Toward Adaptive Privacy-Enhancing Training: A Longitudinal Study of How Personality Shapes Responsiveness to Information Security Awareness Training

Ofir Cohen
Ben-Gurion University of the Negev
cofir@post.bgu.ac.il

Asaf Shabtai
Ben-Gurion University of the Negev
shabtaia@bgu.ac.il

Rami Puzis
Ben-Gurion University of the Negev
puzis@bgu.ac.il

Abstract

Human error remains the primary vector for privacy breaches, with users frequently exposing sensitive personal data through unsafe behaviors, such as granting excessive permissions, neglecting screen locks, or falling for social engineering. Yet, existing information security awareness (ISA) programs often adopt a one-size-fits-all approach, neglecting users' personality traits and risk perception. In this work, we conducted a comprehensive five-week study with 105 participants to evaluate how personality traits and Passive Risk-Taking (PRT) interact with different ISA training methodologies. Prior research has primarily examined direct associations between personality traits and self-reported ISA measures, producing mixed findings and providing limited insight into how individual differences shape responsiveness to ISA training interventions. To address these limitations, we used a longitudinal sensor-based framework that captures real-world security behaviors over time, enabling a more reliable assessment of ISA training outcomes. Our results show that individual differences moderate training efficacy: users with high Agreeableness achieved greater gains through active-risk training (social engineering simulations), whereas those with lower Agreeableness benefited more from passive-risk training targeting omission-based risks. Furthermore, we find that Agreeableness moderates the relationships between both baseline PRT and changes in PRT and subsequent ISA improvements, while Conscientiousness moderates how resource-related PRT translates into ISA gains. Overall, these findings show that personality traits shape how users benefit from different training strategies and influence the alignment between risk reduction and awareness gains, supporting a shift toward adaptive, privacy-enhancing ISA training.

Keywords

Information Security Awareness, Personality Traits, User Study, Privacy, Human Factors

1 Introduction

In the modern digital ecosystem, the individual user is the final guardian of their own privacy. Today, the majority of privacy violations occur through human error, such as granting dangerous permissions, clicking on phishing links that harvest credentials, or failing to secure physical devices [23, 38]. For example, Verizon's 2025 data breach investigation report found that 60% of breaches were caused by human error [43]. Additionally, the Sophos 2024 threat report noted an exponential increase in attacks targeting mobile users, affecting not only individuals but also small businesses, where compromises directly jeopardize the privacy of employees and customers [41]. The vulnerabilities associated with the human factor remain a persistent threat, making information security awareness (ISA) training a key measure to empower individuals to protect their privacy or the privacy of others.

However, most training programs adopt a one-size-fits-all design, delivering uniform content without accounting for individual differences in user personality or risk perception [2, 7, 10, 46]. This lack of personalization limits their ability to influence behavior and reduce privacy vulnerability effectively. While recent work has identified the necessity for tailored training that accounts for diverse learners' profiles, it primarily focused on skill-based personalization [37], leaving the role of personality traits as a predictor of training efficacy largely underexplored.

Unlike skill-based profiles, which reflect a user's current knowledge or technical ability and may change through training, personality traits represent relatively stable psychological tendencies that influence how individuals perceive risk, process information, and respond to persuasive interventions. Among the widely used personality frameworks, the Big Five model characterizes individuals across five continuous dimensions, including traits such as Conscientiousness (self-discipline and impulse control) and Agreeableness, which reflects tendencies toward trust and cooperation. Because these traits shape behavioral decision-making, they may also influence how users respond to different forms of ISA training.

Prior research has explored how individual differences, particularly personality traits and risk-taking tendencies, relate to ISA [7, 13, 15, 35, 39]. However, most prior studies have focused on direct associations between these traits and ISA, often relying on self-reported measures that may introduce subjectivity and reporting biases [33], and have shown limited correlation with real-world security behavior [4]. As a result, existing work provides limited insight into how individual differences influence responsiveness to ISA training interventions. This highlights the need for longitudinal, objective evaluations of ISA that can capture training-driven improvements in real-world security behavior.

To further understand behavioral factors underlying insecure actions, recent work has examined the role of passive risk-taking (PRT), the tendency to incur risk through inaction, as a complementary behavioral construct associated with insecure behavior [3]. However, this relationship has been studied primarily in static settings and treats PRT as a single aggregate construct. As a result, it remains unclear how different factors of PRT contribute to safe

Proceedings on Privacy Enhancing Technologies 2026(4), 1072–1087

https://doi.org/10.56553/popets-2026-0160

behavior and whether changes in specific risk-taking tendencies are meaningfully aligned with improvements in awareness.

In this work, we address these gaps by jointly examining personality traits, PRT, and ISA in a longitudinal, objective setting. Using the ConGISATA framework [9], we monitored university students' real-world cybersecurity behaviors over five weeks while participating in one of three ISA training strategies targeting different forms of risk. ISA was assessed using a combination of sensor-derived behavioral measurements and security challenges spanning multiple cybersecurity domains. In parallel, participants completed assessments of Big Five personality traits and PRT, enabling examination of how individual differences relate to behavioral improvement following training. To obtain a more fine-grained understanding of omission-based risk, we further analyze PRT at the level of its underlying factors rather than as a single aggregate construct.

While our study explores a broad set of relationships between personality, PRT, and ISA, in this paper we focus on the research questions that yield the most robust and actionable insights, with additional exploratory analyses reported in the appendix. The central research questions examined in this work are:

(1) **RQ1:** How do personality traits influence the effectiveness of different ISA training strategies?
(2) **RQ2:** How do personality traits moderate the relationships between PRT and changes in ISA?
    (a) **RQ 2.1:** Which personality traits moderate the relationship between baseline PRT and ISA changes?
    (b) **RQ 2.2:** Which personality traits moderate the relationship between PRT changes and ISA changes?

Our findings show that personality traits shape how users respond to ISA training, with moderation effects varying across both training strategy and PRT factor. Most consistently, Agreeableness emerged as a central moderator of ISA improvement: individuals with higher Agreeableness benefited more from active-risk training, whereas those with lower Agreeableness benefited more from passive-risk training. Agreeableness also moderated the relationship between both baseline PRT and PRT changes and subsequent ISA improvement, particularly for the ethical dimension of PRT.

In addition, Conscientiousness moderated the relationship between resource-related PRT and ISA gains, suggesting that users characterized by stronger self-discipline and impulse control may translate certain risk perceptions into ISA improvement differently than other users.

Together, these findings suggest that ISA improvement is shaped not only by training content, but also by stable individual differences and distinct forms of risk-taking behavior. This highlights the potential value of adaptive ISA training approaches that account for personality-related differences rather than relying solely on uniform interventions.

2 Background

This section provides the theoretical and methodological foundations for our study. We introduce the psychological constructs that inform our personalization approach, review the distinction between active and passive risk-taking, and describe the objective ISA measurement and training framework used in our experiments; related work is discussed separately in Section 3.

2.1 The Big Five Personality Model

The Big Five personality model is one of the most widely accepted frameworks for describing stable individual differences in behavior and cognition [12, 29]. Unlike typological approaches that categorize individuals into discrete types, the Big Five conceptualizes personality as five continuous dimensions, enabling fine-grained and quantitative analysis of behavioral variation. This dimensional structure is particularly suitable for statistical modeling, as it allows trait intensity to be directly associated with differences in decision-making and security-related behaviors. Consequently, the Big Five has become the de facto standard in behavioral and cybersecurity research [13, 15, 28, 39], facilitating comparability with prior work.

The model consists of five core personality traits that capture the most fundamental aspects of an individual's personality. The five traits, often referred to by their acronym, OCEAN, were described by John *et al.* work [19] as follows:

(1) Openness to Experience - "Openness describes the breadth, depth, originality, and complexity of an individual's mental and experiential life."
(2) Conscientiousness - "Conscientiousness describes the socially prescribed impulse control that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks."
(3) Extraversion - "Extraversion implies an energetic approach toward the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality."
(4) Agreeableness - "Agreeableness contrasts a prosocial and communal orientation towards others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty."
(5) Neuroticism - "Neuroticism contrasts emotional even-temperedness and stability with negative emotionality, such as feeling anxious, nervous, sad, and tense."

To measure these traits, we use the Big Five Inventory (BFI) [18, 19], a validated 44-item questionnaire where participants rate agreement with short trait-descriptive statements. The instrument yields continuous scores for each trait, enabling quantitative analysis.

Personality traits are particularly relevant for the design of ISA interventions. While temporal or situational indicators, such as time of day or location, may optimize the timing of an intervention, they provide limited guidance regarding the most effective content or framing strategy [20]. In contrast, personality traits represent stable psychological constructs that shape how individuals process information, perceive risk, and respond to persuasive messages [30]. Leveraging these stable differences enables training systems to move beyond simple temporal nudges toward persuasion profiling, adapting the framing of security warnings and educational content to users' intrinsic motivations and decision-making styles [17, 31]. This motivates our use of personality measures to inform the design and evaluation of ISA training strategies.

2.2 Active Versus Passive Risk-Taking

Risk-taking behavior is generally categorized into two distinct constructs: active and passive risk. Keinan *et al.* [21] conceptualized this dichotomy based on the psychological distinction between acts of commission and acts of omission. Active risk-taking involves the

deliberate initiation of a behavior that introduces danger (commission), such as speeding or gambling. In contrast, passive risk-taking (PRT) refers to the failure to act or the neglect of safety measures that result in exposure to danger (omission), such as failing to wear a seatbelt or ignoring a warning sign. Keinan *et al.*further reported that these two forms of risk are perceived differently, even when the potential consequences are identical. Specifically, they argued that individuals tend to feel less accountable for their omissions than for their commissions.

This lack of perceived responsibility reduces the motivation to act, causing individuals to underestimate threats that require proactive intervention. Consequently, passive risks are often invisible to the observer due to an "attentional gap": a cognitive bias where the absence of an event (e.g., not updating software) captures less attention than a tangible action [22]. However, when personal accountability is emphasized, individuals process information more deeply, bridging this attentional gap and increasing their willingness to engage in preventive behavior.

In the context of information security, Arend *et al.* [3] demonstrated that this psychological distinction is a critical predictor of user vulnerability. Their work found that individuals with high tendencies toward PRT were significantly less likely to adopt secure behaviors. This suggests that cybersecurity negligence is often not a deliberate choice to be unsafe (active risk), but rather a failure to recognize the necessity of action (passive risk), highlighting the need for training interventions that specifically target this "omission bias."

2.3 ISA Assessment: From Self-Reports to Objective Sensor-Based Measurement

2.3.1 Traditional ISA Assessment Methods and the KAB Model. Traditionally, ISA has been measured primarily through self-reported questionnaires and surveys. These instruments are generally grounded in the Knowledge-Attitude-Behavior (KAB) model, which posits that a user's security posture is a function of what they know (Knowledge), their internal dispositions toward security policies (Attitude), and their subsequent actions (Behavior). Under this framework, the primary objective of ISA measurement is to quantify these dimensions to evaluate and predict actual security behavior.

A prominent example of applying this theoretical framework is the Human Aspects of Information Security Questionnaire (HAIS-Q) [32]. HAIS-Q operationalizes the KAB framework by eliciting knowledge, attitude, and (perceived) behavior separately. This granular separation is designed to allow researchers to analyze each component independently and identify specific drivers of risk. However, because HAIS-Q and similar scales like the Security Behavior Intentions Scale (SeBIS) [11] rely on users to self-evaluate, they are limited to measuring security intentions rather than objective actions. Recognizing that earlier scales often failed to accurately predict behavior, recent advancements have introduced higher-coverage instruments. Notably, the Extended Security Behavior Scale (ESBS) [36] utilizes a substantially wider range of security-related factors, demonstrating significantly higher predictive accuracy for actual user behaviors compared to prior psychometric iterations.

Table 1: The focus and sub-focus areas of the ISA taxonomy.

Focus Area	Sub-Focus Area	Abbreviation
Applications	Application Installation	AI
	Application Handling	AH
Browsing & Communication	Browser	B
	Virtual Communication	VC
Communication Channels	Accounts	A
	Networks	N
	Physical Channels	PC
Device	Operating System	OS
	Security Systems	SS

2.3.2 Limitations of Self-Reporting. While these instruments provide a scalable method for gathering data, they suffer from significant methodological limitations. They are prone to 'social desirability bias', where respondents over-report positive behaviors to appear security-conscious [33]. Recent work further shows that this bias is domain-specific in security and privacy contexts, motivating the development of dedicated measurement scales to quantify it [1].

Most critically, while advanced instruments like the ESBS have demonstrated improved predictive capabilities [36], broader empirical comparisons between self-reports and objective data have found weak or non-existent correlations [4]. This suggests that self-reporting primarily captures a user's perception of their security rather than their actual technical vulnerability. To address these gaps and more accurately capture the 'Behavior' component of the KAB model as it occurs in practice, there is an evident need for objective, granular assessment.

2.3.3 The ISA Taxonomy. To overcome limitations of self-reports, we utilize the taxonomy proposed by Bitton *et al.* [5]. This taxonomy classifies ISA into specific, observable criteria based on technological focus areas and three psychological dimensions: knowledge, attitude, and behavior.

The taxonomy is structured hierarchically. Each focus area is divided into sub-focus areas, which are further composed of specific security topics. For instance, the "Browsing and Communications" focus area is divided into "Browser" and "Virtual Communication" sub-focus areas. Under "Virtual Communication," the security topic "Malicious Hyperlinks" intersects with the behavioral dimension to form a specific, measurable criterion: *Does not click on suspicious hyperlinks.*

The framework comprises four focus areas, nine sub-focus areas, and 30 distinct criteria, as detailed in Table 1 and Table 2.

We adopted this taxonomy as the foundation for our work for two primary reasons.

First, it offers objective granularity. It was the only taxonomy identified in our review that translates abstract security concepts

Table 2: Criteria for assessing ISA.

ID	Criterion
AI1	Downloads apps from trusted sources
AI2	Does not install apps that require dangerous permissions
AI3	Does not install apps with a low rating
AI4	Rarely installs apps that require root privileges
AH1	Regularly updates apps
AH2	Rarely clicks on advertisements
AH3	Properly manages running/installed apps
B1	Does not enter malicious domains
B2	Prefers to download files via HTTPS
B3	Does not send sensitive information via HTTP
B4	Does not insert private information into pop-ups
B5	Does not use untrusted certificates
VC1	Does not open messages received from unknown senders
VC2	Does not click on links received from unknown senders
A1	Updates passwords regularly, uses unguessable and diverse passwords, stores passwords safely
A2	Uses two-factor authentication mechanisms
A3	Uses password management services
OS1	Uses an updated OS
OS2	Does not root or jailbreak the device
SS1	Uses embedded security systems
SS2	Uses anti-virus application regularly to scan the device
SS3	Updates security systems
SS4	Operates in accordance with security alerts
SS5	Uses PIN code, pattern, or fingerprint
N1	Does not connect to unencrypted networks
N2	Does not download files on unencrypted networks
N3	Uses VPN services on public networks
N4	Does not transmit private data via unencrypted channels
PC1	Disables connectivity when not in use
PC2	Does not connect the device to unknown devices

into clear, binary-measurable criteria (e.g., a user either installs an app with dangerous permissions, or they do not). This contrasts with other taxonomies that remain at a high conceptual level, making them difficult to operationalize for sensor-based assessment.

Second, although the taxonomy was originally developed in the context of mobile security, it is grounded in broader ISA principles and extends a widely adopted ISA prototype [25]. Importantly, many of its criteria, such as handling untrusted links, protecting sensitive information, or responding appropriately to security warnings, are not inherently mobile-specific and naturally generalize to broader cybersecurity contexts. More specialized criteria, such as rooting or jailbreaking, represent concrete instances of broader security principles such as privilege escalation and unsafe system modification. This broader behavioral structure makes the taxonomy well-suited for evaluating objective ISA across diverse cybersecurity domains while still enabling rigorous and operationalizable measurement [4, 8, 9, 40].

2.4 The Assessment and Training Framework: ConGISATA

To conduct the reliable and detailed ISA assessment required for this study, we utilize the ConGISATA (Continuous Gamified Information Security Awareness Training and Assessment) framework [9]. ConGISATA was chosen as the experimental framework for two primary reasons:

(1) **Objective Behavioral Reliability:** ConGISATA assesses ISA through behavior monitoring. By utilizing mobile sensors to continuously verify adherence to the ISA taxonomy criteria in real-world settings, the framework provides reliable, objective metrics of ISA rather than a subjective reflection of intent.
(2) **Focus on Passive Risk:** As discussed in Section 2.2, passive risks (omissions) are critical yet often overlooked. ConGISATA is distinct in its ability to assess ISA with emphasis on PRT in the cybersecurity domain.

ConGISATA assesses ISA with regard to the two types of risks:

- **Passive ISA Score:** The passive ISA score represents the tendency for cybersecurity PRT. The application uses background sensors to passively monitor adherence to the taxonomy criteria (e.g., checking if the device is rooted or if the screen lock is enabled) without user intervention.
- **Active ISA Score:** The active ISA score represents the tendency for cybersecurity active risk-taking. The app periodically presents simulated social engineering challenges, such as phishing attempts, impersonation scenarios, and permission requests, to assess the user's reaction to risky situations.

The final ISA score is the average of these passive and active metrics, providing a comprehensive "Ground Truth" for our analysis.

Beyond assessment, ConGISATA implements a unique training method designed to address the psychological "attentional gap" associated with passive risks (as discussed in Section 2.2). The framework aims to transform passive risks (omissions) into active ones by employing a penalty mechanism that imposes immediate point deductions for passive failures. This approach forces users to acknowledge their lack of action, thereby increasing accountability and motivation.

Because ConGISATA represents a distinct training approach, we utilize its training module as one of the three training methods examined in this study. This allows us to isolate how effective this unique strategy is across different personality profiles, compared to the other training methods detailed later in Section 4.

3 Related Work

Prior work has shown that the effectiveness of ISA training depends on multiple human and contextual factors rather than a one-size-fits-all design. Several studies highlight that users' current skill levels, entrenched habits, and situational cues substantially influence learning outcomes. For example, recent phishing interventions group users by baseline proficiency rather than demographic characteristics, demonstrating that skill-based tailoring can equalize post-training performance across learners [37]. Complementary findings from organizational psychology emphasize the role of intentional forgetting, suggesting that replacing insecure routines requires actively suppressing obsolete behaviors and their environmental cues rather than merely introducing new knowledge [16]. Similarly, embedded phishing training benefits from timely nudges and reminders, indicating that improvements often arise from behavioral reinforcement in addition to informational content [26].

Collectively, these studies suggest that effective ISA training should account for diverse learner characteristics, including proficiency, habits, and behavioral context. Alongside these dynamic and situational factors, stable individual differences may also shape how users perceive risks and respond to interventions. Among these, personality traits have received growing attention as a potential factor influencing cybersecurity behavior and phishing susceptibility.

In recent years, researchers have increasingly examined the relationship between personality traits and cybersecurity-related behaviors. Most studies have focused on ISA measures collected through self-reported questionnaires [13, 14, 24, 28, 34, 39, 45]. Across these studies, conscientiousness has been associated with higher ISA, while additional and sometimes conflicting associations have been reported for the other traits and risk-taking tendencies.

Other work examined personality in the context of phishing susceptibility using simulated phishing attacks [7, 15, 35]. These studies similarly suggest that personality traits may influence phishing-related decision making, although findings vary across studies and evaluation methods. Notably, Halevi *et al.* [15] reported that conscientiousness, despite being positively associated with self-reported ISA in prior work, did not correspond to improved phishing performance, highlighting the importance of objective behavioral measures in ISA research.

A smaller body of work has examined personality in the context of phishing training interventions. Sumner *et al.* [42] investigated anti-phishing training effectiveness using DISC personality traits and reported differences in phishing susceptibility both before and after training. Their results showed that some personality groups consistently demonstrated higher phishing-detection accuracy and confidence levels, while other groups remained comparatively more susceptible even following training. The authors also observed variation in the magnitude of improvement across personality groups, suggesting that stable individual differences may influence responsiveness to phishing interventions.

Taken together, prior work suggests that personality traits may influence both cybersecurity behavior and responsiveness to training interventions. However, existing studies have primarily focused on phishing-specific tasks or static self-reported ISA measures, providing limited visibility into how individual differences relate to longitudinal behavioral change.

Recent work has also identified PRT as an important behavioral factor underlying insecure behavior [3]. Prior studies found that higher PRT tendencies are associated with lower adoption of secure behaviors, but typically examined this relationship in static settings using cybersecurity intentions or limited proxy behavioral measures collected in short-term experimental contexts. In addition, PRT has generally been treated as a single aggregate construct, leaving it less understood how its underlying factors relate to behavioral change over time.

Motivated by these gaps, our work adopts a broader perspective on ISA. We evaluate longitudinal changes in objective ISA behaviors across multiple cybersecurity domains and training strategies using a combination of challenge-based and sensor-derived measurements. This enables us to examine how personality traits and different dimensions of PRT interact with behavioral improvement following ISA training.

4 Method

In this section, we describe our methodology for investigating the relationship between personality, PRT, and training responsiveness.

4.1 Experimental Design

The experiment follows a three-phase design:

(1) **Pre-training (Week 1):** Baseline measurements are established. Subjects complete personality and PRT questionnaires, and their baseline ISA is assessed.
(2) **Training (Weeks 2-5):** Subjects are randomly assigned to one of three experimental groups. Each group receives a different type of ISA training for four weeks (see Table 3).
(3) **Post-training (End of Week 5):** A final assessment of PRT tendencies is conducted to measure changes.

4.2 Measures

The primary independent variables (IVs) are: (i) training method (three different strategies), (ii) personality traits (Big Five), (iii) time (longitudinal progression across weeks), and (iv) baseline ISA/PRT levels (pre-training). The dependent variables (DVs) are changes in PRT and ISA scores (pre- vs. post-training), as well as weekly ISA performance trajectories. We operationalize these constructs using validated psychological scales and objective behavioral metrics, as detailed below.

4.2.1 Big Five Personality Traits. Subjects' personality traits are assessed using the Big Five Inventory (BFI) [19], as detailed in Section 2.1. The questionnaire is administered only at the start of the experiment (Pre-training). We do not administer a post-training personality test, as personality traits are defined psychologically as stable, enduring constructs that are unlikely to change over a four-week intervention period [6]. Reliability is verified using Cronbach's alpha for each of the five traits. These traits serve as individual-difference predictors and potential moderators of training effectiveness.

4.2.2 Passive Risk-Taking (PRT). To measure the psychological tendency toward passive risk, we use the PRT scale developed by Keinan *et al.* [21] (discussed in Section 2.2). Unlike personality, risk perception is malleable; therefore, this instrument is administered both pre- and post-training to quantify changes in risk perception. The 25-item scale covers three factors: Resource risks (12 items), Medical risks (7 items), and Ethical risks (6 items).

To validate this approach within our sample, internal consistency is assessed for both the individual factors and the overall PRT scale using Cronbach's alpha. Individual mean scores for each factor are also computed for both pre- and post-training data. Pre-training PRT scores are treated as predictors, while pre-post differences are used as outcome measures reflecting changes in PRT tendencies following training.

4.2.3 Active Risk and ISA Assessment. We utilize the ConGISATA framework [9] as discussed in Section 2.4 to capture objective behavioral data continuously throughout the five weeks. Three distinct metrics are computed:

Table 3: Experimental design: Training groups, measures, and assessment tools.

Group	Pre-training Measures			ISA Training Method	ISA Assessment Method		Post-training PRT
	PRT	BFI	Baseline ISA		Active	Passive	
1 - Active Risk	✓	✓	✓	**Active Risk Based:** Personalized based on each participant's failures in social engineering simulations.	ConGISATA's attack simulations	ConGISATA's sensor-based behavior monitoring	✓
2 - Passive Risk	✓	✓	✓	**Passive Risk Based:** Personalized based on each participant's sensor-detected behavioral omissions.			✓
3 - Generic	✓	✓	×	**Generic:** Non-Personalized content delivered on a pre-determined schedule.	None	None	✓

- **Passive ISA Score:** A real-time metric derived from sensor data monitoring compliance with the ISA taxonomy (e.g., OS updates, encryption usage).
- **Active ISA Score:** A performance-based metric derived from the user's success or failure in handling simulated social engineering attacks. This score serves as our ground-truth measure for Active Risk-Taking.
- **Overall ISA:** The average of the passive and active scores.

Learning rates are calculated by tracking the week-over-week change in these scores. These metrics constitute the primary dependent variables for evaluating both overall improvement and longitudinal learning rates.

4.3 Training Interventions

A core component of this study is the comparison of different training strategies (active-risk, passive-risk, and generic). While the strategy was assigned at the group level for experimental comparison, the specific educational content delivered to each participant was selected adaptively at the individual level based on their ongoing performance. The training logic varies as follows:

- **Group 1 - Active Risk Training:** Training content is dynamically selected for each participant based on their active-risk-related performance. Subjects are challenged with social engineering attack simulations, using ConGISATA's active ISA assessment. The teaching modules presented are based on the subjects' performance in these challenges. For example, a subject that fails in a permission attack receives a teaching module about permissions management. This group only uses the ConGISATA ISA assessment module and does not use the training module.
- **Group 2 - Passive Risk Training:** Training content is dynamically selected for each participant based on their passive-risk-related performance. Subjects' behavior is monitored to assess their passive ISA with ConGISATA's passive ISA assessment. The framework identifies the areas in which the subject needs to improve, and teaching modules on that topic are presented to the subject. This group uses ConGISATA for both ISA assessment and ISA training.
- **Group 3 - Generic Training:** This group receives generic training that is not personalized. The teaching modules are presented in a pre-determined, scheduled manner. This group is not using ConGISATA at all, neither for assessment nor training.

4.4 Data Preprocessing, Validation, and Analysis

Internal Consistency Validation: Cronbach's alpha is used to validate the reliability of each Big Five trait and PRT factor for both the pre- and post-intervention questionnaires.

Normality Assessment: Normality is assessed for all variables using the Shapiro-Wilk and Kolmogorov-Smirnov tests. These tests are conducted both across the full sample and separately within each experiment's group to assess the success of randomization. Variables that are not normally distributed within any group are flagged for further consideration.

As part of validating the experimental setup and verifying successful random group assignment, we conduct between-group comparisons on personality traits and PRT factors using both one-way ANOVA and the non-parametric Kruskal-Wallis test. Baseline ISA scores are not compared across groups because the ConGISATA framework standardizes scores within each group (using z-scores) before the training begins. This ensures that subsequent differences reflect the training intervention rather than baseline variability.

Regression Modeling: Linear regression models are fitted to examine the predictive value of the training method and personality traits on ISA improvements. p-values for interaction terms were adjusted using the FDR procedure. Where appropriate, effect sizes (Cohen's d) are computed to quantify the magnitude of the training effects. The ISA and PRT scores and personality trait measures were mean-centered prior to analysis. Centering was applied to facilitate interpretation of the regression coefficients, as the original scales do not include meaningful zero values, making uncentered intercepts and main effects difficult to interpret.

5 Evaluation and Results

In this section, we present the empirical findings of our longitudinal study. We begin by describing participant recruitment and validating the experimental design, including the reliability of the measurements. We then address the primary research questions, while additional exploratory analyses are provided in Appendix A.

5.1 Participants Recruitment and Ethical Principles

We initially recruited 105 undergraduate and graduate students who regularly use smartphones; their ages ranged from 21 to 31 ($M = 25, Mdn = 26$). We aimed for approximately 30 participants per group, a commonly used threshold in experimental research. Under the Central Limit Theorem, samples of this size yield approximately normal sampling distributions of the mean, supporting the use of parametric statistical tests even when the underlying measures are not perfectly normally distributed. Consequently, for our three-group design, we set a recruitment target of at least 90 participants. We recruited 105 participants (35 per group) to exceed this threshold and account for potential withdrawals or technical data loss.

Participation was strictly voluntary. Students were recruited through emails and were explicitly informed, both verbally and in writing, that they could withdraw at any time without any academic or personal repercussions. Participants received monetary compensation for their time. Several students chose to withdraw during the study, and their data were excluded, confirming that withdrawal was feasible and free of consequences.

The study procedures, consent materials, and data handling practices were reviewed and approved by the Institutional Review Board (IRB). The consent form described the nature of the intervention, including the use of messages, app notifications, and alerts.

To safeguard participants' privacy and minimize any associated risks, the study adhered to the following principles:

(1) **Voluntary Participation and Informed Consent**: Participants were informed in advance about the study procedures, the types of data collected, and their right to withdraw at any time without penalty or loss of compensation.
(2) **Data Minimization**: Only non-content metadata was collected and transmitted (e.g., counts of emails containing URLs or binary security indicators such as whether two-factor authentication was enabled). No email or SMS content was stored or transmitted. The data that was collected is detailed in Table 4.
(3) **Data Security**: All data transmissions between the mobile application and the server were encrypted.
(4) **Server Security**: The server was hosted within the university's secure domain, with restricted access and institutional security controls.
(5) **Safety of Attack Simulations**: In phishing simulations, authentication credentials were neither transmitted to the server nor stored.

5.2 Experimental Procedure and Tasks

Upon recruitment, participants were randomly assigned to one of three groups, as detailed in Table 3. Prior to the intervention, all participants filled out the Big Five Inventory (BFI) and the pre-training Passive Risk-Taking (PRT) questionnaire. Participants in Groups 1 and 2 installed the ConGISATA mobile application and had their behavior monitored using the sensors embedded in the app. The criteria and the way they were assessed are presented in Table 4. Group 3 received training via email and did not use any app. The experiment spanned five weeks: a one-week Calibration period (baseline measurement without intervention) followed by a four-week training Phase. At the conclusion of the fifth week, all participants completed the post-training PRT questionnaire to assess shifts in risk perception.

To assess active risk-taking and to calculate the active ISA score, participants in groups 1 and 2 were presented with three types of weekly social engineering attacks. Participants received one challenge of each type per week, resulting in three challenges every week and a total of 15 challenges. Three types of simulated social engineering attacks were used:

Table 4: List of criteria assessed for the experiment.

Criterion	Means of assessment
AI1: Downloads apps from trusted sources	An app was considered trusted if it was downloaded from an official app store (such as Google Play). The score for this criterion is the number of untrusted apps found on the subject's device (lower is better).
AI2: Does not install apps that require dangerous permissions	The score for this criterion is the number of apps on the subject's device which require dangerous permissions, as classified by Android (lower is better).
AI3: Does not install apps with a low rating	We considered a low rating to be less than three and a half stars (out of five) in the Google Play store. The score for this criterion is the number of apps with a low rating found on the subject's device (lower is better).
AH1: Regularly updates apps	Google Play features the last date on which an app was updated. The score for this criterion is the number of apps that are not up-to-date found on the subject's device (lower is better).
AH3: Properly manages running/installed apps	An app is considered unused if the subject did not use the app for more than two weeks. The score for this criterion is the number of unused apps found on the subject's device (lower is better).
B1: Does not enter malicious domains	A domain is considered malicious if Google's safebrowsing API has classified it as such. The score for this criterion is the number of malicious domains the subject has entered in the last seven days (lower is better).
VC1: Does not open messages received from unknown senders	We monitored two message inboxes for each subject - SMS and Gmail's spam inbox. An SMS is considered to be from an unknown sender if the sender of the SMS is not in the subject's contact list. The SMS score is the percentage of how many unknown SMSs the subject has opened in the last seven days. Likewise, the Gmail score is the percentage of emails classified as spam by Gmail that were opened in the last 30 days. The final score for this criterion is the average of the SMS and Gmail scores (lower is better).
VC2: Does not click on links received from unknown senders	We considered an event to be of the 'clicking on links received from unknown senders' type if the following three conditions were met: (1) the subject opened a message from an unknown sender, as defined in VC1, (2) the message that was opened contained a URL, and (3) we also identified a transition between the SMS/Gmail apps and the browser app (Google Chrome), suggesting the subject has clicked on that URL. The score for this criterion is the number of times a subject has clicked on URLs from unknown senders in the last seven days (lower is better).
A2: Uses two-factor authentication mechanisms	A subject was considered to be using two-factor mechanisms if either a two-factor authentication app or an SMS (from the last seven days) indicating two-factor use was found on their device. The score for this criterion is one if the subject uses two-factor mechanisms and otherwise zero (higher is better).
A3: Uses password management services	The subject was considered to be using password management services if a password-managing app was found on their device. The score for this criterion is one if the subject uses password management services and otherwise zero (higher is better).
OS2: Does not root or jailbreak the device	We used a dedicated Android package (rootBeer) that implements various heuristics to determine whether or not a device is rooted. The score for this criterion is one if the subject has not rooted the device and otherwise zero (higher is better).
SS2: Uses anti-virus application regularly to scan the device	The score for this criterion is one if the subject has an anti-virus app installed on the device and otherwise zero (higher is better).
SS5: Uses PIN code, pattern, or fingerprint	A device was considered secured if a lock-screen was enabled. The score for this criterion is one if the subject's device is secured and otherwise zero (higher is better).
N1: Does not connect to unencrypted networks	A network was considered encrypted if a security protocol was enabled (such as WPS, WPA2). The score for this criterion is the number of unencrypted networks the subject has connected to in the last seven days (lower is better).
N3: Uses VPN services on public networks	The subject was considered to be using VPN services if a VPN app was found on their device. The score for this criterion is one if the subject uses VPN services and otherwise zero (higher is better).
PC1: Disables connectivity when not in use	The score for this criterion is the number of times in the last seven days that a connectivity channel (i.e., Bluetooth, Wi-Fi, NFC) was enabled for more than five minutes, without being connected (lower is better).

- **Phishing:** Participants received emails from familiar aliases (e.g., "Student Administration") mimicking university notifications (e.g., grades, password changes). These emails contained links to spoofed, non-HTTPS domains attempting to solicit credentials.
- **Impersonation:** The app generated deceptive push notifications mimicking legitimate apps (e.g., Facebook, Instagram). Clicking the notification launched a fake login screen to test if the user could distinguish the application source.
- **Permission Attacks:** The system simulated runtime permission requests where a benign-looking app (e.g., Calculator) requested unnecessary dangerous permissions (e.g., Camera or SMS access).

5.3 Descriptive Statistics

While 105 participants were recruited, data completeness varied across measures. Participants were excluded from computation of the results based on several criteria; the criteria included technical issues or non-engagement with the framework and incomplete survey submissions. ISA scores for the participants in group 3 were not available, because their training was administered through scheduled email modules rather than the ConGISATA framework, which meant that sensor-based behavior monitoring and the presentation of diverse attack simulations were not possible. Consequently, group 3 subjects were omitted from all ISA-related analyses. After applying these exclusion criteria, 102 participants who completed all three questionnaires (pre-training PRT, post-training PRT, and BFI) were included in longitudinal analyses involving PRT change and personality traits, and 55 participants were included in the analyses involving ISA.

The mean scores and standard deviations for the Big Five personality traits ($N = 102$) were within the expected ranges: extraversion (3.48 ± 0.68), agreeableness (4.00 ± 0.45), conscientiousness (3.83 ± 0.61), neuroticism (2.57 ± 0.65), and openness (3.55 ± 0.56).

The average PRT scores, measured on a 1-to-5 Likert scale, were 3.11 (± 0.56) before training ($N = 105$) and 3.00 (± 0.63) after training ($N = 100$), yielding a mean change of -0.13 (± 0.50) ($N = 100$), indicating a general decrease in participants' willingness to take passive risks after training. It is important to note the inverse scoring of our primary metrics. ISA scores are performance-based (higher is better), whereas PRT scores are risk-based (lower is better). Consequently, a negative correlation between ISA and PRT represents a positive outcome, indicating that as security awareness rises, risk-taking tendencies decline.

In every following analysis, we first assessed the normality of each key variable within each experimental group using the Shapiro-Wilk test. Most personality traits and PRT factors were normally distributed within the groups, with a few exceptions. To ensure that the random assignment to groups resulted in statistically equivalent groups at baseline, we conducted both one-way ANOVA and Kruskal-Wallis tests for all personality traits and pre-training PRT measures. As shown in Table 5, no trait significantly differed between the groups.

5.4 Reliability of Questionnaires

Internal consistency was evaluated using Cronbach's alpha. Personality traits showed acceptable to strong reliability: extraversion ($\alpha = 0.842$), agreeableness (0.674), conscientiousness (0.805), neuroticism (0.808), and openness (0.741). For the overall 25-item PRT scale, internal consistency was $\alpha = 0.679$ pre-training and $\alpha = 0.78$ post-training. Reliability for the individual PRT factors was as follows: resource risks $\alpha = 0.601$ (pre-training) and 0.715 (post-training); medical risks $\alpha = 0.606$ (pre) and 0.709 (post); and ethical risks $\alpha = 0.333$ (pre) and 0.380 (post). While a lower Cronbach's alpha was obtained for the ethical risks factor, this is consistent with the values reported in the original PRT validation study [21].

Table 5: Between-group comparisons of personality traits and pre-training PRT scores.

	One-way ANOVA		Kruskal-Wallis	
Variable	F	p_{FDR}	H	p_{FDR}
Agreeableness	0.71	.871	1.14	.726
Conscientiousness	0.84	.871	1.38	.726
Extraversion	0.87	.871	1.33	.726
Neuroticism	0.43	.871	2.35	.726
Openness	3.30	.368	6.49	.351
Ethical PRT	0.42	.871	2.27	.726
Medical PRT	0.10	.907	0.38	.825
Overall PRT	0.22	.899	1.89	.726
Resource PRT	0.39	.871	0.84	.739

Legend: F: ANOVA F-statistic; H: Kruskal-Wallis H-statistic; p_{FDR}: FDR corrected p-value.

Table 6: Spearman correlation matrix for Big Five personality traits. Significant relationships are bolded.

Trait 1	**Trait 2**	ρ	p_{FDR}
Agreeableness	Conscientiousness	.196	.060
Agreeableness	Extraversion	.201	.060
Agreeableness	Neuroticism	-.228	**.035**
Agreeableness	Openness	.103	.335
Conscientiousness	Extraversion	.254	**.033**
Conscientiousness	Neuroticism	-.318	**.011**
Conscientiousness	Openness	.067	.506
Extraversion	Neuroticism	-.229	**.035**
Extraversion	Openness	.249	**.033**
Neuroticism	Openness	-.245	**.033**

Legend: ρ: Spearman's rank correlation coefficient; p_{FDR}: FDR corrected p-value.

5.4.1 Correlations Between Big Five Personality Traits. To examine the relationships between the Big Five personality traits in our sample, we computed pairwise correlations after filtering out participants who did not complete the questionnaire. Several traits are not normally distributed; therefore, we used Spearman's rank correlation. Table 6 shows the internal correlations between the Big Five traits across all participants in the user study.

As expected, based on the literature, neuroticism was negatively correlated with conscientiousness, openness, and agreeableness; extraversion was positively correlated with both conscientiousness and openness. Overall, the observed correlations largely aligned with known Big Five patterns, serving as validation [44].

Table 7: Regression results examining personality traits as moderators of the relationship between training method and ISA change.

			Main Effects		Interaction (Group × Trait)					
Outcome	**Moderator Trait**	R^2	B_{Grp}	B_{Trt}	B_{Int}	SE	t	95% CI	p_{fdr}	Partial f^2
Active ISA Score Change	Agreeableness	.210	3.355	25.318	-58.972	16.243	-3.63	[−91.58, −26.36]	**.003**	.259
	Conscientiousness	.006	2.773	2.147	-1.418	11.557	-0.12	[−24.62, 21.78]	.903	.000
	Extraversion	.037	2.799	5.603	-13.880	10.435	-1.33	[−34.83, 7.07]	.473	.035
	Neuroticism	.015	3.208	-2.999	7.348	10.019	0.73	[−12.77, 27.46]	.583	.011
	Openness	.026	2.457	2.495	-11.622	12.463	-0.93	[−36.64, 13.40]	.583	.017
Overall ISA Score Change	Agreeableness	.196	3.051	11.056	-29.680	8.843	-3.36	[−47.43, −11.93]	**.008**	.221
	Conscientiousness	.014	3.019	-1.020	0.167	6.215	0.03	[−12.31, 12.64]	.979	.000
	Extraversion	.033	2.693	2.303	-5.939	5.647	-1.05	[−17.27, 5.40]	.496	.022
	Neuroticism	.018	2.996	-0.471	2.953	5.400	0.55	[−7.89, 13.79]	.734	.006
	Openness	.040	2.585	2.763	-7.923	6.677	-1.19	[−21.33, 5.48]	.496	.028
Passive ISA Score Change	Agreeableness	.028	2.747	-3.205	-0.388	7.961	-0.05	[−16.37, 15.59]	.961	.000
	Conscientiousness	.061	3.266	-4.187	1.752	4.965	0.35	[−8.21, 11.72]	.931	.002
	Extraversion	.018	2.587	-0.998	2.003	4.658	0.43	[−7.35, 11.35]	.931	.004
	Neuroticism	.029	2.785	2.057	-1.441	4.397	-0.33	[−10.27, 7.39]	.931	.002
	Openness	.028	2.713	3.031	-4.225	5.503	-0.77	[−15.27, 6.82]	.931	.012

Legend: $\mathbf{R}^2$: coefficient of determination; $\mathbf{B}_{Grp}$ and $\mathbf{B}_{Trt}$: main effects of group and trait; $\mathbf{B}_{Int}$, **SE**, **t**, and **95% CI**: interaction coefficient, standard error, t-statistic, and confidence interval; $\mathbf{p}_{fdr}$: FDR-adjusted p-value; **Partial** $\mathbf{f}^2$: Cohen's effect size. Bold values indicate $\mathbf{p}_{fdr} < .05$.

5.5 RQ1: How Do Personality Traits Moderate the Effectiveness of Different ISA Training Strategies?

To examine the extent to which personality traits moderate the effectiveness of different ISA training interventions, we analyzed the interaction between training methodology (Active vs. Passive) and the Big Five traits in predicting ISA score changes. The analysis included 55 participants who completed the BFI and were assigned to either the Active-risk (Group 1) or Passive-risk (Group 2) groups.

As detailed in Table 7, Agreeableness was found to be a differentiating factor for training effectiveness. For both Overall and Active ISA changes, the interaction between Group and Agreeableness was significant, with a large partial effect size.

The results, illustrated in Figure 1, indicate that the association between training method and ISA gains differed depending on a participant's level of Agreeableness. Specifically, higher Agreeableness was associated with greater gains in the active-risk training (Group 1), whereas lower Agreeableness was associated with greater gains in the passive-risk training (Group 2). These findings suggest that Agreeableness can serve as a differentiating trait for personality-informed training assignments.

Beyond the statistically significant Agreeableness effects, the confidence intervals also revealed several informative non-significant trends. In particular, Extraversion showed a predominantly negative confidence interval ([−34.83, 7.07]) for the interaction effect on Active ISA change, suggesting a possible tendency for individuals with higher Extraversion to benefit more from the active-risk training than from the passive-risk training. Openness exhibited a similar, though weaker and more uncertain, directional trend ([−36.64, 13.40]). In contrast, Conscientiousness showed confidence intervals that were substantially centered around zero, suggesting little evidence for a meaningful moderating effect in this context. Although these effects did not reach statistical significance, the directional structure of the confidence intervals suggests that these trends warrant further examination in future work.

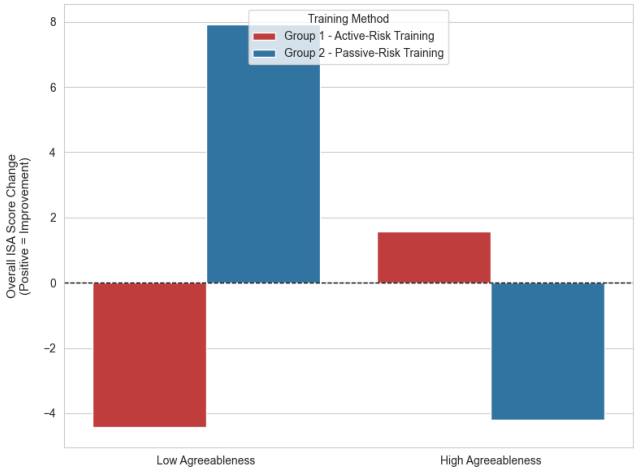

Figure 1: Mean change in Overall ISA scores by Agreeableness level (Median Split). Low-agreeableness users benefit from passive-risk training, while high-agreeableness users show a preference for active-risk training.

5.6 RQ2: How Do Personality Traits Moderate the Relationships Between PRT and Changes in ISA?

To better understand the relationship between omission-based risk tendencies and ISA improvement, we analyze PRT from a longitudinal perspective. Specifically, we examine (i) which personality traits moderate the relationship between baseline PRT and subsequent ISA changes, and (ii) which personality traits moderate the relationship between changes in PRT and changes in ISA over time.

5.6.1 RQ 2.1: Which Personality Traits Moderate the Relationship Between Baseline PRT and ISA Changes? To examine how baseline

PRT relates to changes in ISA after training, we analyzed the association between initial PRT levels and ISA improvement while accounting for individual differences. The analysis included participants who completed the first PRT and were assigned to either the Active-risk (Group 1) or Passive-risk (Group 2) training groups ($N = 55$).

Table 8 presents the regression results across PRT factors and personality traits. The findings suggest that moderation effects depend on the specific factor of PRT, with significant interactions observed only for the Resource PRT factor.

Specifically, significant interaction effects were observed between Resource PRT and both Agreeableness and Conscientiousness in predicting changes in Active ISA. In addition, Conscientiousness significantly moderated the relationship between Resource PRT and Overall ISA change. No significant interaction effects were observed for Medical or Ethical PRT factors.

The direction of these effects suggests that the relationship between baseline PRT and ISA gains depends on the underlying personality trait. For Agreeableness, the negative interaction coefficient indicates that higher levels of this trait attenuate the influence of Resource PRT on Active ISA improvement. In contrast, the positive interaction coefficients for Conscientiousness suggest that higher levels of this trait strengthen the association between Resource PRT and ISA gains.

Beyond the statistically significant findings, the confidence intervals also revealed several informative non-significant trends. In particular, the interaction between Resource PRT and Neuroticism in predicting active ISA change showed a negative confidence interval ($[-34.04, -1.47]$), while the corresponding confidence interval for overall ISA change was largely negative ($[-16.95, 0.78]$), suggesting a potential tendency for higher Neuroticism to attenuate the relationship between Resource PRT and subsequent ISA improvement. Conversely, the interaction between Resource PRT and Agreeableness in predicting passive ISA change exhibited a positive confidence interval ($[0.35, 28.22]$), suggesting that higher Agreeableness may strengthen the relationship between Resource PRT and passive ISA improvement. In contrast, many of the remaining personality traits and PRT factors exhibited confidence intervals that were substantially centered around zero, suggesting little evidence for consistent moderation effects in those cases.

Overall, these results indicate that baseline PRT does not uniformly translate into ISA improvement, but instead interacts with specific personality traits and risk factors.

5.6.2 RQ 2.2: Which Personality Traits Moderate the Relationship Between PRT Changes and ISA Changes? To further examine the relationship between PRT and ISA training, we analyzed whether changes in PRT are associated with corresponding changes in ISA, and how this relationship varies across individuals. The analysis included participants who completed both pre- and post-training PRT assessments, and were assigned to either the Active-risk (Group 1) or Passive-risk (Group 2) training groups ($N = 55$).

Table 9 presents the regression results examining whether personality traits moderate the relationship between changes in PRT factors and changes in ISA outcomes.

Agreeableness emerged as a significant moderator in the relationship between changes in Ethical PRT and changes in both Active and Overall ISA scores. The interaction effects are negative, indicating that the association between reductions in Ethical PRT and improvements in ISA is weaker for individuals with higher levels of Agreeableness.

Importantly, these moderation effects were not observed uniformly across all PRT factors. Rather, significant interactions were only observed within the Ethical PRT sub-construct, while no statistically significant moderation effects were found for Resource or Medical PRT sub-constructs across ISA outcomes.

Beyond the statistically significant findings, the confidence intervals also revealed several informative non-significant trends that were directionally consistent with the broader moderation patterns observed earlier.

In particular, several interaction effects exhibited confidence intervals skewed in the negative direction, suggesting potential attenuation effects in the relationship between reductions in PRT and improvements in ISA. For Active ISA change, the interaction between Medical PRT change and Agreeableness showed a largely negative confidence interval ($[-42.02, 8.28]$), while the interaction between Ethical PRT change and Extraversion exhibited a similar negative trend ($[-36.69, 8.10]$). Likewise, the interaction between Ethical PRT change and Neuroticism in predicting Passive ISA improvement showed a largely negative confidence interval ($[-13.13, 1.72]$). These trends align with the statistically significant negative interaction effects observed for Agreeableness in Ethical PRT, where higher Agreeableness weakened the association between reductions in Ethical PRT and improvements in ISA.

In contrast, the interaction between Ethical PRT change and Extraversion in predicting Passive ISA improvement exhibited a primarily positive confidence interval ($[-2.49, 16.72]$), suggesting that higher Extraversion may strengthen the alignment between reductions in Ethical PRT and improvements in Passive ISA. Many other trait-factor combinations exhibited confidence intervals centered around zero.

Overall, these findings suggest that the alignment between reductions in PRT and improvements in ISA depends on individual differences and that this effect is not uniformly observed across all factors of PRT. The observed confidence-interval trends further suggest that additional personality-specific moderation effects may exist and warrant further examination in future work.

6 Discussion and Conclusions

This study aims to advance the design of human-centric privacy defenses by moving beyond the "one-size-fits-all" paradigm of ISA training. By combining longitudinal, behavior-based measures of ISA with validated psychological constructs, we demonstrate that individual differences, particularly personality traits and PRT tendencies, play a central role in shaping training effectiveness.

RQ1: Personality Informing Training Method Selection. Across our analyses, a pattern emerges in which personality traits shape how users respond to different training methods. Specifically, Agreeableness differentiated responsiveness to the two training methods: higher Agreeableness was associated with greater gains in active-risk training, whereas lower Agreeableness was associated with greater gains in passive-risk training. This pattern is reflected in the regression results, where Agreeableness exhibited a relatively

Table 8: Regression results examining personality traits as moderators of the relationship between PRT factors and ISA change.

				Main Effects			Interaction (PRT × Trait)					
Outcome	**PRT Factor**	**Moderator**	R^2	B_{Grp}	B_{PRT}	B_{Trt}	B_{Int}	SE	t	95% CI	p_{fdr}	Partial f^2
Active ISA Score Change	**Resource PRT**	Agreeableness	.123	6.328	-5.524	-10.338	-38.979	15.569	-2.50	[−70.25, −7.71]	**.045**	.125
		Conscientiousness	.118	3.657	-4.160	2.701	18.805	7.677	2.45	[3.38, 34.23]	**.045**	.120
		Extraversion	.022	2.805	-3.416	-1.166	5.592	7.882	0.71	[−10.24, 21.42]	.602	.010
		Neuroticism	.098	4.653	-5.376	-1.839	-17.753	8.108	-2.19	[−34.04, −1.47]	.055	.096
		Openness	.018	2.436	-2.836	-3.925	-2.976	9.483	-0.31	[−22.02, 16.07]	.755	.002
	Medical PRT	Agreeableness	.019	2.430	-3.001	-2.284	-0.185	10.457	-0.02	[−21.19, 20.82]	.986	.000
		Conscientiousness	.055	3.663	-2.332	2.471	6.521	4.731	1.38	[−2.98, 16.02]	.557	.038
		Extraversion	.048	3.107	-1.530	-1.490	7.360	5.962	1.23	[−4.62, 19.33]	.557	.031
		Neuroticism	.025	2.226	-3.102	-2.170	-2.211	4.351	-0.51	[−10.95, 6.53]	.876	.005
		Openness	.027	1.805	-2.864	-4.089	-2.825	7.301	-0.39	[−17.49, 11.84]	.876	.003
	Ethical PRT	Agreeableness	.028	5.050	2.604	-0.157	14.779	17.083	0.87	[−19.53, 49.09]	.782	.015
		Conscientiousness	.019	2.533	2.255	2.236	4.071	8.288	0.49	[−12.58, 20.72]	.782	.005
		Extraversion	.055	2.406	4.027	0.203	13.050	8.678	1.50	[−4.38, 30.48]	.695	.045
		Neuroticism	.015	3.513	3.438	-1.457	-0.996	6.915	-0.14	[−14.89, 12.89]	.886	.000
		Openness	.021	3.565	2.689	-2.748	-5.111	10.178	-0.50	[−25.55, 15.33]	.782	.005
Overall ISA Score Change	**Resource PRT**	Agreeableness	.069	4.011	-2.865	-5.737	-12.348	8.661	-1.43	[−29.74, 5.05]	.267	.041
		Conscientiousness	.148	3.594	-2.740	-0.399	11.001	4.072	2.70	[2.82, 19.18]	**.047**	.146
		Extraversion	.029	2.685	-2.133	-0.770	2.459	4.238	0.58	[−6.05, 10.97]	.706	.007
		Neuroticism	.083	3.681	-2.946	-0.079	-8.088	4.413	-1.83	[−16.95, 0.78]	.182	.067
		Openness	.025	2.611	-1.821	-1.442	-1.273	5.101	-0.25	[−11.52, 8.97]	.804	.001
	Medical PRT	Agreeableness	.025	2.668	-1.036	-2.447	1.009	5.626	0.18	[−10.29, 12.31]	.902	.001
		Conscientiousness	.064	3.781	-0.664	-0.514	3.896	2.542	1.53	[−1.21, 9.00]	.659	.047
		Extraversion	.022	2.625	-0.877	-0.658	1.408	3.261	0.43	[−5.14, 7.96]	.902	.004
		Neuroticism	.021	2.686	-0.990	-0.096	-0.995	2.353	-0.42	[−5.72, 3.73]	.902	.004
		Openness	.021	2.292	-1.211	-1.200	0.489	3.953	0.12	[−7.45, 8.43]	.902	.000
	Ethical PRT	Agreeableness	.061	4.355	1.930	-1.189	11.052	9.062	1.22	[−7.15, 29.25]	.633	.030
		Conscientiousness	.034	3.154	1.991	-0.732	1.874	4.439	0.42	[−7.04, 10.79]	.914	.004
		Extraversion	.054	2.792	2.745	0.157	5.420	4.688	1.16	[−4.00, 14.84]	.633	.027
		Neuroticism	.028	3.300	2.371	-0.062	-0.402	3.707	-0.11	[−7.85, 7.04]	.914	.000
		Openness	.029	3.149	2.274	-0.674	0.725	5.470	0.13	[−10.26, 11.71]	.914	.000
Passive ISA Score Change	**Resource PRT**	Agreeableness	.109	1.694	-0.205	-1.136	14.283	6.939	2.06	[0.35, 28.22]	.224	.085
		Conscientiousness	.079	3.531	-1.320	-3.500	3.197	3.467	0.92	[−3.77, 10.16]	.902	.017
		Extraversion	.018	2.566	-0.851	-0.374	-0.674	3.490	-0.19	[−7.68, 6.34]	.919	.001
		Neuroticism	.033	2.708	-0.516	1.681	1.578	3.711	0.43	[−5.88, 9.03]	.919	.004
		Openness	.019	2.785	-0.806	1.041	0.430	4.188	0.10	[−7.98, 8.84]	.919	.000
	Medical PRT	Agreeableness	.038	2.906	0.929	-2.610	2.202	4.575	0.48	[−6.99, 11.39]	.790	.005
		Conscientiousness	.071	3.900	1.004	-3.500	1.271	2.073	0.61	[−2.89, 5.43]	.790	.008
		Extraversion	.076	2.143	-0.224	0.174	-4.543	2.596	-1.75	[−9.76, 0.67]	.431	.061
		Neuroticism	.036	3.147	1.122	1.978	0.220	1.912	0.12	[−3.62, 4.06]	.909	.000
		Openness	.049	2.779	0.442	1.688	3.803	3.189	1.19	[−2.60, 10.21]	.597	.028
	Ethical PRT	Agreeableness	.056	3.659	1.256	-2.221	7.324	7.441	0.98	[−7.62, 22.27]	.824	.019
		Conscientiousness	.071	3.775	1.727	-3.699	-0.323	3.564	-0.09	[−7.48, 6.83]	.950	.000
		Extraversion	.032	3.178	1.464	0.112	-2.209	3.881	-0.57	[−10.00, 5.59]	.950	.006
		Neuroticism	.034	3.086	1.305	1.333	0.193	3.026	0.06	[−5.89, 6.27]	.950	.000
		Openness	.072	2.732	1.858	1.401	6.560	4.379	1.50	[−2.24, 15.36]	.702	.045

Legend: $\mathbf{R}^2$: coefficient of determination; $\mathbf{B}_{Grp}$, $\mathbf{B}_{PRT}$, and $\mathbf{B}_{Trt}$: main effects of group, PRT factor, and trait; $\mathbf{B}_{Int}$, **SE**, **t**, and **95% CI**: interaction coefficient, standard error, t-statistic, and confidence interval; $\mathbf{p}_{fdr}$: FDR-adjusted p-value; **Partial f**^{2}: Cohen's effect size. Bold values indicate $\mathbf{p}_{fdr} < .05$.

positive main effect alongside a negative interaction with training group, indicating that increasing Agreeableness shifts the advantage toward the active-risk ISA training method rather than the passive-risk one. A plausible explanation is that highly agreeable individuals tend to be more trusting, which has been linked to increased susceptibility to social engineering attacks such as phishing [7]. As a result, training that explicitly targets adversarial interactions may address a more relevant vulnerability for these users, leading to greater improvements specifically in the Active ISA score. Accordingly, the moderation effects associated with Agreeableness were concentrated primarily in Active ISA and, consequently, in Overall ISA, whereas Passive ISA effects remained comparatively weak and centered near zero. This domain-specific structure suggests that personality traits selectively influence particular forms of ISA adaptation rather than uniformly affecting all domains.

Prior work primarily examined Agreeableness as a predictor of ISA levels. Our results extend this perspective by showing that Agreeableness also differentiates responsiveness to distinct training methods. Our findings also relate to prior work by Sumner *et al.* [42], who reported that agreeable users benefited less from phishing-oriented training interventions relative to users with other dominant traits. While this may initially appear inconsistent with our results, the two findings are not directly contradictory. Agreeable

Table 9: Regression results examining personality traits as moderators of the relationship between changes in PRT factors and ISA change.

				Main Effects			Interaction (PRT × Trait)					
Outcome	**PRT Factor**	**Moderator**	R^2	B_{Grp}	B_{PRT}	B_{Trt}	B_{Int}	SE	t	95% CI	p_{fdr}	Partial f^2
Active ISA Score Change	**Resource PRT Score Change**	Agreeableness	.056	1.987	-9.350	-3.553	15.206	18.963	0.80	[−22.90, 53.31]	.711	.013
		Conscientiousness	.062	1.313	-9.069	1.962	-9.730	10.367	-0.94	[−30.56, 11.10]	.711	.018
		Extraversion	.042	1.737	-8.230	0.165	0.730	8.712	0.08	[−16.78, 18.24]	.934	.000
		Neuroticism	.063	0.477	-7.857	-0.557	9.551	9.791	0.98	[−10.12, 29.23]	.711	.019
		Openness	.052	0.778	-8.693	-4.465	3.844	11.722	0.33	[−19.71, 27.40]	.930	.002
	Medical PRT Score Change	Agreeableness	.041	1.820	1.111	0.213	-16.871	12.515	-1.35	[−42.02, 8.28]	.815	.037
		Conscientiousness	.016	2.390	0.124	0.684	-4.405	6.122	-0.72	[−16.71, 7.90]	.815	.011
		Extraversion	.009	2.935	-0.423	-0.881	-2.863	5.598	-0.51	[−14.11, 8.39]	.815	.005
		Neuroticism	.005	2.925	-0.161	-0.676	0.693	4.098	0.17	[−7.54, 8.93]	.866	.001
		Openness	.014	2.180	-0.103	-2.956	-4.496	9.898	-0.45	[−24.39, 15.39]	.815	.004
	Ethical PRT Score Change	Agreeableness	.186	7.645	-2.815	4.593	-60.320	18.546	-3.25	[−97.59, −23.05]	**.010**	.216
		Conscientiousness	.011	2.454	-2.703	0.816	-2.995	13.520	-0.22	[−30.16, 24.17]	.826	.001
		Extraversion	.041	3.849	-5.099	0.039	-14.295	11.144	-1.28	[−36.69, 8.10]	.514	.034
		Neuroticism	.023	2.448	-3.168	0.221	7.154	8.780	0.82	[−10.49, 24.80]	.566	.014
		Openness	.025	3.200	-4.209	-2.463	-11.940	15.782	-0.76	[−43.66, 19.78]	.566	.012
Overall ISA Score Change	**Resource PRT Score Change**	Agreeableness	.028	2.904	-2.123	-2.948	0.470	10.364	0.05	[−20.36, 21.30]	.964	.000
		Conscientiousness	.032	2.903	-2.252	-0.729	-4.093	5.672	-0.72	[−15.49, 7.30]	.964	.011
		Extraversion	.022	2.793	-2.051	-0.072	-0.651	4.742	-0.14	[−10.18, 8.88]	.964	.000
		Neuroticism	.027	2.553	-1.756	0.507	2.711	5.378	0.50	[−8.10, 13.52]	.964	.005
		Openness	.023	2.562	-2.123	-1.022	0.654	6.410	0.10	[−12.23, 13.54]	.964	.000
	Medical PRT Score Change	Agreeableness	.060	2.583	0.147	-1.466	-9.457	6.677	-1.42	[−22.87, 3.96]	.631	.041
		Conscientiousness	.041	3.003	-0.282	-1.638	-3.773	3.256	-1.16	[−10.32, 2.77]	.631	.027
		Extraversion	.015	3.147	-0.396	-0.300	-0.369	3.008	-0.12	[−6.41, 5.68]	.984	.000
		Neuroticism	.014	3.236	-0.264	0.316	-0.045	2.197	-0.02	[−4.46, 4.37]	.984	.000
		Openness	.019	2.894	-0.299	-0.665	-2.096	5.319	-0.39	[−12.79, 8.59]	.984	.003
	Ethical PRT Score Change	Agreeableness	.180	5.468	-0.526	1.021	-30.885	10.033	-3.08	[−51.05, −10.72]	**.017**	.193
		Conscientiousness	.016	3.456	-0.657	-0.515	1.105	7.266	0.15	[−13.50, 15.71]	.880	.000
		Extraversion	.021	3.278	-1.008	-0.089	-3.589	6.065	-0.59	[−15.78, 8.60]	.880	.007
		Neuroticism	.015	3.122	-0.417	0.461	0.726	4.750	0.15	[−8.82, 10.27]	.880	.000
		Openness	.033	3.450	-1.231	-0.078	-8.146	8.469	-0.96	[−25.16, 8.87]	.852	.019
Passive ISA Score Change	**Resource PRT Score Change**	Agreeableness	.140	3.821	5.104	-2.343	-14.265	7.899	-1.81	[−30.14, 1.61]	.385	.067
		Conscientiousness	.115	4.494	4.566	-3.420	1.545	4.393	0.35	[−7.28, 10.37]	.727	.003
		Extraversion	.078	3.850	4.129	-0.309	-2.031	3.729	-0.55	[−9.53, 5.46]	.727	.006
		Neuroticism	.113	4.628	4.345	1.570	-4.128	4.158	-0.99	[−12.48, 4.23]	.727	.020
		Openness	.090	4.347	4.447	2.422	-2.535	5.012	-0.51	[−12.61, 7.54]	.727	.005
	Medical PRT Score Change	Agreeableness	.042	3.347	-0.817	-3.146	-2.042	5.460	-0.37	[−13.01, 8.93]	.887	.003
		Conscientiousness	.087	3.615	-0.688	-3.961	-3.141	2.575	-1.22	[−8.31, 2.03]	.887	.030
		Extraversion	.040	3.358	-0.370	0.280	2.126	2.405	0.88	[−2.71, 6.96]	.887	.016
		Neuroticism	.038	3.547	-0.367	1.308	-0.782	1.758	-0.45	[−4.31, 2.75]	.887	.004
		Openness	.031	3.609	-0.496	1.626	0.304	4.281	0.07	[−8.30, 8.91]	.944	.000
	Ethical PRT Score Change	Agreeableness	.041	3.291	1.763	-2.551	-1.450	8.786	-0.17	[−19.11, 16.21]	.870	.001
		Conscientiousness	.078	4.458	1.388	-1.846	5.205	5.695	0.91	[−6.24, 16.65]	.609	.017
		Extraversion	.074	2.708	3.084	-0.218	7.117	4.779	1.49	[−2.49, 16.72]	.357	.045
		Neuroticism	.092	3.795	2.334	0.700	-5.703	3.694	-1.54	[−13.13, 1.72]	.357	.049
		Openness	.049	3.701	1.746	2.307	-4.352	6.805	-0.64	[−18.03, 9.32]	.657	.008

Legend: $\mathbf{R}^2$: coefficient of determination; $\mathbf{B}_{Grp}$, $\mathbf{B}_{PRT}$, and $\mathbf{B}_{Trt}$: main effects of group, PRT factor change, and trait; $\mathbf{B}_{Int}$, **SE**, **t**, and **95% CI**: interaction coefficient, standard error, t-statistic, and confidence interval; $\mathbf{p}_{fdr}$: FDR-adjusted p-value; **Partial** $\mathbf{f}^2$: Cohen's effect size. Bold values indicate $\mathbf{p}_{fdr} < .05$.

individuals may still exhibit smaller overall gains than other personality groups while nevertheless benefiting relatively more from active-risk training compared to passive-risk training alternatives. Taken together, these findings suggest that the relationship between personality and ISA training effectiveness depends not only on user traits, but also on the type of training intervention and the behavioral outcomes being measured.

Beyond the statistically significant Agreeableness effects, the confidence intervals also revealed several informative non-significant trends. In particular, Extraversion showed predominantly negative confidence intervals for the interaction effect on Active ISA change, suggesting a similar differentiating pattern as Agreeableness. Openness exhibited a similar, though weaker and more uncertain, directional trend. In contrast, Conscientiousness showed confidence intervals that were substantially centered around zero, suggesting little evidence for a meaningful moderating effect in this context. Although these effects did not reach statistical significance, their directional structure suggests that these trends warrant further examination in future work.

RQ2: Personality, PRT, and ISA Change. Our findings show that personality traits also influence how risk-taking tendencies translate into awareness gains, by moderating the relationships between baseline PRT levels and changes in PRT and ISA improvements.

Both Agreeableness and Conscientiousness moderate the relationship between Resource PRT and ISA gains, but in opposite directions: higher Agreeableness attenuates this relationship, whereas higher Conscientiousness strengthens it. Similarly, Agreeableness moderates the relationship between changes in Ethical PRT and ISA improvement, such that reductions in Ethical PRT are associated with ISA gains primarily among individuals lower in Agreeableness.

Importantly, these findings suggest that training interventions aimed at reducing risk-taking do not necessarily translate directly into improvements in ISA. Rather, this relationship appears to depend on underlying individual differences, highlighting the importance of understanding the mechanisms that tie risk reduction to awareness gains in order to optimize training efficacy.

The structure of these moderation effects also helps explain the divergent interaction patterns observed across ISA outcomes. Agreeableness consistently exhibited stronger effects for Active and Overall ISA, whereas Passive ISA effects were comparatively weak, null or occasionally directionally positive. This aligns with the findings from RQ1, where higher Agreeableness was associated with greater responsiveness to active-risk training.

These findings can be interpreted in light of prior work on personality and risk-taking. Agreeableness has previously been associated with lower engagement in risky behaviors [27], which may explain why highly agreeable individuals show weaker relationships between reductions in PRT and subsequent ISA gains: their baseline tendency toward lower risk-taking may limit the extent to which further reductions translate into measurable awareness improvements. At the same time, the weak directionally positive Passive ISA trend observed for the Agreeableness-Resource PRT interaction may suggest that highly agreeable individuals who nevertheless exhibit elevated resource-related risk-taking represent a less common behavioral profile for whom passive-risk training makes omission-based vulnerabilities more salient. In contrast, higher Conscientiousness strengthened the relationship between Resource PRT and ISA gains. One possible explanation is that, for individuals who are typically organized and diligent, elevated Resource PRT reflects a lack of awareness rather than a stable disposition toward risk-taking. Once training highlights these behaviors as privacy risks, conscientious individuals may be particularly well-positioned to align their behavior with the newly recognized standards.

Our findings also extend prior ISA-personality research, which primarily examined direct associations between personality traits and self-reported ISA. Rather than acting as uniformly positive or negative predictors of awareness, personality traits appear to shape how users operationalize training and risk-related information into behavioral adaptation over time.

Importantly, these moderation effects were concentrated in Resource and Ethical PRT, whereas Medical PRT showed little relationship with ISA outcomes. This differentiation refines prior work that treated PRT as a unified construct. Ethical PRT captures tendencies related to rule adherence and norm compliance, while Resource PRT reflects effort allocation and attention-management behaviors that become particularly relevant during training. In contrast, Medical PRT reflects specific health-related risk-taking that may not meaningfully transfer to cybersecurity and privacy contexts.

Finally, the confidence intervals revealed several informative non-significant trends that were directionally consistent with the broader moderation patterns. In particular, Neuroticism exhibited predominantly negative confidence intervals in interactions involving Resource and Ethical PRT, suggesting a potential attenuation effect in the relationship between reductions in PRT and improvements in ISA. Several additional Agreeableness-related intervals also showed directional skew rather than being centered around zero, particularly for Resource PRT interactions. Taken together, these trends suggest that additional personality-specific moderation effects may exist and warrant future examination.

Conclusion. Taken together, these findings demonstrate that ISA improvement is not solely a function of training content, but also emerges from the interaction between personality traits and distinct forms of risk-taking behavior. These results further indicate that personality traits shape responsiveness to particular training strategies. More broadly, the present work highlights the importance of moving toward adaptive, privacy-enhancing training paradigms that account for both personality traits and the multidimensional nature of risk-taking. By integrating longitudinal behavioral measures with psychological constructs, this study provides evidence that personalization may improve the effectiveness of ISA interventions and offers a foundation for future human-centric privacy defense mechanisms.

Limitations and Future Work. While our study provides valuable insights, it is not without limitations. To address limitations of prior work, this study involves continuous sensor-based monitoring to generate a granular longitudinal dataset of daily behaviors rather than relying on static self-reported measurements, thereby increasing data density.

As a result, the high dimensionality of the collected behavioral and psychometric data created a high statistical threshold that our varied sample sizes were not always powered to surpass. Several effects exhibited directionally informative confidence intervals but did not reach statistical significance.

To reduce multicollinearity and stabilize estimation, personality traits were modeled in separate regression models rather than jointly within a single multivariate framework. Consequently, the analyses do not account for correlations among personality traits. The findings should be interpreted more as evidence that personality-related individual differences play a meaningful role in shaping ISA training responsiveness, even if the precise attribution between highly correlated traits remains imperfect.

Translating these moderation effects into organizational deployment is non-trivial and raises practical, ethical, and regulatory challenges. Accordingly, in future work, we plan to investigate how such systems can be designed, validated, and governed in organizational settings using privacy-enhancing approaches such as encrypted local storage and edge-based computation, to maximize training efficacy to protect users while preserving their privacy.

Acknowledgments

We thank the anonymous reviewers and revision editor for their constructive feedback. In accordance with the policy on AI usage, we disclose that an LLM was utilized during the preparation of this manuscript. The use of the LLM was strictly limited to correcting grammar and checking code syntax. All research contributions, experimental design, analysis, and the core text were made by the human authors. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Laura Marie Abels, Matthew Smith, and Anna-Marie Ortloff. 2025. I never reuse passwords! Development and Validation of a Security and Privacy Social Desirability Scale ({{{{{SP-SDS}}}}}) for end users without a background in computer science. In *Twenty-First Symposium on Usable Privacy and Security (SOUPS 2025)*. 415–434.

[2] Saad Alahmari, Karen Renaud, and Inah Omoronyia. 2022. Moving beyond cyber security awareness and training to engendering security knowledge sharing. *Information Systems and e-Business Management* (2022), 1–36.

[3] Isabel Arend, Asaf Shabtai, Tali Idan, Ruty Keinan, and Yoella Bereby-Meyer. 2020. Passive-and Not Active-Risk Tendencies Predict Cyber Security Behavior. *Computers & Security* (2020), 101929.

[4] Ron Bitton, Kobi Boymgold, Rami Puzis, and Asaf Shabtai. 2020. Evaluating the Information Security Awareness of Smartphone Users. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[5] Ron Bitton, Andrey Finkelshtein, Lior Sidi, Rami Puzis, Lior Rokach, and Asaf Shabtai. 2018. Taxonomy of mobile users' security awareness. *Computers & Security* 73 (2018), 266–293.

[6] Wiebke Bleidorn, Ted Schwaba, Anqing Zheng, Christopher J Hopwood, Susana S Sosa, Brent W Roberts, and DA Briley. 2022. Personality stability and change: A meta-analysis of longitudinal studies. *Psychological bulletin* 148, 7-8 (2022), 588.

[7] Matthew Canham, Clay Posey, and Michael Constantino. 2022. Phish Derby: Shoring the Human Shield Through Gamified Phishing Attacks. In *Frontiers in Education*, Vol. 6. Frontiers, 536.

[8] Ofir Cohen, Gil Ari Agmon, Asaf Shabtai, and Rami Puzis. 2026. LISAA: A Framework for Large Language Model Information Security Awareness Assessment. arXiv:2411.13207 [cs.CR] https://arxiv.org/abs/2411.13207

[9] Ofir Cohen, Ron Bitton, Asaf Shabtai, and Rami Puzis. 2023. ConGISATA: A Framework for Continuous Gamified Information Security Awareness Training and Assessment. In *European Symposium on Research in Computer Security*. Springer, 431–451.

[10] Ersin Dincelli and InduShobha Chengalur-Smith. 2020. Choose your own training adventure: designing a gamified SETA artefact for improving information security and privacy through interactive storytelling. *European Journal of Information Systems* 29, 6 (2020), 669–687.

[11] Serge Egelman and Eyal Peer. 2015. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2873–2882.

[12] Lewis R Goldberg. 1993. The structure of phenotypic personality traits. *American psychologist* 48, 1 (1993), 26.

[13] Margaret Gratian, Sruthi Bandi, Michel Cukier, Josiah Dykstra, and Amy Ginther. 2018. Correlating human traits and cyber security behavior intentions. *computers & security* 73 (2018), 345–358.

[14] Tzipora Halevi, Nasir Memon, James Lewis, Ponnurangam Kumaraguru, Sumit Arora, Nikita Dagar, Fadi Aloul, and Jay Chen. 2016. Cultural and psychological factors in cyber-security. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services*. 318–324.

[15] Tzipora Halevi, Nasir Memon, and Oded Nov. 2015. Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. *Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks (January 2, 2015)* (2015).

[16] Jonas Hielscher, Annette Kluge, Uta Menges, and M Angela Sasse. 2021. "Taking out the trash": Why security behavior change requires intentional forgetting. In *Proceedings of the 2021 New Security Paradigms Workshop*. 108–122.

[17] Jacob B Hirsh, Sonia K Kang, and Galen V Bodenhausen. 2012. Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological science* 23, 6 (2012), 578–581.

[18] O. P. John and R. L. Donahue, E. M. & Kentle. 1991. *The Big Five Inventory—versions 4a and 5*. University of California, Berkeley, Institute of Personality and Social Research.

[19] Oliver P John, Sanjay Srivastava, et al. 1999. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. (1999).

[20] Maurits Kaptein, Panos Markopoulos, Boris De Ruyter, and Emile Aarts. 2015. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies* 77 (2015), 38–51.

[21] Ruty Keinan and Yoella Bereby-Meyer. 2012. " Leaving it to chance"–Passive risk taking in everyday life. *Judgment & Decision Making* 7, 6 (2012).

[22] Ruty Keinan and Yoella Bereby-Meyer. 2017. Perceptions of active versus passive risks, and the effect of personal responsibility. *Personality and Social Psychology Bulletin* 43, 7 (2017), 999–1007.

[23] Ross Kelly. 2017. Almost 90% of cyber attacks are caused by human error or behavior. *ChiefExecutive. net* (2017).

[24] Shelia M Kennison and Eric Chan-Tin. 2020. Taking risks with cybersecurity: Using knowledge and personal characteristics to predict self-reported cybersecurity behaviors. *Frontiers in Psychology* 11 (2020), 546546.

[25] Hennie A Kruger and Wayne D Kearney. 2006. A prototype for assessing information security awareness. *Computers & security* 25, 4 (2006), 289–296.

[26] Daniele Lain, Tarek Jost, Sinisa Matetic, Kari Kostiainen, and Srdjan Capkun. 2024. Content, Nudges and Incentives: A Study on the Effectiveness and Perception of Embedded Phishing Training. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*. 4182–4196.

[27] Marco Lauriola and Joshua Weller. 2018. Personality and risk: Beyond daredevils—risk taking from a temperament perspective. In *Psychological perspectives on risk and risk analysis: theory, models, and applications*. Springer, 3–36.

[28] Agata McCormac, Tara Zwaans, Kathryn Parsons, Dragana Calic, Marcus Butavicius, and Malcolm Pattinson. 2017. Individual differences and information security awareness. *Computers in Human Behavior* 69 (2017), 151–156.

[29] Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of personality* 60, 2 (1992), 175–215.

[30] Rita Orji, Lennart E Nacke, and Chrysanne Di Marco. 2017. Towards personality-driven persuasive health games and gamified systems. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 1015–1027.

[31] Kiemute Oyibo, Rita Orji, and Julita Vassileva. 2017. Investigation of the Influence of Personality Traits on Cialdini's Persuasive Strategies. *PPT@ PERSUASIVE* 2017 (2017), 8–20.

[32] Kathryn Parsons, Dragana Calic, Malcolm Pattinson, Marcus Butavicius, Agata McCormac, and Tara Zwaans. 2017. The human aspects of information security questionnaire (HAIS-Q): two further validation studies. *Computers & Security* 66 (2017), 40–51.

[33] Elissa M Redmiles, Ziyun Zhu, Sean Kross, Dhruv Kuchhal, Tudor Dumitras, and Michelle L Mazurek. 2018. Asking for a friend: Evaluating response biases in security user studies. In *Proceedings of the 2018 acm sigsac conference on computer and communications security*. 1238–1255.

[34] Justin D Russell, Carl F Weems, Irfan Ahmed, and Golden G Richard III. 2017. Self-reported secure and insecure cyber behaviour: factor structure and associations with personality factors. *Journal of Cyber Security Technology* 1, 3-4 (2017), 163–174.

[35] Dawn M Sarno, Maggie W Harris, and Jeffrey Black. 2023. Which phish is captured in the net? Understanding phishing susceptibility and individual differences. *Applied Cognitive Psychology* (2023).

[36] Yukiko Sawaya, Sarah Lu, Takamasa Isohara, and Mahmood Sharif. 2024. A high coverage cybersecurity scale predictive of user behavior. In *33rd USENIX Security Symposium (USENIX Security 24)*. 5503–5520.

[37] Lorin Schöni, Victor Carles, Martin Strohmeier, Peter Mayer, and Verena Zimmermann. 2024. You Know What?-Evaluation of a Personalised Phishing Training Based on Users' Phishing Knowledge and Detection Skills. In *Proceedings of the 2024 European Symposium on Usable Security*. 1–14.

[38] Visahl Samson David Selvam. 2020. Human Error in IT Security. *arXiv preprint arXiv:2005.04163* (2020).

[39] Alexander T Shappie, Charlotte A Dawson, and Scott M Debb. 2020. Personality as a predictor of cybersecurity behavior. *Psychology of Popular Media* 9, 4 (2020), 475.

[40] Adir Solomon, Michael Michaelshvili, Ron Bitton, Bracha Shapira, Lior Rokach, Rami Puzis, and Asaf Shabtai. 2022. Contextual security awareness: A context-based approach for assessing the security awareness of users. *Knowledge-Based Systems* 246 (2022), 108709.

[41] Sophos. 2024. Sophos 2024 Threat Report. https://assets.sophos.com/X24WTUEQ/at/wwf5phjtj9bjvmpqqsbfxc/sophos-2024-threat-report.pdf.

[42] Alex Sumner, Xiaohong Yuan, Mohd Anwar, and Maranda McBride. 2022. Examining factors impacting the effectiveness of anti-phishing trainings. *Journal of Computer Information Systems* 62, 5 (2022), 975–997.

[43] The Verizon DBIR Team. 2025. DBIR: Data Breach Investigations Report. https://www.verizon.com/business/resources/Teeb/reports/2025-dbir-data-breach-investigations-report.pdf

[44] Dimitri Van der Linden, Jan Te Nijenhuis, and Arnold B Bakker. 2010. The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of research in personality* 44, 3 (2010), 315–327.

[45] Monica Whitty, James Doodson, Sadie Creese, and Duncan Hodges. 2015. Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking* 18, 1 (2015), 3–7.
[46] Tienhua Wu, Kuang-You Tien, Wei-Chih Hsu, and Fu-Hsiang Wen. 2021. Assessing the effects of gamification on enhancing information security awareness knowledge. *Applied Sciences* 11, 19 (2021), 9266.

A Additional Exploratory Results

This appendix presents additional analyses examining the relationships between personality traits, baseline ISA, and PRT. These analyses complement the main results but did not yield findings that remained statistically significant after correction for multiple testing. Nevertheless, they provide useful exploratory insights into potential associations. As such, these results are reported for completeness and to inform future research.

A.1 Are Personality Traits Correlated with a Tendency to Take Passive Risks?

To explore baseline associations between personality traits and PRT, we examined the relationships between the Big Five traits and PRT scores across all factors. Only participants who completed both the first PRT questionnaire and the BFI were included in this analysis ($N = 102$).

Given that several variables deviated from normality (Shapiro-Wilk $p < .05$), we used Spearman's rank correlation (ρ) for all pairwise tests. Table 10 provides the full results.

While none of the correlations remained significant after correction, several exploratory patterns emerged. In particular, Agreeableness showed a negative association with Ethical PRT ($\rho = -0.282$, $p_{unc} = .004$), and Conscientiousness showed a negative association with Resource PRT ($\rho = -0.244$, $p_{unc} = .013$). These trends may suggest that individuals higher in Agreeableness and Conscientiousness tend to engage less in certain forms of passive risk-taking.

A.2 Are Personality Traits Correlated with ISA When Measured Using Objective Sensor Data?

To explore whether personality traits are associated with objectively measured ISA using real-world sensors, we examined the relationships between the Big Five traits and baseline passive ISA scores. Only participants with available sensor-based ISA data and completed personality measures were included in this analysis ($N = 55$).

Given deviations from normality, Spearman's rank correlation (ρ) was used for all pairwise tests. Table 11 presents the full results.

No correlations remained significant after correction. However, an exploratory negative association was observed between Extraversion and passive ISA ($\rho = -0.306$, $p_{unc} = .023$), which may suggest that more extraverted individuals exhibit lower baseline ISA as captured by sensor-based measures. This pattern should be further investigated in future work.

Table 10: Spearman correlation matrix between Big Five traits and baseline PRT scores.

PRT Variable	Trait	ρ	p_{unc}	p_{FDR}
Ethical PRT	Agreeableness	-.282	.004	.082
	Conscientiousness	-.098	.325	.500
	Extraversion	-.130	.192	.377
	Neuroticism	.176	.077	.306
	Openness	-.126	.207	.377
Medical PRT	Agreeableness	-.066	.513	.579
	Conscientiousness	-.001	.989	.989
	Extraversion	-.064	.521	.579
	Neuroticism	-.136	.172	.377
	Openness	-.090	.366	.523
Overall PRT	Agreeableness	-.178	.074	.306
	Conscientiousness	-.189	.057	.306
	Extraversion	-.158	.112	.320
	Neuroticism	.051	.608	.640
	Openness	-.110	.270	.450
Resource PRT	Agreeableness	-.126	.206	.377
	Conscientiousness	-.244	.013	.133
	Extraversion	-.164	.100	.320
	Neuroticism	.083	.409	.545
	Openness	-.071	.480	.579

Legend: ρ: Spearman's rank correlation coefficient; p_{unc}: uncorrected p-value; p_{FDR}: FDR corrected p-value.

Table 11: Spearman correlations between Big Five traits and baseline passive ISA scores.

ISA Score	Trait	r	p_{unc}	p_{FDR}
Passive ISA	Agreeableness	-.005	.971	.988
	Conscientiousness	-.151	.270	.451
	Extraversion	-.306	.023	.115
	Neuroticism	.216	.114	.285
	Openness	-.002	.988	.988

Legend: r: Spearman's rank correlation coefficient; p_{unc}: uncorrected p-value; p_{FDR}: FDR-corrected p-value.

A.3 How Do Personality Traits Influence the Rate of Improvement in ISA Over the Course of a Longitudinal Training Program?

To explore whether personality traits are associated with differences in ISA improvement over time, we analyzed week-by-week ISA progression using linear mixed-effects models. Only participants with sufficient longitudinal sensor-based ISA data were included in this analysis ($N = 55$).

We employed models with random intercepts for participants and fixed effects for week, training group, and personality traits. Interaction terms (e.g., $Week \times Trait$) were included to assess whether personality traits moderate the rate of ISA change over time. This

Table 12: Mixed-effects regression results examining personality traits as moderators of longitudinal change in ISA.

		Model Fit		Main Effects				Interaction (Time × Trait)					
Outcome	**Moderator**	R^2_{Marg}	R^2_{Cond}	B_{Grp}	B_{Trt}	B_{Time}	B_{TxG}	B_{Int}	SE	z	95% CI	p_{unc}	p_{fdr}
Passive ISA	Agreeableness	.018	.893	2.537	-4.639	2.621	0.701	-0.593	0.550	-1.08	$[-1.67, 0.48]$	.281	.468
	Conscientiousness	.060	.895	3.654	-6.391	2.911	0.825	-0.727	0.325	-2.24	$[-1.36, -0.09]$	.025	.126
	Extraversion	.076	.892	1.169	-7.237	0.309	0.668	-0.021	0.318	-0.07	$[-0.64, 0.60]$	.948	.948
	Neuroticism	.075	.893	3.304	6.257	-0.721	0.731	0.371	0.278	1.34	$[-0.17, 0.92]$	.182	.454
	Openness	.009	.893	2.522	1.280	-0.111	0.688	0.097	0.381	0.25	$[-0.65, 0.84]$	.800	.948

Legend: $\mathbf{R}^2_{Marg}$ and $\mathbf{R}^2_{Cond}$: marginal and conditional coefficients of determination; $\mathbf{B}_{Grp}$, $\mathbf{B}_{Trt}$, $\mathbf{B}_{Time}$, and $\mathbf{B}_{TxG}$: main effects of group, trait, time, and the time × group interaction; $\mathbf{B}_{Int}$, **SE**, **z**, and **95% CI**: time × trait interaction coefficient, standard error, z-statistic, and confidence interval; $\mathbf{p}_{unc}$ and $\mathbf{p}_{fdr}$: uncorrected and FDR-adjusted p-values.

approach accounts for the nested structure of repeated measures. The results are presented in Table 12.

No interaction effects remained significant after correction. However, an exploratory negative interaction between Conscientiousness and time was observed ($B = -0.727$, $p_{unc} = .025$), which may suggest that individuals higher in Conscientiousness engage more thoroughly with the training material, potentially taking more time to internalize and explore its details. This pattern may indicate that such users could benefit from longer or more sustained training programs, a possibility that should be examined in future work.