023. Manuale Operativo di Design: test di usabilità, https://docs.italia.it/italia/designers-italia/manuale-operativo-design-docs/it/versione-corrente/doc/design-research/test-usabilita.htm
10. Distler, V., Zollinger, M.L., Lallemand, C., Roenne, P.B., Ryan, P.Y.A., Koenig, V.: Security - visible, yet unseen? In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. p. 1–13. CHI '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300835, https://doi.org/10.1145/3290605.3300835

11. Fuglerud, K.S., RØssvoll, T.H.: An evaluation of web-based voting usability and accessibility. Univers. Access Inf. Soc. **11**(4), 359–373 (nov 2012). https://doi.org/10.1007/s10209-011-0253-9, https://doi.org/10.1007/s10209-011-0253-9
12. Greene, K.K., Byrne, M.D., Everett, S.P.: A comparison of usability between voting methods. In: USENIX Workshop on Accurate Electronic Voting Technology (2006), https://api.semanticscholar.org/CorpusID:13641141
13. Hao, F., Wang, S., Bag, S., Procter, R., Shahandashti, S.F., Mehrnezhad, M., Toreini, E., Metere, R., Liu, L.Y.: End-to-end verifiable e-voting trial for polling station voting. IEEE Security & Privacy **18**(6), 6–13 (2020). https://doi.org/10.1109/MSEC.2020.3002728
14. Herrnson, P.S., Niemi, R.G., Hanmer, M.J., Bederson, B.B., Conrad, F.G., Traugott, M.: The importance of usability testing of voting systems. In: 2006 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT 06). USENIX Association, Vancouver, B.C. (Aug 2006), https://www.usenix.org/conference/evt-06/importance-usability-testing-voting-systems
15. International Organization for Standardization. ISO 9241-11:2018, Ergonomics of human-system interaction. Part 11: Usability: Definitions and Concepts (2018), https://www.iso.org/standard/63500.html
16. Juels, A., Catalano, D., Jakobsson, M.: Coercion-resistant electronic elections. In: Towards Trustworthy Elections. Springer (2010). https://doi.org/10.1007/978-3-642-12980-3_2
17. Kulyk, O., Neumann, S., Budurushi, J., Volkamer, M.: Nothing comes for free: How much usability can you sacrifice for security? IEEE Security & Privacy **PP**, 1–1 (06 2017). https://doi.org/10.1109/MSP.2017.265093646
18. Laskowski, S., Yen, J., Autry, M., Cugini, J., Killam, W.: Improving the usability and accessibility of voting systems and products (2004-04-01 2004), https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=150478
19. Longo, R., Morelli, U., Spadafora, C., Tomasi, A.: Adaptation of an i-voting scheme to Italian elections for citizens abroad. In: E-Vote-ID 2022 (10 2022). https://doi.org/https://doi.org/10.15157/diss/027
20. Marky, K., Zimmermann, V., Funk, M., Daubert, J., Bleck, K., Mühlhäuser, M.: Improving the usability and ux of the swiss internet voting interface. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2020)
21. Marky, K., Zollinger, M.L., Funk, M., Ryan, P.Y.A., Mühlhäuser, M.: How to assess the usability metrics of e-voting schemes. In: Financial Cryptography Workshops (2019), https://api.semanticscholar.org/CorpusID:86497210
22. Marky, K., Zollinger, M.L., Roenne, P., Ryan, P.Y.A., Grube, T., Kunze, K.: Investigating usability and user experience of individually verifiable internet voting schemes. ACM Trans. Comput.-Hum. Interact. **28**(5) (sep 2021). https://doi.org/10.1145/3459604
23. Nielsen, J.: Thinking aloud: The #1 usability tool (2012), https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/
24. Yao, Y., Murphy, L.: Remote electronic voting systems: An exploration of voters' perceptions and intention to use. EJIS **16**, 106–120 (04 2007). https://doi.org/10.1057/palgrave.ejis.3000672
25. Zollinger, M.L., Distler, V., Rønne, P., Ryan, P., Lallemand, C., Koenig, V.: User experience design for e-voting: How mental models align with security mechanisms. E-Vote-ID 2019 TalTech Proceedings (2019). https://doi.org/10.13140/RG.2.2.27007.15527