Cultivating a Tech-Safety Mindset using Game-Based Learning for Defending against Technology-Facilitated Abuse

Majed Almansoori
University of Wisconsin–Madison and United Arab Emirates University (UAEU)
malmansoori2@wisc.edu

Chirag Ghosh
Indian Institute of Technology, Kharagpur
cghosh828049@gmail.com

Sarita Singh
Indian Institute of Technology, Kharagpur
sarita01.ar@gmail.com

Rahul Chatterjee
University of Wisconsin–Madison
chatterjee@cs.wisc.edu

Mainack Mondal
Indian Institute of Technology, Kharagpur
mainack@cse.iitkgp.ac.in

Abstract

Technology-facilitated abuse (TFA) has become increasingly common as abusers exploit everyday technologies to monitor and harass others, mainly their intimate partners. Preventing TFA requires not only reactive technical support but proactively cultivating protective mindsets — awareness of personal vulnerability, recognition of threat severity, and confidence to implement defensive strategies — before abuse escalates. Yet no research has developed educational tools grounded in behavior change theory specifically for technology-facilitated abuse prevention. We address this gap with BeSafe, a narrative-driven visual novel game grounded in Protection Motivation Theory (PMT) and designed to shift how users perceive and respond to TFA threats. Through a study with 198 participants across six platform contexts, we assessed both knowledge acquisition and changes in PMT constructs: perceived vulnerability, perceived severity, self-efficacy, and fear arousal. Our results show that BeSafe produced significant knowledge gains alongside meaningful shifts in protection motivation. Participants with prior exposure to online abuse showed substantially greater gains across both knowledge and motivation measures. Many participants reported intentions to review privacy settings and share protective strategies with others, indicating motivation to act on what they learned. Our findings demonstrate that game-based interventions can cultivate digital safety mindsets, offering a scalable, proactive complement to existing reactive support services.

Keywords

Technology-facilitated abuse; Game-based learning; Digital safety; Protection motivation theory; Behavior change; Privacy education

1 Introduction

Abusive intimate partners and family members frequently use technological means to spy on and harass their victims [3, 19, 32–34]. Studies show that 38%–63% of women worldwide experience technology-facilitated abuse (TFA) [65]; 97% of advocacy organizations who support survivors of intimate partner violence (IPV) report their clients being abused using technology [57]. College students are particularly vulnerable; prior studies report that as much as 75% of college students experience TFA perpetrated by a partner or former partner [42, 55].

To support survivors of TFA, Havron et al. [44] proposed tech clinics [21], where survivors can get tailored help from trained technology consultants. However, tech clinics require specialized expertise, substantial resources, and geographic accessibility, which limits their accessibility for all survivors. Moreover, such approaches are predominantly reactive, focusing on supporting victims *after* abuse has occurred. Recently some advocacy organizations use different awareness campaigns to educate users and advocates about TFA. However, TFA requires not just teaching technical risks, but cultivating protective mindsets: awareness of personal vulnerability, recognition of threat severity, and confidence to implement defensive strategies before abuse escalates.

Educational games have shown promise for teaching complex topics and promoting behavior change across diverse domains [10, 15, 47, 59, 78, 81], including domestic violence prevention [13]. Yet despite the growing threat of TFA and the demonstrated potential of game-based learning, no research has developed games specifically designed to cultivate both knowledge and protective mindsets for technology-facilitated abuse prevention.

To address this gap, we developed BeSafe, a narrative-driven visual novel interactive game designed to cultivate digital privacy and safety mindsets for TFA prevention. Visual novel games use branching storylines and character-driven dialogue to immerse players in interactive scenarios. Research has demonstrated their effectiveness for teaching across diverse domains [14, 16, 58], including sensitive topics such as bullying [54]. BeSafe integrates technical education with awareness of abuse patterns while fostering the motivation and confidence that drive behavior change. We ground our approach in Protection Motivation Theory (PMT) [29, 56, 73], which identifies key constructs that predict whether individuals will adopt protective behaviors: perceived severity (recognizing that threats have serious consequences), perceived vulnerability (awareness of personal risk), self-efficacy (confidence in one's ability to respond), and fear arousal (emotional response to potential threats). The game's narrative structure is specifically designed to

Proceedings on Privacy Enhancing Technologies 2026(4), 507–525

https://doi.org/10.56553/popets-2026-0132

engage these constructs in a safe environment, fostering protective mindsets necessary for individuals to recognize warning signs and implement defensive strategies.

To evaluate the effectiveness of this approach, we investigate two research questions:

RQ1: To what extent does game-based learning cultivate protection motivation — including threat awareness, vulnerability recognition, and self-efficacy — for detecting and preventing technology-facilitated abuse (TFA)?

RQ2: What user characteristics and design elements influence the development of protective mindsets and game-based learning effectiveness?

We conducted a controlled user study with 198 participants across three conditions: a game with gamification features (badges, leaderboard, knowledge meter), the same game without these features, and privacy guides covering equivalent content. Privacy guides are reference documents maintained by advocacy organizations such as the Clinic to End Tech Abuse [21] that provide step-by-step instructions for securing accounts and devices; they are designed for quick information retrieval rather than structured learning. We include them as a baseline to validate that the game achieves knowledge transfer, while PMT constructs provide our primary evaluation of whether the game cultivates protective mindsets. Participants were randomly assigned to one of six platform contexts (WhatsApp, Instagram, Snapchat, Gmail, iCloud, or Android) to assess generalizability across technologies.

Our findings demonstrate that BeSafe achieves both successful knowledge transfer and significant protection motivation development. Participants showed improvements in threat awareness, vulnerability recognition, and self-efficacy, with 95.5% reporting engagement and 39.4% explicitly appreciating its learning value. Notably, the game's narrative approach—rather than gamification mechanics—appears to be the primary driver of motivation shifts, with the no-features game producing even higher odds of PMT improvement. Participants with prior exposure to online abuse showed significantly greater gains across both knowledge and motivation measures, suggesting that personal relevance enhances educational effectiveness. Participants expressed strong intentions to implement what they learned, with many describing immediate plans to review privacy settings, check linked devices, and share knowledge with others. These findings demonstrate that game-based interventions can shift TFA prevention from reactive support toward proactive empowerment.

Contributions of this study:

(1) We developed BeSafe, a narrative-driven visual novel game grounded in Protection Motivation Theory, designed to shift users' digital privacy and safety mindsets by teaching detection and mitigation of technology-facilitated surveillance.

(2) Through a user study with 198 participants, we demonstrated that BeSafe fosters protection motivation—including threat awareness, vulnerability recognition, and self-efficacy—while successfully transferring technical knowledge about TFA, making prevention education accessible to both technical and non-technical users.

(3) We identified factors, including personal relevance and lived experience with online abuse, that substantially enhance both knowledge acquisition and motivation development in privacy and safety contexts.

(4) We showed that narrative engagement drives motivation shifts in game-based privacy education—offering design implications for future educational interventions addressing technology-facilitated abuse.

2 Related Work

Technology-facilitated abuse (TFA) is a growing problem where abusers use everyday technologies to harm their partners from anywhere, at any time [9, 65]. Most abusers are UI-bound, relying on user interface features rather than sophisticated technical exploits [33], and research has identified spyware apps and dual-use apps that require minimal technical knowledge [3, 19]. Recent work developed a comprehensive framework for IoT-enabled abuse, identifying four abuse vectors and revealing the severe psychological impact of such surveillance, with participants reporting feelings of constant confinement and loss of freedom [69, 70]. Analysis of available resources demonstrates an information asymmetry where abusers easily find surveillance tools while victims encounter poor, inaccurate advice online [4]. While technology clinics provide valuable individualized support [44], they face significant limitations in reaching broader populations due to requirements for specialized expertise, substantial resources, and geographic accessibility. These interventions are also inherently reactive, providing support only after abuse has occurred.

Protection Motivation Theory and behavior change. Protection Motivation Theory (PMT) provides a framework for understanding how individuals develop motivation to adopt protective behaviors in response to threats [29, 56, 73]. PMT identifies four key constructs that influence behavior change: perceived severity (how serious the threat consequences are), perceived vulnerability (likelihood of experiencing the threat), self-efficacy (confidence in one's ability to enact protective behaviors), and response efficacy (belief that protective actions will be effective) [66].

In security and privacy contexts, PMT has been applied to understand protective behaviors [68]. Self-efficacy has been identified as a particularly strong predictor of security behavior intentions [23], while threat appraisal constructs (severity and vulnerability) significantly influence whether individuals adopt protective measures. Research demonstrates that simply providing information about threats is insufficient—individuals must also perceive personal vulnerability and develop confidence in their ability to respond [73]. This is particularly relevant for TFA, where victims often do not recognize abuse patterns until significant harm has occurred [33], and may lack confidence in their technical abilities to implement protective strategies. Educational interventions that successfully shift PMT constructs are more likely to produce lasting behavior change than those focused solely on knowledge transfer [56]. In

privacy contexts specifically, PMT has been used to predict privacy-protective behaviors, with both threat appraisal and coping appraisal contributing to behavioral intentions [23].

Game-based learning foundations. Game-based learning leverages interactive gameplay to enhance learning outcomes and promote behavior change across diverse educational contexts. Foundational work by Prensky [61] and Gee [37] identified key principles that make games effective learning environments, including immediate feedback, progressive difficulty, and intrinsic motivation. A systematic review of serious games found positive effects on knowledge acquisition, skill development, and attitude change [22]. Meta-analytic evidence demonstrates that serious games produce learning gains comparable to conventional instruction methods while offering advantages for engagement and motivation [77].

Games offer unique affordances for engaging psychological mechanisms relevant to behavior change. Narrative transportation theory [39] suggests that when people become absorbed in a story, they are more likely to adopt the attitudes and beliefs presented, making narrative-driven games particularly effective for attitude change. Emotional engagement through storytelling enhances learning outcomes and retention [60]. Experiential learning theory [49] posits that learning occurs through reflection on direct experience, which games facilitate through interactive scenarios and decision-making. Self-determination theory [67] explains how games foster intrinsic motivation through autonomy (player choice), competence (achievable challenges), and relatedness (connection to characters) [25].

However, game design for behavior change requires careful consideration of engagement versus learning trade-offs. Educational games must balance entertainment value with learning objectives, as excessive focus on either dimension can undermine the other [36]. Effective educational games address these competing demands through thoughtful integration of learning content with engaging mechanics [43].

Games for teaching internet safety. Current cybersecurity educational games like Space Shelter from Google guide players through password management and multi-factor authentication, whereas Google Interland teaches children safe online behavior [71]. For professional audiences, training programs like the Cyber Awareness Challenge provide comprehensive training on topics including social engineering and malicious code detection [28]. Research evaluating cybersecurity games has shown mixed results: while games increase engagement and knowledge retention, their effectiveness at changing actual security behaviors remains understudied [45]. Existing games focus primarily on general internet safety without addressing TFA. The unique challenges of TFA—including relationship dynamics, psychological abuse, and power imbalances—require specialized educational approaches that current cybersecurity games do not provide.

Visual novels. Visual novels are a narrative-driven game genre characterized by branching storylines, character dialogue, and player decision-making, where players progress through text-based scenarios while making choices that influence the narrative outcome [17]. This genre has been increasingly adopted for educational purposes, with research demonstrating its effectiveness for teaching topics ranging from language learning [18, 51, 63] to health education [41, 79]. Visual novels are particularly well-suited for sensitive educational domains because their text-driven format allows players to engage at their own pace, re-read critical information, and process emotionally complex material with appropriate distance [39]. Their emphasis on narrative immersion and decision-making aligns with experiential learning theory [49], as players learn by navigating scenarios and observing the consequences of their choices rather than passively receiving information.

Digital games for violence prevention. Research has demonstrated the potential for prosocial video games to increase empathic responses and prevent intimate partner violence. One example is *Jesse*, a point-and-click role-playing game where players assume different character roles experiencing family violence [13]. Jennifer Ann's Group has sponsored the *Life.Love. Game Design Challenge* since 2008, producing over 70 games focused on teen dating violence prevention [40]. These games primarily focus on recognizing relationship dynamics and physical violence. While valuable, they do not address the technical aspects of technology-facilitated abuse or teach users concrete strategies for detecting and preventing digital surveillance.

Research gaps and study contribution. Despite the growing research on both TFA and game-based learning, significant gaps remain at their intersection. Current games addressing domestic violence focus on general relationship dynamics and physical violence, while security games concentrate on technical threats without considering interpersonal contexts where technology abuse occurs. Most critically, existing approaches do not integrate behavior change theories like PMT with game-based learning design, leaving unclear how games can best cultivate the protective mindsets necessary for TFA prevention. As smart devices and surveillance apps continue to proliferate, there is an urgent need for proactive educational approaches that instill protective mindsets before abuse occurs. To our knowledge, no research has created a proactive educational tool that integrates behavior change theory with game-based learning specifically for TFA prevention. Our research addresses this gap by developing and evaluating an educational game grounded in Protection Motivation Theory and designed to cultivate protective mindsets alongside technical knowledge. Unlike existing clinical approaches that require specialized expertise and infrastructure, our game provides a scalable proactive intervention. We use privacy guides as a baseline to validate knowledge transfer while focusing our evaluation on whether the game's narrative approach fosters the protection motivation that drives behavior change—an outcome that passive educational materials are not designed to achieve.

3 Design and Development of the *BeSafe* Game

We aim to design an interactive game that educates users on detecting and mitigating TFA while cultivating the protective mindsets necessary for behavior change. We named this game BeSafe[1]. The design methodology involves collecting real-world abuse cases and technical guides, constructing narratives grounded in behavior

[1]BeSafe is publicly accessible at https://besafegame.vercel.app

change theory, and storyboarding. We first describe desired learning outcomes from the game, then the data collection process, and finally we present our game design.

Why game-based learning? TFA often goes unrecognized by victims until significant harm has already occurred. Such abuse is widespread and takes various forms, and as discussed in Section 2, numerous efforts have been made to support victims. However, existing interventions focus on providing help after abuse has happened. Our goal is to create a proactive educational tool that raises awareness and cultivates protective mindsets *before* abuse occurs.

Game-based learning is well-suited for this goal. Beyond actively involving users in the learning process and encouraging exploration, problem-solving, and decision-making, games engage psychological mechanisms that drive behavior change. By simulating real-world situations in a safe and controlled environment, games allow users to practice applying knowledge, reflect on outcomes, and build confidence. This approach helps players not only understand key concepts but also develop the threat awareness, vulnerability recognition, and self-efficacy necessary to recognize and respond to technology-facilitated abuse in real life.

Research on game-based learning suggests that engagement is enhanced when the challenge level matches the user's skill level, creating conditions conducive to sustained attention and learning [1, 24]. In our game design, we aim for this balance through progressive difficulty in scenarios, immediate feedback on decisions, and clear goals that maintain user engagement without creating cognitive overload. Games can be accessed across different geographic locations and time constraints, potentially reducing some barriers that limit traditional educational approaches [37]. This scalability is crucial for proactive education, as it allows widespread education before abuse occurs rather than requiring users to seek help after experiencing harm.

We chose not to employ video-based interventions because effective learning requires student engagement and interactivity [26, 35]. Videos primarily represent a passive learning medium where students receive information without opportunities for immediate interaction or decision-making practice [48], which is particularly important for TFA education that requires practical skills. Additionally, video interventions are more resource-intensive and harder to keep updated as privacy, security and safety features change.

Learning outcomes. Our game targets dual outcomes: knowledge acquisition and protection motivation development. For knowledge outcomes, we relied on Bloom's Taxonomy [50], a widely used cognitive and educational framework. Bloom's Taxonomy classifies learning outcomes into six hierarchical levels (Appendix B), with our game targeting the third level: *Application*. Our game is designed to help players apply what they have learned in simplified contexts and extend that knowledge to new situations.

For motivation outcomes, we ground our design in Protection Motivation Theory (PMT). PMT posits that behavior change requires shifts in four key constructs: ❶ *Perceived severity:* understanding that TFA has serious consequences for privacy, safety, and well-being; ❷ *Perceived vulnerability:* recognizing personal risk of experiencing TFA, moving beyond the belief that "it won't happen to me"; ❸ *Self-efficacy:* developing confidence in one's ability to detect surveillance signs and implement protective strategies; ❹ *Fear arousal:* experiencing appropriate emotional concern about potential threats that motivates protective action without causing paralysis. The game's narrative structure is specifically designed to engage these constructs. By following characters through abuse scenarios, players develop empathy and recognize that TFA can happen to anyone (perceived vulnerability). By observing serious consequences of surveillance on victims' lives, players understand the threat's severity. By guiding characters through detection and mitigation steps, players build confidence in their own ability to respond (self-efficacy). By seeing successful outcomes where protective strategies work, players develop belief in the effectiveness of these actions (response efficacy). The emotional investment in characters' stories creates appropriate concern (fear arousal) that motivates attention to protective behaviors.

Based on these frameworks, our key outcomes are: ❶ *understanding* that technology can be leveraged as a tool for abuse across various forms and contexts, ❷ *developing awareness* of personal vulnerability to TFA, ❸ *identifying* signs of online surveillance based on recognized indicators, ❹ *building confidence* in one's ability to navigate and mitigate online surveillance safely and effectively, and ❺ *fostering motivation* to implement protective behaviors.

Forms of abuse. To ensure our game scenarios reflected real-world abuse, we compiled a list of abuse cases from existing research and resources. These sources included studies on spyware and dual-use apps [3, 19], IoT devices [69, 70], as well as work on interventions [32, 44] and resources available to abusers and victims [4, 12, 74]. We also referenced practical resources provided to victims by organizations such as the Clinic to End Tech Abuse (CETA) and the Safety Net Project.

Scope of the intervention. While organizations like CETA and Safety Net Project provide critical clinical support for active TFA survivors, BeSafe is designed for a different role: proactive, general-population education for users who have not yet experienced TFA but use platforms where it occurs. Its goal is to cultivate the threat awareness, vulnerability recognition, and self-efficacy that motivate routine protective behaviors — periodically reviewing linked devices, location-sharing settings, and account access — before abuse occurs. BeSafe deliberately does not address scenarios requiring expert intervention: coercive control with active surveillance, stalkerware removal under threat, or safety planning during separation remain the domain of tech clinics [21, 44]. We scope the intervention this way because the population at risk of TFA vastly exceeds the population currently receiving clinical support, and proactive education and clinical care address complementary needs rather than substituting for one another.

3.1 Game Design

Based on the learning goals and abuse scenarios identified above, we developed an educational story-driven visual novel game using established learning theories and behavior change frameworks. We chose the visual novel format for several reasons. Visual novels rely on narrative, dialogue, and branching choices rather than complex game mechanics, substantially reducing development and maintenance costs compared to other game genres—a critical consideration

given the need to frequently update content as platforms change. Their text-driven format is inherently accessible, requiring no specialized hardware or gaming experience, which broadens reach to non-gaming audiences who are often the primary targets of TFA. Additionally, before investing in more complex game designs, we sought to establish whether narrative-driven game-based learning can effectively cultivate protective mindsets—if a lightweight visual novel achieves significant motivation shifts, this validates the educational approach and provides a foundation for future iterations with richer game mechanics.

Our design process focused on three key elements: learning frameworks that support effective education and motivation development, story structures that engage users while teaching necessary skills and fostering protective mindsets, and design choices that focus on accessibility and psychological safety.

Theoretical Foundations. Our design uses three types of engagement — cognitive, emotional, and motivational — based on frameworks from educational psychology [31]:

- *Cognitive theories:* We used experiential learning theory [49], which emphasizes learning through thinking about doing. In our game, players are placed in scenarios where they must make choices and see consequences, strengthening knowledge through experience.
- *Theories of Emotion:* We used narrative transportation theory [39], which suggests that when people become absorbed in a story, they are more likely to adopt the attitudes and beliefs presented. By including emotional content—such as control within close relationships—players are more likely to relate personally and remember what they learned. This approach is supported by research on empathy in educational contexts [11].
- *Motivational theories:* We applied self-determination theory [67], focusing on three basic psychological needs: autonomy (player choice in decisions), competence (clear feedback and achievable challenges), and relatedness (connection to characters and scenarios). These elements work together to create intrinsic motivation and sustained engagement [25].

Narrative Construction. In our game, each story was created using a three-stage structure: ❶ *Background*: Players are introduced to characters in normal relationship dynamics, establishing trust and emotional investment. This stage demonstrates regular relationship patterns before introducing surveillance elements, helping players recognize the contrast when abuse begins. ❷ *Identification and investigation*: Suspicious behaviors and technology-facilitated abuse emerge gradually. Players learn to identify red flags such as excessive monitoring and privacy violations. The narrative structure allows players to explore evidence and make deductions about the abuse occurring. ❸ *Mitigation*: This stage allows players to help the victim navigate the abuse and address it. Players learn about protective strategies against TFA. These stages are based on trauma-informed principles [62]. Each chapter was based on real abuse cases documented in academic literature, focusing on different types of surveillance including location tracking, social media monitoring, and device compromise.

This narrative structure aligns with Protection Motivation Theory's threat and coping appraisal processes. The background phase establishes baseline expectations, making violations in the identification phase more impactful and increasing perceived severity. By having players discover abuse gradually (rather than being told explicitly), we foster personal vulnerability recognition—players realize they might miss similar signs in their own contexts. The mitigation phase directly builds self-efficacy by demonstrating concrete, achievable protective actions, while successful outcomes reinforce response efficacy. The emotional arc creates appropriate fear arousal that motivates protective behavior without overwhelming players. This balance is critical: excessive fear without adequate coping information can trigger defensive avoidance, where individuals mentally reject threatening messages to protect themselves from anxiety [76]. By pairing threat information with concrete, achievable protective actions, the game ensures that fear arousal enhances rather than undermines motivation.

Trauma-informed design choices. While designing BeSafe, we followed trauma-informed principles [20, 62]. We storyboarded game scenarios using *StoryboardThat*[2] to plan scenes and narrative flow, then implemented them in our custom engine (see Section 3.2). The stories are dialogue-driven, featuring clear text, expressive characters, and simple click-through interactions. Throughout development, we made deliberate design choices to promote psychological safety and support educational effectiveness.

First, players assume the role of observers rather than victims directly experiencing abuse. This perspective creates psychological safety by allowing players to engage with abuse scenarios without identifying as the target of surveillance. Players maintain emotional distance while still learning to recognize warning signs and protective strategies. Second, the narrative structure positions players as supporters helping victims navigate TFA situations, fostering empathy and prosocial engagement. This helper role can generate feelings of agency and positive contribution when dealing with emotionally difficult material, potentially increasing sustained engagement with challenging content through the psychological benefits of altruistic behavior [46, 80]. Third, while each story includes branching to reflect player choice, all paths converge on positive outcomes showing correct protective strategies (Fig. 7). Rather than punishing poor choices with negative consequences, this approach avoids re-traumatizing players who may have personal experience with abuse. Players learn through supported decision-making rather than punishment mechanics, ensuring consistent learning outcomes while maintaining agency and aligning with trauma-informed educational practices [62]. Finally, we selected a cartoon art style rather than photorealistic graphics to create psychological distance from potentially triggering content while maintaining emotional engagement. The stylized approach allows players to process difficult topics without overwhelming realism, following principles of self-distancing used in trauma research [75]. We also excluded audio elements to ensure accessibility across different environments and abilities. Text-based interaction allows players to engage at their own pace, re-read important information, and maintain privacy in shared spaces.

[2] StoryboardThat: http://storyboardthat.com

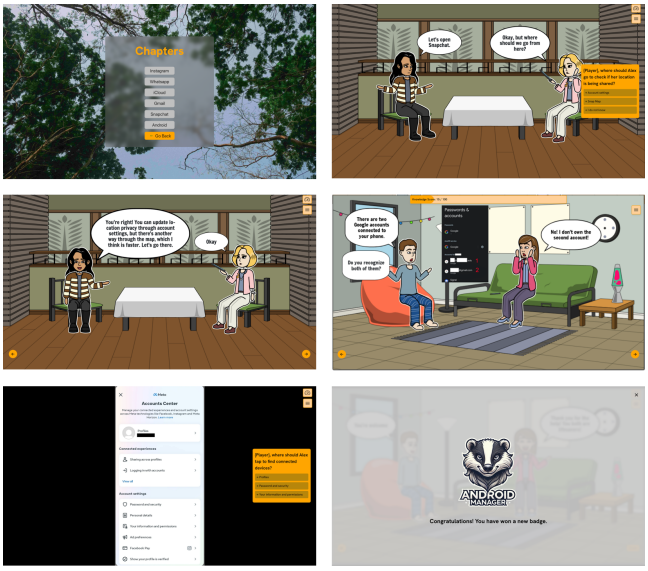

Figure 1: The figure shows key components of the game, including the list of chapters, the interactive story-based gameplay, and quiz questions presented during the story. Some quiz questions include screenshots from platforms, while others do not. Players receive immediate feedback on their answers, and a knowledge meter at the top reflects their progress and correctness. At the end of each chapter, players receive a badge as a reward to incentivize engagement.

3.2 Game Implementation

We built the game as a browser-based application that runs smoothly on both desktop and mobile. Our goal was to make it lightweight, fast, and easy to maintain. Fig. 1 shows the game interface, including the list of chapters and the interactive gameplay.

Scenarios. We designed six scenarios representing common TFA contexts: WhatsApp, Instagram, Gmail, iCloud, and Android all focused on account compromise and how to remove unauthorized devices, while Snapchat focused on location sharing and controlling who receives location updates. These scenarios were designed to span complementary dimensions of TFA rather than reuse a single template across platforms. Our scenarios vary along two axes: the *threat actor* (anonymous attackers, acquaintances, intimate partners, ex-partners, and relatives) and the *attack surface* (account compromise, location sharing, cloud syncing, and device-level compromise). Each scenario also targets a distinct recognition or mitigation skill—for example, distinguishing password change from session removal, recognizing patterns of unexpected proximity, or separating app-layer threats from account-layer threats. Fig. 2 summarizes this design (more details for each story are provided in Appendix C, with additional screenshots in Appendix D).

Visual novel engines. We tested existing visual novel engines like Ren'Py, Narrat, and Dialogic. None provided the level of control and responsiveness we needed, so we built the game using a custom engine with *React*, *TypeScript*, and *Vite*, which provided an efficient and customizable development environment tailored to our needs.

Platform	Threat actor	Attack surface
Instagram	Anonymous	Account compromise
WhatsApp	Acquaintance	Account compromise (physical access)
iCloud	Anonymous	Cloud syncing
Gmail	Relative	Account compromise
Snapchat	Ex-partner	Location sharing exploitation
Android	Intimate partner	Device-level compromise

Figure 2: Scenario design matrix. Each scenario combines a different threat actor with a different attack surface to broaden the range of TFA situations players encounter.

Game features. To support learning and engagement, we implemented features grounded in game-based learning principles:

- *Knowledge Meter*: Tracks player progress by awarding points for correct decisions and completed story sections, providing immediate, visible feedback.
- *Badges*: Players earn badges upon completing stories, serving as positive reinforcement and encouraging replayability.
- *Leaderboard*: Ranks players based on accumulated knowledge points, creating a light sense of competition.
- *Nickname*: Players choose a nickname that personalizes their experience and appears on the leaderboard, without requiring account creation or revealing personal information.

All progress is stored using cookies, so players can return to the game later and continue where they left off.

Deployment. We deployed the game frontend using Vercel, which provides automated builds, CDN distribution, and fast global performance. For backend functionality, we used Amazon EC2 instances. The frontend and backend are decoupled, making it easier to update or expand each component independently. The system uses a modular design that allows new stories and features to be added without major infrastructure changes.

Expert consultation and game pilot testing. We consulted with experts at technology clinics to gather feedback on the educational content and game mechanics. We conducted iterative refinement based on their input, focusing on ensuring that technical content was accurate and that the narrative approach effectively conveyed both technical knowledge and awareness of abuse dynamics. Expert feedback led to refinements in dialogue pacing, clarification of technical terminology, and adjustments to branching logic to ensure all paths provided educational value.

4 Evaluating the *BeSafe* Game

To assess the effectiveness of our game-based intervention for cultivating digital privacy and safety mindsets, we designed a between-subjects study with three conditions: (1) a game with gamification features, (2) the same game without those features, and (3) privacy guides covering the same content.

Knowledge validation. We use privacy guides as a baseline measure to validate that game-based learning successfully transfers

technical information about TFA. Privacy guides—such as those maintained by the Clinic to End Tech Abuse (CETA) [21]—are reference documents designed for quick information retrieval, not structured learning. They serve users who need immediate answers to specific questions, typically during or after a crisis. They are not designed to teach, engage, or cultivate lasting understanding. We include them in our study not as a competing educational intervention but as a baseline to validate that our game achieves the fundamental prerequisite of information transfer: if participants cannot acquire technical knowledge from the game at least as effectively as from reading a reference document, then the game fails at its most basic objective. We selected CETA's guides[3] as our baseline because they are among the most widely used TFA resources actively deployed by tech clinics working directly with victims [44]. Unlike generic online privacy advice, which prior work has shown to be inconsistent and often inaccurate [4], CETA's guides are maintained by researchers and practitioners with sustained engagement in the TFA domain, vetted through clinical use, and updated as platforms evolve. This makes them a meaningful real-world baseline: they represent what a motivated user would encounter today if seeking authoritative guidance on the topics our game covers. For each platform context, we presented participants with the corresponding CETA guide, covering the same content as the game scenario.

Protection motivation assessment. Our primary evaluation focuses on whether the game's narrative approach can shift PMT constructs that drive behavior change. These constructs predict whether individuals will actually implement protective behaviors [66, 73], making them critical outcomes for effective TFA prevention. While knowledge measures tell us what participants learned, PMT constructs reveal whether they developed the psychological foundations necessary for translating knowledge into action. Passive educational materials like privacy guides are designed for information transfer but do not systematically engage the emotional and motivational mechanisms that narrative-driven games can activate [39]. By measuring both knowledge and PMT constructs, we assess whether games offer value beyond information delivery—specifically, whether they cultivate the protective mindsets that motivate behavior change.

Study design. Our study evaluates game-based learning across two dimensions. First, we validate knowledge acquisition by including privacy guides as a baseline, demonstrating that games successfully transfer technical information. Second, we assess protection motivation development through PMT constructs, examining whether the game's narrative approach fosters psychological changes that passive educational materials cannot achieve. Additionally, we investigate whether gamification features (badges, leaderboards, knowledge meter, nicknames) enhance these outcomes compared to narrative alone.

The study follows a 3×6 design where participants are randomly assigned to one of three learning conditions (game with features, game without features, or privacy guides) and one of six application contexts (WhatsApp, Instagram, Snapchat, Gmail, iCloud, or Android). Our primary statistical analyses compare across the three learning conditions (approximately 66 participants per group). The six platform contexts serve as a blocking variable to assess generalizability rather than as a primary comparison, and we test for interaction effects between learning format and platform context.

Recruitment. We recruited participants through Prolific, using a multi-stage screening process. Initial screening assessed participants' interest in learning about online privacy and safety topics using a Likert scale. Only those expressing interest were invited to the main study. For each of the six applications (WhatsApp, Snapchat, Instagram, iCloud, Android, Gmail), we implemented specific screening questions to ensure participants had familiarity with the platform they would be learning about. This approach improved ecological validity by connecting the educational content to participants' actual technology experiences. We ensured that participants could not participate multiple times—each individual could complete the study only once, regardless of which application context they were assigned to. The study was restricted to US residents aged 18 and older, with additional requirements that participants have an approval rating of at least 95% on Prolific. Random assignment to learning group and application context ensured balanced representation across all combinations.

Study procedure. First, we collected basic background information including age, internet usage, social media usage, technical background, education level, and relationship status. Next, participants completed the Big Five personality inventory (10 items) to assess individual differences that might influence learning outcomes [38, 64].

This was followed by baseline measures including PMT assessments of four constructs: 1) *perceived vulnerability*, 2) *perceived severity*, 3) *self-efficacy*, and 4) *fear arousal*. Participants also completed a pre-test knowledge assessment about the specific app they would learn about. Knowledge assessment scores are calculated as the percentage of correctly answered questions on a 0–100 scale.

Then, participants were randomly assigned to one of three learning groups: game with features, game without features, or privacy guides. After completing their assigned educational intervention, participants provided feedback about their learning experience and completed immediate post-measures including the same knowledge assessment and PMT constructs to assess changes in knowledge, threat perceptions, and protective motivation. The participants were also presented with three attention check questions; all participants answered them correctly, suggesting attentive engagement with the survey.

One week later, participants who showed improvement in their knowledge scores were invited to a follow-up survey to assess knowledge retention by repeating the same knowledge assessment a third time. During this follow-up, we also collected information about participants' personal experiences with online interpersonal attacks or their knowledge of others who had experienced abuse. This information was collected at follow-up rather than during the main study to avoid priming effects that might influence initial learning or PMT responses.

Upon completion of the study, participants were compensated at $10 per hour, with the main study taking approximately 15–20 minutes. Participants invited to the one-week follow-up received an additional $1 bonus, as this survey took less than 2 minutes. Our study instrument is provided in Appendix D.

[3] https://ceta.tech.cornell.edu/resources

4.1 Ethical Considerations

This study was reviewed and approved by the university's Institutional Review Board (IRB) as a minimal risk human subject study. Given the sensitive nature of TFA content, we implemented trauma-informed design principles throughout both the game and the study.

Participant safety. Participants received content warnings about surveillance scenarios before beginning the study. The game's narrative deliberately avoided re-traumatization by ensuring all story paths converged on positive outcomes demonstrating effective protective strategies, rather than using punishment mechanics that could reinforce vulnerability or helplessness. Players assumed observer roles rather than victim roles, creating psychological distance from abuse scenarios. Participants were informed of their right to withdraw at any time without penalty.

Data protection. The game required no account creation or personal information beyond research requirements. Participants selected pseudonymous nicknames, and game progress data was stored locally using browser cookies rather than centralized databases. Backend logging captured only interaction patterns (e.g., choices selected, time per screen) and did not collect personally identifiable information. All data was stored on secured Amazon EC2 instances accessible only to authorized research team members, following institutional data protection standards.

Research team wellness. Team members developing content about TFA received appropriate advocacy training and had access to institutional support resources. Team members without such training were excluded from processes involving direct exposure to abuse scenarios.

4.2 Descriptive statistics

Piloting the study. We conducted a pilot study on Prolific with 20 participants to test the complete study protocol and make final adjustments before deploying the full experiment. The pilot revealed minor issues with survey flow and timing estimates, which we corrected before the main study. Pilot participants were not included in the final analysis.

Demographics. Of 454 individuals who completed our screening survey, 352 (77.5%) indicated interest in learning about online abuse and safety topics. Our power analysis indicated that we needed at least 198 participants. We invited all 352 interested individuals to participate via Prolific in batches, with Prolific's recruitment system automatically closing enrollment once 198 participants completed the study. Our sample had a female bias—63.4% of our participants were female. The largest age group was 25–34 years (33%), followed by 35–44 years (24%) and 18–24 years (20%), indicating a primarily young to middle-aged adult sample. Our participants were highly educated, with 49% holding bachelor's degrees, 27% holding graduate degrees, and 23% having completed high school or equivalent. Relationship status varied considerably: 54% were married, 21% were dating, 16% were single, and 8.6% were either divorced, widowed, or separated. Our participants were evenly split across technical background, with 51% having CS/IT background and 49% without. Our participants were long-term internet users: 87.4% had been using the internet for more than 10 years and 11.6% for 5–10 years. The detailed demographics of our participant pool are included in Fig. 3.

Demographic	Total	%
Gender		
Male	72	36.4
Female	126	63.6
Age Bucket		
18–24	40	20.2
25–34	66	33.3
35–44	48	24.2
45–54	26	13.1
55–64	15	7.6
65 or older	3	1.5
Highest Level of Education		
High school degree or equivalent	46	23.2
Bachelor degree	97	49
Graduate degree	55	27.8
CS/IT background		
Yes	100	50.5
No	97	49
Prefer not to answer	1	0.5
Relationship Status		
Single, never dated	32	16.2
Married	106	53.5
Dating / Living together	42	21.2
Divorced	10	5.1
Widowed	4	2
Separated	3	1.5
Prefer Not to Answer	1	0.5
Internet Usage		
Less than a year	1	0.5
3 - 4 years	1	0.5
5 - 10 years	23	11.6
More than 10 years	173	87.4

Figure 3: Participant demographics.

Data analysis. We analyzed quantitative data using statistical tests determined by variable type and sample characteristics, with significance set at $\alpha = 0.05$. For each participant, learning gain was calculated as the difference between pre-test and post-test scores (percentage of correctly answered questions, 0–100 scale), while retention was measured as the change from immediate post-test to one-week follow-up scores. We used one-way ANOVA to compare scores across the three learning conditions (game with features, game without features, privacy guides). We assessed potential interaction effects between learning format and application context to ensure that our findings generalized across platforms. To analyze individual differences in learning performance, we employed Welch's t-tests when comparing two groups (e.g., participants with vs. without technical background) and Pearson correlations for continuous variables such as age.

PMT construct changes were analyzed using binary logistic regression, with PMT improvement as the dependent variable and learning group as the independent variable, reporting odds ratios with 95% confidence intervals to quantify effect size. We operationalized PMT improvement as positive change in any of the four constructs (perceived vulnerability, perceived severity, self-efficacy, or fear arousal) from pre-test to post-test.

Personality trait effects on learning and retention were examined using multiple linear regression models, with Big Five personality

dimensions as independent variables and learning outcomes as dependent variables. We also tested for interaction effects between personality traits and the learning groups. Effect sizes were reported using Cohen's d for t-tests and η^2 for ANOVA analyses.

Open-ended survey responses were analyzed through systematic thematic coding. One researcher developed a codebook by reading participant responses and identifying recurring themes. A second coder independently coded 30% of randomly sampled responses using the codebook. Inter-coder reliability was assessed using Cohen's κ, which ranged from 0.75 to 1.00 across question sets, with a median of 0.83, indicating substantial to excellent agreement according to established benchmarks [30, 53]. The codebook along with secondary analyses are provided in Appendix D.

5 Efficacy of *BeSafe* Game

5.1 Interest in Learning

Before evaluating BeSafe's effectiveness, we examined whether people want to learn about online surveillance proactively. Screening data showed that 352 of 454 individuals (77.5%) expressed interest in learning about online abuse and safety topics. Observed engagement behaviors throughout the study—including quality of feedback responses and task completion rates—provided additional evidence that participants remained actively involved with the educational content beyond initial expressed interest. These findings suggest substantial demand for proactive TFA education.

5.2 Effectiveness of *BeSafe*

To evaluate *BeSafe*'s effectiveness, we first present our primary outcome—whether the game cultivates protection motivation that drives behavior change—then validate that game-based learning successfully transfers technical knowledge.

Game-based learning fosters privacy and safety mindset. Our primary evaluation examined whether game-based learning cultivates the motivation and protective attitudes that drive behavior change. We used binary logistic regression to examine whether the likelihood of improvement in PMT constructs differed across learning formats. We defined PMT improvement as positive change in any of the four constructs from pre-test to post-test, indicating development of protective mindsets.

Results revealed that game-based learning more consistently fostered PMT improvements compared to the privacy guides group. Participants in the game with features group showed higher odds of PMT improvement compared to the privacy guides group (OR = 3.63, $p = 0.018$), while those in the game without features group showed even higher odds (OR = 6.56, $p = 0.005$).

The stronger effect for the no-features game suggests that participants were more responsive to a streamlined narrative experience where attention focused entirely on storyline and decision-making, without additional game elements. This finding indicates that narrative engagement—rather than game mechanics like badges or leaderboards—may be the primary driver of motivation shifts. While knowledge gains demonstrate what participants learned, PMT improvements reveal whether they developed the protective mindsets necessary to implement that knowledge in real-world contexts. The significantly higher odds of PMT improvement in game groups suggest that narrative-driven approaches offer value beyond information transfer—they cultivate the threat awareness and self-efficacy that motivate protective action.

Knowledge transfer validation. We validated that game-based learning successfully transfers technical knowledge about TFA by measuring pre-test to post-test knowledge gains. Privacy guides served as a baseline reference point, demonstrating that our game achieves comparable knowledge outcomes.

Analysis of pre-test and immediate post-test scores revealed statistically significant knowledge gains across all groups. Paired-sample t-tests showed that post-test scores exceeded pre-test scores for the game with features ($p < 0.001$, Cohen's $d = 0.67$), game without features ($p < 0.001$, $d = 0.54$), and privacy guides ($p < 0.001$, $d = 0.71$). All effect sizes reached at least moderate levels, indicating that participants effectively comprehended the learning materials regardless of delivery format.

Excluding participants with perfect pre-test scores (ceiling effect) revealed even stronger learning effects. The no-ceiling analysis showed large effect sizes across all groups: game with features ($d = 1.14$, $\bar{x}_{gain} = 44.7$ points), game without features ($d = 0.89$, $\bar{x}_{gain} = 35.1$ points), and privacy guides ($d = 0.96$, $\bar{x}_{gain} = 38.5$ points). A one-way ANOVA comparing learning gains across the three groups revealed no statistically significant differences ($p = 0.492$, $\eta^2 = 0.007$), and Tukey's HSD post hoc tests confirmed no significant pairwise differences. All educational formats achieved significant within-group learning, validating that game-based approaches successfully transfer technical information about TFA.

Knowledge retention after one week. Retention was assessed as the change in scores from the immediate post-test to the delayed post-test one week later. We note that only participants who showed improvement on the immediate post-test were invited to the follow-up; consequently, these retention findings apply specifically to participants who demonstrated initial learning gains and should be interpreted with this selection criterion in mind.

Among those who completed the follow-up, participants retained most of what they had learned. The game with features group maintained near-complete retention ($\bar{x}_{change} = -1.76$ points, $p = 0.795$, $d = -0.06$). The game without features group exhibited a small, non-significant decline ($\bar{x}_{change} = -11.41$, $p = 0.085$, $d = -0.42$), while the privacy guides group showed a marginal decline ($\bar{x}_{change} = -17.67$, $p = 0.075$, $d = -0.37$). A one-way ANOVA revealed no statistically significant differences in retention between groups ($p = 0.425$, $\eta^2 = 0.03$).

Perceived effectiveness and engagement. Understanding user perceptions is critical because perceived effectiveness influences whether people will engage with educational materials and recommend them to others. Of the 132 game participants, 126 (95.5%) described the game as engaging, and 52 (39.4%) specifically highlighted its educational value. Many reported that it helped them learn new concepts or reinforced existing knowledge, especially due to its interactive and scenario-based format.

The branching storylines, character dialogue, and choice-making were frequently mentioned as engaging ways to present information. One participant noted (P181): *"Since I now understand how to*

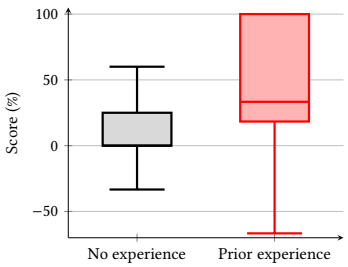

Figure 4: Effects of abuse experience on learning gain.

remove unauthorized accounts, I will confidently apply this knowledge if I or someone else faces the issue". Several described the game as immediately applicable: *"I went to check my account's linked devices right away"* (P46); *"Will now always check on my Snapchat location just to make sure I'm not being tracked at any time"* (P144). Over 100 game responses (107 of 132) described the game as successfully teaching or reinforcing key ideas, while 16 characterized the material as already familiar—suggesting broad educational impact even among participants with some prior knowledge.

Our evaluation reveals that game-based learning achieves two outcomes: (1) successful knowledge transfer, validated through comparison with privacy guides, and (2) cultivation of protection motivation through narrative engagement. While privacy guides effectively convey technical information, the game's narrative approach fosters protective mindsets alongside knowledge acquisition, suggesting particular value for proactive TFA prevention where the goal extends beyond information delivery to developing the threat awareness and self-efficacy that motivate protective action.

5.3 Effects of User Characteristics

Having established that game-based learning produces both knowledge gains and motivation improvements, we examined what factors influenced these outcomes. We analyzed participant characteristics that shaped engagement, comprehension, and development of protective mindsets.

Personal relevance and perceived risk are associated with higher engagement and learning gains. Participants with direct or indirect exposure to online abuse demonstrated markedly higher immediate learning gains compared to those without such exposure. Using Welch's t-test, we found that participants with prior exposure improved by an average of 45.42 points, while those without prior exposure improved by 15.10 points ($p < 0.001$, $d = -0.87$), indicating a large effect (Fig. 4). This pattern held for both knowledge gains and PMT improvements, suggesting that personal relevance is associated with greater cognitive learning and motivation development. However, this advantage did not persist at the one-week follow-up ($p = 0.627$, $d = -0.13$). Qualitative responses reinforced this finding. One participant connected their motivation to a recent personal experience (P187): *"The game grabs my attention because I've actually recently been dealing with a similar issue after a bad break-up"*. Others described increased interest when content addressed their own privacy concerns (P150): *"It definitely informed me on Snapchat and the ways that it could be used against me, as well as how to prevent this from happening"*. These examples suggest that perceived personal risk appears closely linked to engagement with privacy and safety education.

Learning is equally effective across age groups and genders. Age showed no significant relationship with either learning gains ($r = 0.11$, $p = 0.21$) or retention ($r = -0.03$, $p = 0.81$). Similarly, gender differences were not statistically significant for learning gains, with females averaging 37.47 points and males 43.10 points ($p = 0.44$, $d = -0.14$). Gender also showed no effect on retention ($p = 0.56$, $d = -0.15$). These findings suggest that the educational materials were equally effective across different age groups and genders, indicating broad accessibility.

Prior knowledge correlates with learning outcomes. Both technical background and education level showed significant effects on learning gains. Participants with technical background outperformed those without ($p = 0.0044$, $\eta^2 = 0.054$), and education level was also significantly related to learning ($p = 0.012$, $\eta^2 = 0.044$). Among those with technical experience, some expressed confidence in navigating privacy settings independently (P45): *"Most of it is self-explanatory. You could just go into account or privacy and see what shows up"*. Others acknowledged that although the material was familiar, it served as a useful reinforcement (P97): *"I was aware of the information presented in settings already, but it was still helpful"*.

In contrast, participants with less prior experience frequently reported that the game clarified concepts and showed concrete steps they could take—107 of 132 game responses described it as teaching or reinforcing key ideas. One participant explained (P170), *"I already had some of the knowledge and was aware of these features. However, the game did reinforce my knowledge, so it serves as a reminder and refresher to double-check if I find myself in a similar situation"*.

Hoever, neither technical background of participants nor education level showed significant effects on retention ($p = 0.477$ and $p = 0.777$, respectively), suggesting that while these factors enhance initial comprehension, other factors—such as engagement or personal relevance—may play a more central role in sustaining knowledge over time.

Personality traits. Multiple linear regression examining whether the Big Five personality traits affected learning gain was not statistically significant ($R^2 = 0.049$, $p = 0.082$), and no individual traits reached significance, though agreeableness showed a marginal negative trend ($\beta = -0.14$, $p = 0.08$). Interaction effects between learning group and personality traits were also non-significant ($p = 0.349$). For retention, agreeableness was significantly and negatively correlated ($r = -0.17$, $p = 0.017$), indicating that more agreeable participants tended to retain less knowledge one week later. No other traits showed significant correlations. Four separate regression models examining Big Five traits and individual PMT constructs revealed small but non-significant associations ($p > 0.05$), including agreeableness with perceived vulnerability ($\eta^2 = 0.033$) and self-efficacy ($\eta^2 = 0.020$), and extraversion with perceived vulnerability ($\eta^2 = 0.029$) and fear arousal ($\eta^2 = 0.028$).

5.4 Effects of the Game Design Elements

We analyzed how specific features and design elements influenced participants' educational experiences.

Immediate feedback is associated with higher retention. For the groups that played the game, immediate feedback on answers showed a positive relationship with knowledge retention ($r = 0.317$, $p = 0.046$). Participants who reported receiving clear information about their success or failure retained more knowledge one week later. As one participant explained (P14), *"The game gives you immediate feedback of the steps followed in a question. In case you are incorrect you get to know what is the answer to the wrong choice made."* This suggests that timely feedback serves as an important design element for educational effectiveness.

Emotional engagement does not always translate to learning gains. The analysis revealed counterintuitive correlations: participants who reported higher enjoyment without boredom or anxiety ($r = -0.254$, $p = 0.025$) and greater emotional involvement ($r = -0.241$, $p = 0.033$) demonstrated lower immediate learning gains. Several participants distinguished between educational utility and entertainment value (P154): *"It didn't really seem like a fun game I would play but more like a seminar to teach you about account security"*. These participants approached the game with a learning mindset, recognizing the interactive features but having realistic expectations about entertainment value. One noted that *"There are questions that you have to engage in"* (P46), while another commented, *"It really wasn't that engaging but it is an educational game so it kind of teaches you thus grabbing your attention"* (P55).

Participants who reported low entertainment value still demonstrated sustained attention to educational content. As one explained, *"I paid attention because I was reading but it wasn't very interesting"* (P147), while another noted, *"It's fine and informational, just not attention grabbing"* (P115). Even participants who felt overwhelmed, such as one who shared, *"I'm a bit overwhelmed at [the] time and don't want to learn much"* (P11), still engaged sufficiently to extract educational value. Participants also frequently emphasized the game's value as a reinforcement tool (P119): *"It acts as a reminder on how to change security of Gmail accounts"*, and (P108): *"I forgot about reviewing connected devices from time to time"*. Even participants with existing security knowledge found value in the structured review of familiar concepts: *"I think it did a good job at reminding you to be cautious and to occasionally go and check your settings"* (P69). This pattern aligns with the *disfluency effect*, where reduced entertainment value or mild cognitive effort promotes deeper processing and better recall [8, 27]. Highly immersive experiences can encourage focus on game-related goals over learning objectives [1, 24]. Participants who were less emotionally absorbed may have approached the game more deliberately as a learning exercise, leading to higher cognitive investment in the privacy and safety concepts presented.

Narrative elements were associated with sustained engagement. Many participants described being motivated to continue because they were curious about how the story would unfold or how their choices might affect the outcome. Of 132 game participants, 126 described the game as engaging and 52 specifically highlighted its educational value. One participant explained (P175), *"The game presents information in an interactive and engaging way. Instead of just reading or watching a video, I was actively involved in making choices and following a storyline. This made me more curious about what would happen next and helped me focus on the content, even though it was educational"*. Another noted (P145), *"I really enjoyed playing and following along. I felt like I was helping a friend,"* highlighting how the narrative structure fostered personal investment in the learning process.

Game features such as quizzes, badges, and the knowledge meter were also mentioned as helping maintain attention. Participants described how the cartoon-style images and animations drew them into the experience (P3): *"The cartoon images were interesting enough to draw me in at first"*, and how interactive questions created a sense of personal involvement (P117): *"It feels like I am part of the game and there are questions directed to me in the game"*. Others highlighted the opportunity to test their knowledge (P87): *"It gives me a chance to test my knowledge on internet application"*, and the role of quizzes in maintaining their interest (P143): *"The animations and random questions kept me interested in the game development"*. Together, these mechanics and design elements provided immediate feedback, a sense of progression, and active engagement, motivating participants to complete the game. Several participants indicated that the game inspired them to continue exploring online privacy and safety topics afterward.

Participants reported that information density affected comprehension. Content density impacted some participants' ability to engage. One participant noted (P114), *"It seems difficult for someone who is not tech-savvy, to remember all the steps,"* and another (P11) explained, *"I'm a bit overwhelmed at [the] time and don't want to learn much."* In contrast, others valued the step-by-step format (P52): *"The game outlined a simple process to follow and block other devices in case you think your account is being used by someone without your consent."* These responses suggest that progressive disclosure of information, rather than presenting all steps simultaneously, supports comprehension in privacy education contexts.

5.5 Post-Learning Impact of *BeSafe*

Learning gains and motivation improvements are meaningful only if they translate into behavioral intentions. We explored participants' planned actions and anticipated barriers.

Planned actions were driven by personal relevance. A substantial number of participants (91 of 132, 68.9%) reported intentions to apply what they had learned by reviewing or modifying their privacy and security configurations. For example, one participant (P134) stated, *"I will be verifying my Snapchat is still in ghost mode,"* while another (P142) noted, *"I will explore whether the same settings apply to other apps."*

Participants were most inclined to take action when the material addressed situations they personally faced. When examples matched their own online behavior, it prompted immediate plans for change. One participant (P185) shared, *"As someone who handles sensitive data at work, knowing how to secure a device is essential. I'll use some of these steps to check my own and my family's devices"*.

However, 13 participants identified barriers to applying knowledge, including concerns about forgetting specific steps or lacking confidence to navigate unfamiliar platform interfaces.

Knowledge sharing intentions indicate broader impact. Several participants explicitly planned to share what they learned with others. One shared (P118), *"I am looking forward to inform my friends of the discovery I made today. Maybe it will help them be safe"*, while another remarked (P168), *"I want to let my kids play this game!"*. The willingness to share was most often expressed when participants viewed the material as relevant to others' experiences, indicating that relatable content can motivate learners to become advocates for broader online safety awareness.

6 Discussion

Cultivating transferable and resilient protective mindsets. Technology platforms rapidly evolve their features and vary significantly in how they implement privacy controls [2], creating dual challenges for TFA education: procedural knowledge becomes outdated as platforms change, and platform-specific training may not transfer across applications. Our game-based approach addresses both challenges by cultivating protective mindsets. When participants develop threat awareness, vulnerability recognition, and confidence in their ability to investigate privacy settings, they acquire transferable principles that persist across platform changes and different applications. Rather than memorizing specific menu sequences that may become obsolete, participants learn to recognize surveillance patterns, understand what to look for, and navigate unfamiliar interfaces with protective intent.

Participants demonstrated transfer thinking in their intentions, with several explicitly describing plans to apply what they learned beyond their specific platform and noting they now understood *"what to look for"* rather than just memorizing steps. By fostering motivation to investigate and self-efficacy to explore privacy controls, games prepare users to adapt protective strategies as platforms evolve. Future research should examine whether mindset-focused educational interventions produce longer-lasting protective behaviors compared to procedural training, and whether users can successfully identify and mitigate surveillance threats on unfamiliar platforms and after interface changes.

Coverage and the limits of platform-specific instruction. A reasonable concern about any platform-grounded TFA intervention is whether six scenarios can meaningfully address a much larger threat surface. As noted above, prior work analyzing 100 privacy and safety features across 10 social media platforms found substantial variation in feature availability, default settings, and access pathways, much of which is directly relevant to TFA prevention [2]. No realistic curriculum can keep pace with the breadth and rate of change of the TFA threat surface, which spans social media, messaging, cloud services, mobile operating systems, IoT devices, and emerging technologies, with each evolving its interfaces and privacy controls on its own timeline. BeSafe is therefore designed to instill a transferable protective mindset — awareness of personal vulnerability, recognition of surveillance patterns, and confidence to investigate privacy controls — rather than to deliver an exhaustive catalogue of platform-specific procedures. The six scenarios serve as concrete grounding for this mindset, and we expect the mindset itself to transfer to platforms and attack vectors not directly covered. Whether this transfer occurs in practice is an empirical question that future work should address directly.

Industry and academic responsibility. The technology industry has substantial opportunities to improve user education beyond simply providing safety features. Platforms could develop more comprehensive introductory experiences that teach users about privacy controls and potential risks, rather than expecting users to discover these features independently. Such technical improvements alone cannot address the complex interpersonal dynamics involved in online surveillance. Academic institutions training future software engineers also play a role. Research has identified gaps in cybersecurity education, including limited mandatory security coursework at major universities [5], widespread use of insecure programming practices among students and instructors [5], and insufficient coverage of security topics in computer science textbooks [6, 72]. These educational gaps result in students lacking fundamental prerequisites for secure programming [7, 52], with limited exposure to privacy and safety topics that affect end users. Our research demonstrates that both technical and non-technical users can effectively learn about technology-facilitated abuse through accessible educational approaches, suggesting that prevention education need not be limited to specialized audiences.

Challenges of creating and maintaining educational games. Developing educational games for TFA presents challenges beyond typical security and privacy education. Game content must be continuously updated as platforms evolve their interfaces and privacy settings [2] — a scenario teaching users to check linked devices may become inaccurate after a single platform update. Abuse scenarios must be realistic enough to educate yet carefully designed to avoid re-traumatizing players with lived experience [20, 62], requiring ongoing collaboration with domain experts at advocacy organizations. Additionally, abuse tactics themselves evolve as abusers discover new ways to exploit emerging technologies [33], meaning game content must address novel threats, not only document known ones. Our modular architecture — which decouples story content from the game engine — was designed to reduce this maintenance burden by allowing individual scenarios to be updated independently. Future work should explore community-driven approaches to maintaining such games, such as open-sourcing game content to allow advocates, educators, and technically skilled volunteers to contribute updated scenarios as platforms change.

Scenario framing and ecological validity. The narrative structure we adopted—an observer perspective, branching paths converging on positive outcomes, and successful mitigation in a single sitting—was a deliberate trauma-informed design choice [20, 62], but introduces ecological tradeoffs worth acknowledging. Real TFA situations are often ambiguous, escalating, and resistant to clean resolution: victims may not recognize abuse patterns, may not accept help when offered, and protective actions can themselves carry safety risks [33]. Our Android scenario partially addresses this by warning that account removal will be visible to the abuser, but anticipating the full range of possible costs across scenarios is difficult, since the consequences of any protective action depend on

the specific abuser, relationship, and platform context. In real cases, removing a linked device, changing privacy settings, or revoking access can alert an abuser and trigger retaliation, and the "correct" technical action may be contraindicated by safety considerations—particularly during separation or cohabitation. By presenting scenarios where investigation reliably leads to detection and detection reliably leads to safe mitigation, BeSafe may inflate participants' confidence in protective actions in ways that do not match real-world conditions. Future iterations should consider scenarios that introduce safety-aware decision-making, including cases where the technically correct action is not the safest action and where seeking expert support is itself the recommended response. Such scenarios would require careful design to avoid undermining self-efficacy or re-traumatizing players, but would more accurately model the conditions under which protective behaviors are actually deployed.

Alternative explanations for the no-features version advantage. We interpret the higher odds of PMT improvement in the no-features version (OR = 6.56) compared to the with-features version (OR = 3.63) as evidence that narrative engagement, rather than gamification mechanics, drives motivation shifts. Several alternative explanations merit consideration. First, badges, leaderboards, and the knowledge meter introduce additional visual elements and competing goals that may divert attention from narrative content, reducing the depth of processing applied to the abuse scenarios—consistent with the disfluency findings discussed in Section 5.4 and with broader cognitive load research [60]. Second, gamification mechanics designed to feel rewarding (badges, point accumulation) may clash with serious content about surveillance and abuse, subtly undermining the emotional engagement that drives PMT shifts; several participants commented on a perceived tension between the educational content and entertainment framing (P154, P115). Finally, although random assignment balances groups in expectation, the with-features and no-features version may differentially engage participants with different motivational orientations—extrinsic versus intrinsic—so that one version is not globally superior but rather better suited to different sub-populations.

Limitations. Our work has several limitations. Our Prolific-based recruitment required participants to express initial interest in privacy education, potentially excluding individuals with lower baseline motivation who might benefit from such interventions. Our sample was highly educated (77% held bachelor's or graduate degrees) with extensive internet experience (87% using internet for 10+ years), and our findings may not generalize to broader populations. However, we argue that populations with lower digital literacy and less awareness about TFA would likely benefit even more from our intervention, suggesting that the effectiveness presented here may represent a conservative estimate.

Our study relied on self-reported measures of knowledge and behavioral intentions rather than observing actual behaviors. While participants expressed strong intentions to implement protective strategies, we cannot confirm whether these intentions translated into sustained protective behaviors, and longitudinal studies tracking actual privacy setting changes would strengthen these findings. Our one-week follow-up period captured only short-term retention and could not assess long-term knowledge maintenance or behavior change. However, our qualitative analysis partially mitigates this limitation by uncovering concrete post-learning plans, including specific intentions to review linked devices, enable privacy settings, and share knowledge with others.

BeSafe is designed as a proactive, general-population educational tool rather than a clinical resource for active survivors or a comprehensive TFA curriculum. Its scope is bounded in several ways that shape how its findings should be interpreted. First, the six scenarios cover account compromise, location sharing, cloud account compromise, and device-level compromise on widely-used Western consumer platforms; they do not address regional platforms (e.g., WeChat, LINE, Telegram), IoT-mediated abuse [69, 70], or financial and employment platforms increasingly implicated in TFA. Second, our recruitment was restricted to U.S. residents through Prolific, and TFA dynamics, platform availability, and legal recourse vary substantially across jurisdictions. Third, our screening required participants to express prior interest in privacy and safety education, meaning our sample reflects the most receptive segment of an already online research-participant pool. Populations with lower digital literacy, less prior exposure to TFA discourse, or active safety concerns may experience the intervention differently—potentially benefiting more, but also potentially requiring different adaptations than what BeSafe currently provides. We expect the educational effect demonstrated here to be a conservative estimate of what is achievable with similarly designed interventions in less self-selected populations, but this requires direct empirical study rather than extrapolation. Finally, our PMT improvement measure—defined as positive change in any of the four constructs—captures broad motivation development. Future work should explore more precise measurement of construct-level changes.

7 Conclusion

We presented BeSafe, an interactive educational visual-novel game grounded in Protection Motivation Theory and designed to cultivate both knowledge and protective mindsets for TFA detection and prevention. Through a survey study with 198 participants, we demonstrated that game-based learning achieves successful knowledge transfer and significant protection motivation development. Our findings reveal that narrative engagement, rather than gamification mechanics, drives these motivation shifts, suggesting that focused storytelling is more effective than game features for cultivating privacy and safety mindsets in sensitive educational domains. Participants with prior exposure to online abuse showed substantially greater gains, identifying personal relevance as a key design principle for TFA education. Participants expressed concrete intentions to review privacy settings, check linked devices, and share protective strategies with others. BeSafe demonstrates that proactive educational interventions can cultivate the threat awareness and self-efficacy necessary for users to recognize and respond to technology-facilitated abuse before significant harm occurs.

Acknowledgments

We thank the anonymous reviewers for their feedback. We acknowledge the use of Claude Sonnet 4.6 to assist with grammar and typo corrections. We acknowledge funding from Google Academic Research Award (GARA) 2024 grant and NSF #2339679.

References

[1] Wilfried Admiraal, Jantina Huizenga, Sanne Akkerman, and Geert Ten Dam. 2011. The concept of flow in collaborative game-based learning. *Computers in human behavior* 27, 3 (2011), 1185–1194.

[2] Majed Almansoori and Rahul Chatterjee. 2025. Can Social Media Privacy and Safety Features Protect Targets of Interpersonal Attacks? A Systematic Analysis. *Proceedings on Privacy Enhancing Technologies* 2 (2025), 326–343.

[3] Majed Almansoori, Andrea Gallardo, Julio Poveda, Adil Ahmed, and Rahul Chatterjee. 2022. A Global Survey of Android Dual-Use Applications Used in Intimate Partner Surveillance. *Proceedings on Privacy Enhancing Technologies* 4 (2022), 120–139.

[4] Majed Almansoori, Mazharul Islam, Saptarshi Ghosh, Mainack Mondal, and Rahul Chatterjee. 2024. The Web of Abuse: A Comprehensive Analysis of Online Resource in the Context of Technology-Enabled Intimate Partner Surveillance. In *2024 IEEE 9th European Symposium on Security and Privacy (EuroS&P)*. IEEE, 773–789.

[5] Majed Almansoori, Jessica Lam, Elias Fang, Kieran Mulligan, Adalbert Gerald Soosai Raj, and Rahul Chatterjee. 2020. How secure are our computer systems courses?. In *Proceedings of the 2020 ACM conference on international computing education research*. 271–281.

[6] Majed Almansoori, Jessica Lam, Elias Fang, Adalbert Gerald Soosai Raj, and Rahul Chatterjee. 2021. Textbook underflow: Insufficient security discussions in textbooks used for computer systems courses. In *Proceedings of the 52nd ACM technical symposium on computer science education*. 1212–1218.

[7] Majed Almansoori, Jessica Lam, Elias Fang, Adalbert Gerald Soosai Raj, and Rahul Chatterjee. 2023. Towards finding the missing pieces to teach secure programming skills to students. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 973–979.

[8] Adam L Alter, Daniel M Oppenheimer, Nicholas Epley, and Rebecca N Eyre. 2007. Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General* 136, 4 (2007), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

[9] Louis Bailey, Joanne Hulley, Tim Gomersall, Gill Kirkman, Graham Gibbs, and Adele D Jones. 2024. The Networking of Abuse: Intimate Partner Violence and the Use of Social Technologies. *Violence Against Women* 30, 3 (2024), 567–589.

[10] Ugur Bakan and Ufuk Bakan. 2018. Game-based learning studies in education journals: A systematic review of recent trends. *Actualidades Pedagógicas* 72 (2018).

[11] C Daniel Batson. 2009. These things called empathy: Eight related but distinct phenomena. *The social neuroscience of empathy* (2009), 3–15.

[12] Rosanna Bellini, Emily Tseng, Nora McDonald, Rachel Greenstadt, Damon McCoy, Thomas Ristenpart, and Nicola Dell. 2021. " So-called privacy breeds evil" Narrative Justifications for Intimate Partner Surveillance in Online Forums. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27.

[13] Daniel Boduszek, Agata Debowska, Adele D Jones, Mengtong Ma, Donna Smith, Dominic Willmott, Gill Kirkman, and Philip Hyland. 2019. Prosocial Video Game as an Intimate Partner Violence Prevention Tool Among Youth: A Randomised Controlled Trial. *Computers in Human Behavior* 93 (2019), 260–266.

[14] Galina Borschenko, Anna Rubtsova, and Olga Zhelezniakova. 2023. Visual Novels as a Means of Business Communication Skills Development for Computer Science Students. In *International Conference on Professional Culture of the Specialist of the Future*. Springer, 314–322.

[15] Willie Brown, Lei Zhang, Ellis Lawrence, Yuanwei Jin, Ibibia Dabipi, Weiwei Zhu, and Dinesh K Sharma. 2018. Undergraduate engineering education and the game-based learning methods to promote a culture of laboratory safety. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–7.

[16] Sandra Câmara Olim, Ana Andrade, Valentina Nisi, Pedro Campos, and Mara Dionísio. 2025. SeaStory: An Interactive Visual Novel to Raise Awareness About the Monk Seal. In *Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter*. 1–8.

[17] Dani Cavallaro. 2014. *Anime and the visual novel: narrative structure, design and play at the crossroads of animation and computer games*. McFarland.

[18] Anthony Chan. 2021. *Visual novels: video game-style storytelling as a multimodal medium for language learning and instruction*. Ph. D. Dissertation. University of Illinois at Urbana-Champaign.

[19] Rahul Chatterjee, Periwinkle Doerfler, Hadas Orgad, Sam Havron, Jackeline Palmer, Diana Freed, Karen Levy, Nicola Dell, Damon McCoy, and Thomas Ristenpart. 2018. The spyware used in intimate partner violence. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 441–458.

[20] Janet X Chen, Allison McDonald, Yixin Zou, Emily Tseng, Kevin A Roundy, Acar Tamersoy, Florian Schaub, Thomas Ristenpart, and Nicola Dell. 2022. Trauma-informed computing: Towards safer technology experiences for all. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–20.

[21] Clinic to End Tech Abuse. 2021. Computer security and privacy for survivors of intimate partner violence. https://www.ceta.tech.cornell.edu. Online; accessed 21 Feb 2021.

[22] Thomas M Connolly, Elizabeth A Boyle, Ewan MacArthur, Thomas Hainey, and James M Boyle. 2012. A systematic literature review of empirical evidence on computer games and serious games. *Computers & education* 59, 2 (2012), 661–686.

[23] Robert E Crossler, James H Long, Tina M Loraas, and Brad S Trinkle. 2014. Understanding compliance with bring your own device policies utilizing protection motivation theory: Bridging the intention-behavior gap. *Journal of Information Systems* 28, 1 (2014), 209–226.

[24] Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row.

[25] Edward L Deci and Richard M Ryan. 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry* 11, 4 (2000), 227–268.

[26] Louis Deslauriers, Logan S McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin. 2019. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences* 116, 39 (2019), 19251–19257.

[27] Connor Diemand-Yauman, Daniel M Oppenheimer, and Erikka B Vaughan. 2011. Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition* 118, 1 (2011), 114–118. https://doi.org/10.1016/j.cognition.2010.09.012

[28] DoD Cyber Exchange Public. n.d.. *Cyber Awareness Challenge*. https://www.cyber.mil/cyber-awareness-challenge

[29] Fatemeh Estebsari, Zahra Rahimi Khalifehkandi, Marzieh Latifi, Abdollah Farhadinasab, Parvaneh Vasli, and Davoud Mostafaie. 2023. Protection motivation theory and prevention of breast cancer: a systematic review. *Clinical Breast Cancer* 23, 4 (2023), e239–e246.

[30] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. 2013. *Statistical methods for rates and proportions*. john wiley & sons.

[31] Jennifer A. Fredricks, Phyllis C. Blumenfeld, and Alison H. Paris. 2004. School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research* 74, 1 (2004), 59–109.

[32] Diana Freed, Sam Havron, Emily Tseng, Andrea Gallardo, Rahul Chatterjee, Thomas Ristenpart, and Nicola Dell. 2019. " Is my phone hacked?" Analyzing Clinical Computer Security Interventions with Survivors of Intimate Partner Violence. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.

[33] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2018. "A Stalker's Paradise" How Intimate Partner Abusers Exploit Technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[34] Diana Freed, Jackeline Palmer, Diana Elizabeth Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2017. Digital technologies and intimate partner violence: A qualitative analysis with multiple stakeholders. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

[35] Scott Freeman, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences* 111, 23 (2014), 8410–8415.

[36] Rosemary Garris, Robert Ahlers, and James E Driskell. 2017. Games, motivation, and learning: A research and practice model. In *Simulation in aviation training*. Routledge, 475–501.

[37] James Paul Gee. 2003. *What video games have to teach us about learning and literacy*. Palgrave Macmillan.

[38] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

[39] Melanie C Green and Timothy C Brock. 2000. The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology* 79, 5 (2000), 701.

[40] Jennifer Ann's Group. n.d.. Video Games to Prevent Teen Dating Violence. https://jenniferann.org/games.htm. https://jenniferann.org/games.htm

[41] Erwin Gunawan, Citra Fragrantia Theodorea, Boy Muchlis Bachtiar, Ade Fadly H Masse, Gamaliel Juliadi Biring, and Handika Dwi Prasetyo. 2026. Visual novel game as a new breakthrough in dental health education media for stunted children in Indonesia. *Journal of Clinical Pediatric Dentistry* 50, 1 (2026), 245–254.

[42] Naman Gupta, Sanchari Das, Kate Walsh, and Rahul Chatterjee. 2024. A Critical Analysis of the Prevalence of Technology-Facilitated Abuse in US College Students. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 15, 12 pages. https://doi.org/10.1145/3613905.3652036

[43] MP Jacob Habgood and Shaaron E Ainsworth. 2011. Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences* 20, 2 (2011), 169–206.

[44] Sarah Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicolle Dell, and Thomas Ristenpart. 2019. Clinical computer security for victims of intimate partner violence. In *28th USENIX Security Symposium (USENIX Security 19)*. 105–122.

[45] Maurice Hendrix, Ali Al-Sherbaz, and Bloom Victoria. 2016. Game based cyber security training: are serious games suitable for cyber security training? *International Journal of Serious Games* 3, 1 (2016), 53–61.
[46] Anna S Irani. 2018. Positive altruism: helping that benefits both the recipient and giver. (2018).
[47] Shamila Janakiraman, Sunnie Lee Watson, and William R Watson. 2018. Using game-based learning to facilitate attitude change for environmental sustainability. *Journal of Education for Sustainable Development* 12, 2 (2018), 176–185.
[48] Johns Hopkins University Academic Support. 2021. Active Versus Passive Learning. https://academicsupport.jhu.edu/resources/study-aids/active-versus-passive-learning/. https://academicsupport.jhu.edu/resources/study-aids/active-versus-passive-learning/ Accessed: 2024.
[49] David A. Kolb. 1984. *Experiential Learning: Experience as the Source of Learning and Development.* Prentice-Hall.
[50] David R. Krathwohl. 2002. *A Revision of Bloom's Taxonomy: An Overview.* Vol. 41. Taylor & Francis. 212–218 pages.
[51] Kuo-Wei Kyle Lai and Hao-Jan Howard Chen. 2023. A comparative study on the effects of a VR and PC visual novel game on vocabulary learning. *Computer Assisted Language Learning* 36, 3 (2023), 312–345.
[52] Jessica Lam, Elias Fang, Majed Almansoori, Rahul Chatterjee, and Adalbert Gerald Soosai Raj. 2022. Identifying gaps in the secure programming knowledge and skills of students. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1*. 703–709.
[53] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.
[54] Noorlela Marcheta, Malisa Huzaifa, Alifa Salsabila Deshendra, Ariawan Andi Suhandana, Asep Kurniawan, and Muhammad Yusuf Bagus Rasyiidin. 2025. Developing a Visual Novel Game for Bullying Prevention in Adolescents. In *2025 8th International Conference of Computer and Informatics Engineering (IC2IE)*. IEEE, 1–6.
[55] Alison Marganski and Lisa Melander. 2018. Cyber aggression in intimate relationships: The roles of technology, behaviors, and gender. *Partner Abuse* 9, 3 (2018), 253–278.
[56] Robin L Nabi and Jessica Gall Myrick. 2019. Uplifting fear appeals: Considering the role of hope in fear-based persuasive messages. *Health communication* 34, 4 (2019), 463–474.
[57] National Network to End Domestic Violence, Safety Net Project. 2015. A Glimpse From the Field: How Abusers Are Misusing Technology. https://www.nnedv.org/safetynet. Original NNEDV report citing that 97% of U.S. domestic violence programs report survivors experiencing technology-facilitated abuse.
[58] Kristine Øygardslia, Charlotte Lærke Weitze, and Juhyung Shin. 2020. The educational potential of visual novel games: Principles for design. *Replaying Japan* 2, 2 (2020), 123–134.
[59] Stamatios Papadakis and Michail Kalogiannakis. 2019. Evaluating the effectiveness of a game-based learning approach in modifying students' behavioural outcomes and competence, in an introductory programming course. A case study in Greece. *International Journal of Teaching and Case Studies* 10, 3 (2019), 235–250.
[60] Jan L Plass, Bruce D Homer, and Charles K Kinzer. 2015. Foundations of game-based learning. *Educational psychologist* 50, 4 (2015), 258–283.
[61] Marc Prensky. 2001. *Digital game-based learning.* McGraw-Hill.
[62] A Treatment Improvement Protocol. 2014. Trauma-informed care in behavioral health services. *Rockville, USA: Substance Abuse and Mental Health Services Administration* (2014).
[63] Amalia Putri, Nurul Zia Aida, Renindra Atsilah Putri, Tazkia Syifa Arrahmah, and Dewi Kusrini. 2021. A3! visual novel game as an audio-visual media that motivates Japanese language learning. In *Fifth International Conference on Language, Literature, Culture, and Education (ICOLLITE 2021)*. Atlantis Press, 67–74.
[64] Beatrice Rammstedt and Oliver P John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality* 41, 1 (2007), 203–212.
[65] Michaela M Rogers, Colleen Fisher, Parveen Ali, Peter Allmark, and Lisa A Fontes. 2023. Technology-Facilitated Abuse in Intimate Relationships: A Scoping Review. *Trauma, Violence, & Abuse* 24, 4 (2023), 2210–2226.
[66] Ronald W Rogers. 1975. A protection motivation theory of fear appeals and attitude change1. *The journal of psychology* 91, 1 (1975), 93–114.
[67] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 1 (2000), 68–78.
[68] Teodor Sommestad, Jonas Hallberg, Kristoffer Lundholm, and Johan Bengtsson. 2014. Variables influencing information security policy compliance: A systematic review of quantitative studies. *Information Management & Computer Security* 22, 1 (2014), 42–75.
[69] Sophie Stephenson, Majed Almansoori, Pardis Emami-Naeini, and Rahul Chatterjee. 2023. "It's the Equivalent of Feeling Like You're in Jail": Lessons from Firsthand and Secondhand Accounts of IoT-Enabled Intimate Partner Abuse. In *32nd USENIX Security Symposium (USENIX Security 23)*. 105–122.
[70] Sophie Stephenson, Majed Almansoori, Pardis Emami-Naeini, Danny Yuxing Huang, and Rahul Chatterjee. 2023. Abuse Vectors: A Framework for Conceptualizing {IoT-Enabled} Interpersonal Abuse. In *32nd USENIX Security Symposium (USENIX Security 23)*. 69–86.
[71] F.Learning Studio. 2024. Gamified Cyber Security for Training: TOP 5 Best Examples to Watch in 2024. https://flearningstudio.com/gamified-cyber-security/. https://flearningstudio.com/gamified-cyber-security/
[72] Cynthia Taylor and Saheel Sakharkar. 2019. '); DROP TABLE textbooks;– An Argument for SQL Injection Coverage in Database Textbooks. In *Proceedings of the 50th ACM technical symposium on computer science education*. 191–197.
[73] Hsin-yi Sandy Tsai, Mengtian Jiang, Saleem Alhabash, Robert LaRose, Nora J Rifon, and Shelia R Cotten. 2016. Understanding online safety behaviors: A protection motivation theory perspective. *Computers & Security* 59 (2016), 138–150.
[74] Emily Tseng, Rosanna Bellini, Nora McDonald, Matan Danos, Rachel Greenstadt, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2020. The tools and tactics used in intimate partner surveillance: An analysis of online infidelity forums. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*. 1893–1909.
[75] Blair E Wisco, Brian P Marx, Denise M Sloan, Kaitlyn R Gorman, Andrea L Kulish, and Suzanne L Pineles. 2015. Self-distancing from trauma memories reduces physiological but not subjective emotional reactivity among veterans with posttraumatic stress disorder. *Clinical Psychological Science* 3, 6 (2015), 956–963.
[76] Kim Witte. 1992. Putting the fear back into fear appeals: The extended parallel process model. *Communications Monographs* 59, 4 (1992), 329–349.
[77] Pieter Wouters, Christof Van Nimwegen, Herre Van Oostendorp, and Erik D Van Der Spek. 2013. A meta-analysis of the cognitive and motivational effects of serious games. *Journal of educational psychology* 105, 2 (2013), 249.
[78] Jie Chi Yang, Yi Lung Lin, and Yi-Chun Liu. 2017. Effects of locus of control on behavioral intention and learning performance of energy knowledge in game-based learning. *Environmental Education Research* 23, 6 (2017), 886–899.
[79] Langxuan Yin, Lazlo Ring, and Timothy Bickmore. 2012. Using an interactive visual novel to promote patient empowerment through engagement. In *Proceedings of the International Conference on the foundations of digital Games*. 41–48.
[80] Javad Zare. 2023. The impact of L2 learners' altruistic teaching on their task engagement in English essay writing. *System* 118 (2023), 103137.
[81] Chun Zhou, Aurora Occa, Soyoon Kim, and Susan Morgan. 2020. A meta-analysis of narrative game-based interventions for promoting healthy behaviors. *Journal of health communication* 25, 1 (2020), 54–65.

A Appendix

B Bloom's Taxonomy for Learning

Bloom's Taxonomy [50] classifies learning outcomes into six hierarchical levels: ❶ *Remembering* (lowest level): The learner can recall information and concepts. ❷ *Understanding*: The learner can explain acquired information in their own words. ❸ *Applying*: The learner can use learned knowledge in new situations. ❹ *Analyzing*: The learner can identify patterns and draw connections between concepts. ❺ *Evaluating*: The learner can critically assess information and defend opinions based on criteria. ❻ *Creating*: The learner can use their knowledge to produce new ideas.

C Story Scenarios

C.1 Scenario 1: Instagram

Setup. Alex's friend Emma sets up his first Instagram account, which he begins using normally to follow friends and share posts.

Threat and identification. Alex notices that some of his posts have been deleted, friends have been removed, and unfamiliar posts have appeared on his account. He consults his tech-savvy friend John, who diagnoses account compromise. Alex's first instinct is to change his password; John explains why this alone is insufficient and walks through where to find evidence of unauthorized access.

Mitigation. The scenario walks Alex through Instagram's Settings → Account Center → Password and Security flow to review and

remove logged-in devices, with embedded screenshots of the real Instagram and Meta UI.

Pedagogical focus. The scenario distinguishes password change (which prevents future unauthorized access) from session removal (which terminates current unauthorized access)—a distinction that the other compromise scenarios reinforce.

C.2 Scenario 2: WhatsApp

Setup. Alex works in a typical office environment. One morning, he receives a notification on his phone but decides to check it later; when he does, he finds nothing unusual and assumes he had imagined it.

Threat and identification. A coworker later confronts Alex about a WhatsApp message he did not send. Alex initially dismisses the incident as a moment of inattention, then proposes changing his device passcode. His friend Robert clarifies that device-level passcode protection does not remove a WhatsApp session that has already been linked from another device. Together they review the Linked Devices panel and identify an unrecognized macOS session—a device Alex does not own.

Mitigation. Alex removes the unauthorized linked device through WhatsApp's Linked Devices panel, with the corresponding interface elements shown alongside the dialogue.

Pedagogical focus. The scenario teaches that device-level security (passcode) does not address account-level compromise once another device has been linked. It also models the QR-linking attack vector that requires only brief physical access to an unlocked phone.

C.3 Scenario 3: iCloud

Setup. Alex's partner gifts her a new iPad—her first Apple device, which she now uses for everyday tasks.

Threat and identification. Alex notices unfamiliar images appearing in her photo library. She consults her brother John, who works through and rules out several plausible-sounding but incorrect explanations—a software bug in an installed app and a virus—before identifying iCloud account access as the actual vector. Alex asks whether changing her password would resolve the issue; John explains that password change alone does not remove a device whose iCloud trust has already been established.

Mitigation. The scenario walks through Apple ID → iCloud → device list, removing the suspicious device via the "Remove from Account" action. It also covers related cloud hygiene (Find My settings, sync controls), with player agency preserved on whether to disable location features.

Pedagogical focus. The scenario separates malware-style and platform-glitch explanations from account-level compromise and teaches that an unrecognized device signed into one's iCloud account may persist access to synced data—photos, messages, backups—even after a password change.

C.4 Scenario 4: Gmail

Setup. Alex's brother Mike created Alex's Gmail account years ago, and it remains his primary email. Mike teases Alex about pranking him—a setup that will mislead Alex when his account behaves suspiciously.

Threat and identification. Alex first notices that he has stopped receiving expected emails. Afterward, a coworker confronts him about an email he did not send. Alex initially attributes the unfamiliar message to his brother, assuming it is a prank; his friend John corrects this attribution and identifies actual account compromise. Together they review the account's connected devices to confirm.

Mitigation. The scenario walks through `myaccount.google.com` → device management, identifying and signing out unrecognized sessions. The real Google Account UI is embedded.

Pedagogical focus. The scenario trains threat-attribution discipline: examining evidence such as devices and login activity rather than defaulting to the most familiar explanation. It also introduces multiple compromise indicators (unauthorized outgoing actions *and* missing expected activity).

C.5 Scenario 5: Snapchat

Setup. Alex is one month past her divorce and feeling that she is now free of an unhealthy marriage. She attends her sister's wedding and resumes daily routines.

Threat and identification. On her way home from the wedding, Alex's ex-husband appears at a grocery store and attempts to engage her, framing the encounter as coincidence. Alex initially accepts this framing. The next day, the same ex appears at a café where she is meeting a friend—the repetition makes the "coincidence" frame untenable. Alex and her friend Sarah work through how he could have known her location, eventually identifying Snap Map as the surveillance vector.

Mitigation. The scenario presents two valid paths to Snapchat's location privacy controls (Settings menu and direct map access) and walks through enabling Ghost Mode or scoping Last Active Location visibility. The real Snapchat UI is embedded.

Pedagogical focus. The scenario teaches pattern recognition across incidents rather than single-event detection, modeling the gradual recognition process documented in prior TFA research [33]. It also introduces location-mediated stalking as a distinct threat from account compromise.

C.6 Scenario 6: Android

Setup. Alex's partner gives her a new Android phone as a gift—her first Android device, which becomes her primary phone.

Threat and identification. An argument escalates and the partner explicitly declares he will monitor her phone. Alex is left wondering whether her phone has been "bugged." She consults her brother John, who first checks her installed apps for spyware—and finds nothing suspicious. The phone, however, is not safe: in Settings →

Passwords & Accounts, a second Google account that Alex does not recognize is connected to the device, providing the partner persistent access to her synced data.

Mitigation. The scenario walks through removing the unknown account from the device. Notably, the helper warns that account removal will be visible to the account's owner—introducing operational-security considerations under ongoing surveillance. The scenario closes with durable advice (set a passcode, do not leave the phone unobserved). Figure 6 shows panels from this scenario.

Pedagogical focus. The scenario decouples app-layer security ("is the app safe?") from account-layer and device-layer security ("is the device safe?"). It is also the only scenario where the abuser declares the surveillance before it is discovered, complementing Snapchat's inferred-from-pattern detection.

D Supplementary Materials

Game question	Code	#
Do you wish certain features to be added to the game?	Improve feedback and learning support	7
	Improve usability and polish	16
	Add more interactive/gameplay content	31
I want to know more about the knowledge taught	Feels like they know enough and don't need more	26
	Curious and motivated to learn more	93
I will try to apply the knowledge I learned from the game in real world	Application depends on context/relevance	6
	Barriers to applying knowledge	13
	Will apply or already applying knowledge	91
Is there a specific digital security and privacy topic you'd like to learn about?	Online risks & scams	9
	Data & browsing privacy	11
	Social media	12
	Account & device security	21
Did you face any issue while playing the game?	Content issue	1
	Language or clarity problems	1
	Technical/usability issues	4
The game grabs my attention	Learning-related appreciation	52
	Unengaging or unsuitable design	24
	Engaging and enjoyable game	126
The game increases my knowledge	Struggled to retain or understand	2
	Knowledge was already known or obvious	16
	Game-design makes understanding the content easier	19
	Game succesfully teaches or reinforces knowledge	107

Figure 5: Summary of open codes per game question.

Screenshots of the game. We present a few screenshots from one of the stories in Fig. 6 to demonstrate the scenarios used in the game. We also present an example of the questions players encounter and the branching paths that follow based on their choices (Fig. 7).

Survey questions. The survey questions can be found on the following anonymous github: https://anonymous.4open.science/r/BeSafe-7114/

Codebook. We present themes from open-ended responses in Fig. 5 for the game.

Secondary analyses. We used one-way ANOVA to examine the effect of relationship status on learning gain and retention. The analyses showed no statistically significant effects on either learning gain ($p = 0.107$, $\eta^2 = 0.053$) or retention at one week ($p = 0.19$, $\eta^2 = 0.099$). Individuals may have limited awareness of the specific risks involved in online abuse regardless of relationship status, which could diminish the influence of relationship context on engagement with the learning material.

We also examined whether frequency of internet use was associated with learning gains or retention using one-way ANOVA. The analyses showed no statistically significant effects on either learning gain ($p = 0.122$, $\eta^2 = 0.029$) or retention ($p = 0.192$, $\eta^2 = 0.044$). While these findings suggest little relationship based on our current measure, we note that internet usage was assessed using a single question on frequency. Internet usage encompasses more than frequency alone, and a more nuanced understanding—capturing patterns, purposes, and contexts of use—may yield richer insights. Future research should explore how internet usage influences learning, retention, and awareness in greater depth.

E Open Science

The BeSafe game is publicly accessible at https://besafegame.vercel.app/ to enable continued use by educators, advocates, and individuals seeking privacy and safety education. Following publication, we will release the game source code under an open license to facilitate content updates by tech clinics and other organizations.

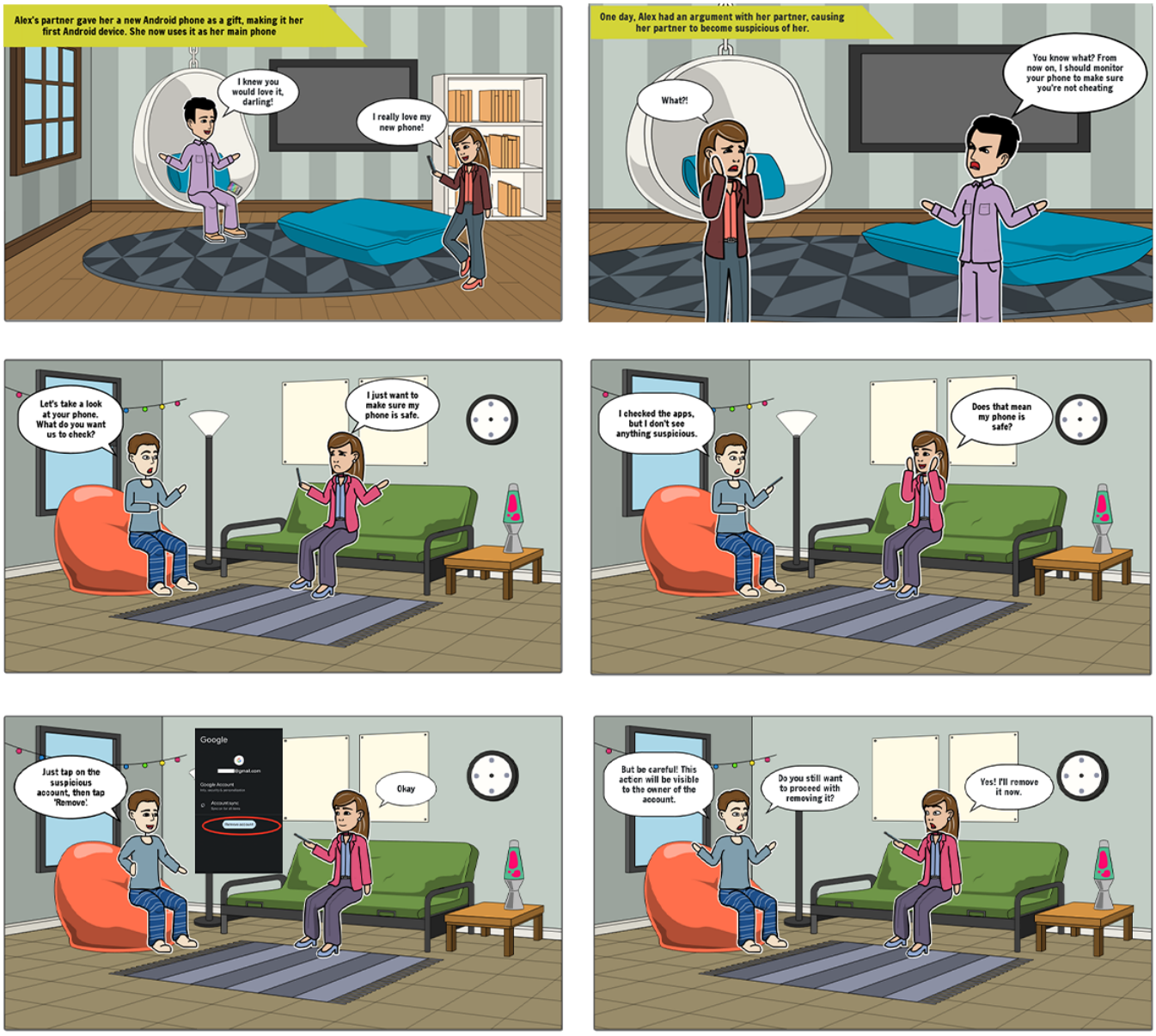

Figure 6: Screenshots from Android's story showing the the background phase (top), identification and investigation phase (center), and mitigation phase (bottom).

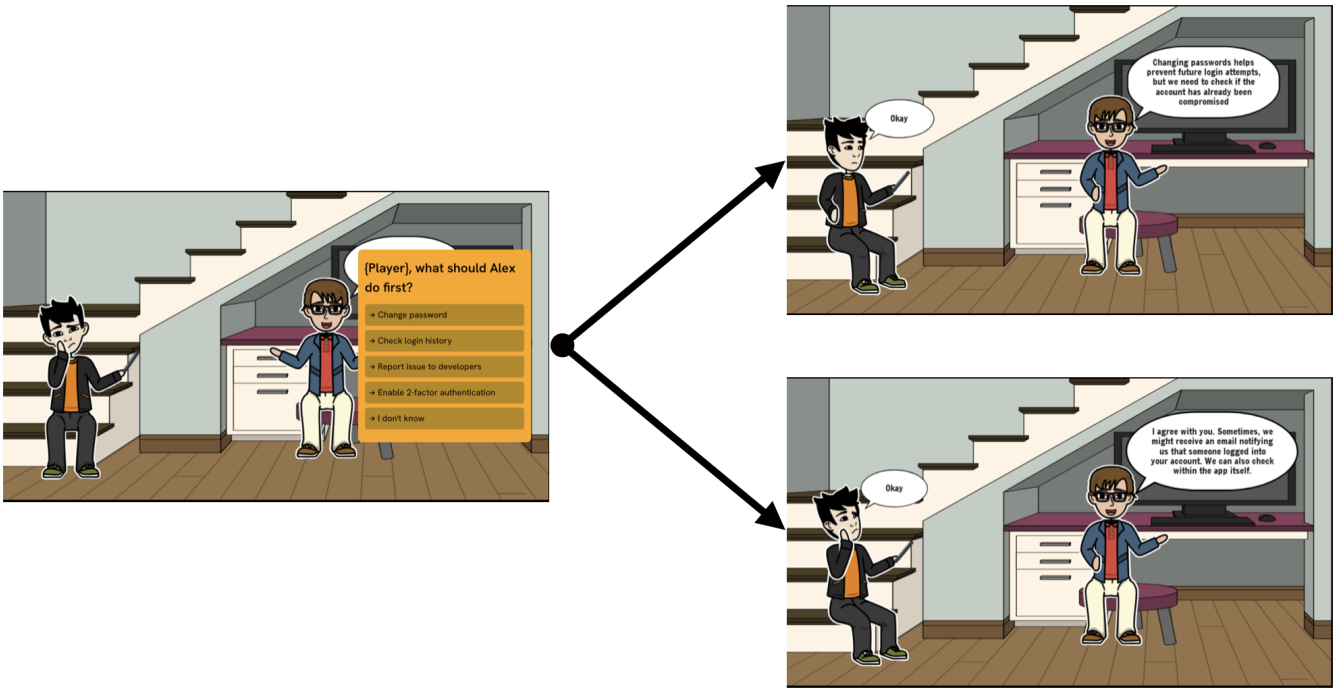

Figure 7: Example of branching gameplay with supportive feedback. On the left, the game presents a decision point where the player selects from multiple courses of action. On the right, two resulting narrative branches illustrate how different choices lead to distinct paths. Notably, even incorrect selections receive constructive feedback designed to guide learning rather than penalize the player.