Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.5559

Efficiency of an educational game for teaching the behavior of distinguishing legitimate emails from phishing attempts

Jasson Marques Fontoura Júnior [**Federal University of Roraima** | ***jassonjr5@gmail.com***]
Marcelo Henrique Oliveira Henklain [**Federal University of Roraima** | ***marcelo.henklain@ufrr.br***]
Felipe Leite Lobo [**Federal University of Roraima** | ***felipe.lobo@ufrr.br***]
Eduardo Luzeiro Feitosa [**Federal University of Amazonas**| ***efeitosa@icomp.ufam.edu.br***]

Universidade Federal de Roraima, Av. Cap. Ene Garcês, 2413 - Aeroporto, Boa Vista - RR, 69310-000, Brazil.

Received: 15 February 2025 • **Accepted:** 21 November 2025 • **Published:** 26 May 2026

Abstract Phishing attacks are on the rise worldwide, with email remaining a highly effective vector. As anti-phishing tools alone are insufficient to prevent cyberattacks, educational measures are essential. Educational games offer a promising yet underexplored approach to cybersecurity training. This study aimed to develop and evaluate the efficiency of the educational game Alerta for teaching the behavior of "distinguishing legitimate emails from phishing attempts". For the two experiments, the results were promising, suggesting preliminary evidence of learning in distinguishing between legitimate and phishing emails, improved self-confidence for most of the participants, and positive indicators of usability and satisfaction. Future studies should continue to refine and test the game with a larger and more diverse sample, as well as incorporate new learning objectives to empower individuals to handle modern phishing techniques.

Keywords: Cybersecurity; Phishing; Educational Games.

1 Introduction

According to APWG reports [APWG, 2023, 2024], phishing attacks increased by 40% globally, reaching over 960,000 incidents in the first quarter of 2024. In Brazil, more than 5,000 fake pages used for phishing have been detected [CERT.br, 2024]. This represents a significant threat to the digital security of individuals and organizations. Of the various vectors for such attacks, emails continue to be highly effective with users, warranting close attention. [Souza and Tanaka, 2023].

Cybersecurity scientists and professionals emphasize the urgency of educational measures to mitigate this threat, as users' ability to identify phishing attempts is crucial for prevention [Rahman *et al*., 2020; Bernardino *et al*., 2023], since anti-phishing tools aren't sufficient to accomplish this goal [Arachchilage and Cole, 2011]. Awareness campaigns and information-exposure-based teaching methods alone have proven not always effective, as users persist in exhibiting unsafe behaviors and suffering their consequences [Wash, 2020; Henderson *et al*., 2024].

The low effectiveness of current educational approaches may stem from the fact that users are exposed only to rules without explicit training [Moreira and Medeiros, 2018; Alanazi *et al*., 2022] in recognizing phishing. Educational games are promising for this type of training because they foster motivation even during repetitive tasks, allow for immediate reinforcement of behaviors that must be learned, and can be utilized by many people simultaneously [Tsutsumi *et al*., 2020; Abreu *et al*., 2024]. However, there are few educational games for teaching cybersecurity described in the scientific literature.

Furthermore, the development of effective educational games has several requirements. For instance, it requires clearly defined learning objectives. Without them, there is a risk that the game may foster engagement without effectively promoting the intended learning that motivated its creation [Gris and De Souza, 2016; Kienen *et al*., 2021]. Another essential requirement for developing an effective educational game is to test it to empirically demonstrate its efficiency [Panosso *et al*., 2015]. In summary, the cybersecurity education literature lacks studies about educational games in general, and, specifically, about those that meet the requirements we have highlighted.

For these reasons, the objective of this study was to develop and evaluate the efficiency of an educational game designed to teach the behavior of "distinguishing legitimate emails from phishing attempts" and contribute to cybersecurity education. After developing three iterations of the game, we sought to answer the following research questions: RQ1. Was the game effective in promoting learning for all players? RQ2. Did players increase their self-confidence in identifying legitimate emails and phishing attempts after playing the game? RQ3. To what extent was the gaming experience enjoyable? RQ4. What was the game's level of usability? Based on behavior-analytic theory, this study assumes that if our game promotes learning, fosters self-confidence, provides an enjoyable experience, and demonstrates usability, then there is favorable evidence of its efficiency.

This study includes sections on theoretical foundations, related work, methods, results, discussion, and conclusion. In the Theoretical Foundation section, the main concepts and theory that underpinned the creation of the game are presented. Related Work explores studies that address research problems similar to those in this study, highlighting the gap in the literature that this work aims to fill. At the end of this section, our research problem is outlined. In the Method, we present

the game as a solution for teaching the behavior of "distinguishing legitimate emails from phishing attempts", along with the strategy for evaluating its efficiency. In the Results and Discussion section, the study findings and the arguments that support its conclusions are described, in addition to the limitations found. Finally, the work ends with a summary of our answers to the research questions, in the Conclusion section.

2 Theoretical Foundation

This section explains the primary form of phishing — email-based attacks — which are among the most common threats in the digital age [Souza and Tanaka, 2023]. Understanding these attacks, their mechanisms, and how user behavior plays a key role in mitigating these risks is crucial. Identifying and responding effectively to these threats is essential to protecting individuals and organizations in an increasingly hostile digital environment.

2.1 Phishing attacks: One of the main threats in cyberspace

According to CERT.br [2024], phishing attacks are frequent in Brazil, posing a serious problem for the economy and the well-being of individuals. Despite awareness campaigns, users continue to engage in unsafe online behaviors, contributing to the success of cyberattacks [Guilherme *et al*., 2021].

Phishing attacks are typically carried out via email and involve the use of social engineering to obtain valuable information or persuade the user to click on a link or open an attachment [Syafitri *et al*., 2022]. Such actions can result in malware infections on the user's computer [Melo *et al*., 2011; Diorio *et al*., 2018] or redirect the user to fake websites where they may be tricked into providing personal information or making purchases. According to Scarfone *et al*. [2008], attackers use false information in emails to impersonate a known company or individual, aiming to capture the user's attention and trust. They may also employ email spoofing techniques to forge email headers, making the messages appear legitimate [Morgado *et al*., 2023].

2.2 Behavioral-analytic theory on the processes of teaching and learning

According to behavior-analytic theory, behavior is a system of interactions between antecedent environment, a person's actions, and subsequent environmental outcomes (typically called "consequent environment") [Skinner, 1981; Cortegoso and Coser, 2023]. "Distinguishing legitimate emails from phishing attempts" is a behavior (specifically, a cybersecurity behavior) and it is our general learning objective in this study. In behavior-analytic theory, the term "teaching objective" is adopted, but we prefer "learning objective" because it emphasize that our main goal is to promote learning. Moreover, it is a term we have observed in the fields of computer science education and computing in education, sometimes appearing as "learning goals". Learning objectives describe the behaviors that must be acquired or improved to enable the learner to effectively manage their social reality and to guide teaching decisions. Teaching is defined as the process of organizing conditions under which learning takes place. For instance, an educational game is one such condition that could foster learning. Learning can be defined as a lasting change in behavioral repertoire. Therefore, teaching is considered efficient when it is evident that, prior to instruction, learners' were completely or to some extent unable to perform a certain behavior (the learning objective) and, as a result of the instruction, acquired or improved it, enabling them to interact more effectively with their environment [Kienen *et al*., 2021].

Learning requires a repeated process in which actions are performed under specific conditions and followed by reinforcement [Moreira and Medeiros, 2018]. However, reinforcers are rarely equally effective for everyone, and their impact depends on how temporally close they are to the action they aim to reinforce [Moreira and Medeiros, 2018]. In this context, digital educational games are particularly useful because they promote engagement, expose learners to problem solving situations that demand the behaviors that compose the learning objectives, and provide immediate and automated feedback stimuli, such as points and badges. Although these stimuli are artificial, they may function as reinforcers [Panosso *et al*., 2015; Tsutsumi *et al*., 2020]. When consistent reinforcement takes place, students learn, tend to feel confident about the learned behavior and satisfied, which means they are willing to continue learning.

2.3 Educational games

Games are artificial systems defined by rules in which players engage voluntarily, with clear indicators of victory and defeat. These systems are characterized by interactivity, immersion, and automatic feedback [Gris and De Souza, 2016]. Educational games share all the characteristics of "games" while also aiming to promote the learning of behaviors. Studies suggest that educational games are effective in fostering learning and are particularly useful for training individuals on a large scale [Panosso *et al*., 2015; Tsutsumi *et al*., 2020].

Based on typical game elements and behavior-analytic theory [Stutz, 2020; Gris and De Souza, 2016; Tsutsumi *et al*., 2020], we adapted the framework of Schell [2015] to guide the creation of our educational game (see **Appendix 1**). The original framework consists of four key components: aesthetics, mechanics, technology, and story. Our version retains aesthetics and technology as they were proposed, incorporates rules related to mechanics (which encompasses the learning objectives), replaces "story" with "narrative" to emphasize that the relationship between what the story is and how it is told is crucial, and places "education" at the center of the framework. Learning objectives are developed through the relationship between mechanics and rules. Mechanics involve the actions the player must perform, while rules specify under what conditions and how these actions should occur to bring the player closer to victory. This relationship within the game — between the situation where action is appropriate, the action itself, and the outcome — is the very learning objective we aim to teach. The narrative represents the way the game's story is conveyed, designed to enhance engagement. Aesthetics play a similar role by fostering interest and immersion.

Finally, technology serves as the medium that enables the game's execution and interaction.

Although adopting frameworks previously tested in other studies is a promising approach, we sought to adapt an existing framework to make it more consistent with the theoretical perspective adopted in this study. We assessed that the modification did not contradict the core logic of the original framework; rather, it made it easier for the research group involved in this study, guided by a behavior-analytic perspective, to conceptualize and apply it to the development of the game.

In this study, we also adopted an iterative design process to develop the game, as it allows for continuous evaluations and improvements through the successive testing of functional prototypes [Panosso *et al*., 2015]. After each development stage, the game is tested to identify areas that need refinement. At the end of this process, which does not have a fixed number of iterations, the educational game should become effective in promoting both learning and engagement.

2.4 The relevance of educational games to cybersecurity education

Learning cybersecurity behaviors can be difficult, as much of the natural reinforcement for any protective behavior lies in the fact that nothing bad happens. This type of stimulation is of little use when a behavior needs to be taught. At the same time, the dangers of cyberspace may be temporally distant and uncertain, making it harder for the consequence of "nothing bad happening" to function as a useful reinforcer to an educator. That is why cybersecurity education is so complex.

In this context, the educator can attempt to bring temporally distant or uncertain events, such as cyberattacks, closer to students' reality in a safe way (e.g., through simulations or games) and can also create artificial reinforcers to promote the learning of cybersecurity behaviors. Among the various teaching strategies available to cybersecurity educators, educational games are among the most promising Henderson *et al*. [2024]. As previously discussed, they allow for repeated practice of the cybersecurity behaviors targeted as learning objectives, within a context in which the dangers of cyberspace can be simulated and brought closer to the student's experience (without real risk), and the behaviors performed can be reinforced in multiple ways within the game environment Arachchilage and Cole [2011]; Farias *et al*. [2019]; Tsutsumi *et al*. [2020].

3 Related Work

We reviewed the literature on cybersecurity education to mitigate phishing attacks and found, as reported by Kovacevic *et al*. [2020], that even computer science students express uncertainty about how to protect themselves from cyberattacks. However, they often do not seek to learn more about cybersecurity, highlighting the need for incentives to encourage everyone to adopt and implement good security practices. Along these lines, Rahman *et al*. [2020] identified in their literature review a lack of cybersecurity training for children and noted that, prior to this, their teachers must also be trained. Bernardino *et al*. [2023] noticed that seniors also need training to use the Internet safely.

Regarding what to teach about cybersecurity, Butavicius *et al*. [2016] demonstrated that people are particularly vulnerable to attacks that exploit authority figures to lend credibility to phishing emails, highlighting the importance of training users to handle such situations. Consequently, in our game, we incorporated threat-based messages that implicitly convey a sense of authority. Wash [2020] also provides valuable guidance on what to teach. He studied the behaviors of IT experts in identifying phishing emails and concluded that it is essential to develop individuals' ability to recognize the signals of such attacks. This behavior became the central focus of our training approach.

Regarding how to teach, we found promising studies that used educational games. For example, Farias *et al*. [2019] taught cybersecurity concepts related to antivirus, firewall, and backup to children and adolescents using the game Self-Protect, achieving positive results, with 97% of participants reporting that the game was useful to cybersecurity education. Although promising, this game do not teach any concepts about phishing, and no tests were conducted to assess whether participants' repertoire changed after playing the game.

Henderson *et al*. [2024], in turn, developed the disPHISH-information Game, designed to train enterprise employees to recognize phishing attacks in various forms, such as email, voice, and SMS. This analog card game is intended for group play, which is noteworthy as it encourages discussion and information sharing. However, its core characteristics — being analog and group-based — may limit its scalability for training larger numbers of people in a shorter time. In this study this game was not tested to evaluate its effectiveness in promoting learning.

Considering digital games, Arachchilage and Cole [2011] developed the Anti-phishing Game, using the Google App Inventor Emulator to teach users how to identify and avoid phishing email URLs. Their study provided a clearer description than Farias *et al*. [2019] of the specific behaviors the game aims to promote. However, similar to Henderson *et al*. [2024], they did not test the game to evaluate its efficiency.

Abreu *et al*. [2024], conversely, developed and tested Encrypta, a game based on minigames designed to teach users how to create strong passwords, detect phishing, identify malicious websites and files, and understand the importance of backups. The game was evaluated by 10 adolescents, 80% of whom found it visually appealing, 90% considered it an effective strategy for teaching cybersecurity concepts, and 80% said they would recommend it. Participants made few errors on the final learning test, suggesting the game was efficient, although the authors did not clearly state the learning objectives.

Based on these studies, we confirmed that teaching cybersecurity is essential [Rajasekharaiah *et al*., 2020] and that technological resources may facilitate large-scale learning. Considering **Table 1**, we also observed that few studies made explicit if and which solid psychological theory of human behavior and learning they were grounded, and learning objectives are rarely stated with precision. Educational games are not always tested, and when they are, the evaluations are often not directly aligned with the learning objectives, focusing

instead on perceived learning. Taken together, these aspects reduces the reproducibility of the studies and the strength of the conclusions. Therefore, we decided to develop and evaluate the effectiveness of an educational game designed to teach the behavior of "distinguishing legitimate emails from phishing attempts", addressing questions related to learning, self-confidence, satisfaction, and usability.

Table 1. Comparison between studies

Study	Phishing	Technology	Learning objectives	Behavior-analytic theory	Learning Test
Farias et al. (2019)	No	Digital	Not clear	No	No
Henderson et al. (2024)	Yes	Analog	Not clear	No	No
Arachchilage and Cole (2016)	Yes	Digital	Yes	No	No
Abreu et al. (2024)	Yes	Digital	Not clear	No	Yes
Present sutdy	Yes	Digital	Yes	Yes	Yes

4 Method

This section describes the methods used in the design, development, and evaluation of our educational game. We outline the structure of the experiments conducted, the computational and pedagogical strategies adopted, and the technological implementation choices.

4.1 Experiment 1

Experiment 1 focused on the initial development and testing of our game through two iterative cycles.

4.1.1 Computational Solution

Based on our framework for educational games, we developed the first two versions of the educational game that we called Alerta (Alert, in English) over two iterations. We prioritized creating the game's mechanics and rules while ensuring its functionality by focusing on the technology. During these iterations, we also crafted a simple narrative to tell the story of Alerta.

The game's narrative is set in the context of a company called "JR Company", which is conducting a recruitment process for a cybersecurity assistant position. The player participates in this selection process with the mission of distinguishing legitimate from phishing emails received by the company's employees. The game consists of one pre-test phase (an initial assessment of the player's prior knowledge about phishing), one post-test phase (a final assessment of the knowledge acquired), and four training phases (each targeting one of the four learning objectives, that we'll describe bellow). Each training phase includes three email differentiation tasks, while the pre-test and post-test each include 12 differentiation tasks. The pre-test is presented to the player as a way for the company to assess their existing knowledge about phishing. The four training phases are introduced as preparatory exercises to ensure all candidates have equal opportunities for success in the selection process. The post-test serves as the final evaluation to determine who will be hired. To win the game, the player must score over 100 points (answering 10 out of 12 questions correctly), which represents the victory condition. The narrative is delivered through text, and the game includes a total of 36 emails — 18 legitimate and 18 phishing — distributed as six of each type (legitimate or phishing) for the pre-test, post-test, and training phases. For accessibility purposes, all emails have been translated into English so that readers can choose to read them either in English or in the original Portuguese. However, it is important to note that the current version of the game is only available in Portuguese. These emails are available upon request to the first author.

The game introduces four characters: Hannah, Lucas, Joana, and Mateus. Mateus, the HR manager, is the first character the player encounters, as he provides the game's instructions. The other characters appear during the training phases, where the player evaluates their information — including their role in the company and the services they use (banking, social media, streaming, and online shopping) — to determine whether the email in their inbox is legitimate or a phishing attempt. The game's progression follows the same sequence for all characters.

Regarding the game's rules and mechanics, we defined four learning objectives as components of the overarching learning objective, that was to "Distinguish legitimate emails from phishing attempts". These objectives are: (1) Identifying whether the email is from a company the user is a client of or from a known person; (2) Verifying whether the email address domain matches those typically used by the sender; (3) Detecting requests for personal information, clicking on links, or downloading attachments; and (4) Recognizing message content that pressures the user to make a quick decision under the threat of a penalty. These objectives we proposed based on what we identified in previous work [Arachchilage and Cole, 2011; Wash, 2020; Abreu *et al*., 2024; Henderson *et al*., 2024] and in the NIST Phishing Scale [Dawkins and Jacobs, 2023] as key and basic behaviors one should exhibit to prevent phishing attacks. We believe future work could rely solely on this task to create the best possible set of learning objectives, following the guidelines of Cortegoso and Coser [2023].

To achieve these objectives, the game mechanics involve clicking buttons to indicate whether the email displayed on the screen is legitimate or contains phishing. During training, players receive immediate feedback: a message confirming a correct response or indicating an error, which enhances the feedback's effectiveness as a reinforcer. Each training phase includes three emails, and each email can be answered up to two times. Players have a maximum of six attempts to correctly evaluate all three emails, otherwise they should repeat the whole phase. An attempt consists of selecting either "legitimate" or "phishing" for one email. To progress to the next phase, players must correctly identify at least two emails in each phase; otherwise, they repeat the phase as many times as necessary. In the pre-test and post-test phases, players can respond only once per email and do not receive feedback on whether their answers are correct.

The feedback provided by the game is limited to indicating only correct or incorrect responses, without explaining why an email is or is not phishing. A potential issue with this design is that a participant might complete a task correctly, but not for the intended reasons. Nonetheless, we believe the likelihood of this occurring was low, as each set of emails was directly linked to a single learning objective and therefore contained only one type of phishing-indicative stimulus. In any case,

this remains a limitation that should be acknowledged.

This game was developed in C# using the Unity 3D engine, chosen for its versatility and ease of use [Unity, 2025]. This allowed us to create an interactive and easily maintainable game. The phase configuration and email data management were implemented using Scriptable Objects. The PhaseConfig class was used to store email data, including the sender, content, and a boolean indicator identifying whether the email is phishing. Importantly, the game was designed to be extensible: new emails or even new scenarios can be easily added through the PhaseConfig script, without requiring changes to the core game logic. We also implemented a field to define the score required to advance to the next phase. The EmailManager class was responsible for managing email display, verifying player responses, and updating the score. The results from each test and training phase are saved locally in a text file. The game's interface was designed to be intuitive. Unread emails are highlighted in red, while read emails turn green. Feedback messages appear as pop-ups that can only be dismissed when the player clicks the corresponding button, demanding what is known as an observation response. This term designates a behavior that needs to be emitted for a desired condition to become available. In the case of the game, the desired condition is to continue playing. To do this, one needs to click on the pop-up window, which forces the player to read the message about whether he got the attempt right or wrong.

4.1.2 Participants

Pilot Study - 1st Iteration: To improve the data collection method, a pilot study was conducted with four senior computer science students (P1 to P4). **Experiment I - 2nd Iteration:** The Experiment I involved 27 participants (P5 to P31) with an average age of 25.81 years (SD = 7.50), ranging from 20 to 49 years old, 63% of whom were men. Most participants had some level of higher education in progress (66.70%), and they were typically students of Computer Science (37.00%), Law (22.21%), or Dentistry (11.10%). For the analysis, we considered only data from participants whose pre-test scores were below 85%, to mitigate ceiling effects. We excluded P5, P9, P12, P15, P17, P18, P22, P25, and P30, whose average pre-test score was 94.44%. The final sample included 18 participants. All participants signed an informed consent form (ICF). Our study was approved by the Research Ethics Committee.

4.1.3 Instruments

We used the following instruments: (1) Sample characterization questionnaire: assessed age, gender, education and course name, in the case of college students or graduates; (2) Learning measure: consisting of 12 e-mail evaluation tasks, pre- and post-test, all making up the Alerta game. We also created a scale to measure the player's level of confidence in their email rating responses, filled in on a sheet. The scale was: "1 - Totally Unsafe", "2 - Partially Unsafe", "3 - Partially Safe", and "4 - Completely Safe"; (3) Post-Study System Usability Questionnaire (PSSUQ): this is a tool adopted in the literature [Vlachogianni and Tselios, 2023], with 19 items in its original version [Lewis *et al*., 1990]. We used the reduced version of Azlan and Junaini [2023] with 16 items, but due to an error, Item 15 — "This system has all the functions and capabilities I expect it to have" — was not administered, which represents a limitation regarding this particular data point. Responses to PSSUQ are given on a Likert scale of agreement, "1 - Strongly Disagree" to "5 - Strongly Agree". We used this instrument, but without previously conducting a cross-cultural adaptation study to Brazilian population, which is another limitation of our study; (4) Satisfaction Test: consisting of 5 items that evaluated the game, also on a Likert scale of agreement. This test was not previously validated, which is also a limitation of our study. The items were: "Overall, I am satisfied with the..." (1) "games's story", (2) "games's narrative", (3) "games's rules", (4) "gameplay", and (5) "game Alerta".

4.1.4 Data collection and analysis procedure

In the pilot study, corresponding to the first iteration of the game development, all participants were asked to sign the ICF and were provided with instructions regarding the game's story and objectives. Based on feedback received from the four participants and our observations, we implemented improvements to the game to develop its second version. We modified the storytelling approach by integrating it into a dialogue with the Human Resources manager. We added visual indicators to show which emails had been read and established a rule requiring at least two correct answers during the training phase to progress to the next level. Additionally, we changed how levels were introduced by including the level identification and reminders about the functionality of each button.

During this first iteration, we obtained evidence of the game's efficiency. None of the participants required all six attempts to correctly identify the emails in the training phases, and the number of attempts decreased across the levels. Moreover, in Phase 1, the average accuracy was 83.34%; in Phase 2, it was 100%; in Phase 3, it was 91.67%; and in Phase 4, it was 100%. These results suggest preliminary evidence that the learning objectives were effectively conveyed during the training, as indicated by the programmed feedback for correct and incorrect responses. When comparing pre- and post-test performance, all four participants showed improvement. The average pre-test accuracy was 68.75%, reaching 100% in the post-test. These findings highlight the game's effectiveness in promoting the learning objectives. Lastly, the usability (Mean = 4.71; SD = 0.11) and satisfaction scores (Mean = 4.75; SD = 0.19) were high, with minimal standard deviations.

For the Experiment 1, corresponding to the second iteration of the game's development, all participants were asked to read and sign the ICF. The objective and story of the game Alerta were then explained. The researcher clarified that while he would not provide answers, he could assist in the event of technical issues. Participants were instructed to focus on the task and were prohibited from interacting with others. They were also informed that, for each email analyzed during the pre- or post-test, they needed to complete a form indicating their confidence level in their responses. Participants were encouraged to complete the game and were advised that they would be asked to evaluate their experience at the end.

After receiving the instructions, all participants began playing Alerta in a controlled laboratory environment. The study was conducted collectively and lasted approximately one hour. During this time, participants completed the pre-test, four game phases, and the post-test. The collected data were analyzed using a mixed-methods approach, emphasizing descriptive statistics and the Wilcoxon signed-rank test to compare pre- and post-test results. The mean accuracy of the training phases that we exhibit was calculated only when the participant got two emails right, and from each of the three e-mails, we considered the first when the participant got it right or the result of the second presentation of that email.

4.2 Experiment 2

Experiment 2 was conducted during the third iteration of the game Alerta and aimed to evaluate the impact of recent improvements in gameplay mechanics, user interface, and educational feedback. This phase built upon the lessons learned from Experiment 1 and introduced new features designed to deepen the learning experience and improve usability.

4.2.1 Computational solution

For the new version of the game, used in the third iteration, improvements were implemented in both functionality and interface design. The key innovation was the feature allowing players to assign a confidence level to each analyzed email using four distinct ratings: totally insecure, partially insecure, partially secure, and totally secure. This addition aims to enhance the educational experience by enabling a more detailed evaluation by the player and providing richer data for future analysis. Furthermore, mechanisms were developed to ensure that the confidence rating selection resets correctly when switching between emails. On the visual side, we made adjustments to optimize the interface, delivering a more intuitive and enjoyable user experience. A key aspect of this redesign was enhancing the resemblance to widely used email services. This decision aimed to make interactions within the game more familiar to users. The visual changes also can help players apply their learning in real-world scenarios more effectively.

Examples of game screens can be found in **Appendix 2** and **Appendix 3**. The code is available on GitHub https://github.com/JassonJr1/alerta_prototype.

4.2.2 Participants

Experiment 2 - 3rd Iteration: Participated 35 people (P32 to P66) with an average age of 22.91 years (SD = 2.36), ranging from 19 to 27 years old, 66.67% of whom were men. Most participants had some level of higher education in progress (80%), and they were typically students of Psychology (57,14%) and Computer Science (25,71%). We only considered the data from participants whose pre-test scores were below 85%. We excluded P34, P39, P40, P41, P43, P44, P45, P49, P56, P57, P58, P59, P60, P61, P62, P63, P64, P65, and P66 whose average pre-test score was 95.18%. Additionally, P33 was excluded after we identified a technical issue with the computer he was using during the test, which led to the loss of pre-test data. All participants signed an informed consent form (ICF). The final sample included 15 participants.

4.2.3 Instruments

The same instruments were adopted from Experiment I. The only change was in the safety assessment which, in this experiment, was included in the game.

4.2.4 Data collection and analysis procedure

In Experiment 2, corresponding to Alerta third iteration, we adopted the same procedure as in Experiment 1.

5 Results and discussion

This section presents and analyzes the results obtained from the two experiments conducted with the educational game Alerta. The analysis focuses on evaluating the game's efficiency, mainly, by examining performance differences between pre- and post-tests.

5.1 Experiment I - 2nd Iteration

Table 2 displays the results of the four training phases. We observed a subtle improvement in participants' performance as they progressed through these phases, with an average accuracy increasing from 90.74% in Phase 1 to 98.15% in Phase 4. Fewer errors occurred, and fewer attempts were needed to advance through the phases. We would like to emphasize that each phase has three emails. An email is only repeated when the participant makes a mistake and a phase is only repeated when the participant makes two mistakes in two of the three emails. Some participants (P11, P16, P19, and P21), who made mistakes, did not use all the available attempts. They were able to advance to the next phase because they had correctly answered at least two emails. It is worth noting that P28's computer shut down while starting the repetition of Phase 1, after not meeting the minimum criterion. Only the pre-test data was stored. According to the game's programming, the participant had to repeat the pre-test, but we did not analyze this data. The participant then repeated Phase 1, which was already necessary.

These data suggests that the four learning objectives were consistently conveyed and reinforced. Similarly to the game by Farias *et al*. [2019], we obtained favorable preliminary evidence that the Alerta training was suitable for individuals with different skills, allowing them to progress through the game phases without difficulty. Alerta advanced over other games by explicitly stating learning objectives and building the game around them [Panosso *et al*., 2015]. With the available data, we were still unable to determine why a participant required more attempts in a phase when compared to others. Hypotheses to be investigated in future studies include: (a) fatigue and/or lack of attention; and (b) the complexity of the behavior being trained.

Table 3 shows the results of the pre and post-tests. The average pre-test performance was 73.61 (SD = 15.46) and increased to 89.35 (SD = 8.95) in the post-test, which was statistically significant and had a large effect size (W = 153.00;

Table 2. Experiment 1 - Results of participants in the 4 phases of the 2nd iteration of the game.

Participant	Phase 1		Phase 2		Phase 3		Phase 4	
	Hits	Attempts	Hits	Attempts	Hits	Attempts	Hits	Attempts
P6	100.00	3	100.00	5	100.00	4	100.00	4
P7	100.00	3	100.00	3	100.00	3	100.00	3
P8	100.00	3	100.00	3	100.00	3	100.00	3
P10	66.67	4	100.00	3	100.00	4	100.00	3
P11	66.67	3	66.67	3	66.67	3	66.67	3
P13	100.00	5	100.00	4	100.00	4	100.00	3
P14	100.00	4	100.00	3	100.00	3	100.00	3
P16	66.67	3	66.67	3	66.67	3	100.00	3
P19	66.67	3	66.67	3	66.67	3	100.00	3
P20	100.00	5	100.00	3	100.00	3	100.00	3
P21	100.00	3	66.67	3	100.00	4	100.00	3
P23	100.00	6	100.00	6	100.00	3	100.00	4
P24	100.00	4	100.00	3	100.00	4	100.00	3
P26	100.00	4	100.00	3	100.00	4	100.00	3
P27	100.00	4	100.00	3	100.00	3	100.00	3
P28	100.00	6	100.00	4	100.00	4	100.00	3
P29	100.00	3	100.00	3	100.00	3	100.00	3
P31	66.67	5	66.67	4	100.00	4	100.00	3
Average	**90.74**	**3.94**	**90.74**	**3.44**	**94.45**	**3.44**	**98.15**	**3.11**
SD	**15.36**	**1.06**	**15.36**	**0.86**	**12.78**	**0.51**	**7.86**	**0.32**

$z = 3.621$; $p < 0.001$; $r = 1.00$; $SE = 0.269$). This result probably stems from the successful training across the four phases and serves as preliminary evidence of the game's efficiency in teaching the behavior of "differentiating legitimate emails from phishing attempts". The pre-test data showed that all participants already had part of this behavior in their repertoire, and P19 did not show improvement (perhaps due to insufficient training, as seen in **Table 2**). However, in general, the game seemed useful in improving this repertoire, which represents a form of learning. Therefore, we found that technologies like Alerta may improve cybersecurity behaviors and possibly mitigate phishing attacks, as required by the scientific community [Rahman *et al*., 2020; Guilherme *et al*., 2021].

Table 4 displays the recorded confidence levels regarding the responses provided in the pre and post-tests. The data for P16 is unavailable because this participant incorrectly completed the instrument, evaluating the confidence inspired by the email itself rather than the confidence in his/her own responses. In the group, we observed an improvement in the average confidence score from 57,35% ($SD = 28,55$) in the pre-test to 69,61% ($SD = 31,17$) in the post-test, accompanied by low standard deviation, suggesting homogeneity in confidence scores. This positive change, although it was not statistically significant ($W = 83.50$; $z = 1.946$; $p = 0.055$; $r = 0.590$; $SE = 0.294$), still suggests the efficiency of Alerta in promoting self-confidence, since the effect size was high. We probably did not observe a statistically significant p-value because the sample size was small.

These data suggest that the Alerta game may be useful in reducing users' doubts about phishing and enhancing their perception of being capable of independently handling the task of identifying phishing emails — something that, at times, even computer science students do not feel confident about [Kovacevic *et al*., 2020]. It is worth noting that P20, P27, P28, P29, and P31 did not show improvement, highlighting the need for further refinement of the game and the data collection process to better understand these cases. In summary, the data on the improvement of the behavior of "distinguishing legitimate emails from phishing attempts" and on self-confidence suggest that the educational game Alerta showed preliminary evidence of efficiency and it may have contributed to the development of the basic elements of critical thinking, as proposed by Wash [2020], which is crucial for individuals to learn how to avoid phishing attacks.

Finally, **Table 5** presents data from the usability and satisfaction perception assessment instruments, both with a maximum score of 5. The average usability score (4.52, $SD = 0.65$) and the average satisfaction score (4.76, $SD = 0.52$) were high, suggesting that the educational game was easy to use and generated positive feelings and perceptions, serving as an indirect measure of the engagement it can promote. These findings suggest that Alerta, in addition to showing preliminary evidence of efficiency in promoting learning and self-confidence, has the potential to generate engagement, thereby fulfilling the basic requirements of an educational game [Tsutsumi *et al*., 2020].

5.2 Experiment 2 - 3rd Iteration

Table 6 presents the performance results of the 15 participants throughout the four phases of the Alerta game, including accuracy rates and the total number of attempts used in each phase. In Phase 1, participants achieved an average accuracy rate of 93.33% ($SD = 13.80$), with an average of 4.00 attempts

Table 3. Experiment 1 - Percentage of correct answers in the pre and post-test of the 2nd iteration of the game.

Participant	% Hits		Difference Pre-Post	Participant	% Hits		Difference Pre-Post
	Pre	Post			Pre	Post	
P6	50.00	75.00	25.00	P20	25.00	75.00	50.00
P7	83.33	91.67	8.33	P21	75.00	91.67	16.67
P8	83.33	100.00	16.67	P23	83.33	91.67	8.33
P10	75.00	100.00	25.00	P24	75.00	83.33	8.33
P11	58.33	75.00	16.67	P26	75.00	83.33	8.33
P13	83.33	91.67	8.33	P27	83.33	91.67	8.33
P14	66.67	100.00	33.33	P28	75.00	83.33	8.33
P16	83.33	91.67	8.33	P29	83.33	100.00	16.67
P19	83.33	83.33	0.00	P31	83.33	100.00	16.67
Pre-test average (SD)			**Post-test average (SD)**				**Average difference (SD)**
73.61 (*SD* = 15.46)			89.35 (*SD* = 8.95)				15.74 (*SD* = 11.75)

Table 4. Experiment 1 - Degree of confidence in responses to the pre and post-test of the 2nd iteration of the game.

Participant	% Confidence		Difference Pre-Post	Participant	% Confidence		Difference Pre-Post
	Pre	Post			Pre	Post	
P6	8.33	66.67	58.33	P20	100.00	0.00	-100.00
P7	50.00	100.00	50.00	P21	58.33	75.00	16.67
P8	66.67	100.00	33.33	P23	75.00	83.33	8.33
P10	58.33	66.67	8.33	P24	83.33	91.67	8.33
P11	0.00	25.00	25.00	P26	66.67	75.00	8.33
P13	33.33	83.33	50.00	P27	50.00	50.00	0.00
P14	50.00	100.00	50.00	P28	25.00	25.00	0.00
P16	---	---	---	P29	100.00	100.00	0.00
P19	83.33	100.00	16.67	P31	66.67	41.67	-25.00
Pre-test average (SD)			**Post-test average (SD)**				**Average difference (SD)**
57.35 (*SD* = 28.55)			69.61 (*SD* = 31.17)				12.25 (*SD* = 36.69)

Table 5. Experiment 1 - Usability and satisfaction results from the 2nd iteration of the game.

Participant	Scores		Participant	Scores	
	PSSUQ	Satisfaction		PSSUQ	Satisfaction
P6	4.87	5.00	P20	4.80	5.00
P7	5.00	5.00	P21	4.93	5.00
P8	5.00	5.00	P23	4.80	4.80
P10	4.60	5.00	P24	4.93	5.00
P11	4.67	5.00	P26	4.87	5.00
P13	3.00	4.00	P27	4.40	4.80
P14	3.60	4.60	P28	3.00	3.00
P16	4.33	4.40	P29	5.00	5.00
P19	4.53	5.00	P31	5.00	5.00
Average Usability: 4.52 (*SD* = 0,65)			**Average Satisfaction:** 4.76 (*SD* = 0.52)		

(*SD* = 0.85). From Phase 2 onward, all participants achieved 100% accuracy. In Phase 2, the average number of attempts decreased to 3.53 (*SD* = 0.64). In Phase 3, this average further decreased to 3.40 (*SD* = 0.63) and in Phase 4 to 3.27 (*SD* = 0.59), indicating continued improvement in participant performance.

These findings also suggest that as participants progressed through the phases, they not only improved their accuracy rates but also reduced the number of attempts needed to complete each stage, providing preliminary evidence of progressive and consistent learning. These data replicates our findings in Experiment 1, expanding the empirical base of evi-

Table 6. Experiment 2 - Results of participants in the 4 phases of the 3rd iteration of the game.

Participant	Phase 1		Phase 2		Phase 3		Phase 4	
	Hits	Attempts	Hits	Attempts	Hits	Attempts	Hits	Attempts
P6	100.00	3.00	100.00	3.00	100.00	3.00	100.00	3.00
P7	100.00	6.00	100.00	5.00	100.00	3.00	100.00	5.00
P8	66.67	4.00	100.00	4.00	100.00	3.00	100.00	3.00
P10	100.00	5.00	100.00	3.00	100.00	3.00	100.00	3.00
P11	100.00	3.00	100.00	4.00	100.00	3.00	100.00	3.00
P13	100.00	5.00	100.00	3.00	100.00	5.00	100.00	3.00
P14	100.00	3.00	100.00	4.00	100.00	4.00	100.00	3.00
P16	100.00	4.00	100.00	3.00	100.00	3.00	100.00	3.00
P19	66.67	4.00	100.00	4.00	100.00	4.00	100.00	4.00
P20	100.00	4.00	100.00	3.00	100.00	3.00	100.00	3.00
P21	100.00	3.00	100.00	3.00	100.00	3.00	100.00	3.00
P23	100.00	4.00	100.00	3.00	100.00	4.00	100.00	3.00
P24	66.67	4.00	100.00	4.00	100.00	4.00	100.00	4.00
P26	100.00	4.00	100.00	3.00	100.00	3.00	100.00	3.00
P27	100.00	4.00	100.00	4.00	100.00	3.00	100.00	3.00
Average	**93.33**	**4.00**	**100.00**	**3.53**	**100.00**	**3.40**	**100.00**	**3.27**
SD	**13.80**	**0.85**	**0.00**	**0.64**	**0.00**	**0.63**	**0.00**	**0.59**

dence in favor of our educational game.

Table 7 presents the performance results of the participants selected for analysis in the third iteration of the Alerta game. In the pre-test, the average accuracy was 80.00% (*SD* = 6.14), ranging from 66.67 to 83.33, indicating moderate initial performance with room for improvement. After training, the average accuracy in the post-test increased to 91.67% (*SD* = 8.33), ranging from 75.00 to 100.00, reflecting an average improvement of 11.67 percentage points (*SD* = 9.34). This difference was statistically significant and had a large effect size ($W = 78.00$, $z = 3.059$, $p = 0.002$, $r = 1.00$; $SE = 0.269$).

In particular, six participants (P32, P38, P42, P46, P50, and P51) achieved 100% accuracy in the post-test, suggesting complete development of the trained behaviors. We found three participants, P48, P53, and P54, that had no gains, and six participants, P35, P36, P37, P47, P52, and P55, that showed lower gains (8.33 percentage points). All the other participants obtained more than 15 percentage points gain. These results further reinforce the efficiency of the Alerta training, particularly when focusing on participants with greater potential for improvement.

Table 8 shows the results related to the confidence level perceived by participants in their responses during the pre and post-tests. In the pre-test, the average confidence level was 72.22% (*SD* = 24,53). In the post-test, we observed an increase in the average confidence level, reaching 82.78% (*SD* = 28.43). This increase reflects a gain in confidence after completing the training phases, with notable highlights for P38 and P52 who showed an improvement of more than 40 percentage points. The average gain was 10.56 (*SD* = 25.29) percentage points, indicating a low to moderate improvement in participants' confidence levels. Again, we did not found a statistically sound difference between pre and post-test, but did find a high effect size ($W = 53.50$, $z = 1.823$, $p = 0.075$, $r = 0.621$; $SE = 0.269$). In Experiment 2, the sample was even smaller compared to Experiment 1. The small sample size may have been the reason why the p-value was not significant.

We noted that the integration of this functionality of collecting data about confidence into the game may have contributed to a smoother experience and a clearer measure of confidence levels when compared to Experiment 1. Our hypothesis is that by increasing the number of participants in future studies, the statistically significant effect may become clearer, since the effect size is already large. In any case, we need to develop more data collection instruments so that we can identify why some participants did not improve their self-confidence, even though they improved their performance.

Finally, **Table 9** presents the results obtained from the evaluation of Usability and Satisfaction concerning the game after the third iteration. These data were collected through questionnaires administered at the end of the phases.

The participants provided positive evaluations of the game's usability (*Mean* = 4.75, *SD* = 0.27). Overall satisfaction achieved an average of 4.89 (*SD* = 0.21), exceeding expectations and demonstrating the technical and pedagogical success of the game, with several participants ($n = 11$) awarding the maximum score.

5.3 General Discussion

The results obtained in Experiments 1 and 2 suggest that Alerta may be efficient in promoting learning (behavioral change) regarding the behavior of "distinguishing legitimate emails from phishing attempts". As noted in previous research, traditional cybersecurity awareness methods, such as informational campaigns, have shown limited efficiency in significantly reducing unsafe behaviors [Alanazi *et al*., 2022; Guilherme *et al*., 2021]. In contrast, digital educational games provide an engaging and interactive learning environment that can reinforce safe behaviors through immediate feedback and repetitive practice [Panosso *et al*., 2015; Farias *et al*., 2019; Tsutsumi *et al*., 2020].

Table 7. Experiment 2 - Percentage of correct answers in the pre and post-test of the 3rd iteration of the game.

Participant	% Hits		Difference Pre-Post
	Pre	Post	
P32	75.00	100.00	25.00
P35	66.67	75.00	8.33
P36	83.33	91.67	8.33
P37	83.33	91.67	8.33
P38	83.33	100.00	16.67
P42	66.67	100.00	33.33
P46	83.33	100.00	16.67
P47	83.33	91.67	8.33
P48	83.33	83.33	0.00
P50	83.33	100.00	16.67
P51	83.33	100.00	16.67
P52	75.00	83.33	8.33
P53	83.33	83.33	0.00
P54	83.33	83.33	0.00
P55	83.33	91.67	8.33
Average	**80.00**	**91.67**	**11.67**
SD	**6.14**	**8.33**	**9.34**

Table 8. Experiment 2 - Degree of confidence in responses to the pre and post-test of the 3rd iteration of the game.

Participant	% Confidence		Difference Pre-Post
	Pre	Post	
P32	83.33	100.00	16.67
P35	16.67	50.00	33.33
P36	91.67	100.00	8.33
P37	83.33	83.33	0.00
P38	58.33	100.00	41.67
P42	75.00	16.67	-58.33
P46	66.67	83.33	16.67
P47	75.00	91.67	16.67
P48	66.67	91.67	25.00
P50	100.00	100.00	0.00
P51	33.33	25.00	-8.33
P52	50.00	100.00	50.00
P53	83.33	100.00	16.67
P54	100.00	100.00	0.00
P55	100.00	100.00	0.00
Average	**72.22**	**82.78**	**10.56**
SD	**24.53**	**28.43**	**25.29**

We believe that our game's mechanics, which required players to analyze emails and that delivered immediate feedback upon players' response, was responsible for the learning outcomes. Aligned with behavior-analytic learning principles [Skinner, 1981; Moreira and Medeiros, 2018], Alerta selects and strengthens the desired behaviors through well-structured reinforcement contingencies programmed in the context of the game.

The comparison between pre and post-test results in both experiments also suggests some improvement in participants' self-confidence concerning their ability to distinguish legitimate emails from phishing attempts. Our finding corroborates the idea that structured educational interventions can enhance users' trust in their decision-making abilities regarding cyber-

Table 9. Experiment 2 - Usability and satisfaction results from the 3rd iteration of the game.

Participant	Scores	
	Usability	Satisfaction
P32	4.80	5.00
P35	4.13	5.00
P36	5.00	5.00
P37	4.87	5.00
P38	4.87	4.80
P42	4.67	4.40
P46	5.00	5.00
P47	4.87	5.00
P48	4.47	5.00
P50	5.00	5.00
P51	5.00	5.00
P52	4.40	4.80
P53	4.40	4.40
P54	4.93	5.00
P55	4.80	5.00
Average	**4.75**	**4.89**
SD	**0.27**	**0.21**

security threats [Kovacevic *et al*., 2020]. Combined, learning how to identify phishing and better self-confidence in this ability, are key to reduce society susceptibility to cyberattacks, and we found, as expected by theory [Moreira and Medeiros, 2018; Kienen *et al*., 2021; Wash, 2020], that explicit training could help to boost both.

The usability and satisfaction scores suggested that Alerta may successfully combines ease of use with an engaging experience. These factors are critical for educational games, as highlighted by [Tsutsumi *et al*., 2020], who argued that usability directly influences engagement and, consequently, the efficiency of learning. We noted a slight difference between the scores of usability and satisfaction in Experiments 1 and 2, favoring the modifications implemented in the Alerta third iteration. Thus, the third iteration seems to have been a more integrated, intuitive, and satisfying experience for participants.

Compared to traditional cybersecurity training methods, such as cybersecurity awareness campaigns, which have a limited impact on behavior [Henderson *et al*., 2024], Alerta presents advantages in terms of scalability and efficiency. Additionally, Alerta can replicate real world scenarios because it incorporates different phishing tactics such as authority-based messages [Butavicius *et al*., 2016], which further enhance its pedagogical relevance.

Together, the data on learning, self-confidence, usability, and satisfaction suggest that the Alerta game shows promise applications to teach people how to deal with phishing. However, there are limitations in our study that should be considered: (1) the small number of participants with low variability in characteristics and cybersecurity repertoires, as well as generally high accuracy rates in the pre-test. For future studies, we suggest collecting data from a more diverse participant pool with less than 60% accuracy in the pre-test to allow for a more robust analysis of the game's effectiveness; (2) the narrow scope of the learning objectives covered by the game, making it important to include more complex objectives that empower people to handle more challenging phishing techniques, such as spear phishing; (3) the use of instruments without robust psychometric evidence, which requires specific studies to construct and adapt measures for research and intervention in the field of cybersecurity education; (4) the absence of instruments that would allow a more precise assessment of the reasons why some participants improved their performance more than others; (5) the absence of feedback explaining why a response was correct or incorrect, which may have prevented participants from accurately discriminating what made an email a phishing attempt.

Despite these limitations, the results appear to have been promising and, in line with the iterative game design approach, we are encouraged to continue refining and evaluating our game in future versions. Additionally, we suggest that the forth iteration of the game development include the following improvements: (1) implement improvements in aesthetics and narrative to make the game more engaging and attractive to players, encouraging them to want to play; (2) improve usability to accommodate individuals with reading difficulties; (3) enhance the feedback provided during the game to increase its reinforcing value, and its content in order to show why a response is correct or incorrect; and (4) adopt instruments with robust favorable psychometric evidence. With these improvements, we hope that Alerta will become an even

more efficient educational tool.

6 Conclusion

Our objective was to evaluate the efficiency of the Alerta educational game in teaching the behavior of "distinguishing legitimate emails from phishing attempts". Through three iterations of prototype game development, focusing on rules, mechanics, and technology, we observed promising results. Participants were able to progress through the training phases without difficulty and showed subtle improvements in performance. A comparison between the pre and post-test revealed preliminary evidence of learning the proposed objectives for the game, as well as the development of self-confidence (for some participants) in their ability to recognize phishing attempts. Finally, we observed that the game was perceived as easy to use and provided a satisfying experience. These results align with our initial proposal, which was to create a game through which people could learn relevant behaviors in a comfortable and enjoyable way, making it simple for them to download and play the game without external assistance.

In summary, our study advances previous investigations [Arachchilage and Cole, 2011; Farias *et al*., 2019; Henderson *et al*., 2024; Abreu *et al*., 2024] by clearly grounding its approach in a psychological theory of human behavior and learning. Based on this theory, we defined the learning objectives of the game, which guided the design of both pre- and post-test assessment items — our main indicators of learning. These objectives focused specifically on phishing detection, a key concern in cybersecurity education that is not always addressed or carefully considered in the reviewed studies [Farias *et al*., 2019; Bernardino *et al*., 2023; Abreu *et al*., 2024]. We selected four learning objectives and developed three emails to assess them at each stage of the game (pre-test, training, and post-test), representing a level of methodological clarity and standardization not commonly found in the literature, which supports replication. Finally, we conducted two tests of our game, contributing with empirical data to advance the literature, as not all reviewed studies included an evaluation of their approach [Arachchilage and Cole, 2011; Henderson *et al*., 2024].

We suggest that future studies build on this research, addressing the limitations we have outlined. In addition to our suggestions for the fourth iteration of the game, we recommended data collection with a more diverse group of participants and expand the scope of the learning objectives developed through the game, so that it can help society mitigate phishing attacks, which are becoming increasingly frequent and sophisticated. We also consider it promising to test the game with specific populations, such as the elderly people. We acknowledge that this study still has an exploratory character, and much remains to be improved before a truly robust tool can be developed. Nonetheless, we hope to have contributed to the advancement of the field and to have encouraged further research on cybersecurity education.

Declarations

Authors' Contributions

JMFJ, MHOH, FLL, and ELF contributed to the conception, methodology, investigation, formal analysis, and writing of this study. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES-PROEX) - Finance Code 001. This work was partially supported by Amazonas State Research Support Foundation - FAPEAM - through the POSGRAD project 2024/2025.

Availability of data and materials

The instruments generated and/or analyzed during the current study are available upon request to the first author.

AI Use Statement. In this manuscript, the ChatGPT was used to assist in translating scientific text into English. All final editorial decisions and content validations were made solely by the authors, who retain full responsibility for the accuracy and integrity of the work.

References

Abreu, J., Miani, R., and Nomura, S. (2024). "encrypta: A missão da liga dos robôs" - um jogo educacional para aprendizagem em cibersegurança. In *Anais do XXIII Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 1327–1338, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbgames.2024.240133.

Alanazi, M., Freeman, M., and Tootell, H. (2022). Exploring the factors that influence the cybersecurity behaviors of young adults. *Computers in Human Behavior*, 136:1–14. DOI: 10.1016/j.chb.2022.107376.

APWG (2023). Phishing activity trends reports. Available at:`https://bit.ly/4cEA85K`.

APWG (2024). Phishing activity trends reports. Available at:`https://bit.ly/4dC6GgW`.

Arachchilage, N. A. G. and Cole, M. (2011). Design a mobile game for home computer users to prevent from "phishing attacks". pages 485–489. DOI: 10.1109/i-Society18435.2011.5978543.

Azlan, Z. H. Z. and Junaini, S. N. (2023). Erudite survivor: Usability testing of a gamification-based mobile app for disaster awareness among children. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 31(3):290298. Acesso em: 27 dez. 2024. DOI: 10.37934/araset.31.3.290298.

Bernardino, I., Bidarra, J., Baptista, R., and Mamede, H. (2023). Desenvolvimento do jogo sério web segura: Estudo

de um caso orientado para públicos seniores. *Rotura – Revista de Comunicação, Cultura e Artes*, 3(1):74–101. DOI: 10.34623/ewxe-ek97.

Butavicius, M., Parsons, K., Pattinson, M., and McCormac, A. (2016). Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *arXiv:1606.00887*. DOI: 10.48550/ARXIV.1606.00887.

CERT.br (2024). Páginas falsas utilizadas em tentativas de phishing. Available at:`https://stats.cert.br/phishing/`.

Cortegoso, A. L. and Coser, D. S. (2023). *Elaboração de programas de ensino: Material autoinstrutivo*. EdUFSCar. Book.

Dawkins, S. and Jacobs, J. (2023). Nist phish scale user guide. DOI: 10.6028/NIST.TN.2276.

Diorio, R. F., Serafim, E., Alves, K. R., and Meira, M. C. (2018). Segurança da informação e de sistemas computacionais: Um estudo prático sobre ataques utilizando malwares. In *Anais SULCOMP*. Available at:`https://bit.ly/45Oj62J`.

Farias, F. L. d. O., Medeiros, N. A. A. d., Rocha, S. L. d., Medeiros, D. F. d., Nóbrega, E. C. d., Burlamaqui, A., and Madeira, C. (2019). Self protect: Um jogo para auxílio no ensino de conceitos relacionados a segurança na internet para crianças e adolescentes. In *Anais do XXV Workshop de Informática na Escola*, pages 246–255. DOI: 10.5753/cbie.wie.2019.246.

Gris, G. and De Souza, S. R. (2016). Digital educational games and model of network relations: Development and evaluating of the physical prototype of korsan game. *Perspectivas em Análise do Comportamento*, 7(1):114–132. DOI: 10.18761/pac.2016.003.

Guilherme, L. P., Ferreira, M. F., Da Fonseca, G. M., and Lazarin, N. M. (2021). Uma breve noção sobre o comportamento dos internautas em relação à segurança na rede. In *Escola Regional de Sistemas de Informação do Rio de Janeiro (ERSI-RJ)*, pages 1–7. DOI: 10.5753/ersirj.2021.16972.

Henderson, N., Pallett, H., Linden, S., Montanarini, J., and Buckley, O. (2024). The disphishinformation game: Creating a serious game to fight phishing using blended design approaches. In *AHFE 2024 International Conference*, volume 127, USA. AHFE International. DOI: 10.54941/ahfe1004774.

Kienen, N., Panosso, M. G., Nery, A. G. S., Waku, I., and Carmo, J. S. (2021). Contextualização sobre a programação de condições para desenvolvimento de comportamentos (pcdc): Uma experiência brasileira. *Perspectivas em Análise do Comportamento*, 12(2):360–390. DOI: 10.18761/PAC.2021.jul110.

Kovacevic, A., Putnik, N., and Toskovic, O. (2020). Factors related to cyber security behavior. *IEEE Access*, 8:125140–125148. DOI: 10.1109/access.2020.3007867.

Lewis, J. R., Rieman, J. A., and Wharton, C. (1990). Integrated office software benchmarks: A case study. In *Proceedings of the CHI '90 Conference on Human Factors in Computing Systems*. ACM. Conference Paper.

Melo, L. P., Amaral, D. M., Sakakibara, F., De Almeida, A. R., De Sousa Jr, R. T., and Nascimento, A. (2011). Análise de malware: Investigação de códigos maliciosos através de uma abordagem prática. In *Minicursos do XI Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais*, pages 9–52. DOI: 10.5753/sbc.9559.1.1.

Moreira, B. M. and Medeiros, A. C. (2018). *Princípios básicos de análise do comportamento*. Artmed. Book.

Morgado, E. M., Távora, C. G., Lima, A. C. P. F. d., Albino, J. P., and Bucchianico, I. (2023). Caso de cyber fraud por telefone no brasil e a inteligência artificial. In *Inteligência Artificial e suas Aplicações Interdisciplinares*, page 113–126. Editora e-Publicar. DOI: 10.47402/ed.ep.c202321007201.

Panosso, M. G., Souza, S. R., and Haydu, V. B. (2015). Características atribuídas a jogos educativos: Uma interpretação analítico-comportamental. *Revista Quadrimestral da Associação Brasileira de Psicologia Escolar e Educacional*, 19(2):233–241. DOI: 10.1590/2175-3539/2015/0192821.

Rahman, N. A. A., Sairi, I. H., Zizi, N. A. M., and Khalid, F. (2020). The importance of cybersecurity education in school. *International Journal of Information and Education Technology*, 10(5):378–382. DOI: 10.18178/ijiet.2020.10.5.1393.

Rajasekharaiah, K. M., Dule, C. S., and Sudarshan, E. (2020). Cyber security challenges and its emerging trends on latest technologies. *IOP Conference Series: Materials Science and Engineering*, 981(2):022062. DOI: 10.1088/1757-899X/981/2/022062.

Scarfone, K. A., Souppaya, M. P., Cody, A., and Orebaugh, A. D. (2008). Technical guide to information security testing and assessment. DOI: 10.6028/nist.sp.800-115.

Schell, J. (2015). *The art of game design: A book of lenses*. CRC Press. DOI: 10.5860/choice.46-3898.

Skinner, B. F. (1981). *Ciência e comportamento humano*. Martins Fontes, São Paulo. Book.

Souza, L. C. and Tanaka, S. S. (2023). Estudo sobre ataques de phishing e suas técnicas de defesa. *Revista Terra & Cultura: Cadernos de Ensino e Pesquisa*, 39(Especial):90–95. Available at:`https://bit.ly/4cOisV2`.

Stutz, D. (2020). Regras do jogo: Uma análise de seus tipos e relacionamentos. In *XIX SBGames – Recife – PE – Brazil, November 7th – 10th, 2020*, pages 164–167. SBC – Proceedings of SBGames. DOI: 10.22533/at.ed.01421030518.

Syafitri, W., Shukur, Z., Mokhtar, U. A., Sulaiman, R., and Ibrahim, M. A. (2022). Social engineering attacks prevention: A systematic literature review. *IEEE Access*, 10:39325–39343. DOI: 10.1109/access.2022.3162594.

Tsutsumi, M. M. A., Goulart, P. R. K., Silva Júnior, M. D., Haydu, V. B., and Jimenéz, □L. d. O. (2020). Avaliação de jogos educativos no ensino de conteúdos acadêmicos: Uma revisão sistemática da literatura. *Revista Portuguesa de Educação*, 33(1):38–55. DOI: 10.21814/rpe.19130.

Unity (2025). Unity documentation. Available at:`https://docs.unity.com/`.

Vlachogianni, P. and Tselios, N. (2023). Perceived usability evaluation of educational technology using the post-study system usability questionnaire (pssuq): A systematic review. *Sustainability*, 15(17):12954. DOI: 10.3390/su151712954.

Wash, R. (2020). How experts detect phishing scam emails.

Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2):1–28. DOI: 10.1145/3415231.

7 Appendix

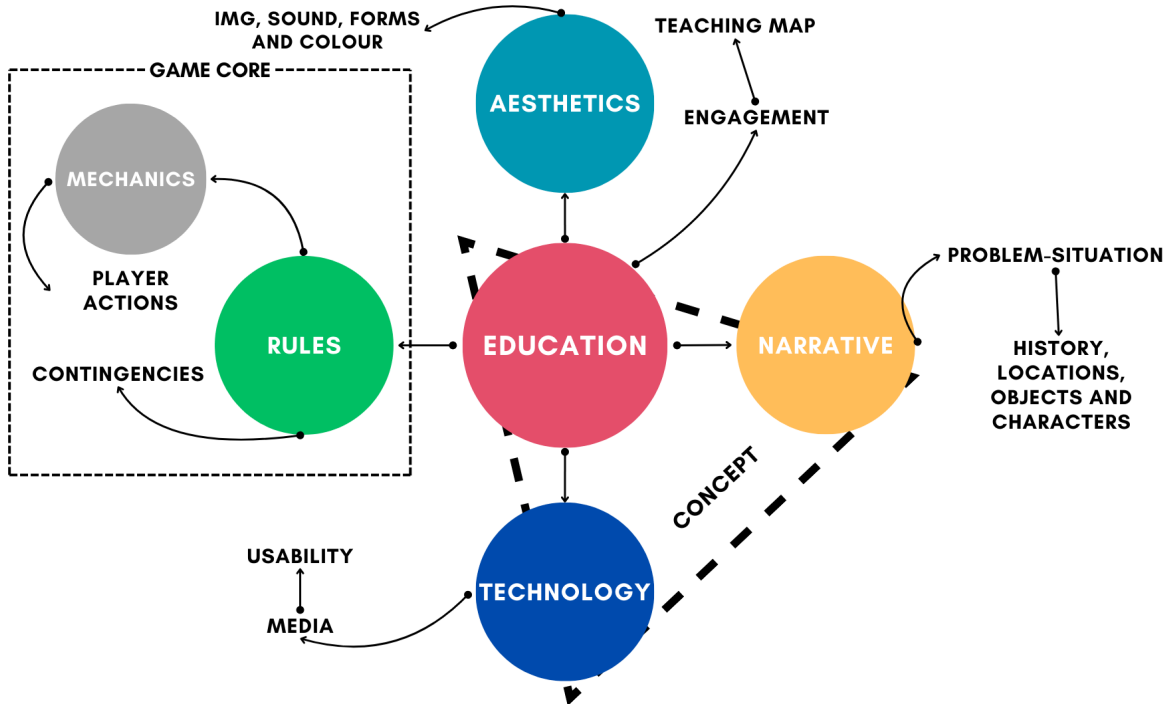
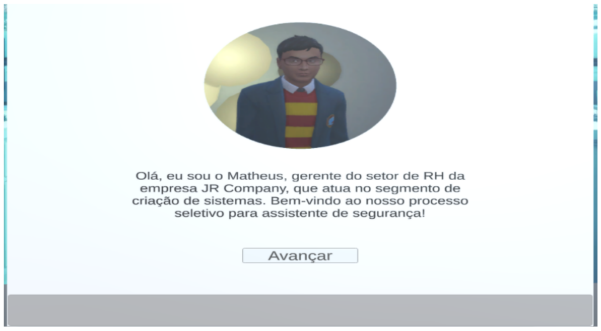


Figure 1. Framework for developing educational games.
Source: The authors.

Main Menu Screen

Alerta game start screen, where the player can begin the experience by clicking the “Play” button.

Instruction Screen



Introduction screen featuring the character Matheus, HR manager at JR Company, who welcomes the player and presents the context of the fictional selection process.

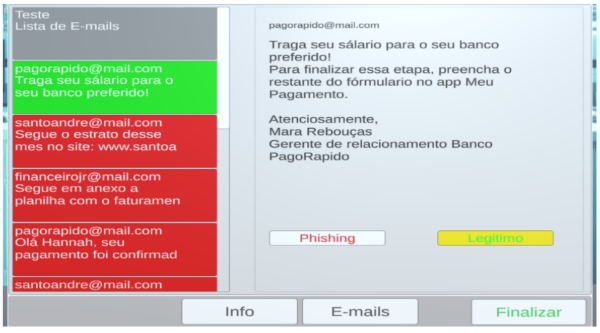
Default Stage Home Screen

Initial screen for every stage, where the user can choose to view personal information or access the inbox to analyze emails.

Character Info Screen

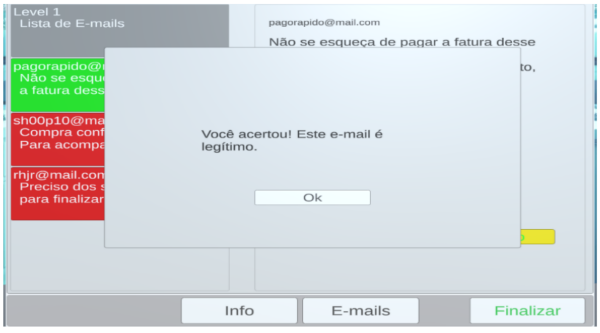
Screen showing the profile of the character Hannah, containing personal and digital information used during the phishing identification challenges in the game.

Inbox and Email Screen



Email analysis interface. The user must classify each message as either "Phishing" or "Legitimate" based on its content and sender.

Feedback Screen



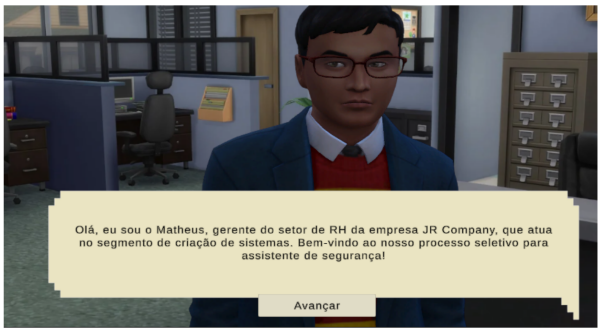
Feedback screen confirming the correct or wrong classification of a email by the user.

Figure 2. Main screens with description of the Alert game (1st and 2nd Iteration)
Source: The authors.

Main Menu Screen

Alerta game start screen, where the player can begin the experience by clicking the “Play” button.

Instruction Screen

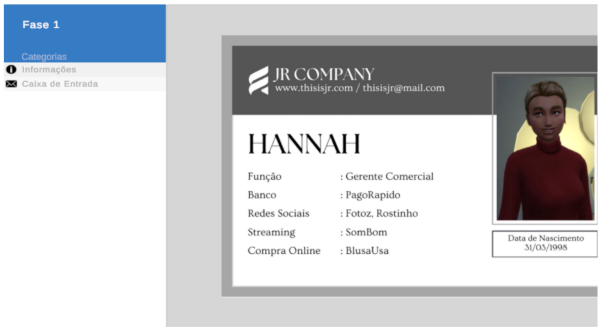


Introduction screen featuring the character Matheus, HR manager at JR Company, who welcomes the player and presents the context of the fictional selection process.

Default Stage Home Screen

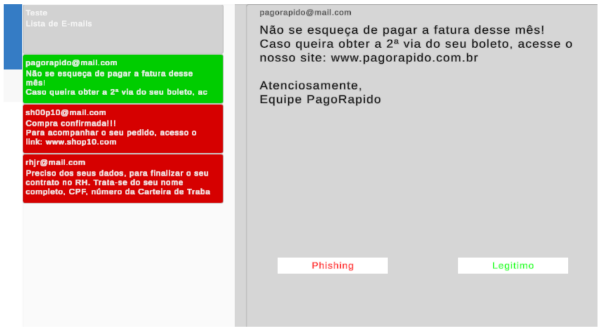
Initial screen for every stage, where the user can choose to view personal information or access the inbox to analyze emails.

Character Info Screen



Screen showing the profile of the character Hannah, containing personal and digital information used during the phishing identification challenges in the game.

Inbox and Email Screen



Email analysis interface. The user must classify each message as either "Phishing" or "Legitimate" based on its content and sender.

Feedback Screen

Feedback screen confirming the correct or wrong classification of a email by the user.

Figure 3. Main screens with description of the Alert game (3rd Iteration)
Source: The authors.